Skolkovo Institute of Science and Technology

DEEP LEARNING FOR REMOTE SENSING OF ENVIRONMENT AND LAND COVER ANALYSIS

Doctoral Thesis
by

Svetlana Illarionova

Doctoral Program in Computational and Data Science and Engineering

Supervisor
Full Professor, Ivan Oseledets

Moscow - 2023

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate (Svetlana Illarionova)

Supervisor (Prof. Ivan Oseledets)

Abstract

Estimation of terrestrial carbon balance is one of the key tasks in understanding and prognosis of climate change impacts and the development of tools and policies according to carbon mitigation and adaptation strategies. The forest ecosystems are one of the major pools of carbon stocks affected by controversial processes influencing carbon stability. Monitoring forest ecosystems is a key to proper inventorying of resources and planning their sustainable use. Development of reliable and up-to-date systems for environmental monitoring and analysis on both local and global scales is crucial for optimal forest management, carbon offset projects, accurate predictions of system changes under different land-use and climate scenarios. In this thesis, we discuss the state-of-the-art computer vision techniques applicable to the most important aspects of forest studies through remote sensing observations. Although there is a wide availability of remote sensing data and various machine learning algorithm to process this data, certain questions of efficient remote sensing pipelines development remains open. We proposed advanced approaches to address the most occurring tasks such as forest areas mapping, tree species classification, canopy height estimation. The goal of this study is to achieve higher quality for environmental characteristics prediction based on novel deep learning approaches using more available and less expensive satellite data. It involves dealing with data imbalance, weakly markup, specific labelled data limitations, and model transferring to new geographical regions.

The present work includes the following steps:

- Precision forest mask estimation;
- Dominant forest species classification;
- Canopy height model prediction;
- Artificial satellite band generation.

Publications

1. Svetlana Illarionova, Dmitrii Shadrin, Polina Tregubova, Vladimir Ignatiev, Albert Efimov, Ivan Oseledets, and Evgeny Burnaev. A survey of computer vision techniques for forest characterization and carbon monitoring tasks. Remote Sensing, 14(22):5861, 2022.

2. Svetlana Illarionova, Dmitrii Shadrin, Vladimir Ignatiev, Sergey Shayakhmetov, Alexey Trekin, and Ivan Oseledets. Augmentation-based methodology for enhancement of trees map detalization on a large scale. Remote Sensing, 14(9):2281, 2022.

3. Svetlana Illarionova, Dmitrii Shadrin, Alexey Trekin, Vladimir Ignatiev, and Ivan Oseledets. Generation of the NIR spectral band for satellite images with convolutional neural networks. Sensors, 21(16):5646, 2021.

4. Svetlana Illarionova, Alexey Trekin, Vladimir Ignatiev, and Ivan Oseledets. Tree species mapping on Sentinel-2 satellite imagery with weakly supervised classification and object-wise sampling. Forests, 12(10):1413, 2021.

5. Svetlana Illarionova, Alexey Trekin, Vladimir Ignatiev, and Ivan Oseledets. Neural- based hierarchical approach for detailed dominant forest species classification by multispectral satellite imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14:18101820, 2020.

6. Svetlana Illarionova, Dmitrii Shadrin, Vladimir Ignatiev, Sergey Shayakhmetov, Alexey Trekin, and Ivan Oseledets. Estimation of the canopy height model from multispectral satellite imagery with convolutional neural networks. IEEE Access, 10:3411634132, 2022.

7. Svetlana Illarionova, Sergey Nesteruk, Dmitrii Shadrin, Vladimir Ignatiev, Mariia Pukalchik, and Ivan Oseledets. Object-based augmentation for building semantic segmentation: Ventura and Santa-Rosa case study. In 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), pages 16591668. IEEE, 2021.

8. Svetlana Illarionova, Sergey Nesteruk, Dmitrii Shadrin, Vladimir Ignatiev, Maria Pukalchik, and Ivan Oseledets. Mixchannel: Advanced augmentation for multispectral satellite images. Remote Sensing, 13(11):2181, 2021.

9. Sergey Nesteruk, Svetlana Illarionova, Timur Akhtyamov, Dmitrii Shadrin, Andrey Somov, Mariia Pukalchik, and Ivan Oseledets. Xtremeaugment: Getting more from your data through combination of image collection and image augmentation. IEEE Access, 10:2401024028, 2022.

Acknowledgments

I would like to thank my supervisor Prof. Ivan Oseledets. I am delighted to be a part of the research group headed by such an eminent scientist. Scientific discussions of applied problems helped me to investigate new study topics. Comments and relevant feedback made it possible to improve my skills in conducting research and writing manuscripts. I realize that much remains to be studied, but I am happy to be on this way.

Let me thank all Skoltech professors, my colleagues and group mates who taught me during my PhD study. It is a great opportunity to work with all of them. I am also grateful to my research co-advisor Dr. Dmitrii Shadrin who kindly shared the experience in scientific work. I would like to thank Prof. Andrey Somov, PhD student Sergey Nesteruk, and Dr. Vladimir Ignatiev for fruitful discussions and consultations on the research topics.

I am really grateful to my parents who support me in becoming a member of the computer science community.

Contents

List of Figures

List of Tables

Abbreviations

ARVI	Atmospherically Resistant Vegetation Index
BAI	Burned Area Index
CNN	Convolutional neural network
CV	Computer vision
DL	Deep learning
EVI	Enhanced Vegetation Index
FPN	Functional pyramid network
GHG	Greenhouse gas
kNN	k Nearest Neighbor
NBR	Normalised Burn Ratio
NBRT	Normalised Burn Ratio Thermal
NDMI	Normalized Difference Moisture Index
NDVI	Normalised Difference Vegetation Index
NDWI	Normalized Difference Water Index
NIR	Near-infrared
OA	Overall accuracy
OBA	Object-based augmentation
RF	Random forest
RS	Remote sensing
LSWI	Land Surface Water Index
ML	Machine learning
SAVI	Soil Adjusted Vegetation Index
SVM	Support vector machines
SWIR	Short-wave infrared reflectance
VCI	Vegetation Condition Index
UAV	Unmanned aerial vehicle

Chapter 1

Introduction

The present Thesis tackles the task of improving existing methods for quantitative assessment of vegetation cover characteristics. The precise and up-to-date vegetation variables estimation is vital for proper environmental studies, in particular, for carbon stocks monitoring and analysis. Recently, the main uncertainty of both local and global estimations is a matter of shortage in relevant and accurate vegetation parameters. On the local scale, errors occur due to insufficient amounts of high quality reference data. On the global scale, the cause of mistakes is diverse environmental conditions.

The Thesis aims at addressing aforementioned challenges. It comprises studies that cover computer vision techniques for the key aspects of forestry analysis. Among these tasks, there are forest mask estimation, forest species classification, and canopy height model estimation. The understanding of these forestry variables is crucial for substantial environmental analysis involving global climate changes mon-

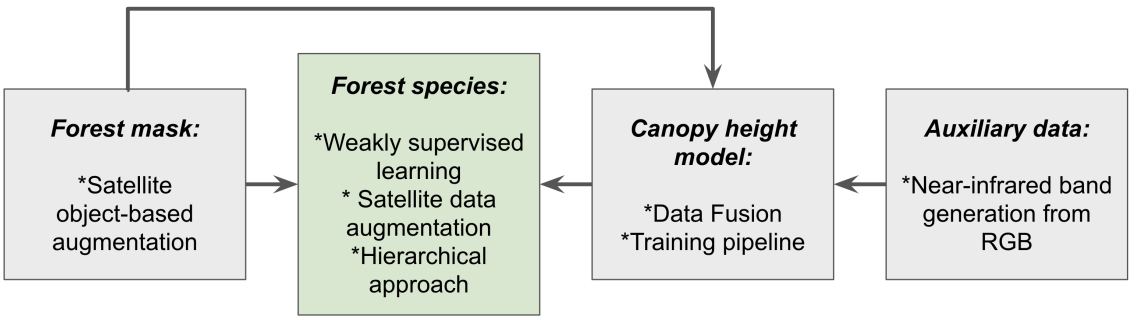


Figure 1-1: Objectives and plan of the study.

itoring. Recently, many remote sensing data sources and computer vision techniques are available for different research purposes. However, there are still particular limitations that provoke further adjustment of approaches applicable to forestry tasks. Deep learning algorithms strongly depend on the amount of high-quality labeled data. Forest inventory data can be out of date, while forestry data collection is a time-consuming procedure that often requires expert knowledge for some tasks. For instance, only an expert can create a precise manual markup with vegetation characteristics based on satellite imagery (distinguish forest species, estimate age). And in most times, data collection cannot be managed without ground-based measurements when it is the only option to distinguish particular forestry variables. Therefore, labeled data limitations are accompanied by a diversity of target objects representation. For example, tree species of different ages vary drastically in spectral range. The proper classification can be affected by environmental conditions, i.e. the surrounding vegetation in the region of interest. Moreover, forest inventory data usually has its own limitations and specificity.

Another challenge in the remote sensing domain is posed by quality-cost ratio. The quality of remote sensing data is defined by its spatial resolution (meters per pixel) and spectral range. Data cost depends on the coverage area. For example, using an unmanned aerial vehicle (UAV) to observe the entire planet is inapplicable. Therefore, one of the currently important points is how to get required data properties artificially.

Aforementioned challenges inspired us to propose approaches less demanding to specific remote sensing labeled data and sensors. Forest species classification has a pivotal role in this study, as it is one of the dominant forest characteristics. We examined how forest species classification can be improved using remote sensing imagery and computer vision algorithms (Figure 1-1). The primary task involves forest mask estimation for different regions using limited precisely labeled data. Then, we propose different approaches to deal with weakly labeled markup and highly imbalanced classes. Additionally, feature generation, such as canopy height model creation, supplied the study with less expensive remote sensing data.

Overall, by using the proposed approaches, we managed to reduce the required

amount of labeled data, to achieve high generalization for new territories and diverse vegetation types. We believe that precise forestry variables will help to understand better climate changes processes.

Chapter 2

Literature review

2.1 Introduction

Climate change adaptation and mitigation policy make relevant the development tools for estimation and monitoring flows of greenhouse gases (GHG). Such accounting of ecosystem balances helps to understand and alter trends of GHG emissions. As for now, a more accurate inventory of carbon stocks and sources is a subject of ongoing discussions. It aims at reducing the uncertainty of carbon balance estimations and their prognosis in different economic and climate change scenarios. It includes clarifying user and social choices in the decision-making process. Improvements of on-site measurement techniques along with scaling of the accounting systems (models) have lead to a more detailed level for a better understanding of the carbon cycle [Peters, 2018, Treat et al., 2018, Tharammal et al., 2019, Santoro et al., 2021].

Among a variety of natural and artificial ecosystems, forests show mostly predominant sequestration of carbon from the atmosphere. Mostly negative net GHG fluxes characterize territories under forests, as a result, gross carbon removals exceeded gross emissions around the world [Harris et al., 2021]. At the same time, in the presence of disturbances, CO_2 emission increases due to a release of the carbon retained in the ecosystem [Seddon et al., 2020]. The list of main disturbing events includes, in general, the change of forests to other land use types, the use of forest resources for materials and energy due to harvesting, the occurrence of fires, fall-

outs, change of water regimes, change of the community structure due to pathogens and invasion outbreaks, and development of deadwood. Those are considered to be managed to maintain sustainable long-term use of natural resources and receive climate benefits [Pingoud et al., 2018, Ontl et al., 2020].

As for now, most of the forest monitoring, management, and planning needs at different spatio-temporal scales can be covered by the use of remote sensing data (RS) [Bourgoin et al., 2018, Kangas et al., 2018, Gao et al., 2020b, Lechner et al., 2020]. Among these tasks are estimation of forest structural and functional diversity, productivity assessment, catching the degradation processes and their patterns, deforestation detection, and analysis, and others. RS data includes both orbital and unmanned aerial vehicle (UAV) observations. For instance, Sentinel-2 mission can provide multispectral information, while the Global Ecosystem Dynamics Investigation (GEDI) mission provides laser measurements. Both of them can be used for carbon cycle studies. The detailed information about orbital missions is presented in Section 2.3. To date, many countries have already included remotely sensed earth observations in their forest inventories within national procedures. However, only 10 to 30% of this information, depending on the data type (satellite images and airborne photography, respectively), is considered for inventory completion [Barrett et al., 2016]. Such data are, in fact, the primary source of information for observing large territories or locations that are hard to access. Currently, there is a strong demand for detailed information about the sources, sinks, and transport of CO_2, as well as about their change under different influencing factors [Janssens-Maenhout et al., 2020, Schepaschenko et al., 2021]. It can be expected that the broad involvement of the RS data into routine protocols of monitoring of natural and managed ecosystems is merely the matter of time [Gschwantner et al., 2022]. Thus, techniques for analyzing RS data are also under development, while their operational integration is an essential part of system knowledge progress [Gao et al., 2020b].

We can notice the steady development of machine and deep learning, improvement of computational resources, along with public availability of the Earth nearly big data (diverse remotely sensed records obtained by a plethora of sensors at various spatial and temporal resolutions). It can be used for the tasks related to the

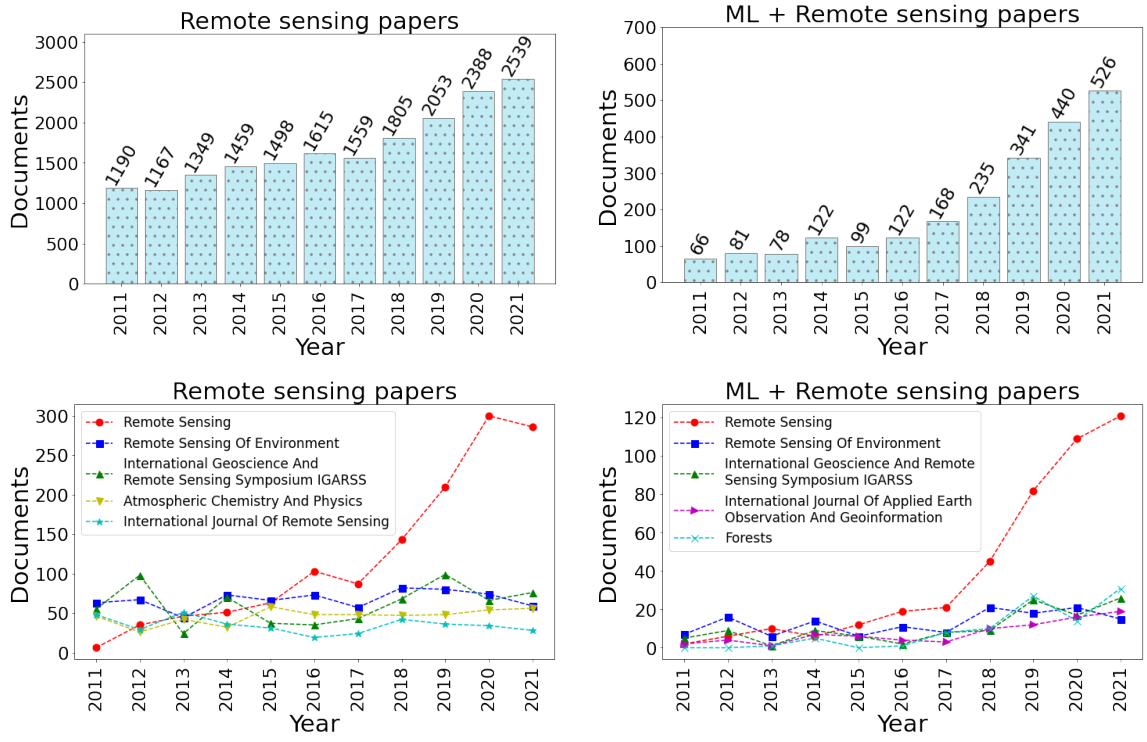
tackling of land-atmosphere interactions, in particular, applicable to the forest areas [Salcedo-Sanz et al., 2020]. In this regard, machine learning algorithms and computer vision (CV) are of great practical and scientific interest. In what follows, by a CV we mean all methods for image processing, and, specifically, classical machine learning methods and deep learning methods based on neural networks. CV techniques are recognized as a powerful tool capable of capturing information from the data of different domains, both photo, and video, and of handling target tasks at different scales. CV algorithms combine the usability and potential for automatization, determined by transparent algorithms' pipelines. By using the standardized list of metrics of can tune the performance and evaluate the model quality. At the same time, it is worth noticing the capability of CV algorithms to integrate expert knowledge during the training procedure [Diez et al., 2021, Spencer Jr et al., 2019]. The sufficient advantage of this group of modeling and analysis methods is its potential to overcome main limitations related to the lack or incompleteness of the data [Chen et al., 2019a, Shorten and Khoshgoftaar, 2019].

There are a number of surveys covering different aspects of forestry studies published in recent years. Particular forest properties estimation, such as aboveground forest biomass, has been surveyed in [Tsitsi, 2016]. It was summarized that RS in forest aboveground biomass estimation is a perspective alternative to conventional ground-based approaches. Since the publication of this survey publishing in 2016, new RS data sources have become available and widely used, and there has also been a drastic rise in machine learning and deep learning, and their implementation in environmental studies. The following surveys were focused on carbon stocks and carbon cycle, highlighting the most commonly used RS data sources [Xiao et al., 2019, Rodríguez-Veiga et al., 2017]. In [Gao et al., 2020b], another forestry problem was observed, namely, forest degradation focused on the used data and its important properties. In turn, in the current survey, we aggregated information from recent studies related mostly to the CV algorithms application for the exact forestry problems: forest mask, tree species, and forest resources estimation. We chose exactly these forest properties, as they are one of the core components for forestry analysis and have a significant impact on carbon monitoring and various environmental

tasks [Xiao et al., 2019]. Currently, there are a lot of data and algorithms that allow one to solve numerous problems, including those related to obtaining forest characteristics based on satellite data. Due to the wide variety of data sources, their specifics, and algorithms, it is difficult for a novice researcher to find a suitable approach that combines the use of certain data and algorithms that would give a good result shortly. In order to have an understanding of the available data and algorithms that best solve the particular problem, taking into account the specifics of the problem being solved, we provide this survey. It covers the most popular data sources and widely used methods, as they are both of high value for accurate RS solutions. It will allow researchers to effectively select a set of suitable algorithms and data sources that will solve a specific problem.

2.2 Review methodology

Interest in remote sensing of the environment, namely, forest characteristics estimation, has been constantly growing during the last decade, as shown in Figure 2-1. To collect year wise statistics, we used two sets of words as keywords in the "article title, abstract, and keywords" in the Scopus database search system. The first set of words specifies remote sensing research domain, including words "remote sensing", "UAV", and certain widely used satellites names. The second set of words concretizes a specific forestry properties and tasks such as "tree mapping", "growing stock volume", "age", "forest species", etc. The "AND" Boolean operator united these two sets of words, while within each set, the "OR" operator was applied. The search resulted in over 18000 publications from 2011 to 2021 year. The search excluded subject areas such as Medicine, Social Science, etc. In Figure 2-1, "ML + Remote sensing" refers to intersection of the previous search results with a set of words specifying artificial intelligence algorithms such as "machine learning", "deep learning", "neural networks", and names of widely used algorithms. The search resulted in over 2200 documents from 2011 to 2021 year. There is a solid growth in the number of publications considering artificial intelligence since 2015 year. Comparing the search results for machine learning applications for different RS forest tasks, we can notice that the



(a) (b)

Figure 2-1: Year wise publication of remote sensing papers for forest characteristics extraction: number of publications per year; the most popular journals according to the number of publications. The data was retrieved from the Scopus database. (a) General search that include remote sensing for forest tasks keywords; (b) Intersection of general remote sensing search for forest tasks results with ML-specific keywords.

most frequently encountered task is forest resources estimation such as aboveground biomass, growing stock, standing volume (over 900 documents from 2011 to 2021 years). Forest species classification using artificial intelligence ranks second in the search results (over 800 documents). Classical machine learning algorithms (such as Random forest, Gradient boosting, etc.) occur three times as often as deep learning algorithms. Among RS data, Landsat is referred with machine learning algorithms in over 400 publications. Sentinel data was mentioned in over 340 papers for forest tasks using machine learning techniques, while WorldView data was mentioned in over 100 publications. The detailed information about data sources and tasks is presented below.

The literature analysis is performed using recently published studies from peer-reviewed journals included in Scopus scientific database. Google Scholar database additionally supported the search. Due to the rapid development of the data science discipline, the survey timeline was limited, starting from the year 2017 to 2022. An exception is applied to the research which is fundamental or was a pioneer to the topic, according to the citation level and earlier years of publishing. For each discussed topic, we used relevant keywords, for example, "aboveground biomass" AND "forestry" AND "remote sensing". Results of such requests were compared with search query after adding the words "computer vision", which was not obligatory used firstly because of the frequent association of the phrase with neural networks only. Then search results were manually examined with and without sorting by the citation level, and relevant works were chosen for detailed analysis. Zero citations were acceptable in the case of specific (influential) journals relevant to research topics of earth science, environmental science, and environmental monitoring and 2022 year of publishing. However, to catch the general context, for the most part, publications with more than 20 citations were considered. Our search was limited by the keywords' combinations in the title and abstract, the final publishing stage, English language. We mainly considered research articles to make conclusions about the applicability and efficiency of the algorithms. However, our analysis also includes relevant comprehensive reviews' references providing general trend analysis within the topic of climate change mitigation actions and data sources.

2.3 Remote sensing data and spectral indices for forest analysis

2.3.1 Sources of remote sensing data

An essential part of developing a vegetation analysis methodology is the informed choice of one or another data source. We refer the reader to the latest extensive surveys dedicated to the descriptions of the common RS platforms and sensor combinations applied to the problem of vegetation analysis [Calders et al., 2020, Gao et al., 2020b, Lechner et al., 2020, Zeng et al., 2022], while noting that this is a rapidly evolving field. In the present study, we provide the list of main characteristics of currently most commonly used instruments of particular importance for forest-related tasks at different scales (Table 2.1).

Table 2.1: Commonly used instruments of remote sensing data acquisition and distribution, and their characteristics aggregated from [Salcedo-Sanz et al., 2020, Tang et al., 2019, Zhang et al., 2019b, Rostami et al., 2022, Stych et al., 2019, Deigele et al., 2020, Lakyda et al., 2019] and missions' technical websites [NASA, JAXA, NASA and the U.S. Geological Survey, European and the Space Agency, MAXAR, Planet, Airbus]

Mission	Sensor	Spatial resolution	Temporal resolution	Distribution of data
Terra MODIS	Multispectral, 36 bands	250 m, 500 m, 1 km	1-2 days	Open and free basis
ALOS PALSAR/ ALOS-2 PALSAR-2	Synthetic Aperture Radar, L-band	From detailed (1-3 m) to low (60-100 m) depending on the acquisition mode and processing level	14 days	On request/commercial use/ALOS Palsar 1-free

Landsat-8/9	Multispectral 8 bands, panchromatic band, and thermal infrared 2 bands	Multispectral: 30 m, Panchromatic: 15 m, Thermal Infrared Sensor: 100 m	16 days (the combined Landsat 8 and 9 revisit time is 8 days)	Open and free basis
Sentinel-1	Synthetic aperture radar, C-band	From detailed (1.5 x 3.6 m) to medium (20-40 m) depending on the acquisition mode and the processing level	Mission closed (during operating time - 3 days on the Equator, <1 day at the Arctic, 1-3 days in Europe and Canada)	Historical data is open and free basis
Sentinel-2	Multispectral, 13 bands	10, 20, 60 m depending on the band range	5 and 10 days for single and combined constella-tion revisit	Open and free basis
WorldView-1	panchromatic band	panchromatic: 0.5 m	1.7 days	Commercial use

WorldView-2,3	Multispectral – 8 bands, panchromatic band	Multispectral:1.84 m, panchromatic: 0.46 m	Up to 1.1 days	Commercial use
WorldView-4	Multispectral – 4 bands, panchromatic band	Multispectral:1.24 m, panchromatic: 0.31 m	mission closed (during operating time < 1 day)	Commercial use (archive)
GeoEye - 1	Multispectral – 4 bands, panchromatic band	Multispectral:1.64 m, panchromatic: 0.41 m	1.7 days	Commercial use
PlanetScope	Multispectral – 4 bands, from 2019 additional 4 bands	3.7-4.1 m resampled to 3 m	1 day	On request/commercial use
SPOT-6,-7	Multispectral – 4 bands, panchromatic band	Multispectral: 6 m, panchromatic: 1.5 m	1 to 5 days	On request/commercial use
Pleiades	Multispectral – 4 bands, panchromatic band	Multispectral: 2 m, panchromatic: 0.5 m	1 day	Commercial use
RapidEye	Multispectral – 5 bands	6.5 m, resampled to 5 m	1 day	Commercial use

When choosing a data source for research, various details are taken into account: data availability, survey repeatability, spatial resolution, sensor type, sensor specifications, range of spectral channels, etc. Describing RS data, one can distinguish different spatial, temporal and spectral resolution, while the "spatial" meaning is

the most frequent case. In the case of spatial resolution, we refer to the precision classification as coarse (low), medium and fine (high) [Chen et al., 2019b, Xiao et al., 2019]. Thus, low resolution corresponds to the data of pixel size of more than 30 m per pixel, medium resolution corresponds to the data of pixel size from 10 to 30 m, and high resolution, to the size of less than 5 m.

Many RS missions are capable of providing up-to-date and diverse information about the object or phenomenon under consideration. Each approach has advantages and limitations, determined by the detection conditions and the ratio between spatial resolution, revisiting time, or cost. Active sensors such as radar are independent of the weather conditions and do not rely on the sun as a source of illumination, and so can provide the data regardless of the day or night conditions. On the contrary, passive multispectral and hyperspectral sensors require solar radiation. Additionally, the coarser the image obtained from the satellite, the more often it is taken. To combine the frequency of low-resolution imaging with more details of the other available data or to restore the information that was lost due to unsuitable conditions, different machine learning fusion techniques were implemented [Salcedo-Sanz et al., 2020].

For every particular experiment, one can use data from one of the available for required time and region satellites. The number of available bands in different satellites may vary. It is possible to use all bands available in the chosen satellite, select, or combine (to obtain spectral indices) some of them. A special case of the auxiliary use is the panchromatic channel. Due to its wide band, it gains more light; therefore, it has higher spatial resolution. It makes it possible to adjust the resolution of satellite imagery through pan-sharpening techniques [Javan et al., 2021].

One of the advantages of platform-distributed RS data is its availability and easy-to-use web and scripting interfaces for collecting the data by end-users. It can be downloaded, free of charge or for payment, from data-aggregating platforms, provided raw data, as well as pre-processed or converted into various valuable derivatives such as spectral indices or reflectance bottom of the atmosphere (BOA). This allows the user to generate sets of images for efficient training of machine learning algorithms for different regions and relevant dates for the study. For exploring resources

and performance capabilities, we refer the reader to, e.g., Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center (LAADS DAAC) Tools and Services collection (https://ladsweb.modaps.eosdis.nasa.gov/tools-and-services/), Copernicus Open Access Hub (https://scihub.copernicus.eu), the Planet Platform (https://www.planet.com/products/platform/).

2.3.2 Popular spectral indices applied for forest monitoring research

Remote sensing data is rich in information, and in the case of multispectral sources, separate bands can be mathematically transformed and combined. Such composites, namely spectral indices, can be used to catch specific patterns necessary for the most common tasks of forest carbon monitoring. Among these tasks are forest area estimation, tree stand composition classification, change or anomaly detection, and others that will be covered further in the following sections.

There are dozens of different spectral indices, and many of them have different modifications [Xue and Su, 2017, Zeng et al., 2022]. Here we discuss some of the most frequently utilized. One of the most popular spectral indices is the group of vegetation indices, lead by the Normalized Difference Vegetation Index (NDVI), based on the near-infrared (NIR) and red reflectance bands. NDVI has derivatives, one of which is the Vegetation Condition Index (VCI), based on the minimum and maximum values of the NDVI for a given period. For the study of atmospheric effects, the Atmospherically Resistant Vegetation Index (ARVI) can be used. Although NDVI is a commonly used choice to analyze vegetation cover, it is affected by a saturation problem for densely forested areas [Tesfaye and Awoke, 2021]. The main reason is that if there is a total leaves cover in the high vegetation period (peak of the vegetation period) leaves are not able to absorb red light, so the reflectance of red light will increase. Moreover, intensity of NIR will also increase. According to equation (2.1), the calculated NDVI will be underestimated. To address the saturation problem, Enhanced Vegetation Index (EVI) can be used. It is more accurate in areas with high vegetation and considers both soil and atmospheric effects.

Another possible choice for densely forested areas is indices based on a Red-edge spectral band: Normalized Difference Red-edge (NDRE), the Modified Simple Ratio (MSR) Red-edge index, and Chlorophyll Index (CI) Red-edge. However, only some satellites have sensors for Red-edge measurements. The Red-edge band is available, for instance, in the satellite systems such as Sentinel-2, WorldView-2 and 3, and RapidEye. The indices are computed using the following equations:

$$NDVI = \frac{NIR - Red}{NIR + Red}, \tag{2.1}$$

$$NDRE = \frac{NIR - RedEdge}{NIR + RedEdge}, \tag{2.2}$$

where *NIR* is the near-infrared spectral band, *Red* is the red spectral band, *RedEdge* is the red-edge spectral band,

$$VCI = \frac{NDVI_{i,p,j} - NDVI_{min_{i,p,j}}}{NDVI_{max_{i,p,j}} + NDVI_{min_{i,p,j}}}, \tag{2.3}$$

where i is the pixel, p is the period, j is the year,

$$ARVI = \frac{NIR - 1 * (Red - Blue)}{NIR + 1 * (Red - Blue)}, \tag{2.4}$$

where *Blue* is the blue spectral band,

$$EVI = \frac{2.5 * (NIR - Red)}{6 * Red - 7.5 * Blue + 1}, \tag{2.5}$$

$$CI_{RedEdge} = \frac{NIR}{RedEdge} - 1, \tag{2.6}$$

$$MSR_{RedEdge} = \frac{NIR/RedEdge - 1}{\sqrt{NIR/RedEdge + 1}}. \tag{2.7}$$

The aforementioned indices are widely used as input data (separately and along with initial bands) to detect the vegetation cover among other different land cover types [Pflugmacher et al., 2019], to distinguish between different plant species [Immitzer et al., 2019], to evaluate plant target characteristics such as productivity and

mortality [Rogers et al., 2018, Dang et al., 2019], or to detect insect defoliation [Marx and Kleinschmit, 2017].

In the field of forest monitoring, one of the most relevant topics is the fire occurrence and spread detection and mitigation [Anderegg et al., 2020]. Thus, in addition to described indices, other spectral combinations, mostly based on short-wave infrared reflectance (SWIR), are of common use. For example, one can distinguish the Normalized Burn Ratio (NBR) and derivative Normalized Burn Ratio Thermal (NBRT), Burned Area Index (BAI) for the purposes of the assessment of fire severity and burn area detection [Tran et al., 2018]. The indices are defined by the equations:

$$NBR = \frac{NIR - SWIR}{NIR + SWIR}, \tag{2.8}$$

where $SWIR$ is the shortwave infrared band,

$$NBRT = \frac{NIR - SWIR * TIR}{NIR + SWIR * TIR}, \tag{2.9}$$

where TIR is the thermal band,

$$BAI = \frac{1}{(0.1 + Red)^2 + (0.06 + NIR)^2}. \tag{2.10}$$

To consider environmental characteristics and to use it as background for terrestrial and aquatic or coastal forest ecosystems, the following indices are used in addition: Soil Adjusted Vegetation Index (SAVI) allowing to correct soil brightness [Hislop et al., 2018], the Normalized Difference Water Index (NDWI), also known as the Land Surface Water Index (LSWI), the Normalized Difference Moisture Index (NDMI) [Hislop et al., 2018, Zaimes et al., 2019]. Such indices are also employed for track damages other than fire such as, e.g., pathogens outbreaks [Huang et al., 2019a]. SAVI and NDWI are computed using equations:

$$SAVI = \frac{((NIR - Red) * (1 + L)}{NIR + Red + L}, \tag{2.11}$$

where L is the soil factor, ranging from 0 to 1, which corresponds to dense vegetation and no vegetation, respectively, while 0.5 is considered default for the most land

cover types,

$$NDWI/LSWI/NDMI = \frac{NIR - SWIR}{NIR + SWIR}. \tag{2.12}$$

Use of spectral indices has certain limitations that should be taken into account. Such indices are applicable for the work with RS data in general and include atmospheric effect, the possibility of significant difference between index values in case of different data sources [Huang et al., 2021], season dependence, in complete accordance with the objects' features [Cunliffe et al., 2020]. However, spectral indices provide a valuable source of information with careful pre-processing including appropriate atmospheric correction according to the data source, topographic correction, and understanding the uncertainties along with the availability of the actual measurements (label data). Depending on the goals and study object characteristics, new indices can be proposed based on the previously not used band combination sequences [Jia, 2019]. For the time series, index pattern derivatives can be used such as standard deviation, kurtosis, and skewness [Rogers et al., 2018]. For the recognition and modeling tasks, indices are usually used in combination with each other. Hyperspectral data can also be aggregated into the indices [Marrs and Ni-Meister, 2019].

2.4 Computer vision algorithms

In this section, we describe widely useful supervised algorithms for RS data, in particular, for forest tasks. We discuss both classical machine learning and deep learning algorithms with their specifics, learning process details, and intuition behind them.

Semantic segmentation is a machine-learning problem for which the algorithm learns to determine the class of each pixel using training samples. A feature description characterizes each target object. There is a matching between the input image pixels and ground truth image pixels, which is supposed to be a mask of a perfectly segmented image. The model is aimed at reducing the difference between the prediction and reference markup according to a given quality metric. For in-

stance, in the case of forest mapping, two classes are considered: the forest cover and the areas without forest. Below we describe the specifics and differences between the classical machine learning algorithms and the DL methods, schematically shown in Figure 2-2. Although for CNN algorithms, task definition as a semantic segmentation is more conventional; for classical ML approaches, the task is usually defined as a pixel-oriented classification or regression. It means that CNNs work with pixels and their surrounding area (neighbor pixels). ML algorithms typically work with individual pixels independently.

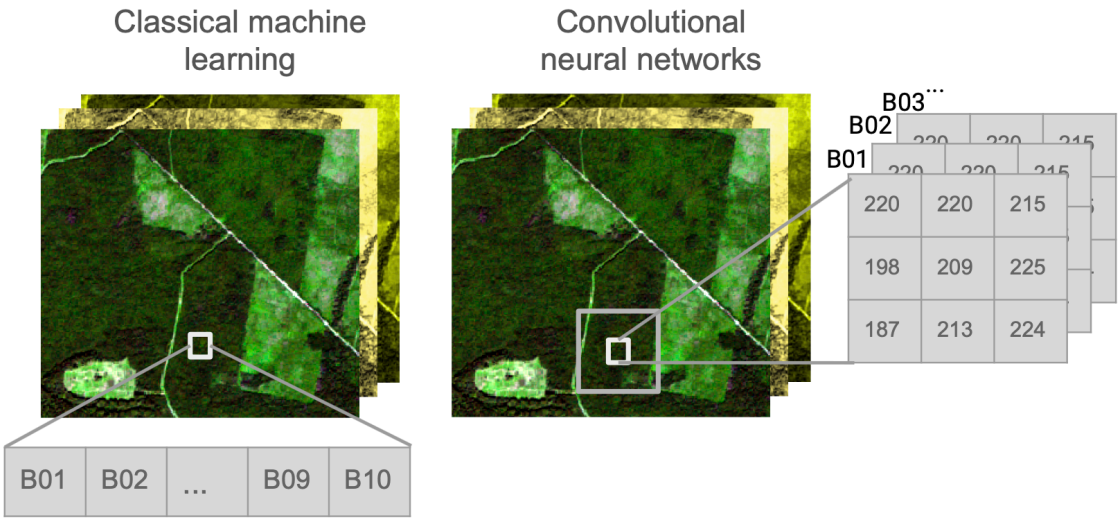


Figure 2-2: Difference between classical machine learning and deep learning algorithms.

2.4.1 Classical machine learning algorithms

To solve various tasks using RS data, one of the most effective and popular methods of classical machine learning is the random forest method (Random Forest, RF), which combines the approaches of an ensemble (a composition) of algorithms, namely, decision trees and the method of random subspaces. This method is widely applicable, for example, for solving problems of classification of forest-forming trees, as well as for solving regression problems, but it is not limited only to these tasks (described in Section 2.8). The ensemble is performed over multiple trees trained on different data subsamples, which helps one to avoid the overfitting issues occur-

ring when only one decision tree is used. The resulting class or value prediction is made by averaging over all trees or choosing a class that is predicted by most of the trees. An important limitation of using a forest of trees compared to a standard decision tree is the interpretability of the results. Namely, a random forest of trees itself is much harder to interpret. The main parameters configured in the RF algorithm are as follows: the number of trees that determine the complexity of the algorithm; the number of features for splitting selection; the maximum depth of the trees responsible for the retraining and accuracy of the model; the criterion by which the homogeneity (entropy) of each leaf in the tree will be evaluated; the minimum number of objects at which splitting is performed, with a decrease in this parameter, the quality of training increases, but the training time also increases. One of the advantages of the RF algorithm is the speed of its learning process and ease of use, meaning that the algorithm is already implemented mostly with open-source programming languages and data analysis interfaces. For example, the Python Scikit-learn library [Pedregosa et al., 2011b] has an implementation that allows the user to quickly tune the hyperparameters and train and test the model.

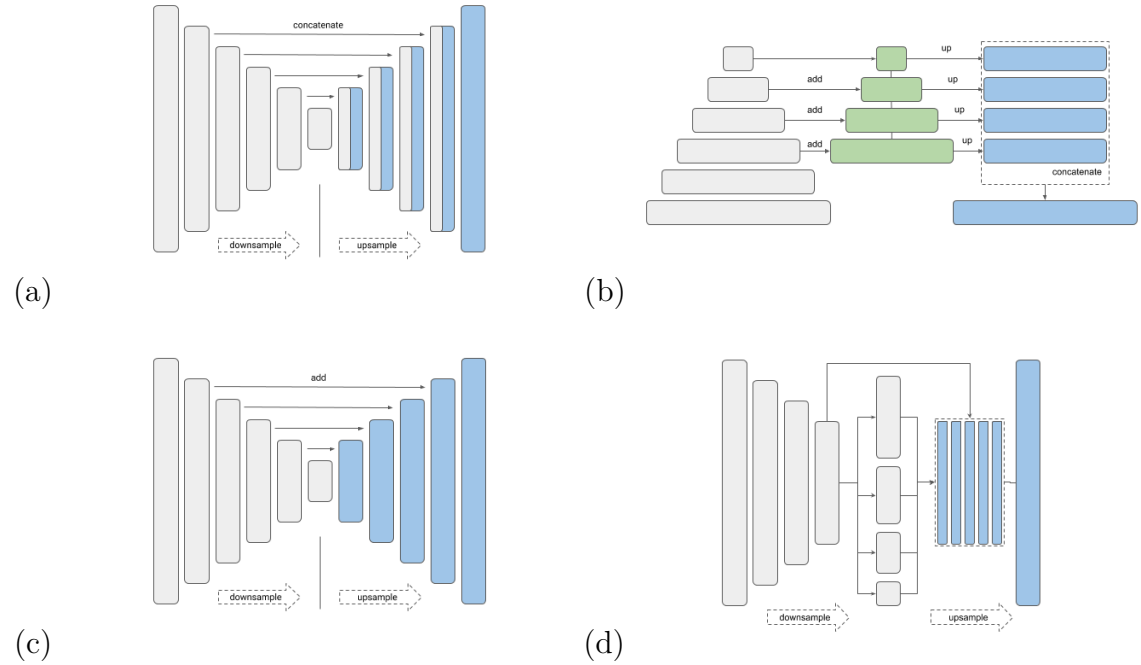
Another effective method of classical machine learning capable of dealing with CV tasks for RS data analysis is the Support Vector Machine (SVM). This is a class of algorithms characterized by the use of kernels (including nonlinear ones) and the absence of local minima; they are aimed at solving both classification and regression problems. In the case of classification problems, the optimal hyperplane is determined, which provides the best separation of classes. SVM requires parameters to be tuned at the implementation, of the main are the kernel type and its hyperparameters. One of the most popular kernel type is the Gaussian kernel (rbf), in which the C and γ parameters are configured for the misclassification penalty and the width of the kernel. It is necessary to vary the above parameters to obtain better accuracy and avoid over-fitting. Support Vector Regression (SVR) is based on the same approach as the SVM for the classification task, specifically, error minimization at determining the separating hyperplane for class extraction with a few slight differences.

The k-nearest neighbor (KNN) algorithm is a frequent choice for RS problems

because of its simplicity and high interpretability [Altman, 1992]. It is a non-parametric supervised learning algorithm that commonly considers the Euclidean distance between an observed sample and its neighbors to make a prediction. The most similar data points are located closer to each other. Therefore, the class of a new data point can be estimated by voting the k most close points as the most frequently observed class. The number of neighbors (k) that participate in voting is defined empirically and depends on a particular task. The KNN algorithm can also be used for a regression problem. So, the output of the algorithms for the observed data point is the averaged target value for all k nearest neighbors.

The Gradient Boosting algorithm is also widely used in environmental studies both for regression and classification tasks. The boosting technique is an efficient ensemble approach when the model is built sequentially using weak learners [Friedman, 2002]. Although gradient boosting can be based on different learners, the most common choice is decision trees. Each weak learner aims to minimize the error of the previous learner, being highly correlated with the negative gradient of the loss function of the previously assembled trees [Natekin and Knoll, 2013]. XGBoost ("Extreme Gradient Boosting") algorithm is an adjustment of gradient boosting over trees that are based on the usage of a more powerful regularization technique to decrease over-fitting [Chen and Guestrin, 2016a]. XGBoost supports parallelization within each tree, creating new branches independently, which makes the algorithm faster.

To adjust the quality of machine learning algorithms, one can apply the following approaches. For instance, the principal components analysis (PCA) is a dimensionality reduction method that creates a smaller dataset from a large amount of features preserving important information. This linear unsupervised statistical transformation was successfully applied to RS multispectral and hyperspectral data [Uddin et al., 2021]. Another approach to reduce feature space is to use the RF algorithm for feature selection and then to train another machine learning algorithm with selected features. However, correlated features should be excluded as their importance might be underestimated. Another important option for model performance adjustment is optimal parameters selection. To optimize machine learning param-



(a) (b)

(c) (d)

Figure 2-3: CNN architectures [Yakubovskiy, 2022]: (a) U-Net; (b) FPN; (c) LinkNet; (d) PSPNet.

eters, one can leverage various optimizations tools, such as Optuna [Optuna], or scikit-optimize [scikit optimize] (including Bayesian optimization).

2.4.2 Deep learning algorithms

One of the main and frequently used architectures of CNNs for RS image processing, including the forest mask segmentation, is the U-Net architecture [Ronneberger et al., 2015]. The schematic layout of various U-Net layers is shown in Figure 2-3 (a).

The architecture comprises two parts, forming a "U"-shape. The first part includes several convolution layers (parameters can be configured) responsible for feature selection. For a convolution operation, it is typical to use a 3-sized convolution kernel, followed by a nonlinear ReLU and Max-pooling layer for dimensionality reduction. The second part uses layers that convert a feature map from a compressed space into the initial dimension (deconvolution). Also, in the second part of the deconvolution process, there is an attachment of relevant maps of features obtained during the roll-up. As a result, the neural networks output produces an image mask

of the same size as the input image, where each pixel corresponds to a particular class. Training the neural network, namely, picking up values in the kernels, is effected by using the backpropagation method of the error. The neural network's weight (trainable) parameters are updated iteratively based on calculated error. The error is computed using different loss functions such as cross-entropy function:

$$Loss = \frac{\sum_{i=1}^{N} \sum_{k=1}^{C} y_{ik} * \log \hat{y}_{ik}}{N}, \tag{2.13}$$

where N is the number of pixels, C is the number of target classes, $\hat{y}$ is the probability predicted by the model that a pixel belongs to a particular target class, y is the ground truth label of membership of a pixel in a class (0 or 1).

The corresponding gradients for all layers and weights in the neural network are then updated.

In addition, one can configure the importance of each class with weight functions so that the neural network can work with an unbalanced number of target classes.

Another efficient architecture for solving the problems of segmentation and mask selection is also worth mentioning — here we mean the Functional Pyramid Network (FPN) [Lin et al., 2017]. The schematic diagram (architecture) of FPN is shown in Figure 2-3 (b). This architecture has two parts, one from the bottom-up (convolution) and another from the top-down (deconvolution). One of the key features of this architecture is the simultaneous utilization of features of different resolutions and levels. The lower semantic weight (the lower generalizing ability) has high-resolution features, and the higher semantic weight has low resolution features. The lateral connections between the two paths make it possible to eliminate the problem of signal attenuation. As a result, it becomes possible to process the detailed information obtained at the bottom of the first pyramid and semantically significant features obtained at the top of the first pyramid.

Other neural network architectures relevant for RS tasks include FCN [Long et al., 2015], DeepLab [Chen et al., 2017b], LinkNet [Chaurasia and Culurciello, 2017] (Figure 2-3 (c)), PSPNet [Zhao et al., 2017] (Figure 2-3 (d)).

All mentioned architectures have the same prediction pipeline. Input data is

passed from the first layer to the following layers. On each layer, input signal is transformed depending on weights of artificial neurons and then fed to activation function. Thus, non-linear data separation is enabled [Forstmaier et al., 2020].

The neural network parameters as well as approach for its training takes a separate vital place in the development of effective algorithms for solving RS problems, in particular to characterize vegetation with the possibility of further conversion to carbon stock. The training parameters include the number of training epochs, the number of steps in each epoch, the size of the batch (sub-sample), and the size of the images that form the batch. The choice of these parameters affects the ultimate result of neural network predictions. Parameters monitoring and analysis during the training allow one to choose an optimal moment of training process termination and to avoid overfitting. Also, the convergence rate of the algorithm depends on the value of the step in learning (learning rate) and the choice of the optimizer. In many studies, it is proposed to use optimizers like SGD [Robbins and Monro, 1951], Adam [Kingma and Ba, 2014], RMSProp [Hinton and Swersky, 2012]. The determination of the stopping moment for neural network training is often based on such indicators as the "plateau". The "plateau" effect means that the validation samples accuracy does not increase during several epochs. Also, among the configurable parameters of the neural network, it is worth mentioning the importance of the CNN's optimal size (depth). The depth choice depends on the problem to be solved and the amount of training data. Therefore, the number of layers and neurons is task-specific. For instance, for a small dataset, it is preferable to use a model with fewer learning parameters. However, if the dataset is large, common architectures with large amount of parameters are required. Such neural networks often do not fit a single GPU. To address this limitation, different approaches and strategies can be applied in order to train large neural networks effectively [J. Gusak and Beaumont, 2022]. To deal with overfitting on small datasets and to enhance model generalization, the dropout technique is usually implemented. It involves discarding some randomly selected nodes (both from input and hidden layers) during each training iteration. It reduces co-adaptation between neurons. During test time, dropout is not implemented, but weights are adjusted by the used training dropout

ratio. However, some dropout modifications support their application during the test phase such as Monte Carlo dropout [Abdar et al., 2021].

2.5 Evaluation metrics

In this section, we describe the most commonly used metrics to evaluate ML and DL models. The input and output data are looked upon here as raster images (a more conventional representation for DL algorithms). However, the described metrics applied to evaluate ML algorithms that work with individual spatial points and satellite features as tabular data.

2.5.1 Classification

To assess the per-pixel or region prediction quality of machine learning algorithms, the following inputs are used:

1. per-pixel mask of the target classes, ground truth;
2. per-pixel predicted mask with target classes.

Masks are in the raster format; the value of pixels belonging to the background is 0, and belonging to an object of the target class is 1 or more for a multiclass case. Therefore, when we have only two classes (target class and background), the mask has a Boolean representation. For instance, in the case of forest mask: areas covered by forest vegetation are marked with value 1, areas of other types have label 0. To calculate the prediction quality, True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values are considered. True Positive is the number of correctly classified pixels of a given class; False Positive is the number of pixels classified as a given class while, in fact, being of another class; True Negative is the number of correctly classified pixels of another class; False Negative is the number of pixels of a given class, missed by the method. One can estimate the model quality based on the ratio between correctly classified objects and all objects representing the study area. This commonly used metric is Accuracy:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}. \tag{2.14}$$

To evaluate the performance of neural networks for semantic segmentation or classical machine learning models, one can also apply F1-score, that is widely used in RS tasks [Kattenborn et al., 2021a]. While the Accuracy metric is a good choice in the case of balanced classes, the F1-score is capable of effectively assessing the prediction quality for imbalanced classes. A high Accuracy score can be obtained for highly imbalanced data by assigning the majority class's label to all observations. F1-score is computed using the following equations:

$$\begin{gathered} Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, \\ F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} = \frac{2 * Precision * Recall}{Precision + Recall}. \end{gathered} \tag{2.15}$$

Another popular metric for semantic segmentation tasks is IoU (intersect over union). F1-score and IoU are positively correlated metrics. However, F1-score is the harmonic mean, and IoU is closer to the minimum value between Precision and Recall. The equation for IoU is

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} = \frac{Precision * Recall}{Precision + Recall - Precision * Recall}. \tag{2.16}$$

The area under the curve (AUC) from the receiver operating characteristic (ROC) also helps to assess the quality of developed algorithms. True positive rate (TPR) and False positive rate (FPR) are estimated in order to build the ROC curve for different decision thresholds. We assume the model outputs certainty that the object belongs to the positive class. Therefore, these thresholds determine objects belonging to the positive class. AUC is the area for all possible decision thresholds for TPR and FPR combinations. It shows the model ability to range predictions correctly. TPR and FPR are computed using the following equations:

$$\text{TPR} = \frac{TP}{TP + FN}, \text{FPR} = \frac{FP}{FP + TN}. \tag{2.17}$$

2.5.2 Regression

One can distinguish the following metrics from the most common metrics in regression tasks: Mean absolute error (MAE), Mean square error (MSE), Root mean square error (RMSE), Coefficient of determination (R^2), Mean absolute percentage error (MAPE), Mean bias error (MBE). Although MAE, MSE, RMSE, and MAPE aim to estimate how close the model prediction is to the actual values, they have differences. Depending on the task, they can be effectively combined for deeper model results analysis. Intuitive interpretation is indispensable for various practical forestry tasks. While MAE provides error in the original unit of measure of actual target values, MAPE is commonly used to assess the error in percentages for more straightforward competitive analysis. Comparing MAPE and MBE with MAE, MSE, and RMSE, we can notice that only MAPE and MBE metrics take into account the position of the actual target and predicted values, i.e., a switching between these values leads to different results. MBE makes it possible to understand the model tendency for under- or overestimation of the target values as it can be both positive and negative. R^2 shows the relation between the total variance explained by the model and the total variance in actual target data. The metrics are computed using the following equations:

$$MAE = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N}, \tag{2.18}$$

$$MSE = \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}, \tag{2.19}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}}, \tag{2.20}$$

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right|, \tag{2.21}$$

$$MBE = \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)}{N}, \tag{2.22}$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y}_i)^2}, \tag{2.23}$$

where $\hat{y}_i$ is the predicted value of the ith object, y_i is the true value of ith object, N is the amount of objects (pixels), $\overline{y}_i$ is the mean value for all objects, SS_{res} is the sum of residual squares, SS_{tot} is total sum of squares.

2.6 Forest mask estimation on the remote sensing data

One of the initial steps in environmental studies based on RS data is the forest mask estimation. One can extract the required vegetation properties within such a mask, for instance, tree species, age, or canopy height. Another strongly related task to forest mask estimation is the deforestation problem, as it directly affect forest boundaries. The approaches to solve these two tasks are often quite similar. Selection of optimal data type and algorithm should always take into account the specifics of the problem to be solved. When ML methods are applied for these tasks one usually considers the semantic segmentation problem. According to the study requirements, different data sources can be used for this task. Therefore, both low, medium, and high spatial resolutions cover various cases with their advantages and disadvantages.

2.6.1 Use of data of different spatial resolution

Low spatial resolution

Low spatial resolution is recommended for regional and national assessments of forest cover characteristics. One of the popular sources of such data is imagery of the MODIS apparatus. Time series usage based on MODIS data with spatial resolution of 500 m in pixels has been successfully implemented in [Hansen et al., 2008] to assess changes in forest cover in Brazil. For the same task of vegetation changes monitoring from MODIS images, the authors [Huang and Friedl, 2014] demonstrated accurate results comparable to maps based on Landsat satellite data on a regional scale. The

approach for rapid forest degradation assessment was proposed in [Morton et al., 2005]. Another commonly used data source for vegetation monitoring is ALOS PALSAR. In [Qin et al., 2015], it was proposed to use PALSAR radiometric data with a spatial resolution of 50 m in combination with MODIS multi-temporal data to get a forest cover map outside China.

Medium spatial resolution

Medium spatial resolution data are helpful for detailed forest mask segmentation. High revisiting time, public availability of data, and spatial resolution of up to 10 m per pixel make Sentinel-2 imagery a promising data source for many purposes such as forest mask estimation. In [Fernandez-Carrillo et al., 2020], Sentinel-2 imagery were used for assessing forest masks in Europe. Another source of multispectral data for forest plots is Landsat imagery. The effectiveness of Landsat and Sentinel imagery for forest degradation was demonstrated in [Mondal et al., 2020]. The use of Sentinel-2 and Landsat data combination was recommended for tropical forest disturbance estimation [Chen et al., 2021b]. In [Ganz et al., 2020], the authors created a forest cover map for the territory of Germany and assessed the developed approach by comparing the generated map with National forest inventory data.

It was shown that Sentinel-2 data provide an additional spectral information enriching aerial photography data for better predictions. Illegal logging drastically affects the state of the environment. Therefore, ERS is applied for operational monitoring with the aim at recognizing and preventing illegal logging. Medium spatial resolution satellite imagery is a suitable data source for logging detection because of the extensive coverage areas and rapid revisit time. Studies on illegal logging recognition using both multispectral and radar data were presented in [Pacheco-Pascagaza et al., 2022, Bullock et al., 2022]. In [Khovratovich et al., 2020], the authors proposed a method for forest logging detection in Russia based on Landsat imagery. Time series are also considered in forest monitoring tasks on the medium spatial resolution satellite data. In [Giannetti et al., 2021], time series based on Sentinel-2 imagery was used to assess the damage caused by a windstorms in Italy. Forest degradation was also considered in [Fernandez-Carrillo et al., 2020]. For precise

annual spatial distributions analysis, time series was implemented in [Zhang et al., 2022b], where a robust mapping approach based on Sentinel-2 data was provided.

One can use open access maps and tools for vegetation area estimation and supplementary materials extraction such as cloud masks. Sentinel-2 provides a pixel classification map based on Level-1C data that includes the following classes: cloud, cloud shadows, vegetation, soils/deserts, water, snow, etc. Spatial resolution of scene classification map is 20m [SentinelHub]. Pan-European High-Resolution Layers (HRL) is another useful tool for environmental studies, in particular, for forest cover estimation [Layers]. HRL is based on Sentinel-1 and Sentinel-2 satellite data. Tree cover density, dominant leaf type, and forest-type products are available for the reference year 2018 in 10m spatial resolution.

High spatial resolution

High spatial resolution data are helpful when a more detailed forest mask is required, including separation individual trees, small tree groups, and small plots in a forest with meadows and tracks. In low or medium spatial resolution images, it is impossible to recognize with a high accuracy such details as an individual tree: an individual pixel covers an area exceeding 100 sq.m. To address this problem, one can use satellite images of high spatial resolution: WorldView, Spot, RapidEye, Planet (see Table 2.1). Mapping eucalypts trees were performed using high-resolution satellite data in [Abutaleb et al., 2021], where WorldView-2 imagery usage provided a better accuracy than Spot-7 multispectral data. In [Wagner et al., 2018], a method based on satellite data of very high spatial resolution for allocating individual crowns was proposed. A map with individual trees is helpful for detailed forest cover analysis. To assess forest degradation and forest cover change, WorldView data were used in [Wagner et al., 2020]. In [Aquino et al., 2021], an effective methodology was proposed for detecting illegal logging on small plots for the forests of Peru and Gabon. In [Zhang et al., 2021b], an approach using high-resolution data from RapidEye to monitor land cover changes (and, in particular, forest areas) was put forward. The deforestation problem was also considered using RapidEye data in [KWON et al., 2021]. The high spatial and temporal resolution of Planet images were utilized with

LiDAR measurements to create the model for estimating the top-of-canopy height of tropical forests in Peru [Csillik et al., 2020]. Images obtained from PlanetScope nanosatellite constellation were used to create a high resolution (1 m) map representing tree cover in African drylands [Reiner et al., 2021], where a possibility to detect trees outside forests was shown.

Use of data from unmanned aerial vehicle

The very high spatial resolution provides a significantly better texture feature extraction than the medium spatial resolution. Therefore, unmanned aerial vehicle (UAV) data are often used for environmental remote sensing studies. Masking the forest with such data effectively used in assessing the state and environmental changes. To evaluate the effect of forest fires, UAV data has been successfully applied in [Yeom et al., 2019]. The approach was based on using only RGB channels and has been tested for forest ecosystems in the Republic of Korea. In [Ocer et al., 2020], a method for detecting and counting individual trees in images of different scales was proposed. To detect individual trees, UAV data have been successfully applied to mixed conifer forests [Mohan et al., 2017]. Combining data of various resolutions and spectral ranges, one can enrich a dataset with valuable features and achieve better prediction quality. In [Singh and Kushwaha, 2021], the authors proposed using UAV data and photogrammetry as part of the overall research methodology and Sentinel-1 data to assess forest degradation. However, a severe drawback of UAV data usage for large-scale studies is the time and cost-consuming of its collecting.

2.6.2 Computer vision algorithms for forest mask estimation. Specifics and limitations of the approach

Vegetation indices based on satellite spectral channels were suggested in many studies. For example, a methodology for assessing forest degradation based on the LAI index analysis using the MODIS data was given in [Richardson, 1981] almost 40 years ago. Since then, vegetation indexes have been used in various studies as a simplest computer vision approach. In [Othman et al., 2018], the NDVI index the

(normalized vegetation index) was shown as being applicable to the forest degradation assessing task for the tropical forests of Malaysia. A s ignificant drawback of such approach was the requirement of a threshold choice for various satellite data, environment conditions, and seasons. Therefore, its reliability cannot be sufficient for precise analysis.

Classical machine learning methods are aimed at automatization of the forest mapping process. It requires less labeled data and computing capacity to train a model. Classical machine learning algorithms were compared in [Vega Isuhuaylas et al., 2018] for land cover classes separation, including forest areas. The authors reported better results for RF and SVM than for kNN. However, RF and SVM showed close results, with AUC values 0.81 and 0.79. In [Xia et al., 2018b], the SVM method, in combination with a submerged mangrove recognition index, was proposed to map mangrove forests with an overall accuracy of 94%. In [Dabija et al., 2021], SVM with an RBF kernel function outperformed the RF algorithm in the CORINE land cover classification task. For forest area separation, an F1-score was found to be larger than 0.9.

To solve the problem of forest mapping, one of the most common approaches is based on deep learning methods, namely convolutional neural networks (CNNs). The forest mask segmentation involves identification of the pixels belonging to a forest class. It is an example of a binary semantic segmentation task. For forest species classification (Section 2.7), the main difference is that a CNN predicts one of several classes for each image pixel. The major advantage of using a CNN over classical machine learning methods is that it takes into account spatial characteristics. When assessing a pixel label, CNN uses spectral information from the local area of the processed image. This provides a more accurate estimation of forest masks due to the forest spatial structure that the CNN also learns. The principal limitation of deep learning methods is the need for a large amount of labeled data to train the model. In addition, training neural networks usually requires a lot of time and computing resources.

One of the widespread CNN architectures for forest mask segmentation is the U-Net architecture. U-Net was implemented for very high spatial resolution in [Ko-

rznikov et al., 2021]. For the medium spatial resolution of Sentinel-2 data (10 m per pixel), a modified U-Net with attention mechanism shows high performance [John and Zhang, 2022]. The authors declared the advantage of U-Net architecture versus ResNet and FCN for experiments with different locations using RGB bands and RGB plus NIR. An example of application and comparison of U-Net, DeepLabv3+, FPN, PSPNet, and LinkNet architectures in Brazils Eucalyptus Forest mapping task on medium spatial resolution data (Sentinel-2) was shown in [da Costa et al., 2021]. The best result with IoU of 76.57 using DeepLabv3+ with the Efficient-net-b7 backbone was achieved.

To deal with limited labeled data and adjust CNN model performance, one can apply transfer learning techniques. In transfer learning, the pre-trained model is adopted for new tasks and data specificity. In [Ahmed et al., 2021], transfer learning was used for forest mapping with subsequent fine-tuning of the model over the target forest domain. The proposed approach enables one to extract features from unlabeled data and using them for progressive unsupervised CNN training.

2.7 Forest-forming species classification on remote sensing data

After estimation the tree cover area as a forest mask, the next important step in forest taxation is determining tree species. This is especially relevant for large territories and locations, which are challenging to access [Schepaschenko et al., 2021]. In terms of CV, achieving the goal of tree type classification is also based on the solution of the image semantic segmentation task. Although the determination of tree species includes mostly more than two classes, the approach remains the same. Each image pixel needs to be labeled according to the class based on the test data for the algorithm training.

The most commonly used metric for estimation of the quality of tree species prediction from image data is the F1-score. Just as was described earlier in the case of forest mask estimation, the evaluation of the F1-score is carried out for each class individually. The closer the resulting value is to 1, the more similar are the

prediction and reference labels.

2.7.1 Use of data of different spatial resolution

The use of the data of different resolutions is determined by the problem that needs to be solved via the knowledge of the tree species composition across the area of interest. In general, at different spatial resolutions, estimates of tree species composition may be required for purposes of mapping large and difficult-to-access areas for biomass and carbon estimation [Grabska et al., 2020], to access succession trends on disturbed regions for a better understanding of carbon accumulation patterns [Reyes-Palomeque et al., 2021], to link climate effects with forest management activities [Majasalmi et al., 2018], for tree mortality monitoring and capturing of its patterns at different scales [Rogers et al., 2018, Koontz et al., 2021], and natural and urban ecosystem assessment [Wang et al., 2018].

Low spatial resolution

Similarly to forest mask determination from low spatial resolution data, Terra MODIS satellite imagery is a common choice of satellite data for forest species classification. An approach to determining the dominant species using the MODIS sensor data with calculation of vegetation indices from multi-temporal images was proposed in [Waring et al., 2006]. Also, the use of low spatial resolution data for solving a similar problem was proposed in [Buermann et al., 2008], and [Fu et al., 2010]. However, it was shown that coarse-scale satellite data might not capture many of the target processes, e.g., degradation development [Mondal et al., 2020], so recent low-resolution data are often used for obtaining more general, aggregating characteristics. Such information can be considered as a distribution of a set of unique surface characteristics reflecting environmental conditions similarly, and mostly represented by land cover type classification [Sulla-Menashe et al., 2019], temporal dynamics of the distribution of derivatives such as vegetation indices [Cano et al., 2017], or plant functional types [Srinet et al., 2020]. Combination of low spatial resolution data with more detailed imagery, e.g., MODIS data together with Landsat satellite data, as was shown in [Zhang et al., 2017], is a current trend.

Medium spatial resolution

A more detailed forest species determination can be achieved using the data of medium spatial resolution, for example, obtained from Landsat and Sentinel missions.

The authors in [Immitzer et al., 2016] suggested using Sentinel-2 data to identify tree species in central Europe. In [Wessel et al., 2018], an approach based on a combination of ML algorithms was also presented for classification of tree species in German forests. Another approach based on an application of linear discriminant analysis to medium spatial resolution images was proposed in [Mngadi et al., 2021]. Sentinel-2 data was also proposed for solving the problem of identification of forest species based on a series of images for different dates [Immitzer et al., 2019]. In [Immitzer et al., 2019], the authors succeeded to increase overall accuracy from 72.9% to 85.7% by using of the multi-temporal analysis. Radar data can adjust multispectral-based predictions, it was shown for Sentinel-1 and 2 data in the task of forest and plantation mapping and stand ages prediction [Spracklen and Spracklen, 2021]. For better understanding of forest properties and patterns, one can use hyperspectral RS data. As an example, Hyperion instrument on board the Earth Observing-1 (EO-1) spacecraft with 30 m spatial resolution provides 220 spectral bands for diverse environmental studies. Forest properties can be also effectively discriminated using the new hyperspectral Precursore IperSpettrale della Missione Applicativa (PRISMA) sensor, launched in 2019 and providing spatial resolution of 30 m [Agency]. These hyperspectral data are also accompanied by 5 m panchromatic band. PRISMA data usage showed high results compared to Sentinel-2 for forest categories classification [Vangi et al., 2021]. Although it is a promising RS data source, there are at present a few studies considering its usage for vegetation analysis compared to more conventional data such as Sentinel-1 and -2, Landsat-7 etc [Shaik et al., 2021].

High spatial resolution

High spatial resolution data allow one to operate not only with the spectral description of the object under study but also with its textural and spatial charac-

teristics. For example, the WorldView-2 panchromatic channel with a resolution of about 0.5 m (depending on the geographic latitude of the survey) can be effectively used to determine the shape of a tree crown. It in turn increases the likelihood of correct classification of tree species. In addition to a more detailed information from satellites providing high spatial resolution images, the possibility of using time series, as in the case of data of lower spatial resolution, is also an advantage of the approach. For example, the authors in [He et al., 2019b] successfully implemented multi-temporal WorldView images for forest hardwoods classification. In [Ferreira et al., 2019], tree species were classified for tropical forests based on 16 high-resolution WorldView-3 bands. One of the advantages of the WorldView-3 mission is the new SWIR sensing capabilities. Mangrove species classification study was conducted for WorldView-2 data in [Jiang et al., 2021], where the overall accuracy 95.89% was achieved. Although WorldView-2 and UAV data provided high results individually, their combination allows one to extract the most relevant features for classification.

Use of data from unmanned aerial vehicle

Hyperspectral and multispectral airborne images are known as a significant source of data for determining forest inventory characteristics [Shinzato et al., 2016, Sothe et al., 2019, Cao and Zhang, 2020]. The use of data-rich in both spectral and spatial features can handle the recognition of multiple tree species even in the case of complex terrain. Such approach provides an efficient classification on a small data set in the presence of many classes. Also, it is often proposed to use a combination of these data with LiDAR measurements [Zhang et al., 2020]. An example of multicopter UAVs with spatial resolution less than 2 cm was described in [Schiefer et al., 2020]. The study area covered 51 ha in Germany, which was sufficient for the representative analysis.

2.7.2 Computer vision algorithms for classifying forest-forming species types. Specifics and limitations of the approach

RF algorithm has demonstrated the ability to classify forest species, for example, when obtaining a forest map in Wuhan, China [Liu et al., 2018]. The RF algorithm can be used as part of a hierarchical tree type classification methodology. In the first stage, it is possible to provide classification according to vegetative indices such as NDVI and RBI (Ratio Blue Index), then classify forest areas and tree types using RF. SVM is broadly used for forest type classification in [Cao et al., 2018, Sothe et al., 2019]. A combination of LiDAR and hyperspectral data was used in [Yang et al., 2019], where SVM outperformed other classical machine learning methods with respect to the OA metric for species classification. However, in [Jiang et al., 2021], better results were achieved for RF than for SVM algorithm, with the best OA of 95.89% for species classification.

Both classical machine learning and deep learning algorithms can handle a single image and a sequence of images covering the same region. For instance, in [Persson et al., 2018a], all available images were combined to train an RF model and to predict four forest species with OA of 88.2%. One of the limitations of using a series of multispectral satellite images is the occurrence of cloud-contaminated images, which can corrupt predictions.

Similarly to solving forest area segmentation tasks, deep learning methods (CNNs, specifically) can be used for determination of forest species types. The main difference from the forest segmentation task is that several classes of pixels are predicted, corresponding to classes of tree species. We provide more details about the adjustable parameters of neural networks and their features in Section 2.4. An essential step in classification of tree species is determining the crown shape, for which high spatial resolution images are needed. Thus, when working with high spatial resolution images, a CNN makes it possible to create an optimal feature space that characterizes various forms of crowns, leading to a more accurate classification. For instance, in [Onishi and Ise, 2021], the developed CNN was shown as being capable

of classifying tree species based on biological structures such as foliage shapes and branching patterns. It is essential when just RGB bands are used, and different forest species may have the same colors. Three of seven considered classes were classified with OA over 90%. Due to very high spatial resolution of UAV, approach was capable of individual tree mapping. In [Cao and Zhang, 2020], U-Net architecture was modified for forest species classification ___combined U-Net with the feature ___ extraction network ResNet. The OA was equal to 87%, which is higher than the initial model results. Another architecture improvement is described in [Qi et al., 2022], where a class imbalance problem was addressed. The approach involves jigsaw resampling strategy to create a balanced training dataset. New training samples with the size of 128*128 pixels are combined from smaller patches with the of 32*32 pixels where each small patch cover a single tree species. Proposed approach improved the baseline from 66% to 80% (quality is measured as the proportion of correctly classified pixels to total pixels). The high-resolution data provides significant features for a CNN model and facilitates its accurate predictions when a sufficient amount of data (over 51 ha with spatial resolution less than 2 cm) is available [Schiefer et al., 2020]. Different tile size and spatial resolution were examined. It was shown, that large tile size is preferable in case of sufficient amount of training data. The best model with optimal tile size and spatial resolution achieved OA of 89% and mean F1-score of 73%. It is also possible to use approaches that combine data from several sources to provide better accuracy. RS images can be supplemented with phenological parameters and forest stand structure data. Although such features can be extracted from forest inventory data, another approach is to train a model to predict it.

2.8 Forest resources estimation on remote sensing data

In this section, we discuss the following forest variables: aboveground biomass, standing volume, growing stock. The definition of aboveground biomass (AGB) is the aboveground standing dry mass of live or dead matter from tree or shrub (woody)

life forms [Wilkes et al., 2018]. We refer to growing stock as "volume of living and standing stems over a specified land area that includes the stem volumes from stump height to the stem top and the bark but excludes the branches" [Gschwantner et al., 2022]. The standing volume is defined as "the volume of standing trees, living or dead, above stump measured over bark to the top. It includes all trees regardless of diameter, tops of stems, large branches and dead trees lying on the ground which can still be used for fibre or fuel" [NATIONS, 1992]. These variables have a strong relationship and considered as quantity measurements of forest and its derivatives. An assessment of forest resources helps to effectively determine the forest carbon stock. Therefore, such forest attributes estimation using RS data is an important area of machine learning methods application.

The problem of aboveground biomass, timber volume, and growing stock estimation is often solved as a regression problem in the following way. The regression task for RS data is a machine learning task, where the model is trained to assign some real value to each pixel of the resulting digital map of the target territory. A machine learning model uses a training set to determine the relationship between the feature description of objects and the target value. Thus, just as in the semantic segmentation problem, the ground truth image with the reference markup is used. During the training procedure, a model reduces the difference between the prediction and the reference values according to the chosen quality metric.

2.8.1 Use of data of different spatial resolution

Low spatial resolution

To obtain timber volume estimation on a large scale, it is often proposed to use MODIS sensor data. Approaches for determination of forest biomass are presented in [Fu et al., 2019, Zhang et al., 2019c, Gao et al., 2020a]. The data effectiveness was verified for regional changes monitoring and supplemented forest inventory data for ecological assessment. Despite the possibility of a large spatial coverage supported by this approach, for some practical problems, more detailed maps are required. Therefore, one can consider higher spatial resolution data.

Medium spatial resolution

When it is necessary to estimate timber volume over a large area with greater details, a common choice of RS data is medium spatial resolution. For example, this type of data can be received from Sentinel and Landsat satellites. The potential of using Sentinel-2 data to determine growing stock volume for the territory of Italy was demonstrated in [Mura et al., 2018] where the prediction quality based on Sentinel-2 data was shown to be better than that for Landsat images in 37.5% of cases and for RapidEye images in 62.5% of cases, even though the resolution of the RapidEye satellite is significantly higher than that of Sentinel-2. In [Rees et al., 2021], Sentinel-2 data was shown as being capable of determining growing stock volume in Russia. Also, Sentinel images were used in [Nink et al., 2015] to map the timber volume in the coniferous forests of Norway. The relevance of using these data was also confirmed in other works on determining the biomass and stock volume in various territories [Malhi et al., 2022] and [Hu et al., 2020].

Another useful instrument for environmental analysis that deserves additional consideration is the Global Ecosystem Dynamics Investigation (GEDI). It is the first spaceborne lidar with a footprint resolution of 25 m. One of its goal is to provide a better understanding of the aboveground carbon balance of the tropical and temperate forests [Dubayah et al., 2020]. It can accompany other RS data for enhanced biomass mapping and help to estimate aboveground carbon change.

High spatial resolution

Commonly, high spatial resolution data is used when it is required to estimate timber volume down to a single tree. In the actual studies on this topic, it is recommended to use WorldView satellite images with a resolution of about 2 m for a spectral range of channels from 396 nm to 1043 nm and sub-meter resolution for the panchromatic channel. An example of using WorldView-2 stereo images was demonstrated in [Straub et al., 2013], where high-resolution data and LiDAR measurements were compared in the problem of assessing the timber stock for the forest area in Germany. In [Vastaranta et al., 2018], panchromatic WorldView-2 stereo-imagery is considered together with a digital elevation model derived from

airborne laser scanning. Using WorldView imagery for different geographic regions was also confirmed by a study of Turkish forests in [Günlü et al., 2021]. Forest standing biomass was estimated and used to assess forest productivity in [Dube et al., 2018] based on WorldView-2 data. The authors evaluated the importance of different bands and vegetation indices and highlighted the Red-edge band significance. Spot-5 is another source of high-resolution data for aboveground biomass estimation [Muhd-Ekhzarizal et al., 2018].

Use of data from unmanned aerial vehicle

UAV data are selected for land cover surveys in cases where very detailed timber volume estimation is required. The use of UAVs makes it possible to analyze the characteristics of an individual tree by constructing a more informative feature description of the vegetation cover with a resolution of up to several centimeters per pixel. The approach to determining the timber volume based on UAV images and photogrammetry was tested with success in [Gülci et al., 2021]. One well-established approach to forest growing stock volume estimation is based on using satellite imagery in combination with UAV data [Puliti et al., 2018]. This approach's advantage is combining the spectral features obtained from the satellite with highly detailed textural features. In [Puliti et al., 2020b], an approach to replace ground-based measurements for growing stock volume estimation with UAV data was used with good results. At the same time, ground-based measurement data were used in this research only to assess the quality of algorithm predictions. In [Tuominen et al., 2017], data with spatial resolution of less than 10 cm per pixel were used to determine the stand volume. In [Hernando et al., 2019], images with the same spatial resolution were used to estimate forest biomass. It was presented the effective use of UAV data for tree stem assessment in [Hyyppä et al., 2020], [Iizuka et al., 2020], and [Yrttimaa et al., 2020]. Not only images can be used for forest analysis. Point cloud obtained from UAV can also be considered in voxel-based representation for further computer vision algorithms application, as shown in [Hyyppä et al., 2020]. In [Iizuka et al., 2020], dense points cloud derived from multicopter is used to extract significant characteristics for stem volume prediction using machine learning

algorithms. For instance, they estimated the height of the forested area by subtracting the digital terrain model (DTM) from the digital surface model (DSM). DTM was obtained from terrestrial laser scanning, while an unmanned aerial system was utilized to get DSM.

Although UAV provides very-high resolution data, one of the significant limitations of UAV-based approaches compared to satellite data is the relative laboriousness of obtaining such data on extensive areas.

2.8.2 Computer vision algorithms for the task of forest resources estimation. Specifics and limitations of the approach

In many studies, it was demonstrated the effectiveness of the linear regression algorithm in the problem of timber stock evaluation. The advantage of this approach is the ease of implementation and use. In addition, an important characteristic is the interpretability of the results. The work [Popescu et al., 2003] proposed to use linear regression to estimate the diameter of tree crowns from UAV data. An approach based on multiple linear regression was presented in [Hawryło and Wężyk, 2018]. The described method makes it possible to determine the stock of plantations on pine plots using Sentinel-2 images and aerial photography data. Different RS data sources and spatial resolution make it important to preserve the same data georeference. Ground Control Points (GCPs) were used to calculate UAV's camera orientation and set a correct georeferencing. Prediction of growing stock using a linear regression algorithm based on Landsat-7 images is demonstrated in [Mohammadi et al., 2011]. Both vegetation indices and linear regression were implemented in [Muhd-Ekhzarizal et al., 2018].

It is also proposed to use the Random forest regression (RFR) algorithm for timber stock estimation. The approach based on ultra-high spatial resolution data is described in [Iizuka et al., 2020], where various RS measurements were considered. The methodology includes a stratified random sampling of training examples and algorithm parameter optimization. The parameters used in the RFR algorithm are

described in more detail in Section 2.4. Besides the problem of determining the stock of wood, the problem of estimating the stock of carbon can be also directly solved using RS data and RFR algorithm. This approach was tested for mangrove forests in [Li et al., 2019b]. In this research, various forest cover characteristics were used to assess the stock: tree species, height, and textural features. Vegetation indices based on UAV spectral data were also used to form the feature space. The most significant features were selected based on the Boruta algorithm [Jayathunga et al., 2019]. It is also important for UAV-derived multispectral data to conduct a reflectance calibration of cameras to support accurate temporal analyses because digital numbers are affected by the atmospheric and illumination conditions and cannot be considered as quantitative values [Crusiol et al., 2020].

SVR is another relevant approach for stem volume estimation [Iizuka et al., 2020]. An approach for biomass estimation using the SVR algorithm with a radial basis function (RBF) kernel was proposed in [Navarro et al., 2019]. As it was shown earlier in Section 2.4, it is important to find the optimal parameters of the algorithm, which can have a significant effect on the final accuracy. In [Navarro et al., 2019], the kernel parameters were selected using the grid search method. The feature space was formed based on various RS data sources: Sentinel-1 radar data, Sentinel-2 multispectral images with 10 vegetation indices obtained on their basis, and UAV photogrammetry data. The use of the SVR algorithm for biomass estimation was also proposed and showed effective in other studies [Gleason and Im, 2012, Shao and Zhang, 2016].

Above-ground biomass estimation with the use of CNNs was examined in [Dong et al., 2020]. The prediction results, as measured by R2, were found to be equal to 0.943. The aboveground carbon density of forests can be estimated directly using RS data and a CNN model, as was demonstrated in [Zhang et al., 2022a], where a CNN model was shown to perform better than classical machine learning algorithms. In [Balazs et al., 2022], a CNN-based approach yields RMSE of 20.3% for the volume of growing stock estimation using airborne laser scanning. Although CNN is highly promising for such studies, no strong difference between the k-NN and CNN performance was observed. It was suggested that additional data should

be utilized to reveal the full potential of CNN models.

For more accurate growing stock volume estimation on the limited dataset size, a deep neural network with transfer learning was implemented in [Astola et al., 2021]; this approach allowed the authors to minimize the amount of ground-based measurements over different areas in Finland.

2.9 Discussion

Based on the current trends in development of satellite imagery and data processing algorithms, we expect the following trends in this domain. First of all, more availability of high quality data and build-in services for data processing that are provided by the space companies. This data will be easy to use even for inexperienced users. Satellite constellations will have better revisit time and coverage allowing near real time observation of ground cover. Also high resolution multispectral imagery will be wider applicable, giving important information about investigated objects including forests. Developments of special augmentation techniques and few short learning algorithms will allow us to detect and make quantitative assessment of rare ground objects and events. In this section, we provide more details about current limitations and future works.

2.9.1 Forest carbon disturbing events

Improved forest management in terms of carbon offsetting is based on carbon sequestration from the atmosphere. Precisely, it means the storage on a long-time basis of more carbon compared to the regional baseline in the ecosystem considering land-use practices, maintaining existing forests, and increasing total forest coverage, while decreasing mortality [vonHedemann et al., 2020, Kaarakka et al., 2021]. On both large scales and in the case of small forest landowners and land rent, this means enhancing carbon pools, thereby reducing emissions caused by different processes of GHG into the atmosphere. At the same time, the above-ground biomass of living trees is considered the most dynamic carbon pool affected by the plethora of factors of distinct nature [Fahey et al., 2010]. Such forest carbon disturbing factors include

the development of areas inundated with water and changes related to them and soil hydrologic cycle in general [Cooper et al., 2019], the occurrence of deadwood due to the influence of biotic and abiotic events [Seibold et al., 2021], wildfires and harvesting [Kirdyanov et al., 2020, Anderegg et al., 2020]. Detection, attribution, and monitoring of such occurrences can be covered using RS techniques. In this way, CV approaches should also be considered for fully and semi-automated solutions development, while a wide range of stakeholders can use such solutions to plan and implement climate change mitigation strategies based on nature preservation actions.

Studying flooded areas in terms of CV is accompanied by multi-challenging tasks. Among the majors ones we mention the following tasks: detection of flooded territories themselves and changes catching [Ballanti et al., 2017]; distinction between different types and classes of flooded lands [DeLancey et al., 2019]; estimation of biomass and potential to CO_2 sequestration [Dronova et al., 2021]; fusion of data of different domains to catch emission patterns and enhance accounting [Bansal et al., 2018, Gerlein-Safdi et al., 2021]. Such research is based on the solution of segmentation and classification tasks. Broad range of tools for these tasks involves conventional unsupervised and supervised ML algorithms such as RF, SVM, XGBoost, random walker segmentation, different types of neural networks (mostly deep CNNs) variations of edge detection, and others. In [Rezaee et al., 2018], the performance of CNN, AlexNet was compared with classic RF to distinct and map different wetland types, including bog, fen, marsh, swamp, and also shallow water, and deep water along with urban areas and upland. In this study, RapidEye multispectral imagery and a small number of input features were used. CNN was shown to overperform RF, catching both the dominant wetland classes and detailed spatial distribution of all studied land cover classes, with showed overall accuracy and Kappa coefficient of 94.82 % and 0.93, respectively. In [Mahdianpari et al., 2020], RF, as declared a computationally efficient and easily adjustable algorithm, was applied to multi-year summer composites of Sentinel-1 and Sentinel-2 images. Wetland spatial distribution was mapped, considering wetland classes across Canada, covering an area of approximately one billion hectares. The model accuracy varied from 74% to 84% in

different territories.

Similarly to wetland research, studying and monitoring wildfire events are comprehensive and consist of the following main aspects: early fire and smoke detection; estimation of fire severity and spread; fire behavior analysis and prediction; detection and estimation of post-fire territories. Forest fires are extremely hazardous to both natural ecosystems and humans, destroying habitat areas, negatively affecting agriculture, and accompanying significant emissions of retained carbon. Thus, related monitoring and detection technologies are rapidly developing, so, for instance, several satellites with low spatial resolution but short revisiting time already have fire detection sensors onboard [Jain et al., 2020, Bouguettaya et al., 2022]. Combination of UAV-based RS with CV techniques, based explicitly on CNN, including previously discussed architectures such as U-Net, DeepLab, and other deep learning architectures such as, e.g., GAN, LSTM, is an effective tool for wildfire monitoring. It is extremely useful for firefighting actions and capable of catching early fire in reduced time and more safely, comparing with ground inspections [Bouguettaya et al., 2022]. Such solutions can provide real-time monitoring but require powerful hardware. An original Burnt-Net inspired by U-Net architecture was used for the development of an end-to-end solution for post-fire tracking and management. It was utilized to map burned areas on Sentinel-2 images across different countries, including Cyprus, Turkey, Greece, France, Portugal, and Spain, showing high robustness and mean accuracy of more than 97% by overall accuracy [Seydi et al., 2022]. In [Brown et al., 2018], Maximum Likelihood, SVMs classifiers, and two multi-index methods were compared for mapping burnt area. Burn severity was also assessed using SVMs and one hidden-layer NN on Sentinel 1,2 images on the study location in Portugal. According to the results obtained, SVMs showed the highest accuracy for both burnt area mapping and burn severity levels estimation, with achieved an overall accuracy of 94.8% and 77.9%, respectively.

Deadwood represents essential carbon stock while simultaneously a significant contributor to carbon dioxide emission and one of the major forest biodiversity loci [Bujoczek et al., 2021]. The development of deadwood can be a consequence of the natural course of things or triggered by biotic and abiotic factors such as pest or

pathogen outbreaks, changes in hydrologic regime due to climatic shifts, and windstorms [Karelin et al., 2017, Cours et al., 2021]. Numerous studies are dedicated to find a difference between target object (deadwood occurred due to a specific reason) and other nontarget objects, or, for example, between damaged trees at different stages of factor influence, existing together and displaying similar spectral signatures [Safonova et al., 2019, Zielewska-Büttner et al., 2020]. For instance, in [Esse et al., 2022], it is recommended to apply Neural Net with standard backpropagation and SVM among other supervised approaches for the deadwood detection in the case of Chilean Central-Patagonian Forests using high-resolution multi-spectral data (RGB+NIR) with best algorithm performance of 98%. In [Briechle et al., 2021], an approach based on CNNs fusion of Lidar and multispectral data was applied for 3-D tree type classification along with dead trees, showing overall accuracy of more than 90% for all classes. At the same time, it was noted that the use of lidar-based data slightly increased the overall accuracy. The proposed comprehensive solution facilitates fast model convergence, as was pointed out even for datasets with a limited number of samples due to the applied transfer learning technique.

2.9.2 Data and labeling limitations

Training an accurate and robust computer vision model requires representative data that cover many possible scenes and are obtained under different illumination conditions. Training models with many parameters on non-representative dataset with low number of samples could lead to model overfitting. The use of models with small parameters that could be trained on a small member of parameters does not allow one to obtain acceptable accuracy and generalization. For a recent comprehensive analysis of overfitting and underfitting reasons in machine learning applications for different domains, see [Roelofs et al., 2019].

Computer vision models for processing RS data are not an exception. A large amount of well-annotated spatial data is required to train algorithms [Pasquarella et al., 2018]. Moreover, there are many additional issues that appear due to the complexity of the data collection procedure. It is difficult and time-consuming to collect and directly label the amount of representative RS data. The principal im-

pediments are weather conditions (clouds) and satellite (sensor) revisit time [Notti et al., 2018], [Misra et al., 2020]. Thus, expanding the dataset with additional useful and reasonable data is vital. One way to solve this problem is to generate image samples from the obtained data. The most common approach for generating new image samples is augmentation. Several typical augmentation approaches are widely used in different domains, starting from classical augmentations, which include geometrical, and color augmentations, and finishing with application of ML techniques for augmentation [Buslaev et al., 2020a]. Nevertheless, new approaches are in high demand and still appearing [Khalifa et al., 2021]. However, there are many restrictions in applying augmentation techniques for RS data because images may have a complex structure [Sun et al., 2021]. For example, the relative locations of objects should be meaningful after the creation of the new image sample. That is why it is important to carefully tune the parameters of augmentation when applying even standard augmentations carefully. However, there are some new advanced augmentation approaches that take into account the specifics of RS domain.

The other limitation in the use of RS data in computer vision models is the involved labeling procedure. Only an expert can create a precise manual markup with vegetation characteristics based on these images (distinguish forest species, age, etc.). Ground-based measurements also have particular limitations. For instance, forest inventory data can be out of date. It also has some specificity in its organization. Information is often available for individual stands that are not necessarily homogeneous. Therefore, the dominant forest species (and other characteristics) are estimated in various tasks. It leads to some mismatches in training data. As a result, CV methods in environmental studies aim to work with invalid markup in particular cases. It is essential to develop a methodology for automatic improvement of RS data labeling. One popular approach is the weakly supervised learning, which is considered a fundamental problem in machine learning [Ahn et al., 2019b]. For land cover mapping and, in particular, forest areas, weakly supervised segmentation was suggested in [Schmitt et al., 2020]. In [Tang et al., 2021], the problem of weakly supervised pixel-level mapping to predict tree species was addressed.

To address spatial and temporal limitations in concrete environmental and forestry

tasks, a combination of Sentinel-2 and Sentinel-3 data can be used. In [Guzinski et al., 2020], it was applied for evapotranspiration estimation. Although, there is a high importance of thermal features obtained from Sentinel-3, their spatial resolution requires adjustment. Sharpened high-resolution thermal data usage was suggested as a promising approach for environmental studies.

Particular uncertainty in data for forestry tasks is connected with markup acquisition. This process requires field measurements that are conducted according to special regulations. The way how data are obtained affects the model performance and the expected range of errors. Therefore, it is crucial to understand the origin of the used data for forestry tasks. In the Russian forestry regulation, three main forest taxation categories are considered. The first category contains forest stands with a total area approximately from 3 to 15 ha. The second and the third forest inventory categories consider stands with areas from 16 to 35 ha and from 100 to 150 ha, respectively. There are different approaches for inventory data retrieval such as eye-measuring, eye-measuring and enumerating, aerial image interpretation. The more accurate taxation approaches are applied for territories with rapid forest exploitation. Each forest inventory category has an acceptable measurement error for the further data usage in forestry management events.

2.9.3 Visual transformers as state-of-the-art CV algorithms relevant for forest taxation problem

Visual transformer-based approaches, which have appeared relatively recently, have also been used for dealing with problems of classification on environmental RS data [Bazi et al., 2021, Zhang et al., 2021a] and change detection [Chen et al., 2021a]. These approaches can also be applied to forest characteristics assessment. Transformer approaches are currently the most advanced models. These approaches use multi-purpose attention mechanisms as the main building block for obtaining long-term contextual information and links between pixels in images rather than standard layers. In the first step, the analyzed images are divided into groups and then transformed into a sequence by constructing a new feature space. The

resulting sequence is then fed to several attention layers to form the final new presentation. The first sequence of tokens is used in the classification layer at the classification stage. The detailed description of self-attention mechanisms and pretraining procedures in visual transformers are described in [Khan et al., 2022]. One of the essential advantages of transformers is the possibility of compressing the network and removing half of the layers while remaining sufficiently accurate classification [Bazi et al., 2021]. Experimental results from various environmental RS data image datasets [Bazi et al., 2021, Zhang et al., 2021a, Chen et al., 2021a] demonstrate the potency of transformers compared to other methods.

2.10 Conclusion

The present survey discusses the key aspects of forestry analysis based on RS data and computer vision techniques. The study was focused on the particular forestry problems such as estimation of forested areas, tree species classification, and forest resources evaluation. These tasks are highly valuable for meaningful environmental analysis involving carbon stock monitoring and global climate changes. In these tasks, we aimed to emphasize both algorithms and data importance. Although various satellite missions and UAV-based approaches support effective solutions, the main current limitation is a lack of high-quality reference data for artificial intelligence algorithms. Also, it has been shown that data source and algorithm choice strongly depend on the objective of the study, as temporal/spatial resolution and cost may vary drastically. For large-scale analysis, satellite-based approaches are more preferable because of broader coverage, while for more detailed measurements, UAV-based approaches allow one to achieve the required results. Various RS data combination and advanced computer vision techniques such as few-shot learning, transfer learning, weakly supervised learning, visual transformers, augmentations techniques show promising perspectives for further environmental studies. At the same time, physical nature of the observed environmental objects should be taken into account both during the data acquisition, processing for computer vision algorithms, or vegetation indices implementation.

Chapter 3

Augmentation-based Methodology for Enhancement of Trees Map Detalization on a Large Scale

3.1 Introduction

Artificial intelligence has already been successfully applied to solve various practical problems, in particular tasks related to the automatization of sensing processes and increasing their precision [Cheng and Yu, 2020, Shan et al., 2021]. With the appearance of new technologies that allow high-quality imaging data to be obtained, the amount of collected imaging data has increased; this leads to demands for the development of effective tools for image data processing. One of the industrial and scientific domains that requires such tools is remote sensing [Yu et al., 2021, Angelini et al., 2017]. Remote sensing data is widely used in various environmental studies that include measuring of the carbon footprint, for which it is crucial to obtain precise forest masks, boundaries of agriculture fields, type of crops, etc. Computer vision algorithms, in particular convolutional neural networks (CNN), can automatically process this data. The vital information such as environmental and vegetation state [Kattenborn et al., 2021b], forest inventory characteristics [Illarionova et al., 2020], and agriculture crop yield [Nevavuori et al., 2019] can be effectively extracted by CNNs. Commonly, the first step in environmental studies is obtaining forest

masks [Hirschmugl et al., 2020, Li et al., 2020a]. The existing satellite-based approaches for obtaining forest masks work well for vast territories where it is not important to detect and quantify small details. The usual spatial resolution for open-access landcover maps is more than 10 m [Malinowski et al., 2020]. Using such datasets, it is possible to create forest masks with sufficient accuracy on a large scale and make an adequate assessments of forest reserves. However, current approaches usually are not intended to detect small details such as individual trees, groups of trees, or meadows. Moreover, commonly used metrics for accuracy assessment of automatically generated forest masks do not take into account these small details in an adequate manner, for the following reason. Separate trees or groups of trees represent a tiny proportion of the target forest class, which is why the impact of detection accuracy for small objects on the overall metrics is low. Thus, a high prediction score for the entire territory does not necessarily mean high performance on small details.

For particular tasks, it is essential to obtain a detailed forest mask that approximates areas of forest. One such task is the monitoring of protected zones or natural reserves, where the territory of interest is too narrow and each small group of trees has to be taken into account [Flores-Martínez et al., 2019, Thomas et al., 2021]. In [Malkoç et al., 2021], the authors showed the importance of trees outside forests for ecosystem functions and ways to improve assessments based on aerial stereo images. In such cases, unmanned aerial vehicles (UAV) or aerial photography are usually used to obtain higher detail [Qiu et al., 2018]. Obtaining the detailed forest mask on large scales is quite a challenging task, and it is common to merge data with different resolutions and from different sources [DAmico et al., 2021]. The main limitation of the UAV-based approach is its cost and the difficulty of its implementation for vast territories on a country-wide scale [Otero et al., 2018]. Another datasource is satellite imagery with high spatial resolution, such as WorldView, Spot, RapidEye, and Planet. These data sources are often used to detect the crown of individual trees, which in turn can be considered in a detailed forest mask. WorldView images were used for forest cover estimation in [Karlson et al., 2014, Wagner et al., 2020], while RapidEye data were considered in [Marx and Tetteh,

2017, Miettinen et al., 2014]. However, these data are more expensive than low or medium spatial resolution satellite images.

The example of low resolution data for making large scale estimations of forest masks is described in [Hansen et al., 2008], where the authors used image data collected by the MODIS mission, which has a resolution of 250 m [mod, Accessed: 20 November 2021] (Moderate Resolution Imaging Spectroradiometer). Using a medium resolution (10–30 m) is most frequent because of the availability of open-access data and comprehensive frameworks for data processing. For example, in [Fernandez-Carrillo et al., 2019] the authors show an approach for forest mask creation over European forests using optical Sentinel-2 data. In [Mondal et al., 2020], the authors monitor forest degradation in South Asian forest ecosystems by implementing Sentinel-2 and Landsat imagery. Deforestation monitoring tasks using data from Sentinel-1, PALSAR-2 and Landsat data are discussed in [Reiche et al., 2018]. The data fusion and preprocessing techniques for aerial and Sentinel-2 data are shown in [Ganz et al., 2020], where the authors calculated the forest cover map for German territory and showed the accuracy of their proposed method by comparison with National forest inventory data. One major problem is deforestation connected to illegal logging. In [Pałaś and Zawadzki, 2020, Bragagnolo et al., 2021], the authors propose and validate approaches for deforestation monitoring using Sentinel-2 data. A time series of images can be used for environmental monitoring and planning of sustainable management. In [Chen et al., 2017a], the authors showed the potential of using, time series of Landsat and Sentinel-1A SAR images to identify and map mangrove forests. Time series of images can be used to detect forest degradation caused by natural reasons, anthropogenically-influenced climate change, damage by insects, etc. [Fernandez-Carrillo et al., 2020].

Machine learning models depend drastically on the data quality and its amount. In many cases, using more data allows the model to reveal hidden patterns deeper and achieve better prediction accuracy [Sun et al., 2017]. However, gathering of a high-quality labeled dataset is a time-consuming and expensive process [Paton, 2019]. Moreover, it is not always possible to obtain additional data: in many tasks, unique or rare objects are considered [Nesteruk et al., 2021] or access to the ob-

jects is restricted [Huang et al., 2019b]. In other tasks, we should gather data rapidly [Shadrin et al., 2020]. The following tasks are among such challenges: operational damage assessment in emergency situations [Novikov et al., 2018], medical image classification [Masquelin et al., 2021]. There are different approaches to address dataset limitations: pseudo labeling, special architectures development, transfer learning [Barz and Denzler, 2020, Bullock et al., 2019, Ng et al., 2015, Zhang et al., 2015]. Another standard method to address this issue is image augmentation. Augmentation means applying transformations (such as flip, rotate, scale, change brightness and contrast) to the original images to increase useful samples that allow training more robust algorithms [Buslaev et al., 2020b].

The lack of labeled data for particular remote sensing tasks makes it crucial to generate more training samples artificially and prevent overfitting [Zhu et al., 2017b]. Data augmentation is especially important to enhance the efficiency of deep learning applications in remote sensing [Ma et al., 2019]. This work aims to propose an object-based augmentation (OBA) pipeline for the semantic segmentation task that works with high-resolution georeferenced satellite images. Naming our augmentation methodology object-based augmentation (OBA), we imply that this technique targets separate objects instead of whole images. The idea behind the approach is to crop objects from original images using their masks and pasting them to a new background. This method is studied in the general domain [Ghiasi et al., 2021, Zhou, 2019, Zoph et al., 2020], but we are the first to study its effectiveness in remote sensing applications. For this purpose, we adopt the method to work with geospatial data formats and experiment with case-specific features (such as objects' shadows and large study area size). In our approach, every object and background can be augmented independently to increase the variability of training images; shadows for pasted objects also can be added artificially. We show that our approach is superior to the classic image-based methods in the remote sensing domain despite its simplicity. In [Illarionova et al., 2021b], we have previously shown this approach application for the building segmentation task, while in present Chapter, we focus on the OBA technique application for the forest mask estimation.

In this study, we propose a neural network-based approach for predicting the

detailed forest mask using Basemap RGB images. We use a small dataset with detailed labelling of individual trees to fine-tune a CNN model that was initially trained on a large dataset with less accurate labels (masks) for individual trees or groups of trees. The novelty of our study includes the implementation of the OBA technique for new training sample generation. This approach increases the amount of training data significantly and allows for the creation of physically meaningful data samples, which is important in remote sensing data analysis. The main contributions of this study are:

- We propose a novel for remote sensing domain simple and efficient augmentation scheme called OBA that improves CNN model generalization for satellite images;
- We propose and validate a pipeline for detailed forest mask segmentation using CNN;
- We provide an open-access tool for detailed forest mask segmentation that can be used for environmental studies, which is available in an SAAS platform through the link provided [Mapflow.ai, Accessed: 10 Febuary 2022].

The Chapter is organized as follows: Section 3.2 describes the characteristics of the datasets used in the present study and the methodology of the proposed solution and validation approach; Section 3.3 shows the obtained detailed tree maps and compares them with the baseline maps; Section 3.4 presents concluding remarks and plans for the possible future development of the developed methodology.

This research of forest mask begins the Thesis as a preliminary study for forest characteristics extraction. To define more specific forest properties, an exact forest area is required. This study can be considered independently for various environmental tasks or can accompany the methods described in the next Chapters.

3.2 Materials and Methods

In this study, we considered two datasets. The first was large and lacking in precision markup for small details, while the second represented a smaller area with each

individual tree presented in markup.

3.2.1 Large Dataset

For the large dataset, we collected data covering more than 500.000 hectares. The study area was located in the Republic of Tatarstan, Russia. There were about 45% hectares of forest and 65% hectares covered by other landcover types (lawns, fields, etc.) and manmade objects (roads and buildings). We used a cloud-free composite orthophotomap provided by mapbox [Mapbox, Accessed: 2020-06-17] via tile-based map service. The imagery was derived from different satellite images obtained by the WorldView satellites series, consisting of three pansharpened spectral channels (RGB). The spatial resolution was about 0.5 m per pixel, depending on the observation latitude. All images were taken during the summer period in 2018. The manual markup for this region was produced based on the aforementioned images. It was first presented in a vector GEOJSON format (as polygons coordinates), then converted into georeferenced rasters (binary masks) with spatial resolutions equal to the satellite data resolution. The study area was split into training, validation, and testing regions at a respective proportion of 70%, 15%, 15%.

3.2.2 Detailed Small Dataset

We used a high-quality small dataset with precision individual tree masks for an area in Dagestan, Russia. The environmental conditions differ from the large dataset territory in that sandy surfaces partially cover the area. The manual markup was performed for satellite images from the mapbox basemap service, with an acquisition date in the summer period of the 2020. The spatial resolution properties of satellite images were the same as for the large dataset. The entire area was 4.000 hectares, of which approximately 40% was forest cover. The final forest mask was presented in both raster (binary mask) and vector (polygon coordinates) format. The number of individual trees in the training dataset with an area smaller than 300 pixels was 6387. The test area included more than 2000 individual trees. Each subset was represented by the individual image and area.

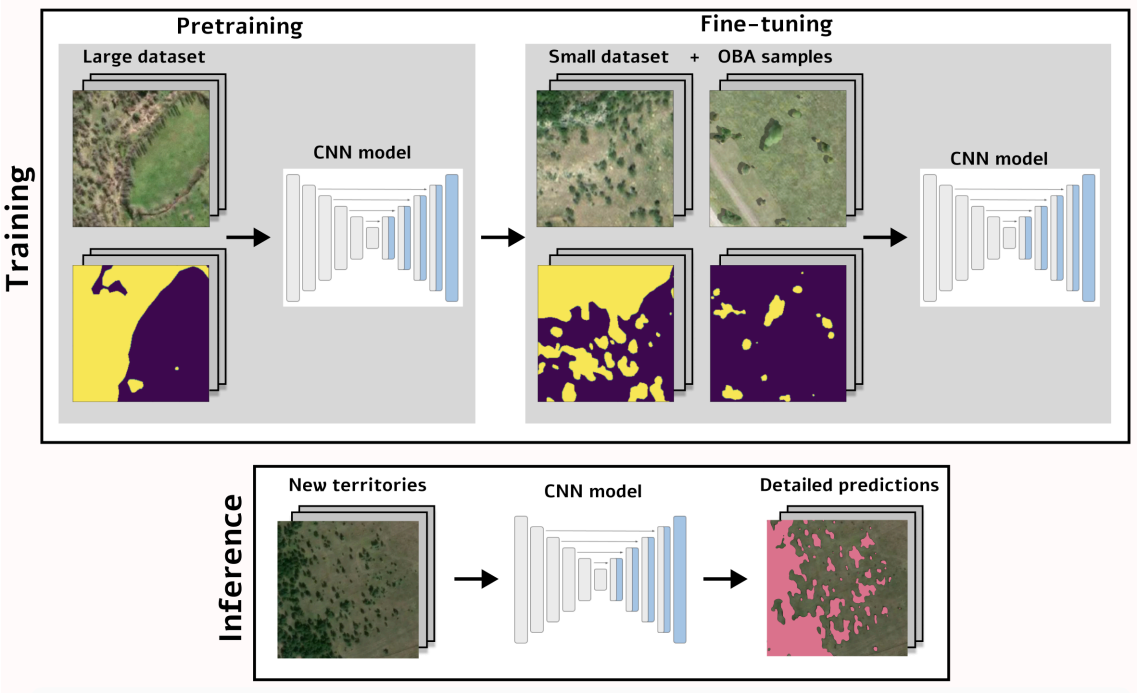


Figure 3-1: Proposed pipeline for CNN model training.

Our solution for forest segmentation included two consecutive steps, which are shown in Figure 3-1. The first was model training on the large dataset in order to learn important feature representation. Then, the model was fine-tuned on the smaller and more detailed dataset that was preprocessed with an object-based augmentation technique.

3.2.3 Baseline Forest Segmentation

For the baseline forest segmentation, we used a large dataset. Training samples were cropped randomly from the entire study territory. Standard color and geometrical transformations (random rotation, brightness, contrast, saturation adjustment, etc.) were implemented for each sample. A neural network was trained to identify pixels belonging to the class "forest" by minimizing the binary cross-entropy loss function

$$L(y, \hat{y}) = -\frac{1}{N}\sum_{i=1}^{N} y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i), \tag{3.1}$$

where N is the number of target mask pixels, y is the target mask, and $\hat{y}$ is the model prediction. For the baseline forest segmentation we used the following imple-

mentations of the CNNs: UNet [Ronneberger et al., 2015], FPN [Lin et al., 2017], and DeepLab [Chen et al., 2017b]. The details of CNN training are discussed in Section 3.2.6. Model implementation was based on the repository in [Yakubovskiy, 2022].

3.2.4 Object-Based Augmentation

After baseline training, we fine-tuned the model using a small dataset and the following augmentation approach.

The object-based augmentation approach has been previously proposed for the remote sensing domain in [Illarionova et al., 2021b] for solving segmentation tasks. For the forest segmentation problem, we provided the following augmentation scheme, the algoritmic implementation of which can be found at the following link [Illarionova, 2021]. The initial detailed markup included both large areas and individual tree masks. In the first stage, we created a list of individual trees selected by the area according to the threshold. The threshold was established empirically and was equal to 300 pixels. Selected individual trees were ascribed IDs associated with coordinates and instance masks. During the augmentation step, the object's ID was selected. The object (individual tree) was cropped according to its boundary. Then, shadows were added to make the generated sample more realistic. The footprint of an object was used to add a shadow. The contrast and saturation of shadows were varied in order to extend the variability of the training instances. Moreover, each individual tree could be augmented using classical color and geometrical transformations. For this task, the Albumentations package [Buslaev et al., 2020b] was leveraged. The cropped and transformed individual trees were then merged with a new background. The background was randomly selected from the initial satellite image or from new images from another geographical location. The main requirement for the background crop was the absence of the target objects. The selected background patch was augmented using geometrical and color transformations. The final step of new training sample generation was background and target object merging. A number of objects was selected randomly for each patch from a predefined range. The maximum number was defined empirically and set to 30 according to the patch

size and target object size. Intersection between the objects was restricted. It is possible that a neural network can fit exactly against generated data and lose essential properties of the original images. To avoid this, we used generated samples with a probability of 0.4 and original samples with a probability of 0.6. Both the original and generated samples were prepared during the training time and did not require extra memory to store patches. Examples of the generated and original samples are presented in Figure 3-2.

The OBA approach was compared with two alternative approaches, namely, classical augmentation (random rotation, brightness, contrast, saturation adjustment) as described in [Illarionova et al., 2021b] (Simple_augm) and training without any image transformations (Baseline_no_augm).

The difference between the general and remote sensing domains often relates to image size in a dataset. The average image resolution in ImageNet dataset is $469 * 387$ pixels, while in many remote sensing datasets image is significantly larger. Images in DOTA dataset have size about $4000 * 4000$ pixels and may contain large-size images with only a handful of small instances [Xia et al., 2018a]. Image size for the remote sensing domain often depends on the study area scale. A single satellite image can cover an entire city or a large county. Moreover, target objects in remote sensing tasks usually have dramatically lower density (as in the beforementioned DOTA dataset) in comparison to general domain images. It is necessary to split an initial image into crops that a CNN model can accept for training. Therefore, sampling strategy is crucial for the remote sensing domain as simple image partition into tiles is unproductive for large study areas [Xu et al., 2020]. Our framework supports an efficient sampling strategy that uses objects coordinates to crop training patches within large georeferenced images. Object-wise sampling was performed for all experiments with model fine-tuning on the small dataset, as this is a more powerful sampling technique for spatially distributed data in the remote sensing domain, especially in the case of target objects with coordinates [Illarionova et al., 2021d]. In this approach, instead of cropping random patches from an image the target objects' IDs were selected and then cropped according to their coordinates. This allowed us to form a training batch for a convolutional neural network with

more valuable instances when target objects such as individual trees were rare in the study area and unevenly distributed. Object-wise sampling was alternated with a classical random cropping in order not to preserve only small objects in the training data. The probability of the object-wise sampling was set to 0.8.

The entire new sample generating process is conducted during model training. It aimed to ensure greater diversity without memory restrictions related to additional sample storage. Therefore, all functions for object-based augmentation were implemented into the data-loader and generator. New generated samples are also alternated with original samples.

In summary, OBA includes the following options:

- Shadows addition (length and intensity may vary);
- Objects number per crop selection;
- Selection of base color and geometrical transformations probability;
- Background images selection;
- Selection of original and generated samples mixing probability.

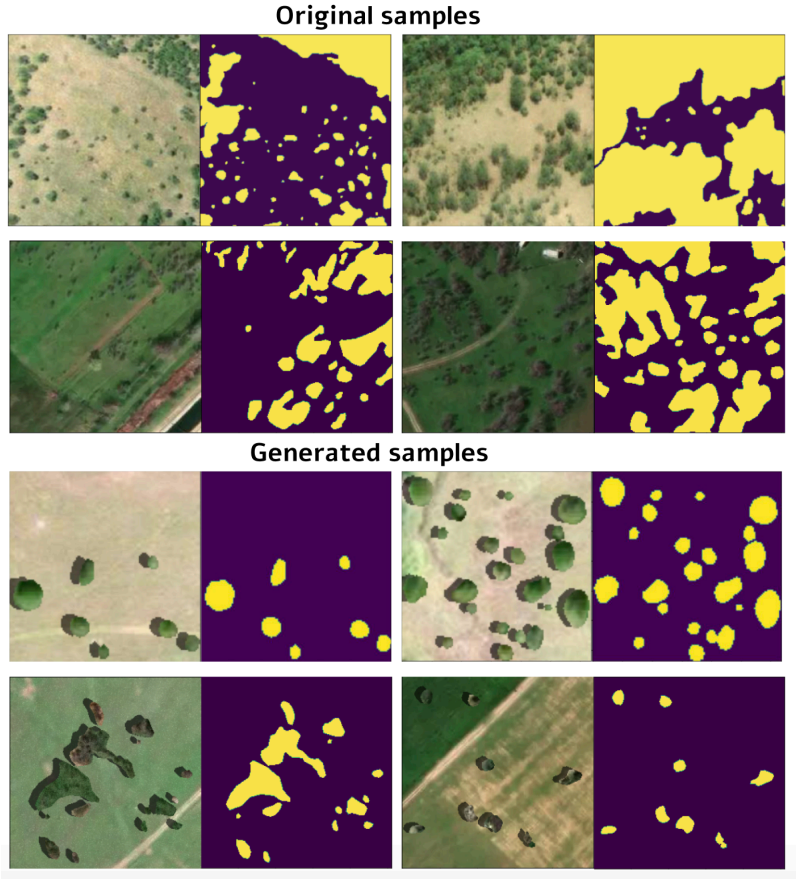


Figure 3-2: Examples of original and generated samples and tree masks. In the generated samples, new various backgrounds were used to achieve greater diversity and to combine trees images and masks from different areas. Artificially added shadows provide more realistic images associated with semantic segmentation masks.

3.2.5 Different Dataset Size

We considered the following subsets of the training dataset in order to evaluate the effect of the dataset's size on the prediction quality: the entire dataset size, 2/3, and 1/3 of the entire training dataset. The chosen subset was used to train the model, while the testing area was permanent and the same for all experiments. We analyzed three and two different dataset splits for the experiments with 1/3 and 2/3 of the entire dataset size, respectively. The final results were defined as an average for each training subset.

3.2.6 Experimental Setup

For the baseline model, we considered the following convolutional neural network architectures: U-Net [Ronneberger et al., 2015], DeepLab [Chen et al., 2017b], and FPN [Lin et al., 2017] with Inception [Szegedy et al., 2017] encoder. Each experiment was run with the same training parameters. The batch size was equal to 20, and the patch size was set to $256 * 256$ pixels. There were 20 epochs with 200 steps. For each epoch, there were 4000 random patches (with size $256 * 256$ pixels) obtained using object-wise sampling or classical random cropping from the training areas. After each epoch, the validation score was estimated. Early stopping was employed after the model reached the plateau with patience 5 epochs. According to the validation score, the best model was then considered in order to compute metrics in the test area. The RMSprop optimizer was used, with a learning rate of 0.001. All experiments used Keras [Keras, Accessed: 20 November 2021].

According to the previous stage, the best model among all considered architectures was employed for fine-tuning on the small dataset. The same training parameters (patch and batch sizes, training epochs number, etc.) were employed. As distinct from the first stage experiments, model weights were already pretrained on the large dataset. Therefore, the model was trained to solve individual tree segmentation and detailed forest mask prediction, which is a more complicated task.

3.2.7 Evaluation

To evaluate the performance of the proposed models we used the general *F1-score*, which is widely used in remote sensing tasks [Kattenborn et al., 2021b]. This allowed us to assess prediction quality for the entire test area.

To assess the quality of our model, we also estimated the average IoU between the predicted masks and the ground truth masks. In order to predict the forest mask, we used test images as an input for the trained neural network. As an output, a neural network predicts a binary mask, which we compared with the labelled binary mask and used to calculate different metrics. For example, prediction quality (*F1-score*) was calculated for each image in the test set. The overall prediction quality

stated for each of the models is the average of the *F1-scores* for all images in the test set.

3.3 Results and Discussion

Figure 3-3 shows the result of the implementation of the first methodological step, namely, the performance on the test data of the model trained on the large dataset (U-Net). It should be noted that the overall model performance is appropriate in terms of metrics (see Table 3.1) as well as by visual comparison; moreover, test images were taken from different parts of the world and represent complex environments. According to the experimental results before fine-tuning, all three considered architectures have almost the same results. Standard deviations for them are about 0.004. Therefore, we can not indicate the statistically superiority of an exact algorithm. For the further experiments we used U-Net architecture. The model performance could be improved for better detection of stand-alone trees.

The results of the model performance after fine-tuning on the small dataset (for which implementation of object-based augmentation was implemented) is shown in Figure 3-4. For the training procedure, we generated about 72,000 patches for the first dataset and about 28,000 patches for the second dataset. From Figure 3-4, it can be clearly seen that the predicted forest mask (see Figure 3-4d) is very similar to the ground truth (see Figure 3-4b), and the separate trees are much better detected compared to the baseline model (see Figure 3-4c). The $F1$ scores for the baseline and improved models are presented in Table 3.2, with the best score being $F1 = 0.929$. The prediction quality for the initial large dataset using the fine-tuned model improved ($F1 = 0.971$). Moreover, it should be noticed from Table 3.2 that when using 1/3 of the whole dataset and implementing OBA in the training procedure the $F1 = 0.913$, which is higher than when using the whole dataset for training the baseline model ($F1 = 0.888$). It is important to mention that standard deviation for the experiment with OBA using the entire training dataset and evaluating small test dataset equals to 0.004 (F1-score is 0.929), while the experiment with simple augmentations for the same data shows standard deviation

of 0.005 (F1-score is 0.888). It supports the significance of the results and shows that our proposed approach is highly relevant in view of the limited amount of high quality labelled remote sensing data.

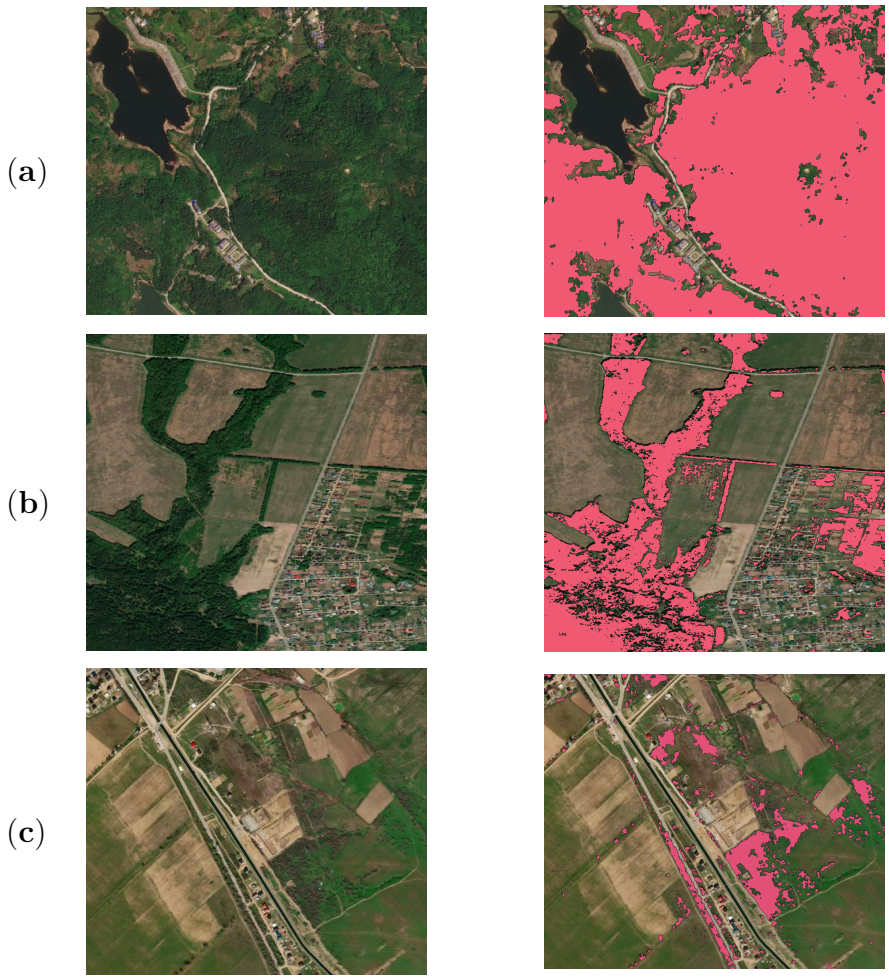


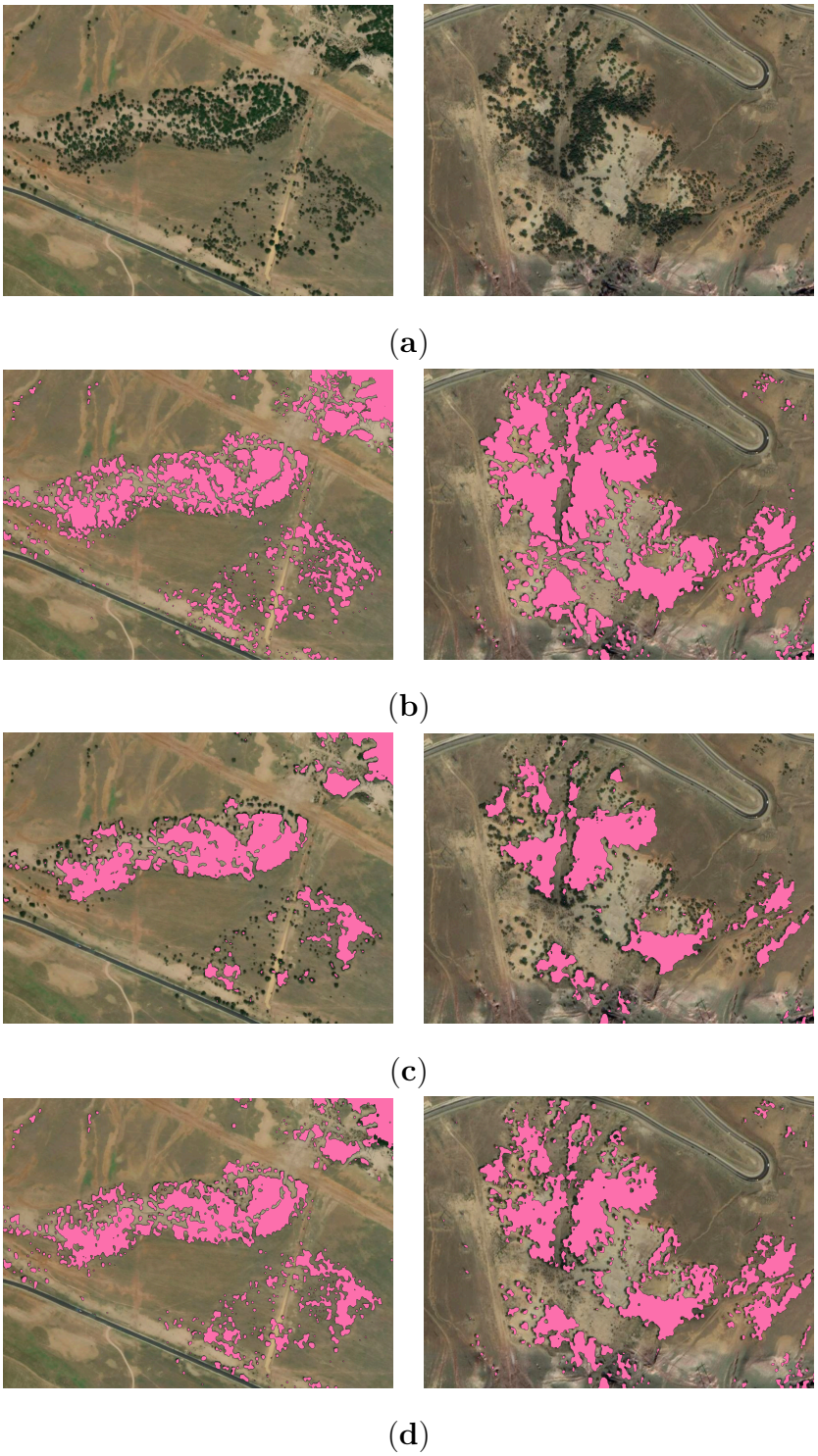
Figure 3-3: Raw images (**left**) and predictions (**right**) for different territories: (**a**) Baoting Li and Miao Autonomous County, Hainan, China, 18°29′24.0″N 109°35′24.0″E; (**b**) Zelenodolsky District, Republic of Tatarstan, Russia, 55°55′48.0″N 48°44′24.0″E; (**c**) Republic of Dagestan, Russia, 43°01′09.1″N 47°19′28.2″E.

The proposed approach allows us to obtain more precise results than the forest cover masks available through OSM serves for particular areas. Examples of generated maps are presented in Figures 3-5 and 3-6. Available open-access forest masks should be updated regularly to include both newly cultivated forests and tree felling. Although trees within built-up areas can be missed in open-access maps, they are crucial for environmental analysis.

The obtained results confirmed the potential of the OBA approach for environmental studies. One of the promising direction for future study is applying precision forest mask for more accurate deforestation analysis. It can be also used for forest species classification as this task usually requires forest boundaries.

A detailed forest mask can be combined with other landcover classes and man-made objects such as the building segmentation task discussed in [Illarionova et al., 2021b]. A promising extension of this research could be the implementation of visual transformers [Wu et al., 2020a] for solving segmentation tasks using remote sensing data. The wide potential of implementing a similar augmentation approach coupled with special image collection techniques for synthetic data generation to improve neural network performance has been shown in a recent study [Nesteruk et al., 2022]. In this study involving segmentation of damage to apples, the authors improved the *F1-score* by up to 4% compared with common augmentation techniques. The authors used DeepLab as the base model for comparing different augmentation techniques. Despite the demonstrated strength of the proposed method, we should take into account its limitations in processing natural scene images. We should carefully use the different types of trees in order to mix them and create the new scene, and trees and scenes should be taken from approximately one period of the year.

Basemap image use makes this approach cost-effective, and high spatial resolution provides significant features for the CNN-based model at the same time. Therefore, this data type is quite competitive with multispectral satellite images which have wider spectral range at lower spatial resolutions. The OBA approach for small precise datasets can be studied for multispectral images to solve other challenges combining RGB bands and vegetation indexes. For instance, NDVI (Normalized Difference Vegetation Index) for deforestation problems was implemented in [Skole et al., 2021].



(**a**)

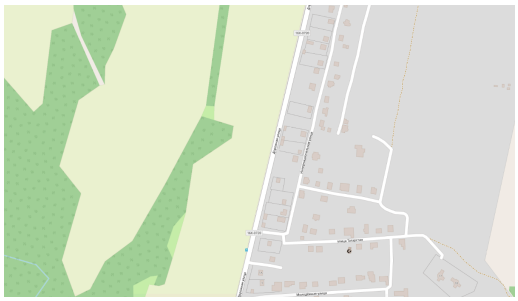
(**b**)

(**c**)

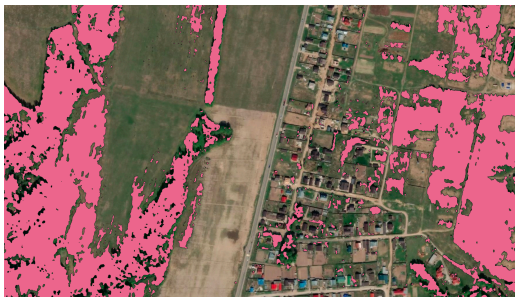
(**d**)

Figure 3-4: Forest segmentation results for Republic of Tatarstan test territories: (**a**) input image; (**b**) ground truth; (**c**) small dataset fine-tuned without OBA; (**d**) small dataset fine-tuned with OBA.

(**a**)



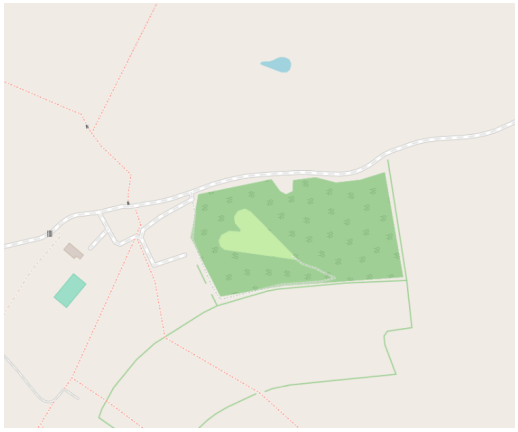
(**b**)



(**c**)

Figure 3-5: Input image from new region outside training site, Zelenodolsky District, Republic of Tatarstan, Russia, 55°55′48.0″N 48°44′24.0″E (composite orthophotomap provided by Mapbox, acquisition date: 20 March 2022) (**a**); Open Street Map (acquisition date: 20 March 2022) (**b**); forest segmentation results of the final CNN model fine-tuned with OBA (**c**).

(**a**)



(**b**)

(**c**)

Figure 3-6: Input image from new region outside training site, Wickwar, England, 51°36′26.7″N, 2°23′17.1″W (composite orthophotomap provided by Google, acquisition date: 30 April 2022) (**a**); Open Street Map (acquisition date: 30 April 2022) (**b**); forest segmentation results of the final CNN model fine-tuned with OBA (**c**).

Table 3.1: Forest segmentation results for Baseline model on two datasets.

	Large dataset			*Small dataset*		
	Precision	Recall	F1	Precision	Recall	F1
U-Net	0.965	0.963	0.964	0.862	0.851	0.856
FPN	0.961	0.958	0.959	0.856	0.849	0.852
DeepLab	0.963	0.962	0.962	0.856	0.854	0.855

Table 3.2: Augmentation approaches comparison for different training set size on the small dataset using fine-tuned U-Net with Inception encoder (F1-score for the test areas from small dataset and large dataset).

	Baseline_no_augm			*Simple_augm*			*OBA*		
Training set size	1/3	2/3	1	1/3	2/3	1	1/3	2/3	1
	F1-score								
Small dataset test	0.861	0.866	0.871	0.867	0.875	0.888	0.913	0.921	0.929
Large dataset test	0.956	0.959	0.962	0.964	0.965	0.967	0.966	0.969	0.971
	Precision								
Small dataset test	0.863	0.865	0.872	0.869	0.877	0.889	0.915	0.922	0.931
Large dataset test	0.955	0.961	0.965	0.965	0.966	0.969	0.964	0.972	0.973
	Recall								
Small dataset test	0.86	0.867	0.871	0.866	0.873	0.887	0.911	0.921	0.928
Large dataset test	0.957	0.958	0.959	0.963	0.964	0.965	0.968	0.967	0.97
	IoU								
Small dataset test	0.754	0.761	0.768	0.774	0.783	0.799	0.851	0.856	0.867
Large dataset test	0.835	0.847	0.856	0.878	0.884	0.891	0.895	0.899	0.912

3.4 Conclusions

High-resolution detailed forest masks are essential for environmental studies. However, in practice such maps are not available for large country-scale territories. Here, we have presented a novel pipeline for forest mask creation using very high spatial resolution basemap RGB images. CNN training included an object-based augmentation approach to achieve more accurate predictions of individual trees and small groups of trees. The created map showed high quality and detalization on various test territories, including in Russia and China. Model prediction showed robustness for regions with complex environmental structures. The provided approach aimed to minimize the need for labeled training data. For the test area used in this study, the *F1-score* for small details was 0.929 compared with a score of 0.856 for the baseline approach. The created forest mask is now available for large-scale and precise environmental studies as part of the open-access platform. As a possible evolution of the current study, we are planning to implement automated selection of hyperparameters and thresholds for augmentation techniques and to use our approach for solving further tree classification tasks.

Chapter 4

Neural-Based Hierarchical Approach for Detailed Dominant Forest Species Classification by Multispectral Satellite Imagery

4.1 Introduction

Algorithmic analysis of remote sensing data allows for solving a wide range of tasks that previously required high professional skills and were time-consuming. One of these challenges is forest species classification, which is commonly considered a dominant species classification problem. A forest dominant species is the one that includes the majority of the timber stock of the stand. Forest management depends on this as a primary characteristic.

The industrial approach to the forest inventory still consists of several methods, including manual and partly automated satellite mapping, LIDAR data analysis, and ground-based surveys. Since the beginning of computer vision method development, many works have aimed to replace some stages with automatic remote sensing imagery analysis.

It is challenging to compare the performance of different methods proposed in

the papers due to their region-specificity and evaluation data inaccessibility for other researchers. Therefore, a comparison of the declared metrics cannot often explain what is better, and a literature survey is mostly qualitative. The presented work partly addresses this issue, as we provided the training markup and the images' IDs to compare achieved results in future studies.

A common choice of remote sensing data is medium-resolution multispectral satellite imagery (Landsat or Sentinel), which is freely available and has a good revisit time. This allows researchers to obtain images for any region of interest with relative ease. The multispectral channels in visible and infra-red wavelengths provide a good deal of information about surface reflection properties. This data type is used in many research works both for single satellite images [Immitzer et al., 2016, Wessel et al., 2018, Mngadi et al., 2019], and time series [Immitzer et al., 2019, Sheeren et al., 2016b]. Although it makes it possible to automatically process the data for vast territories and with decent accuracy, it does not produce high-resolution semantic maps, which can be useful for the precise estimation of timber stock.

A significant number of works cover the usage of airborne multispectral or hyperspectral sensing for forestry inventory classification [Kozoderov and Dmitriev, 2018, Kozoderov et al., 2017, Shinzato et al., 2017, Dalponte et al., 2012], and many of these works leverage a combination with LIDAR scans. It allows for evaluating different forest biomass components [Hernando et al., 2019] and estimating timber stock [Tuominen et al., 2017]. In [Naidoo et al., 2012], they addressed the challenge of savanna tree species classification in South Africa. The basic premise is the heterogeneous nature of the considered region. Therefore, tree height was utilized as structural information to make classification more robust. However, this is not suitable for the preliminary large-area examination due to the high costs of the data and the need for expeditions to the area of interest for imagery acquisition.

A significant source in terms of information depth and availability is very high spatial resolution satellite imagery such as WorldView satellite data (about 2 m spatial resolution). In [He et al., 2019b], they classified deciduous-dominated forest species through three-seasonal WorldView images. In [Immitzer et al., 2012], they

leveraged a single Worldview high-resolution satellite image for species and age classification. The scope of the work included both object- and pixel-based approaches. Sunlit areas of tree crowns presented dataset objects. For such a polygon, a particular species class was ascribed, keeping each object's homogeneity. Moreover, only instances of approximately the same age were chosen for the study, making the samples within a class less diverse. In [Ke and Quackenbush, 2007], they used QuickBird images (a 2.44 m spatial resolution) to classify forest species. Still, a relatively small number of works have given preference to such high-resolution satellite data instead of UAV (unmanned aerial vehicles) images.

Although classical machine learning methods such as Support Vector Machine [Cortes and Vapnik, 1995] and Random Forest [Breiman, 2001] are used in many remote sensing classification studies [Naidoo et al., 2012, Immitzer et al., 2012, Guo et al., 2018b, Belgiu and Drăguţ, 2016, Sheeren et al., 2016a], other works consider newer approaches. In recent years, convolutional neural networks (CNNs) have become a principal method for many computer vision problems, including image classification, segmentation, and object detection. CNNs are applicable in different spheres, and the remote sensing area is no exception [Li et al., 2018b, Dong et al., 2019]. Deep neural networks showed accurate results in the task of deciduous and coniferous classification [Li et al., 2018b], [Hamraz, 2018] and other forest inventory characteristic estimation [Ayrey and Hayes, 2018] using LIDAR sensing data.

Hierarchical problem decomposition can often be implemented in various applied tasks of a particular nature containing sub-classes. It has performed successfully in medical problems [Shen et al., 2019, Huang et al., 2013]. In [Dimitrovski et al., 2012], they implemented a hierarchical multi-label classification for diatom images using a single predictive clustering tree. Just a few studies considered the hierarchical approach for forest species classification [Gerylo et al., 1998, Ahmed et al., 2017]. However, in these works, UAV or airborne data was used with a spatial resolution higher than 0.3 m. The classification approaches were maximum likelihood classification techniques and object-based image classification [Blaschke, 2010], respectively. Thus, all considered forest species classification studies based on satellite images rely exactly on the classical multi-class classification approach [Dalponte

et al., 2012, Immitzer et al., 2012, Ke and Quackenbush, 2007, Sheeren et al., 2016a].

The goal of the presented work is to enhance the spatial detail of dominant forest species estimation using the high-resolution WorldView satellite imagery (2 m per pixel). We have chosen this kind of remote sensing data because it can combine the high availability of satellite imagery (though it is not as high as with moderate resolution) and the spatial precision of aerial imaging. In contrast with most of the work in this area, we did not only concentrate on homogeneous forest stands of approximately the same age. Thus, we aimed to provide a more robust solution applicable to real-life conditions.

We aimed to make the following contribution:

- to improve forest species multi-class image segmentation by splitting the problem into a hierarchy of binary segmentation problems;
- to study the forest height maps usefulness as supplementary data for the forest species classification problem;
- to prepare an open-source dataset for the dominant species segmentation problem — the lack of relevant markup causes obstacles in this sphere of study, so open-access data are crucial.

This Chapter is considered as a core of the Thesis. Forest tree species is one of primer parameters for forest ecosystem analysis. The forest species classification inspires further studies on its refinement in the next Chapters.

4.2 Dataset

4.2.1 Study area

The dataset for this work was created using ground-based observations of Leningrad Oblast of Russia during the 2018 year (Figure 4-1). The total area is around 20000 hectares. The coordinates of this region are between 33°42′ and 33°76′ longitude and between 60°78′ and 61°01′ latitude. The region's climate is humid. The coldest

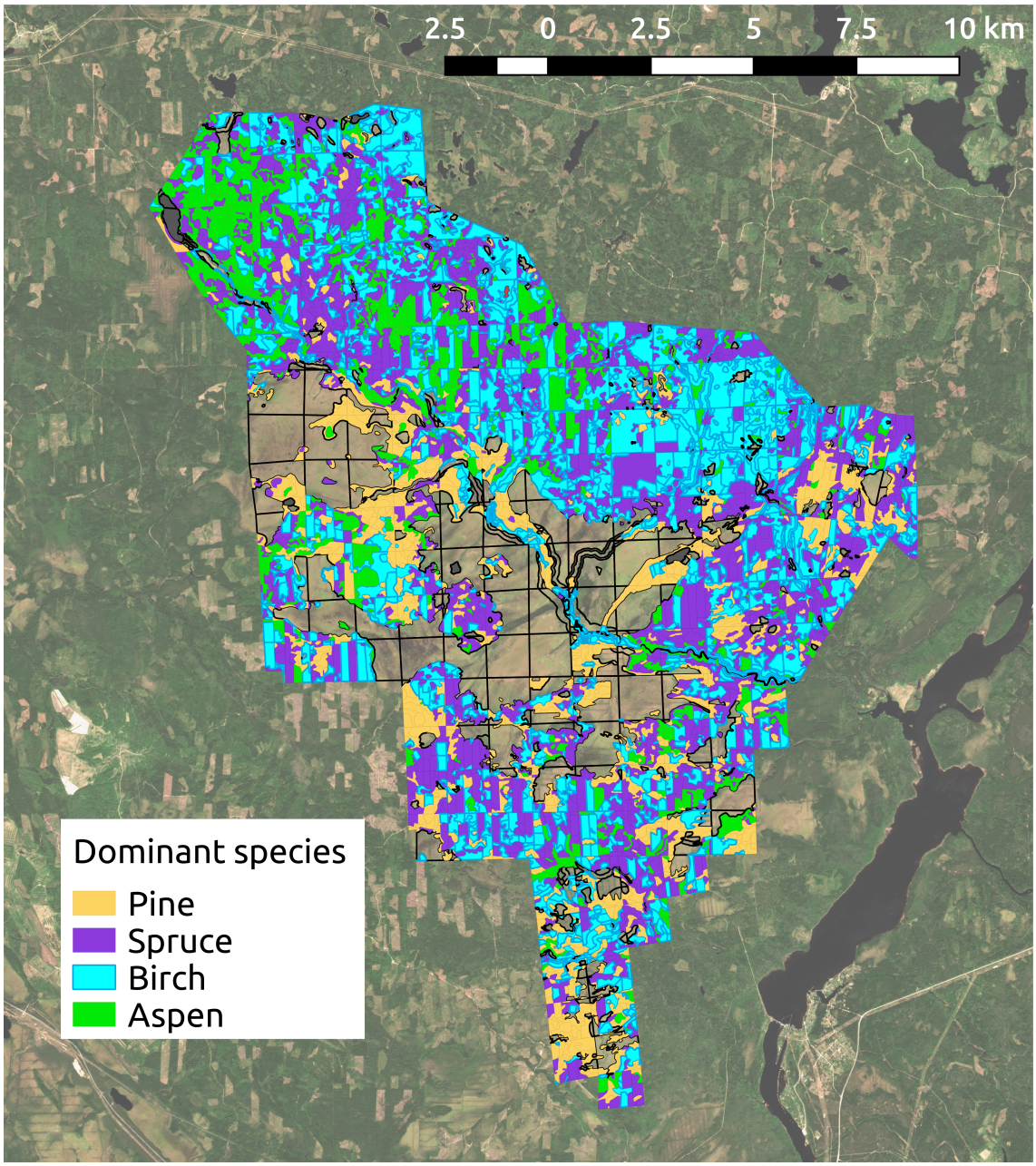


Figure 4-1: Classes markup of study area.

day of the year is in February, with a temperature between $13°F$ and $24°F$ [WeatherSpark, 2020]. The topography is flat. The vegetation cover is mixed and includes deciduous and conifer tree species.

4.2.2 Reference data

The study area was split into small regions representing individual forest stands. The term "forest stand" in forest inventory instructions defines a contiguous forested area sufficiently uniform in essential characteristics to distinguish it from adjacent communities. Each stand is described by several aspects; the most important for this work are the following:

- forest composition, i.e. the percentage of each tree species, denoted with a stride of 10 percent of the relative timber volume (the composition is given in percentage points, each representing 10% of the total timber volume);
- average tree height for each of the primary forest components in the forest composition;
- average tree age for each of the primary forest components in the forest composition.

The rest of the parameters leveraged for the forest analysis are not considered in the current research.

The dominant species is the one that has the highest percentage, and it is the target value that we want to evaluate in this work. Of course, there are situations when two or more forest species have the same or a similar percentage. This case is defined when the difference between the dominant and the second species is not greater than 1 percentage point, and these stands are treated as "mixed forests". The composition of mixed forests is beyond the current study scope, so such stands were excluded from both training and test sets.

The dataset contains forest stands with 4 classes of dominant species: 38% spruce (*Picea spp.*), 14% aspen (*Pópulus tremula spp.*), 26% birch (*Betula spp.*), and 22% pine (*Pínus spp.*) (see Table 4.1). The rest of the study area species are

Table 4.1: Dataset statistics for individual regions.

	area, ha	percentage
aspen	1270.1	14%
birch	2407.7	26%
spruce	3567.2	38%
pine	2063.8	22%

Table 4.2: WorldView images.

	Image ID	Date	Off-nadir angle
0	10300100812E1700	29.07.2018	21
1	1030010081253D00	29.07.2018	29
2	10300100828A7D00	19.07.2018	26
3	103001008067D100	19.07.2018	22
4	1030010080790B00	18.07.2018	22
5	10300110829C9600	18.07.2018	32
6	103001007DCF9400	12.05.2018	14
7	103001007ECC6B00	12.05.2018	18

less distributed and do not compose the stands as a dominant species. It is worth noting that the "dominant species" in forestry does not exactly match the biological term "species" and is connected mostly with the timber class and quality. In this research, the existing forest inventory standards were followed, and this inventory does not distinguish between species within a genus and treats the whole genus as a single class.

4.2.3 Satellite data

WorldView 2 and 3 multispectral imagery with eight spectral bands was downloaded from GBDX [GBDX, Accessed: 2020] in a standard Level 2 format. Product of this level includes radiometric corrections, sensor corrections, geometric corrections, atmospheric compensation, and is processed using a course elevation model. The spatial resolution was about 2 m per-pixel. The central wavelengths of the bands

Table 4.3: Sentinel images.

Image	ID	date
0	L1C_T36VWN_A007126_20180718T092026	18.07.18
1	L1C_T36VWN_A016206_20180730T090554	30.07.18

were: Band 1: Coastal Blue, 427 nm; Band 2: Blue, 478 nm; Band 3: Green, 546 nm; Band 4: Yellow, 608 nm; Band 5: Red, 659 nm; Band 6: Red Edge, 724 nm; Band 7: Near Infrared I, 833 nm; Band 8: Near Infrared II, 949 nm. Sentinel imagery with 13 spectral bands and a spatial resolution of about 10 m per-pixel was downloaded from SentinelHub [Sentinel-Hub]. We considered 13 bands with the following central wavelengths: Band 1: Coastal aerosol, 442.7 nm; Band 2: Blue, 492.4 nm; Band 3: Green, 559.8 nm; Band 4: Red, 664.6 nm; Band 5: Red-edge I (R-edge I), 704.1 nm; Band 6: Red-edge II (R-edge II), 740.5 nm; Band 7: Red-edge III (R-edge III), 782.8 nm; Band 8: Near infrared (NIR), 832.8 nm; Band 8A: Narrow Near infrared (NNIR), 864.7 nm; Band 9: Water vapour, 945.1 nm; Band 10: SWIR Cirrus, 1373.5 nm; Band 11: Shortwave infrared-1 (SWIR1), 1613.7 nm; Band 12: Shortwave infrared-2 (SWIR2), 2202.4 nm. Images were pre-processed with the Sen2Cor package for atmospheric correction (level L2A Bottom of Atmosphere (BoA) reflectance). All images were from the high vegetation period from May to August. Image acquisition dates and catalogue IDs are presented in Tables 6.1 and 8.1. Normalization procedure for satellite data is described in Section 4.3.3.

Dataset consists of geo-referenced satellite images in the format of 8-bit TIFF files and forestry inventory data converted into raster per-pixel masks for each class.

The additional challenge was posed by the temporal mismatch between imagery and markup. Current forest inventory information is sparsely available. Thus, some forest areas were felled after the ground-based observations. To deal with this, we utilized a previously trained neural network that performs forest segmentation. It produces an up-to-date forest mask for the images and excludes the derived non-forested areas from the training and validation sets. We additionally cleaned the test set manually.

4.3 Methods

4.3.1 Problem definition

As described in Section 7.2.1, we treated an individual forest stand as a homogeneous region with a common characteristic within its area. The aim was to develop a method that could produce high-resolution semantic maps outlining forest stands. Thus, the problem was formulated as image segmentation: to assign a species class to every pixel in the image. The background classes were excluded from the dataset before training and did not appear at the test time. The following fact complicated the problem. Forest stands can have inconsistency and include visible parts of the non-dominant species. These parts should be segmented as a separate stand of another dominant species, but the training data do not support it, as the markup is completely stand-wise.

4.3.2 Neural networks for image segmentation

As the most recent computer vision advances are connected with the novel neural network architectures, it is vital to select a suitable one for the given task and available computational resources. Since the task was formulated as a multi-class image segmentation problem, a fully convolutional architecture was considered, such as U-Net [Ronneberger et al., 2015] or a Feature Pyramid Network (FPN) [Lin et al., 2017]. Both of them show good image segmentation performance, including remote sensing data, with FPN being more suitable for multi-class segmentation. These architectures are constructed in an encoder-decoder fashion with skip connections, which allows us to use various convolutional encoders. Modern architectures outperformed the original VGG encoder used in [Ronneberger et al., 2015], so the first variant was ResNet [He et al., 2016], used by [Lin et al., 2017]. As counterparts, we used Inception-ResNet-v2 [Szegedy et al., 2017] and EfficientNet [Tan and Le, 2019] as one of the most resent and advanced architectures, showing state-of-the-art results at the ImageNet benchmark [Deng et al., 2009]. To comply with computational resource restrictions, the model size was limited to ResNet-34 and EfficientNet-B3 cor-

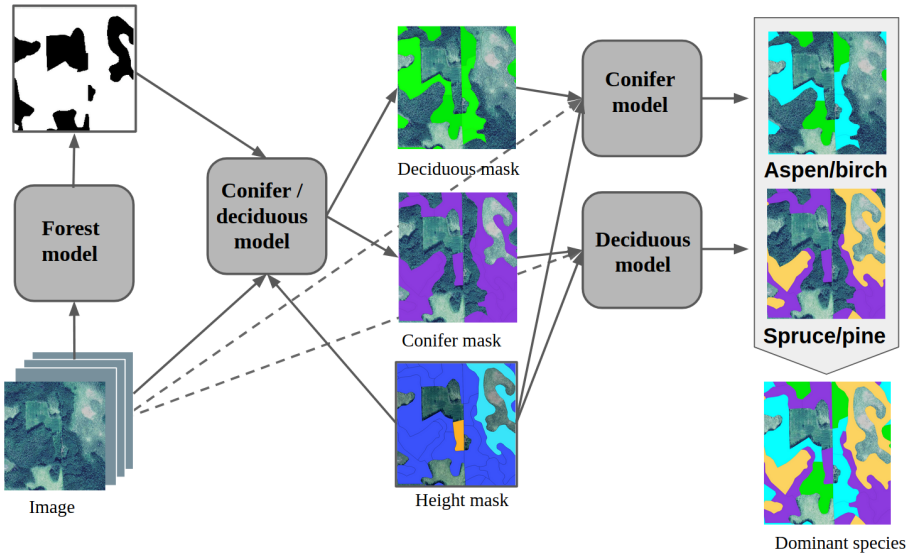


Figure 4-2: Hierarchical model structure.

respondingly. The models' architecture implementation was based on [Yakubovskiy, 2022].

4.3.3 Image preprocessing

As Sentinel images were contrast-enhanced and had a value range of $[0 : 255]$ in each channel, they were scaled as

$$I' = I/255, \tag{4.1}$$

where I and I' are intensities before and after the normalization.

To ensure relative brightness uniformity for different images, we performed minimum-maximum brightness normalization to the range $[0, 1]$, as in [Jayalakshmi and Santhakumaran, 2011].

The WorldView images have a wider dynamic range, different for each channel, so contrast enhancement was also included in the scaling formula to suppress the darkest and the brightest regions that lie beyond three standard deviations from the mean value:

$$m = max(0, mean(I) - 3 * std(I)), \tag{4.2}$$

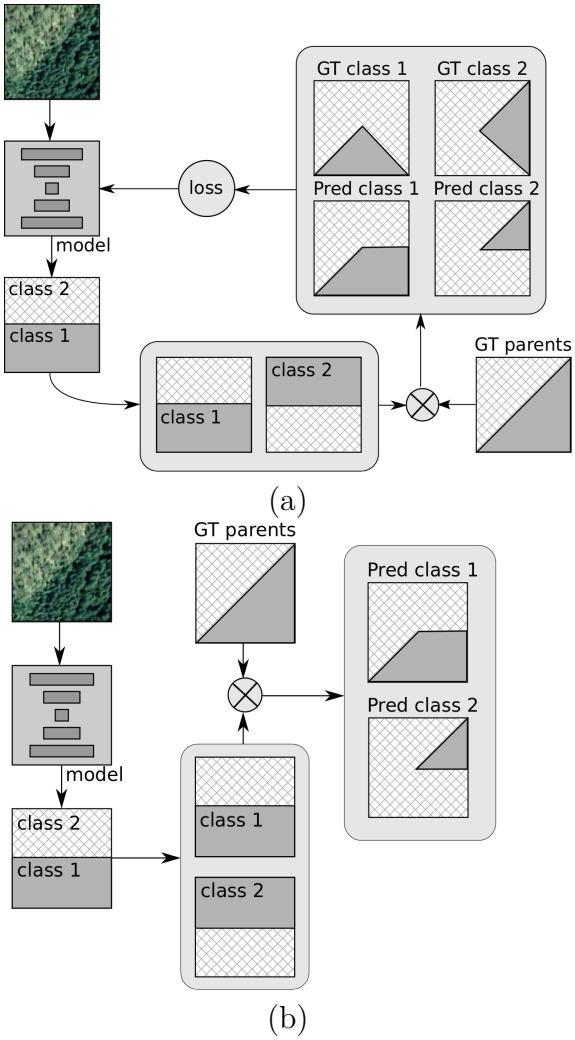


Figure 4-3: The data flow through a level of the hierarchical process: (A) model training, (B) inference.

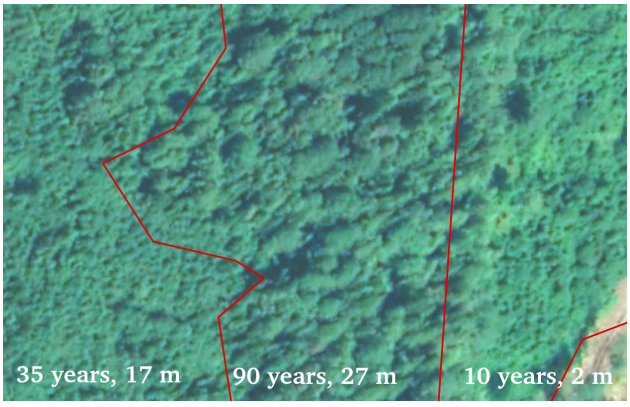


Figure 4-4: Example of age and height variance within one species.

$$M = min(max(I), mean(I) + 3 * std(I)), \tag{4.3}$$

$$I' = (I - m)/(M - m), \tag{4.4}$$

where $mean, std$ are the mean and standard deviation of the image. In equations 8.2, 8.3, we calculate m and M (minimum and maximum of the preserved dynamic range). In equation 8.4, values are scaled to 0 and 255 linearly. The values outside the [m,M] range are clipped. The standardization of the imagery according to the whole dataset statistics proves profitable for the neural network training compared to a simple scaling of the entire value range [Pal and Sudeep, 2016].

There are two ways to compute mean and standard deviation values: for all channels simultaneously or individually for each band. The advantage of the first type is that ratios between the channel values stay constant, which might be necessary for a more in-depth nature processes evaluation. On the other hand, when statistics within each channel are computed, connections between the same channels of different images are more robust, and it can be useful for algorithm adaptability.

4.3.4 Dataset augmentation

The dataset augmentation is a common technique that can improve the robustness of the neural network. In the considered case, the spatial transforms were applied to the training images with 50% probability: rotation with a 90-degree step, a vertical and horizontal flip, and a zoom-in and -out within 20%.

4.3.5 Oversampling

To handle the class imbalance, we added extra weights for the smaller classes during loss computation. For this variant, weighted cross-entropy (WCE) was computed, and optimal weights were estimated according to the class distribution in the training set.

The other problem was that the label for the dominant species property was the same, whether there were 50% or 90%. Still, the former represented a more “dirty” markup for the segmentation, as about half of the pixels represented non-dominant species. We managed to enforce the training on more clean samples by increasing

the probability of the samples with a higher dominant species percentage. Species distribution is provided for each forest region in Table 4.4.

4.3.6 Problem decomposition

The baseline approach used multi-class segmentation, where the output layer of the neural network had a number of outputs that was equal to the number of classes. The argmax (arguments of the maximum) of these values was treated as a class label for a pixel.

The approach modification was based on the fact that forest species classification has an explicit hierarchy: classes are divided into coniferous and deciduous tree species. Therefore, it was reasonable to decompose the problem. The hierarchical solution represented the multi-class segmentation as a set of binary segmentation problems. The multi-class segmentation map was obtained by consistently applying the method and aggregating the results (see Fig. 4-2).

The stages scheme of the hierarchical segmentation process is depicted in Fig. 4-3. We used the "parent" data obtained from the previous stages of the processing at each step. For example, to segment coniferous and deciduous forest stands, the forest mask was utilized to exclude non-forest regions from the observation. During the model training, this "parent" data was used as a mask for the loss function computation. The training loss was calculated within the parent class areas only because, for the same example, there was no need to rely on the non-forested regions to distinguish between the forest types. During the inference, the result of the binary segmentation was multiplied by the "parent" mask.

We also compared this approach with "one versus all" classification, where a set of separate neural network models is trained to predict just one class. All predictions are then aggregated, and the most likely label is ascribed to each pixel.

4.3.7 Height data

It is worth noticing that a part of the intro-class variance is connected with the forest height or age, with a high correlation. The same forest species at different ages shows

Table 4.4: Dataset statistics for individual regions (dominated species by threshold), area in ha.

threshold	pine	birch	aspen	spruce
0.5	2063.8	2407.7	1270.1	3567.2
0.6	1781.9	951.9	659.6	2390.6
0.7	1540.9	463.5	235.8	1350.2
0.8	1234.3	178.7	84.2	643.7

different patterns (see Fig. 4-4). As the height data could be obtained from separate sources, we studied the height data's effect on the dominant species classification. The input data was extracted from the same forest inventory characteristics used for training, and it was used as an additional raster band in the network input. This modification also contributed to the method performance in both multi-class tasks and binary segmentation cases.

4.4 Experiments

4.4.1 Training

The training of all the neural network models was performed on a PC with GTX-1080Ti GPUs.

The batch size varied from 16 to 30 depending on the architecture's memory restrictions.

During the binary segmentation models' training within the hierarchical segmentation approach, only two particular classes of the current stage were taken into account. Accordingly, the loss function was calculated only over the part of the image corresponding to the parent class of the current stage, as is shown in Fig. 4-3 A. The total loss for a training batch was normalized to the parent class area in the batch.

The final model combined all these approaches.

4.4.2 Medium resolution data

The same experiments were performed using widely spread in the forest inventory tasks Sentinel-2 data to compare the selected data to other possible sources.

The base model used 13 bands of Sentinel imagery at a spatial resolution from 10 to 60 m. This data is available for free download. The model was trained in the same manner as a model without height for Worldview data. The image crop size was reduced in batch from 256 to 64 to, by giving the field of view the same size, make the training procedure as similar as possible.

The dataset was split into training, validation, and test sets in the following proportion: 0.7, 0.15, and 0.15. The validation set was used to choose the best neural network parameters and architecture.

F1-score was utilized to measure the segmentation quality and compare the method variants, for the individual classes and averaged over all the classes.

F1-score was computed only for regions covered by species with a domination of more than 0.5, which was described in Section 7.2.1. When the optimal in terms of the validation dataset architecture for each task had been found, the final models were evaluated using the test set of the images, which did not overlap with the training or validation sets. We also used confusion matrices, as this is a commonly considered accuracy assessment approach in remote sensing image classification [Foody, 2002].

4.5 Results and discussion

4.5.1 Hierarchical decomposition

We compared hierarchical decomposition with two commonly used image semantic segmentation approaches: multi-class classification and “one versus all.” All studies were conducted both for WorldView and Sentinel images to assess the proposed method using different data sources. The results of multi-class classification and hierarchical decomposition before aggregation are reported in Tables 4.7. As shown in Tables 4.11, 4.10, which have the aggregated results, the hierarchical approach

Table 4.5: Results for multiclass classification without height (F1-score) for WorldView and Sentinel (baseline) on validation. Bold numbers — the best score (the corresponding model was chosen for the final results aggregation). Incept — Inceptionresnetv2. Standard deviation is presented for average F1-score.

	Unet + Resnet34	Unet + EfficientNet	Unet + Incept	FPN + Resnet34	FPN + EfficientNet	FPN + Incept
WorldView						
aspen	0.39	0.26	0.385	0.35	0.2	0.35
birch	0.79	0.548	0.781	0.76	0.18	0.71
spruce	0.759	0.743	0.754	0.75	0.68	0.76
pine	0.868	0.847	0.859	0.87	0.81	0.859
average	**0.702** ± 0.005	0.599 ± 0.004	0.695 ± 0.007	0.682 ± 0.005	0.47 ± 0.006	0.66 ± 0.004
Sentinel						
aspen	0.367	0.356	0.417	0.361	0.219	0.372
birch	0.713	0.687	0.738	0.694	0.258	0.681
spruce	0.717	0.708	0.658	0.721	0.669	0.722
pine	0.841	0.845	0.83	0.853	0.813	0.845
average	0.659 ± 0.004	0.649 ± 0.005	**0.66** ± 0.004	0.657 ± 0.005	0.489 ± 0.006	0.655 ± 0.005

Table 4.6: Results for multiclass classification with height (F1-score) for WorldView and Sentinel on validation. Bold numbers — the best score (the corresponding model was chosen for the final results aggregation). Incept — Inceptionresnetv2. Standard deviation is presented for average F1-score.

	Unet + Resnet34	Unet + EfficientNet	Unet + Incept	FPN + Resnet34	FPN + EfficientNet	FPN + Incept
WorldView						
aspen	0.38	0.43	0.39	0.42	0.38	0.39
birch	0.78	0.80	0.79	0.79	0.80	0.79
spruce	0.8	0.78	0.76	0.79	0.77	0.74
pine	0.87	0.87	0.85	0.85	0.82	0.84
average	0.707 ± 0.004	**0.72** ± 0.004	0.697 ± 0.005	0.712 ± 0.005	0.692 ± 0.006	0.69 ± 0.004
Sentinel						
aspen	0.426	0.414	0.415	0.419	0.371	0.474
birch	0.726	0.733	0.712	0.726	0.748	0.772
spruce	0.745	0.75	0.736	0.753	0.763	0.78
pine	0.837	0.851	0.847	0.844	0.846	0.864
average	0.68 ± 0.006	0.687 ± 0.005	0.677 ± 0.004	0.685 ± 0.006	0.682 ± 0.007	**0.72** ± 0.005

allows us to improve model performance in terms of the F1-score for WorldView from 0.716 to 0.836 and for Sentinel from 0.668 to 0.77. "One versus all" classification

Table 4.7: Hierarchical classification with height (h+) and without height (h-) data for WorldView and Sentinel on validation before the results aggregation (F1-score) . Blue numbers — the best score for models without height, bold numbers — the best score for models with height (the corresponding models were chosen for the final results aggregation). Incept — Inceptionresnetv2.

	Unet Resnet34		FPN Resnet34		Unet EfficientNet		FPN EfficientNet		Unet Incept		FPN Incept	
	WorldView											
species	h-	h+	h-	h+	h-	h+	h-	h+	h-	h+	h-	h+
aspen / birch	**0.76**	**0.78**	0.69	0.75	0.67	0.74	0.64	0.66	0.69	0.72	0.7	0.7
pine / spruce	**0.92**	0.926	0.9	**0.93**	0.918	0.92	0.88	0.91	0.9	0.92	0.89	0.91
conifer / deciduous	0.86	**0.91**	**0.86**	0.9	0.86	0.9	0.85	0.9	0.86	0.9	0.86	0.9
	Sentinel											
aspen / birch	0.68	0.68	0.63	0.66	**0.69**	0.66	0.58	0.63	0.69	0.7	0.67	**0.72**
pine / spruce	0.89	0.9	0.88	0.92	0.89	**0.92**	**0.9**	0.92	0.88	0.91	0.9	0.92
conifer / deciduous	0.83	0.87	**0.84**	0.86	0.83	0.86	0.82	0.85	0.81	0.87	0.83	**0.87**

Table 4.8: Hierarchical approach (1) in comparison with "one versus all" classification and (2) on test data (both approaches use height data) from the WorldView data. Standard deviation is presented for average F1-score.

	1	2
aspen	0.72	0.75
birch	0.75	0.48
spruce	0.94	0.71
pine	0.92	0.82
average	0.836	0.69
	± 0.007	± 0.006

Table 4.9: Hierarchical approach (1) in comparison with "one versus all" classification and (2) on test data (both approaches use height data) from the Sentinel data. Standard deviation is presented for average F1-score.

	1	2
aspen	0.79	0.46
birch	0.586	0.56
spruce	0.93	0.75
pine	0.789	0.88
average	0.77	0.667
	± 0.005	± 0.006

Table 4.10: Final aggregated results (F1-score) for WorldView test data. Standard deviation is presented for average F1-score.

	hierarchy + height	hierarchy	multi-class + height	multi-class
aspen	0.721	0.714	0.773	0.39
birch	0.751	0.649	0.469	0.796
spruce	0.947	0.954	0.764	0.759
pine	0.925	0.87	0.851	0.869
average	**0.836**	0.797	0.716	0.703
	± 0.007	± 0.005	± 0.003	± 0.005

also shows lower results than those of the hierarchical decomposition depicted in Tables 4.8 and 4.9. For WorldView, there is decline in quality in the F1-score from 0.836 to 0, while for Sentinel, that decline is from 0.77 to 0.667. There is no significant difference between multi-class and "one versus all" classification. For WorldView, the difference is 0.716 and 0.69; for Sentinel, it is 0.668 and 0.667. Confusion matrices for WorldView and Sentinel data are shown in Fig. 4-5. The WorldView prediction quality is higher than that of Sentinel. Moreover, for the

Table 4.11: Final aggregated results (F1-score) for Sentinel test data. Standard deviation is presented for average F1-score.

	hierarchy + height	hierarchy	multi-class + height	multi-class
aspen	0.79	0.612	0.608	0.586
birch	0.586	0.527	0.441	0.274
spruce	0.93	0.943	0.766	0.692
pine	0.789	0.792	0.855	0.791
average	**0.77** ± 0.005	0.72 ± 0.006	0.668 ± 0.007	0.58 ± 0.006

Table 4.12: Oversampling effect on the WorldView validation images (F1-score). (1) All stands with a dominant species content larger than 50% are used. (2) Special thresholds are defined for each class (0.7 for spruce and pine, 0.6 for birch, and 0.5 for aspen). Standard deviation is presented for average F1-score.

	1	2
aspen	0.371	0.39
birch	0.746	0.79
spruce	0.732	0.759
pine	0.751	0.868
average	0.65 ± 0.005	0.667 ± 0.004

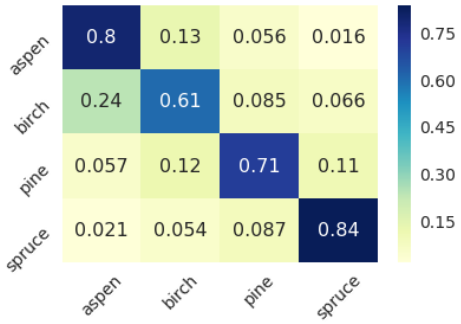
WorldView imagery, coniferous and deciduous sub-classes are less often ascribed to the wrong parent class.

One of the important issues of the hierarchical approach is that, for each classification task, the most suitable neural network architecture can be chosen.

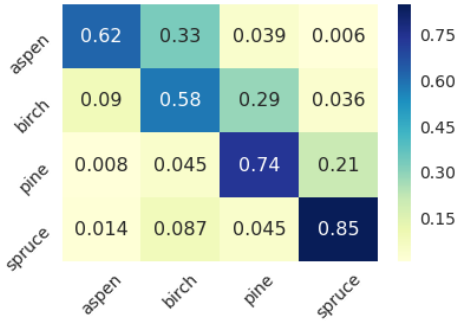
As is shown in Tables 4.10 and 4.11, the accuracy of the classification of aspen and birch became more adequate, and the final performance is more satisfying in the context of available markup.

The proposed work approach is only applicable when a hierarchy of classes is established. However, this approach can yield better results, as shown by utilizing the semantic connections between classes. It also helps to reduce computational costs in the case of a high number of classes (a binary logarithm instead of a linear one).

The computational overhead from the use of four models in the hierarchical approach instead of two in the multi-class baseline is not crucial since the problem is neither real-time nor addressed to the mobile devices.



(a) WorldView.



(b) Sentinel.

Figure 4-5: Confusion matrices for the best aggregated hierarchical models with height data: (a) WorldView data, (b) Sentinel data.

4.5.2 Supplementary height data

Aggregated results for experiments with height data are presented in Tables 4.10 and 4.11. For the multi-class approach and hierarchical decomposition, height data usage improves model performance. WorldView hierarchical decomposition enhances the quality from 0.797 to 0.836. In multi-class classification, the F1-score without height is 0.703; with height, it is 0.716. The same trend is observed for the Sentinel data. Hierarchical decomposition with height improves the quality from 0.72 to 0.77; for multi-class classification, the scores are 0.58 and 0.668, respectively.

A sample of the test region with the ground truth markup and the predictions of the final hierarchical model with height supplementary data is presented in Fig. 4-6 and 4-7, which show a significant intersection between real classes and the artificially estimated classes. Experiments with both high and medium resolution data confirmed the reliability of the chosen strategy.

4.5.3 Architecture selection

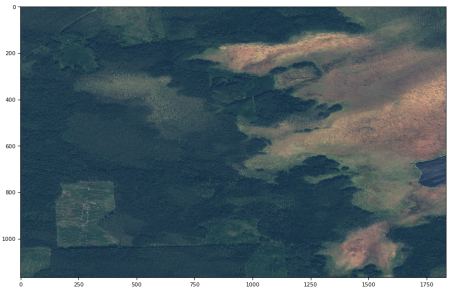
We compared six neural network architectures (U-Net with Resnte34 encoder, U-Net with EfficientNet encoder, U-Net with Inceptionresnetv2 encoder, FPN with Resnte34 encoder, FPN with EfficientNet encoder, and FPN with Inceptionresnetv2 encoder) for each of the classification tasks in the hierarchical decomposition and the multi-class approaches. Results are presented in Tables 4.7. Aggregated predictions were computed for the best models in each category. The batch size was limited by the available memory properties and was reduced for larger models for the WorldView data with a crop size of $256 * 256$ pixels. For Sentinel, the crop size was smaller ($64 * 64$ pixels); therefore, the batch size was the same for all experiments. The best models for WorldView are the smaller ones (U-Net with Resnet34 encoder and FPN with the same encoder). However, for Sentinel experiments, the best architecture is considerably different. Model performance is affected by data amount and structure. Finding a universal architecture is beyond the scope of this study, but experimental results indicate that architecture searching is advisable. Both the WorldView and Sentinel studies show that the correct architecture for each task can adjust classification quality, although classification pipeline and auxiliary data are also of high importance in such applied tasks. Therefore, we assume that these two points should be taken into account to develop a robust computer vision model for environmental tasks.

4.5.4 Augmentation and oversampling

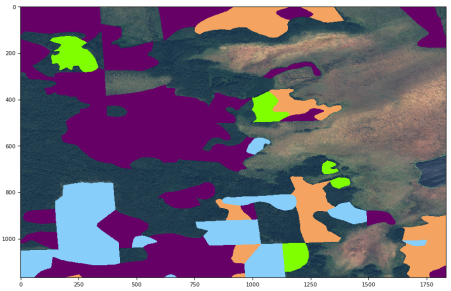
For all models, we implemented geometrical augmentations. This allowed us to achieve a higher diversity in the training dataset. As augmentation in neural network training is well-studied, we assessed its contribution to the classification quality for only one architecture and one classification solution: the U-Net with Resnet34 encoder in the multi-class problem definition, with WorldView images, and without supplementary height data. The F1-score without augmentation during training is 0.67 (for validation augmented data), while the augmentation procedure increases the quality to 0.7 (for the same validation augmented data). This effect is explained

by the fact that a neural network treats any geometrical transformation as a new training sample.

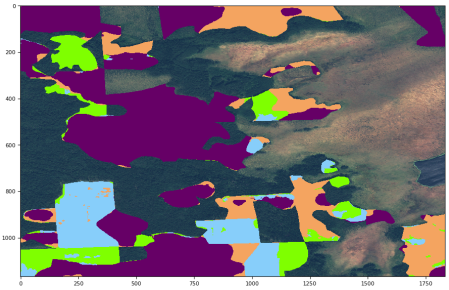
We conducted class oversampling according to the thresholds defined in Table 4.4. Two strategies were compared: 1) A dataset of forest stands was formed with a dominant species content of more than 50%, and 2) a special threshold was defined for each class (0.7 for spruce and pine, 0.6 for birch, and 0.5 for aspen). The averaged results for a multi-class approach with WorldView images and without height data are presented in Table 4.12. It shows that such an oversampling can increase model performance.



(a) Input image.



(b) Ground truth.



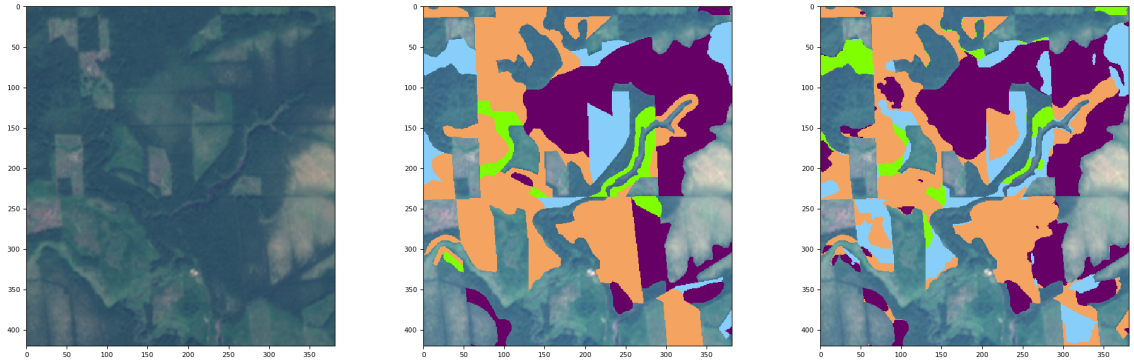
(c) Prediction.

Figure 4-6: A sample of the WorldView imagery for the test area.

4.6 Conclusions

We studied the applicability of the neural networks for the automatic extraction of forest inventory characteristics from satellite imagery and concentrated on the dominant species classification problem. We present the following contributions:

- We provide a labeled dataset for dominant species classification, covering a part of Leningrad Oblast, Russia.



(a) Input image. (b) Ground truth. (c) Prediction.

Figure 4-7: A sample of the Sentinel imagery for the test area.

- We developed a hierarchical pipeline for the neural network segmentation, which allows outperforming the basic network approach in the multi-class image segmentation problem. Applicability and relevance of our solution were proved on two data sources: Sentinel and WorldView satellites.

- We investigated the effect of the supplementary height data, which increases the accuracy significantly.

This approach can be extended to other forest inventory problems and can be improved by a better training markup, both of which we are going to pursue in future work. Moreover, the results in this study are limited to dominant species classification only. However, in future research, we are going to cover mixed forest cases, which will fall entirely into the hierarchical segmentation scheme. The other goal is to add more forest inventory characteristics, which can also be estimated from the satellite imagery.

Chapter 5

Tree Species Mapping on Sentinel-2 Satellite Imagery with Weakly Supervised Classification and Object-Wise Sampling

5.1 Introduction

Many ecological and forest management studies are based on knowledge about tree species within a region of interest. Such knowledge can be used for the precise analysis of natural conditions [Lindenmayer et al., 2000], the development of ecological models [Franklin et al., 2018], and for conservation and restoration decision-making [Wallace and Clarkson, 2019]. Tree species information can be leveraged for timber volume [Hill et al., 2018, Bont et al., 2020] and biomass estimation [Pandey et al., 2019] accompanied by other characteristics, such as tree age and height, crown width.

A commonly used approach for forest type data gathering is field-based measurement, which has the obvious drawbacks of acquisition cost and difficulty. Many studies are now focused on the automatization of land-cover survey through the use of remote sensing-derived data. This approach is more preferable when analysing

vast territories. For instance, the creation of large-scale maps has been described in [Persson et al., 2017, Lei et al., 2016]. For such tasks, both low spatial resolution and high resolution data can be used. Examples of frequently leveraged data sources with resolution lower than 30 m is Landsat satellite imagery [Pasquarella et al., 2018, Gudex-Cross et al., 2017]. Promising results have been shown in studies, both for single image and time-series data [Stoian et al., 2019, Nguyen et al., 2018, Campos-Taberner et al., 2020]. Nevertheless, some tasks require more precise data with higher resolution. Multispectral images with high resolution strive to provide more thorough land-cover analysis.

Recently, image classification algorithms have demonstrated high prediction accuracy in a variety of applied tasks. Algorithms based on machine learning methods are now commonly used for land-cover mapping—particularly for forest species prediction—using satellite imagery. Classical methods, such as Random Forest [Breiman, 2001], Support Vector Machine [Cortes and Vapnik, 1995], and Linear Regression, usually work with feature vectors, where each value corresponds to some spectral band or combination of bands (in the case of vegetation indices) [Hamedianfar and Barakat A. Gibril, 2019, Chen et al., 2018]. Deep neural network approaches have proved to be more capable for many land-cover tasks [Kussul et al., 2016, Mahdianpari et al., 2018, Illarionova et al., 2021c]. In [DeLancey et al., 2020], a CNN was compared with XGBoost [Chen and Guestrin, 2016b]. In [Sun et al., 2019], a CNN approach was examined for tree mapping, through the use of airplane-based RGB and LiDAR data. In [Illarionova et al., 2020], neural-based hierarchical approach was implemented to improve forest species classification.

In contrast with typical image classification tasks (such as in the Imagenet data set), land-cover tasks involve spatial data. Vast study regions are usually supplied, with a reference map covering the entire area. Classes within this area may not be evenly distributed in many cases [Xia et al., 2018a]. Moreover, classes of vegetation types of land-cover are often imbalanced within the study region. In many works, the analysed territory can be covered by a single satellite tile (e.g., the size for Sentinel-2 is $100 * 100$ km^2). Therefore, researchers need to choose both how to select the training and validation regions and how to organize the training procedure to deal

with imbalanced classes and a spatial distribution that is usually far from uniform. Sampling approach is vital for the remote sensing domain as simple image partition into tiles is ineffective for vast territories [Xu et al., 2020]. The training procedure depends on whether we use a pixel-wise [Trisasongko et al., 2017] or object-wise approach [Hamedianfar and Barakat A. Gibril, 2019, Gudex-Cross et al., 2017]. In a pixel-wise approach, each pixel is ascribed a particular class label and the goal is to predict this label using a feature description of the pixel. In an object-wise approach, a set of pixels is considered as a single object. In some classical machine learning methods, a combination of the two approaches has also been considered [Chen et al., 2018]. An alternative approach to classical pixel- or object-wise has been provided in [Sun et al., 2019] for a CNN tree classification task using airplane-based data. During the described patch-wise training procedure, the model strove to predict one label for a whole input image of size $64 * 64$ pixels. However, for some semantic segmentation tasks with lower spatial resolution, the input image can include pixels with different labels and, therefore, the aforementioned approach is not always applicable. The same issue was faced in [Mahdianpari et al., 2018], where patch-wise approach was implemented for CNN for a land-cover classification task using RapidEye satellite imagery. Some patches with mixed labels were excluded, in order to solve the problem. In our study, we aim to provide sampling approach for medium resolution satellite imagery for forest species classification. In contrast to [Sun et al., 2019], we focus on the particular area within a patch and do not exclude from training patches with mixed labels as in [Mahdianpari et al., 2018].

Another important issue is markup limitations. Field-based measurements are commonly used as reference data. Vast territories are often split into small aggregated areas comprised of groups of trees called individual stands. These stands are not necessarily homogeneous but, in some cases, the percentages of different tree species within the stand is available. The location of the non-dominant trees is unknown. In such cases, machine learning algorithms are often trained to predict the dominant class even for regions with mixed forest species [Abdollahnejad et al., 2017], or just areas with a single dominant tree species are selected [Knauer et al., 2019]. This raises the issue of weak markup adjustment. Among weakly supervi-

sion tasks, this one belongs to inexact supervision when only coarse-grained labels are given [Zhou, 2017]. Weakly supervised images occur both in the general domain [Guo et al., 2018a, Ahn et al., 2019a] and in specific tasks such as medical images segmentation [Xu et al., 2019]. These studies involve new neural network architectures or frameworks development to decrease requirements for labor-intensive data labeling. In the remote sensing domain weakly supervised learning was also considered in different tasks such as cropland segmentation using low spatial resolution satellite data [Wang et al., 2020], cloud detection through high resolution images [Li et al., 2020c], and detection of red-attacked trees with very high resolution areal images [Qiao et al., 2020]. However, in the field of forest species classification, the weak markup problem requires additional analysis according to data specificity (both satellite and field-based). In this study, we propose a CNN-based approach to extract more homogeneous areas from the traditional forest inventory data that includes only species' content within stands and does not provide each species' location. We focus on semantic segmentation problem using high resolution multispectral satellite data. The approach is particularly based on the Co-teaching paradigm presented in [Han et al., 2018] where two neural networks are trained, and small-loss instances are selected as clean data for image classification task. In contrast, we split the data adjustment and training process into two separate stages and implement this pipeline for the semantic segmentation task.

In this study, we aim to explore a deep neural network approach for forest type classification in Russian boreal forests using Sentinel-2 images. We set the following objectives:

- to develop a novel approach for forest species classification using convolutional neural networks (CNN) combining pixel- and object-wise approaches during the training procedure, and compare it with a typically used approach for semantic segmentation; and
- to provide a strategy for weak markup improvements and examine forest type classification both as a problem of (a) dominant class estimation for non-homogeneous individual stands and (b) more precise homogeneous classifica-

tion.

This study extends the previous Chapter on forest species classification. It shows forest inventory data specificity in more details and provides ideas how to take it into account. We focus on tree species estimation, while the proposed approaches are also applicable for other forest parameters prediction.

5.2 Materials and methods

5.2.1 Study Site

The study was conducted in the Russian boreal forests of Leningrad Oblast. The coordinates of these regions are between 33°42′ and 33°76′ longitude and between 60°78′ and 61°01′ latitude (Figure 8-1). The vegetation cover is mixed and includes deciduous and conifer tree species. The main species are pine, spruce, aspen, and birch. The climate in the region is humid. An average daily high temperature in the vegetation period (from May to August) is above 15 řC. The rain period usually lasts for 7 months (from April to November). From September to May, it is snowy (or rain mixed with snow). Throughout the course of the year, the region is generally cloudy (with the clearer periods during the summer time, when the probability of a clear sky is about 20%).

5.2.2 Reference Data

Reference data was previously reported in [Illarionova et al., 2020]. It was collected by field-based measurements carried out in July-August 2018. The methodology of data gathering corresponded to the official Russian inventory regulation [reg, 2012]. In accordance, the study area was split into individual stands with the following characteristics: polygonal coordinates, a certain percentage of each tree species, average age, and height within the stand. The distribution of stand sizes is presented in Figure 5-2. The majority of polygons had their longest side length between 100 and 600 m. Although the percentage for each stand was defined, the spatial distribution within the stand was unknown. The number of individual stands

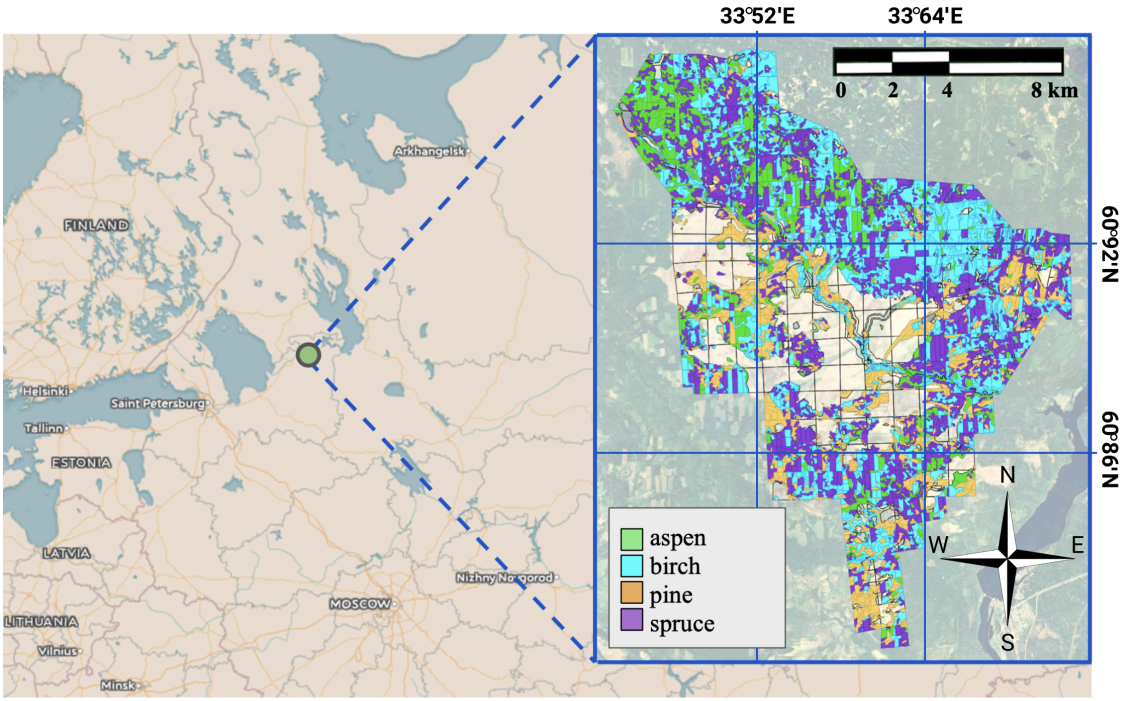


Figure 5-1: Region of interest. Enhanced RGB bands of Sentinel-2 image (tile id is L2A_T36VWN_A010343_20170615T090713) are shown.

Table 5.1: Dataset statistics

	Training Individual stands	Test Individual stands	All Individual stands	Area (ha)
aspen	520	205	725	2298
birch	1143	501	1644	4165
pine	1569	726	2295	3620
spruce	1087	450	1537	6315

with particular dominant tree species (larger than 50% within the stand) is shown in Figure 5-3 and in Table 6.3. The vast majority of individual stands consisted of mixed species; for instance, there were less than 100 stands of pure (not mixed) birch type. Example of mixed individual stands are presented in Figure 5-4.

5.2.3 Satellite Data

For optical multispectral imagery, we acquired Sentinel-2 data. This data is available for free download in L1C format from EarthExplorer USGS. Tiles IDs and acquisition dates are presented in Table 8.1. In this study, we considered only summer images. High cloud cover imposes limits on data for this northern region. There-

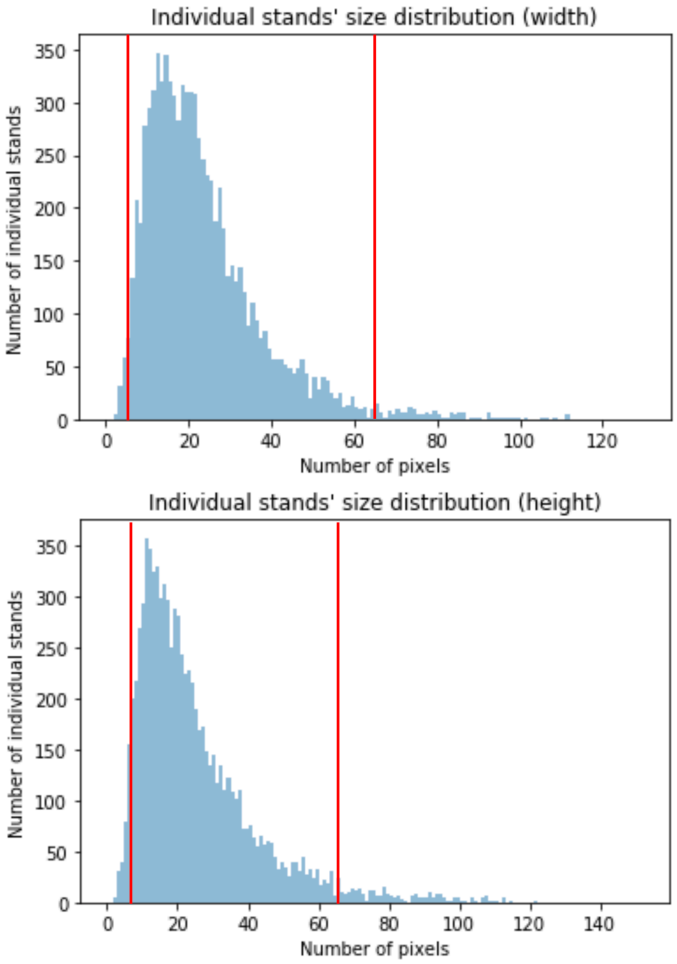


Figure 5-2: Size distribution of individual stands within the study area. Polygons with a side larger than 64 pixels or smaller than 8 pixels were eliminated.

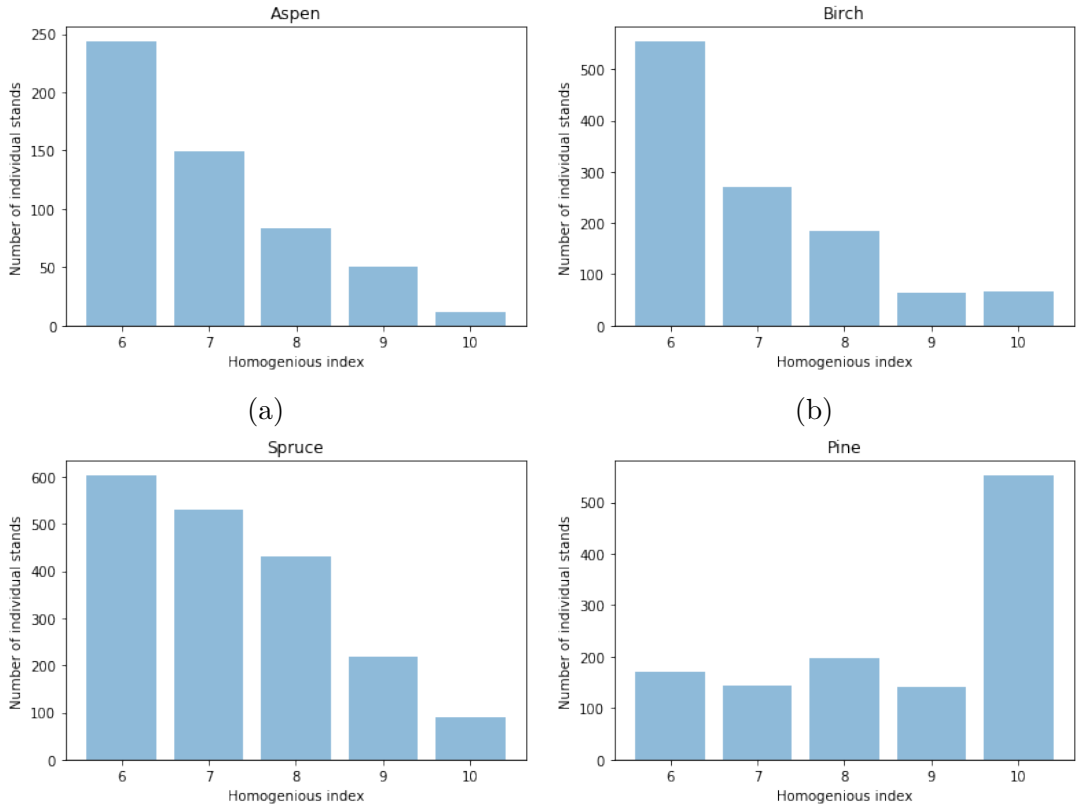


Figure 5-3: Distribution of classes.

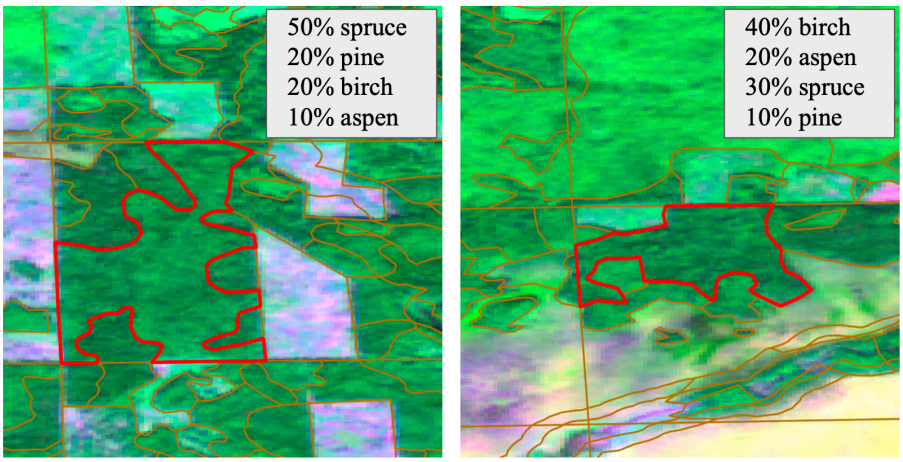


Figure 5-4: Composite of B12, B08, B04 Sentinel-2 bands. Example of mixed individual stands (red polygon) with percentages of species.

fore, only two summer images from different years but of the comparable summer period were used to create the training dataset. Images acquired in other summer dates did not provide a sufficient amount of clear areas without clouds. There were no significant forest cover changes between survey time and image acquisition time; therefore, both images are relevant for the study. 10 bands of the following wavelengths were used: Band 2: Blue, 458–523 nm; Band 3: Green, 543–578 nm; Band 4: Red, 650–680 nm; Band 5: Red-edge I (R-edge I), 698–713 nm; Band 6: Red-edge II (R-edge II), 733–748 nm; Band 7: Red-edge III (R-edge III), 773–793 nm; Band 8: Near infrared (NIR), 785–900 nm; Band 8A: Narrow Near infrared (NNIR), 855–875 nm; Band 11: Shortwave infrared-1 (SWIR1), 1566–1651 nm; Band 12: Shortwave infrared-2 (SWIR2), 2100–2280 nm). Images were pre-processed with the Sen2Cor package for atmospheric correction. Although, Sen2Cor package provides a cloud and shadow map, which can be used to eliminate irrelevant pixels, we selected cloudless images for the study. The obtained data were in L2A format, including values of Bottom-Of-Atmosphere (BOA) reflectances. For CNN-based tasks, image values are often brought to the interval from 0 to 1 [Vaddi and Manoharan, 2020, Debella-Gilo and Gjertsen, 2021]. Therefore, pixel values were mapped to the interval $[0, 1]$ through division by 10000 (the maximum physical surface reflectance value for Sentinel-2 in level L2A) and clipping to 0 and 1. We used bands with a spatial resolution of 10 m per pixel ($B02, B03, B04, B08$ bands) and 20 m per pixel ($B05, B06, B07, B11, B12, B8A$ bands), adjusted to 10 m by Nearest Neighbor interpolation [Persson et al., 2018a]. Each image covered the entire study area, and images were considered separately without any spatial averaging (the same as in [Astola et al., 2019]).

5.2.4 Organizing Samples for Classification

Four tree species were considered: aspen, birch, spruce, and pine. We also considered the 'conifer' class as a combination of spruce and pine, and the 'deciduous' class as a combination of aspen and birch. As a sample for the further analysis, we chose individual stands. There was no information on the spatial distribution of tree species within an individual stand. Therefore, we defined the label for each stand

Table 5.2: Sentinel-2 images from USGS. Wavelength values corresponding to each band: Band 2: Blue, 458-523 nm; Band 3: Green, 543-578 nm; Band 4: Red, 650-680 nm; Band 5: Red-edge I (R-edge I), 698-713 nm; Band 6: Red-edge II (R-edge II), 733-748 nm; Band 7: Red-edge III (R-edge III), 773-793 nm; Band 8: Near infrared (NIR), 785-900 nm; Band 8A: Narrow Near infrared (NNIR), 855-875 nm; Band 11: Shortwave infrared-1 (SWIR1), 1566-1651 nm; Band 12: Shortwave infrared-2 (SWIR2), 2100-2280 nm)

Tile ID	Date	Cloud coverage	10 m bands	20 m bands	Level of processing
L1C_T36VWN_A010343_20170615T090713	2017.06.15	0	2, 3, 4, 8	5, 6, 7, 8A, 11, 12	L1C
L1C_T36VWN_A016206_20180730T090554	2018.07.30	0	2, 3, 4, 8	5, 6, 7, 8A, 11, 12	L1C

as the dominant tree species within it, if the stand contained more than 50% of this forest type (the same approach was described in [Abdollahnejad et al., 2017]). For conifer and deciduous classes, we summed the percentages for spruce and pine, and for aspen and birch, respectively. The described sample definition assumed that the markup had some pre-defined uncertainty for non-homogeneous stands. However, it provided information necessary to the dominant species classification task. Thus, for each sample in the data set, we know the label of the dominant forest type, the percentage of secondary types (if any), and an ascribed polygon in a multispectral satellite image.

For the experiment of training procedure adjustment, we selected 8 test regions of about 450 ha each (Figure 5-5). For the experiment of weak markup improvement, 30% of samples were selected randomly for test. Samples outside test regions were split into train and validation sets randomly, in a ratio of 7:3, following the constraint of no occurrence of the same individual stand in both validation and training sets. For each polygon it can be more than one sample depending on the images' number covering the polygon. Non-overlapping parts of the same satellite image could appear in both the training and test sets.

5.2.5 Forest Species Classification

Instead of typical multi-class classification, we used an hierarchical approach described in [Illarionova et al., 2020]. The task of four-species prediction was split into three tasks: (a) classification of conifer and deciduous; (b) classification of birch and aspen; and (c) classification of spruce and pine. The final results followed from the intersection between the predicted mask of birch and aspen and the predicted deciduous mask (with a similar approach followed in the conifer case). Such an hierarchical approach allows for solving each task independently and ensuring greater control over experiment at each step.

For the forest type classification, we implemented a deep neural network approach, which have been widely used for image classification and segmentation tasks when spatial characteristics are important in the remote sensing domain [Zhang et al., 2019a, Song et al., 2019, Kattenborn et al., 2021b]. At the input of such a

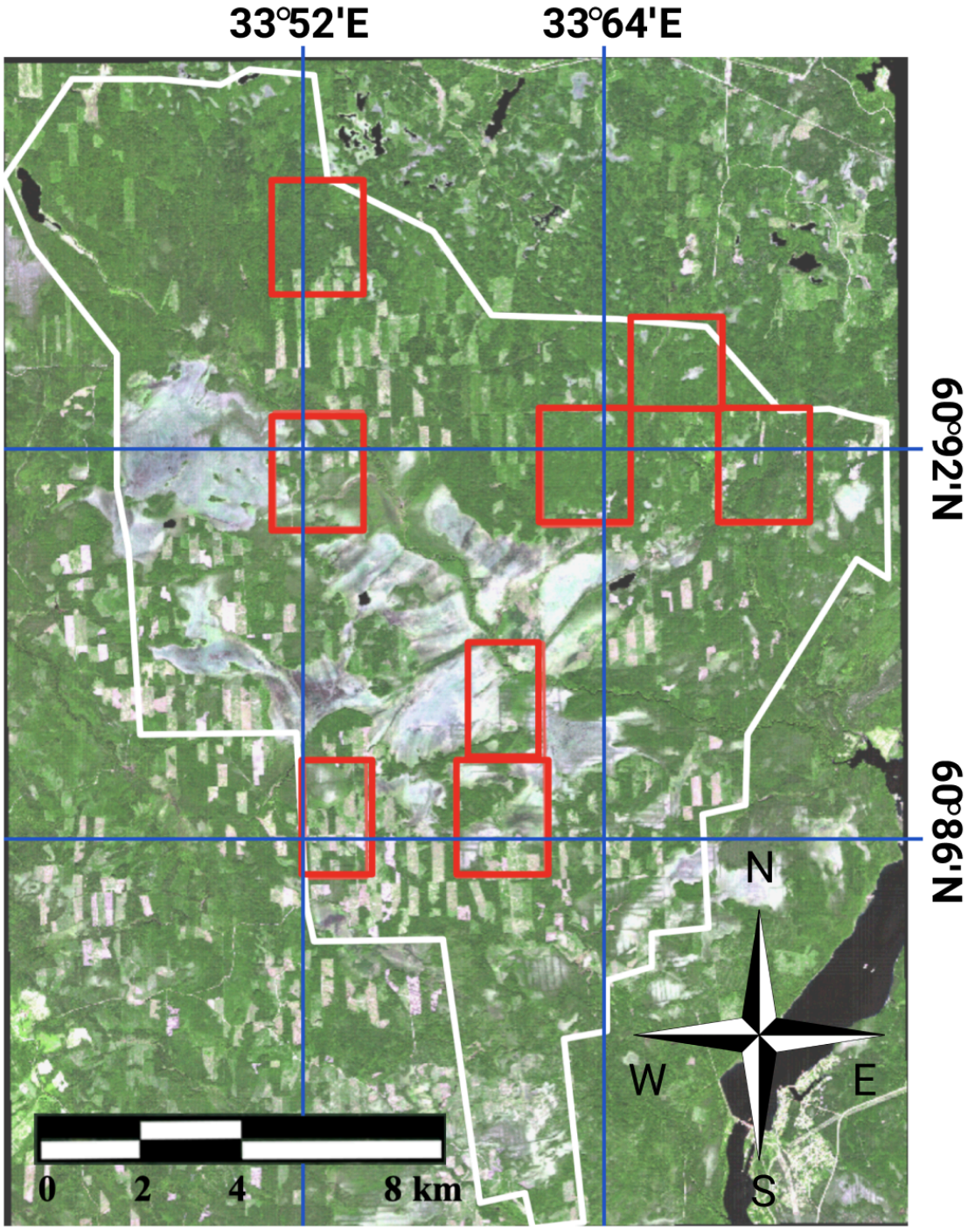


Figure 5-5: The whole study area (white polygon). Test regions (red polygons). Enhanced RGB bands of Sentinel-2 image are shown.

neural network, there is usually a combination of spectral bands. The output of the semantic segmentation model is a map, where each pixel is ascribed a particular class label. During the training procedure, a model is forced to correctly predict as many pixel labels as possible by observing random image patches with pre-defined size. This is achieved through the implementation of a particular loss function. The loss is computed for each step of neural network training, when all images patches from one batch have been processed. For our study, we implemented the categorical cross entropy per-pixel loss function.

In this loss, all pixels in the scene are taken into account. Therefore, if the classes are highly unbalanced, a model rarely observes pixels labeled as the smaller class. This results in poor performance of the model for a less represented class. A common solution is using a larger penalty for errors on the smaller class samples, such as in the weighted categorical cross entropy:

$$\text{Weighted Loss} = -\frac{\sum_{i=1}^{N}\sum_{k=1}^{C}(y_{ik} * \log \hat{y}_{ik}) * weights(y_{ik})}{N}, \tag{5.1}$$

where $\hat{y}_{ik}$—predicted probability of the i-th pixel to belong to the k-th class, y_{ik}—ground truth value for the i-th pixel (1 if the pixel belongs to the k-th class), N—number of not masked pixels, C—number of classes.

Another issue that should be taken into account is that samples of particular classes may not be evenly distributed across the study region. This means that random selection of image patches in batch can lead to a situation where samples concentrated in one area may be seldom observed.

To tackle this problem, we modified the classical sampling approach for semantic segmentation with CNN, as described in the next section.

5.2.6 Object-Wise Sampling Approach

We replaced the commonly used batch creation approach. The sample content was taken into account, instead of simply using random patch selection. The choice of patch size was governed by the relevant size of polygons. As we eliminated polygons with sides less than 80 m and larger than 640 m, the patch size was selected as

$64 * 64$ pixels. The number of patches per batch was set to 128. Although we considered two classes, the general approach is also applicable for more classes. For each class, we picked the same number of polygons and cut patches around these polygons to create the batch. As the polygon size could vary in the defined range, the patch crop could also differ for the same polygon. The only demand was that the polygon's bounding box should be within the patch boundary. The patch was also geometrically augmented, in order to provide more variability during the training procedure. We implemented random rotate, mirror, and flip operations.

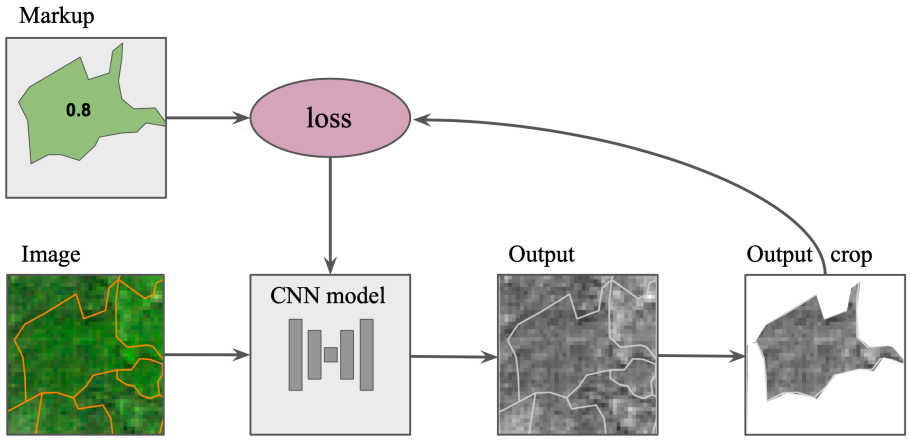


Figure 5-6: The object-wise semantic segmentation approach. The model produces the map where the probability of a class is recorded at each pixel. Loss is computed just for masked area of the polygon. The percentage of dominant class is also can be taken into consideration (in the example, the dominant species percentage for the individual stand is 0.8).

The next step was loss computation. The approach is described in the Figure 5-6. For this purpose, we used polygon mask. Patch has dimension $Patch\_Rows$, $Patch\_Columns$, $Number\_of\_classes$. The patch mask contains non-zero values for pixels within the polygon's area and for the appropriate correct class. Despite the fact that individual stands are not often homogeneous, all pixels within one stand were ascribed the same label. The loss was computed for this area. There can be an available markup for other pixels within the patch, but this was not considered. The main reason for this is that it can affect the balance of classes.

We compared this approach with the commonly used per-pixel semantic segmentation approach, for which the batch was randomly formed and an extra penalty for mistakes in the smaller class was added (Figure 5-7). In this approach, for calculation of the weighted categorical cross-entropy loss, all pixels within the patch were considered. The weights were set proportionally to the amount of each class represented.

5.2.7 Weak Markup

Another adjustment was aimed at addressing weak markup. It includes two stages, as shown in Figure 8-3. The first stage was as follows. The aforementioned reference

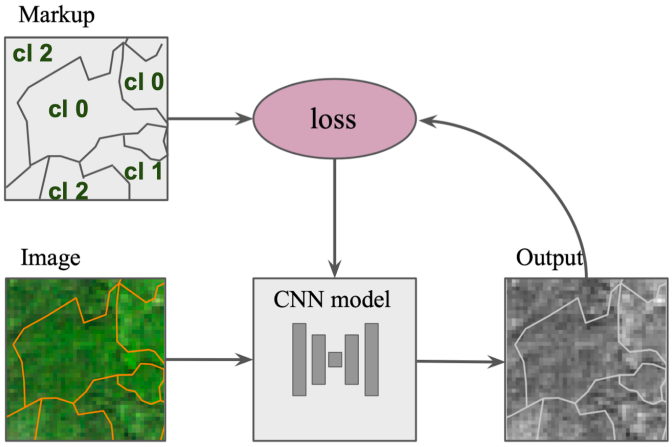


Figure 5-7: The commonly used per-pixel semantic segmentation approach. The model produces the map where the probability of a class is recorded at each pixel. Loss is computed for the entire patch. The patch includes stands with different dominant species (class 0, class 1, etc.)

data consisted of the percentage of each class within the individual stand. We took this knowledge during the loss computation. The loss was calculated for each individual stand and multiplied by the dominant species percentage. For example, for a stand that consisted 60% of conifers and 40% of deciduous trees, the penalty will be 0.6. If the percentage is higher, then the penalty becomes stricter. For a homogeneous stand, all pixels have the maximum loss weights. Therefore, in Equation 5.1, *weights* were defined as the dominant species percentage. When the learning curve started to change less rapidly and could achieve sufficient results on the validation set (after about 15 epochs), we changed the training data set. We eliminated all individual stands with percentage less than 90%. Thus, for a few epochs (about 2 epochs), the model observes more pure data. Obviously, such a model will perform poorly, in terms of the initial dominant species problem statement. However, at the same time, it will not strive to label deciduous trees within a conifer individual stand as conifer trees (as for case with 60% conifer and 40% deciduous). Then, we used this model to predict conifer and deciduous species both for training and validation regions. The first stage of markup adjustment results was the intersection between the predicted mask and initial dominant species markup. It was assumed that the map acquired in this way contained less pixels of minor (i.e., non-dominant) classes.

The next stage of the weak markup study was the implementation of the newly

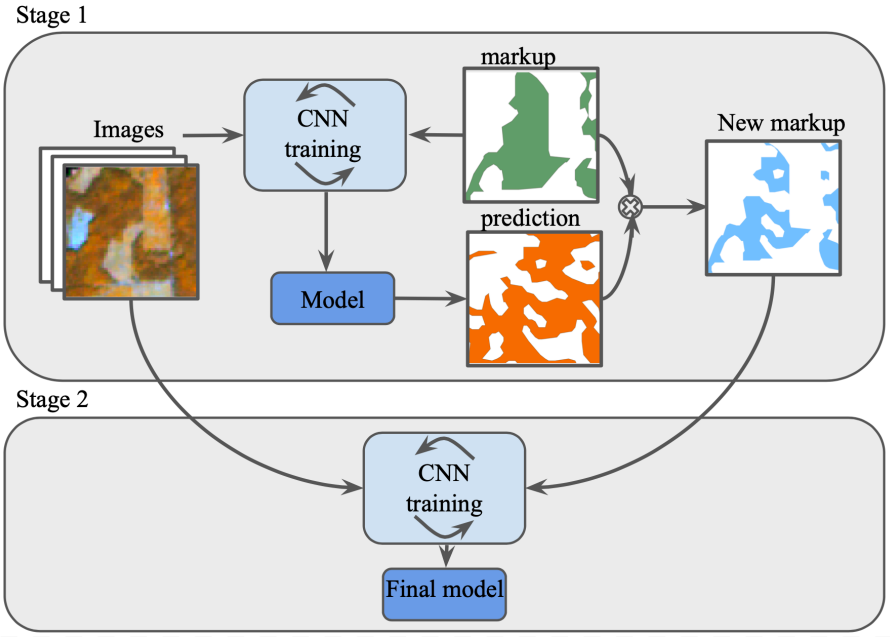


Figure 5-8: Markup adjustment strategy.

obtained markup in further training. We intersected the new conifer mask with the initial spruce and pine dominant masks, and the same for the deciduous classes. The goal of this intersection was to reduce the number of deciduous pixels within individual stands dominated by pine and spruce, and vice versa. For this study, we created the validation data set only from homogeneous individual stands.

5.2.8 Experimental Setup

For all experiments, the U-Net architecture [Ronneberger et al., 2015] with ResNet [He et al., 2016] encoder was used, as it has been shown to successfully perform in popular image classification tasks both in general and remote sensing domains [Kattenborn et al., 2021b]. The model implementation referred to [Yakubovskiy, 2022]. It used Keras with Tensorflow backend. For model training, a PC with GTX-1080Ti GPUs was used. The batch size was 128 patches, where each patch had size of 64∗64 pixels. The batch size was chosen accord-ing to GPU memory limitations. There were 100–200 steps per epoch and the number of epochs varied from 10 to 30, depending on the size of classes. Similar results reproduction was achieved by fixing a random seed for pseudo-random number generator for all training methods.

To assess the classification quality, we considered F1-score. In the one case, we

evaluated the number of correctly predicted individual stands. To this end, per-pixel predictions within stands were aggregated and the dominant class was defined for each stand. Based on reference and predicted stand labels, the amounts of true positive, false positive, and false negative samples were estimated. In the second case, we evaluated the F1-score in a per-pixel manner.

A CNN model for each experiment was trained five times with different random seeds, and then results were averaged. Standard deviation was computed.

5.3 Results

5.3.1 Sampling Approach For Species Classification

We compared a typical sampling procedure for forest species semantic segmentation with our modified one. The results are presented in Table 5.3. For all classes, the object-wise sampling approach performed better. The average F1-score before aggregation was improved from 0.8 to 0.85. The final aggregated results were obtained by multiplying the predicted conifer binary mask with spruce and pine masks and multiplying the predicted deciduous mask with aspen and birch masks. Aggregated results for four forest classes are shown in Table 5.6. The object-wise sampling approach allows us to improve segmentation's F1-score from 0.68 to 0.74. The larger difference between the two methods was for the birch and aspen classes. The reason for this is that these classes were the most difficult to distinguish due to imbalance. The proposed approach leads to a more balanced training samples choice.

Standard deviation was computed for averaged F1-score of different model training running. It shows that achieved results are relevant for further forest analysis studies.

Table 5.3: Forest types classification using different sampling procedure (per-pixel F1-score)

	aspen / birch	pine / spruce	conifer / deciduous	average
Simple sampling	0.48 / 0.88	0.91 / 0.88	0.81 / 0.85	0.8 $\pm$ 0.003
Modified sampling	0.63 / 0.91	0.94 / 0.88	0.85 / 0.87	0.85 $\pm$ 0.004

Table 5.4: Conifer and deciduous classification (average score) using source markup and updated markup.

	Per-pixel F1-score	Per-stand F1-score
Source markup	0.827	0.851
Updated markup	0.769	0.854

5.3.2 Markup Adjustment

We conducted experiments aimed to improve conifer and deciduous markup. Some areas were eliminated by the model predictions intersected with the initial dominant species map. It aims to leave only homogeneous areas with conifer or deciduous trees. The per-pixel metric is intended to label all pixels even within inhomogeneous individual stand as the dominant class type. Therefore, at this stage of the task, the goal was not to improve the per-pixel score. The average per-pixel F1-score for conifer and deciduous classification became 0.76, in comparison with the previously achieved 0.82 (Table 5.4). However, we aimed to preserve the score per individual stands than the per-pixel one. The score of dominant classification per individual stands was still approximately at the same level (F1-score 0.85). It means that the model was trained to ignore pixels of non-dominant classes within the stand. For the further assessment, homogeneous stands were considered.

The obtained map was then used for species classification. We compared the model trained on the source markup and that trained on updated one. Their performances were assessed on homogeneous individual stands for four species from the test set. The results are presented in Table 5.5. Although we eliminated pixels from the (non-homogeneous) training set, the model performed better than when using the larger training data of weaker quality. It allowed us to improve the average F1-score for four species from 0.74 to 0.76 compared with initial markup usage (Table 5.6). The results confirmed the benefit of the proposed approach.

Example of the final predictions using both modified sampling approach and adjusted markup is presented in Figure 5-9.

Table 5.5: Forest types classification for more homogeneous individual stands (per-pixel F1-metric) using source markup and updated markup. Results on test samples.

	aspen/ birch	pine / spruce	average
Source markup	0.77 / 0.9	0.94 / 0.88	0.87 ± 0.003
Updated markup	0.79 / 0.91	0.95 / 0.9	0.89 ± 0.002

Table 5.6: Final aggregated results for forest types classification using modified sampling procedure and markup adjustment (F1-score)

	aspen	birch	pine	spruce	average
Simple sampling procedure	0.42	0.72	0.84	0.74	0.68 ± 0.007
Modified sampling procedure	0.6	0.8	0.81	0.74	0.74 ± 0.004
Modified sampling procedure with new markup	0.62	0.83	0.82	0.76	0.76 ± 0.005

5.4 Discussion

5.4.1 Sampling Approach for Species Classification

The analysis showed that the sampling procedure is highly essential for the forest species classification task. The same approach can be implemented for other problems where maps of vast territories are used and some classes are distributed not uniformly. The proposed object-wise sampling approach for CNN leads to better results than the commonly used approach where patches are chosen randomly within the entire satellite image.

It is worth mentioning the reason why a classical patch-wise approach was not considered suitable for our problem. It implies that we can choose the patch size small enough to include just the pixels of one class. However, in our case, there are two obstacles to implement this. The first being that individual stands are not of rectangular shape and, therefore, the patch size must be rather small. The other point is that individual stands are not homogeneous and we do not know the spatial distributions within stands. Therefore, a random small patch within an individual stand may turn out to, in fact, be a set of pixels of a minor class. This makes the approach described in [Mahdianpari et al., 2018] inappropriate in the presented case.

Another alternative approach to classical pixel- and object-wise classification for

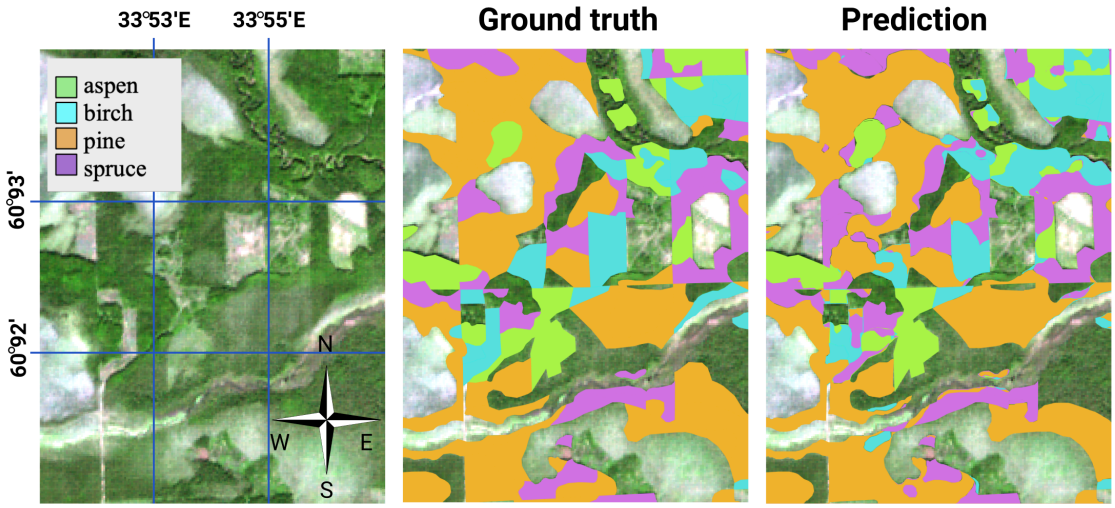


Figure 5-9: Sentinel-2 RGB image. Final predictions using modified sampling approach and adjustment markup.

remote sensing applications (e.g., airplane-based) has been discussed in [Sun et al., 2019]. It should be noted that, despite the apparent similarity of airplane and satellite-derived remote sensing data, they have substantive differences. The main difference is spatial resolution. The relevant observation field can vary by 100 times (e.g., 0.1 m for UAV and 10 m for satellite images). Thus, the approach have to be modified.

5.4.2 Markup Adjustment

Clear markup is essential for remote sensing tasks. In some cases, non-homogeneous areas are excluded from training set [Knauer et al., 2019]. Another approach is to use plots with different species and ascribed it by the dominant species class [Abdollahnejad et al., 2017]. It is reasonable to move further in the direction of an automatic markup adjustment, in order to make the data clearer without extra manual labeling. The next step of the study can be label adjustment for all classes, not only for conifer and deciduous. The weighted loss function adjustment can also be considered to improve homogeneous areas detection.

Weakly supervised learning is now applied in different remote sensing tasks. They vary by the target objects and remote sensing data properties such as spatial resolution and spectral bands number. In our study, we focus on 10 m spatial

resolution and 10 multispectral bands. In cases of very high spatial resolution and just RGB bands such as in [Qiao et al., 2020] markup constraints differ significantly. Particular tasks also pose some limitations and additional opportunities for a weakly supervised learning approach [Li et al., 2020c]. Therefore, remote sensing datasets can differ drastically from such datasets as MNIST or CIFAR considered in [Han et al., 2018]. Another difference is that the forest species classification is considered as a semantic segmentation task instead of an image classification task, such as in the case of noisy labels problem in [Han et al., 2018].

Markup adjustment can be also studied in the case with machine learning algorithms instead of neural network based such as methods described in [Xia et al., 2020, Ha et al., 2020, Pham et al., 2018].

The main error source in such land cover tasks is diversity within each forest species. Spectral characteristics vary drastically for different tree age and depend on environmental conditions. Therefore, markup adjustment and optimal sampling choice are promising approaches to improve model performance. Another error source is mixed border pixels of neighboring individual stands. In the case of 10 m spatial resolution, even for homogeneous forest stands, spectral characteristics on the border can be affected by other species outside this stand. A possible approach to address this problem for homogeneous stands is to consider just inner pixels remote from the border.

One of the potential limitations is the time and computational cost for markup adjustment model training. In our case, we used the same CNN architecture to perform this stage. We trained the model for markup adjustment and the final segmentation model sequentially. In future studies, an alternative approach can be developed and implemented to perform markup adjustment on the fly for remote sensing tasks.

In this study, we considered forest species classification. However, the proposed approach can be transferred in future studies for other tasks where samples are grouped, and for a group, label distribution is known. The described approach is also applicable for other neural network architectures. Therefore, experiments with new state-of-the-art architectures can be conducted using the same method. Both the

sampling and markup adjustment approaches are transferable to new satellite data sources. We considered multispectral Sentinel-2 imagery with a spatial resolution of 10 m. However, it can also be implemented for high-resolution multispectral data such as WorldView or just RGB images such as base maps.

Vegetation indices are significant for environmental tasks as they provide relevant surface characteristics. Therefore, they are widely used as features for classical machine learning methods. However, in the case of deep neural networks, it is assumed that neural networks can learn non-linear connections between raw input data and use prior information for more general characteristics extraction. In our study, we considered only multispectral satellite bands. However, future studies might include vegetation indices or supplementary materials such as digital elevation or canopy height models to achieve higher results and reduce training time.

It is promising to study different augmentation techniques combined with improved markup and the object-wise sampling approach. For example, the object-based augmentation described in [Illarionova et al., 2021b] can further be implemented to create more variable training samples with different homogeneous stands.

Precise forest species classification can also be implemented in ecological and environmental studies, as large forest patches have been proved to affect human health [Kim et al., 2020]. Detailed forest characteristics can be helpful for such analysis.

5.5 Conclusions

The sampling approach and ground truth markup quality are crucial in forestry tasks involving remote sensing data. In this study, we analyzed the potential of combining CNN and Sentinel-2 images for the task of forest species classification using weak markup with non-homogeneous individual stands. During the first stage, a CNN was trained to find the homogeneous areas within each stand, providing a more accurate markup. During the second stage, the final model was trained to predict four forest species. This markup adjustment allows us to increase F1-score from 0.74 to 0.76 compared to the initial markup. The experiment confirms the opportunity of finding

weak labels and shows promising results for further classification enhancement. We also proposed the CNN-based sampling approach for spatial data in forest species classification. The proposed modification outperformed the prediction quality of a commonly used per-pixel semantic segmentation model (the average F1-metric was increased from 0.68 to 0.74). The described pipeline helps to address the issue of highly imbalanced and not evenly distributed classes. The provided training strategy can help solve forest species classification tasks more precisely, even when the reference data has significant limitations. Further study for other vast territories is promising, and the proposed sampling technique seems to be beneficial in such spatial studies.

Chapter 6

Estimation of the Canopy Height Model from Multispectral Satellite Imagery with Convolutional Neural Networks

6.1 Introduction

Canopy height model (CHM) estimation has a long history, but advances in computer vision and satellite sensing technologies have opened new opportunities in this area. The height can be effectively utilized in different applications and broadens the surface's two-dimensional representation in the visible spectrum. There are both natural [Thomas et al., 2018, Trier et al., 2018, Huang et al., 2017, Zhang et al., 2018, Illarionova et al., 2020] and anthropogenic [Mou and Zhu, 2018, Trekin et al., 2020] objects of landcover to be explored. The present study is focused on natural types of landcover, especially wild forest areas. Landcover height characteristics can be used in various applications, such as biomass evaluation [Wu et al., 2016, Sadeghi et al., 2018, Gwenzi et al., 2017], improving the accuracy of tree species classification [Sasaki et al., 2012, Dalponte et al., 2012], and correlated vegetation properties extraction [Majasalmi et al., 2018].

There are three frequently reported sources of canopy height information: 1) field-based measurements; 2) Unmanned Aerial Vehicle (UAV)-based approaches; and 3) satellite remote sensing data. All aforementioned approaches have advantages and limitations connected with acquisition time and cost (Figure 6-1). The first data source is forest inventory documents, usually treated as field-based observations. They are available for some regions and useful in addressing forest owners', governmental, and independent organizations' needs [Venturas et al., 2021]. However, these data do not cover all regions of practical interest [Haakana et al., 2017]. Furthermore, such data actualization is time-consuming and cost intensive in difficult-to-access areas. An alternative approach is to use remote sensing data.

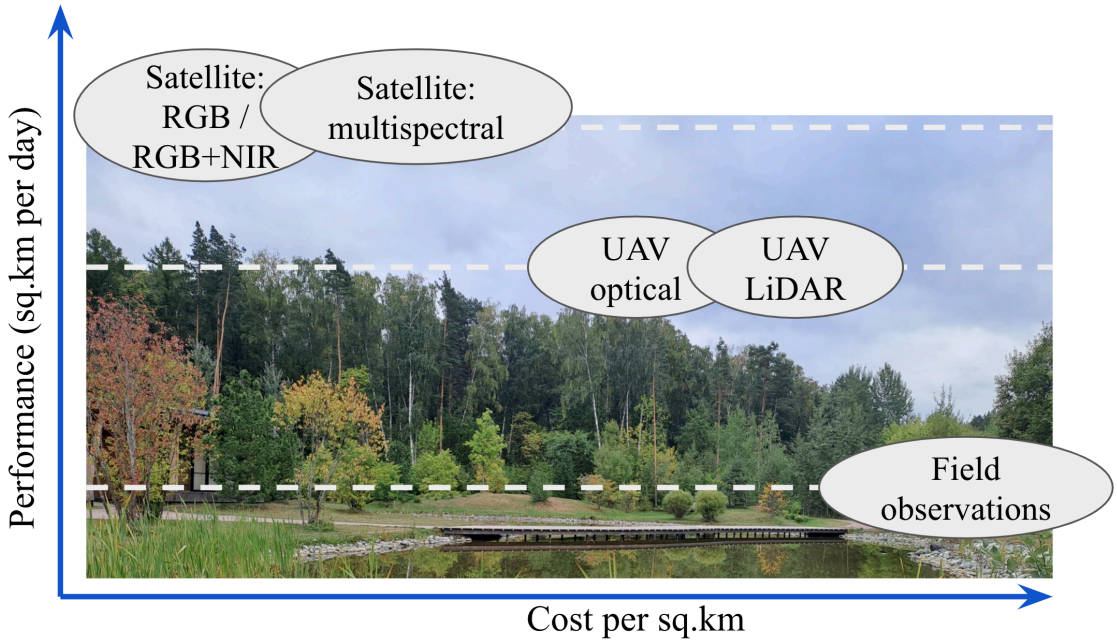


Figure 6-1: Cost comparison of different forest height measurement approaches (diagram is not to scale).

The remote sensing approach draws on both active and passive sensing technologies. During active sensing such as Light Detection and Ranging (LiDAR) measurements, the sensor measures time between the emitted light time and its return time to estimate the distance of an object (a surface). This technology allows digital elevation models to be produced. Passive remote sensing measures radiation that is emitted or reflected by the object in different spectral wavelengths. Spectral bands obtained this way can be used for future analysis and to calculate the height

value in landcover extraction.

A common approach builds on UAV assessment. A UAV with LiDAR sensors is a powerful tool for forest height estimation. It obtains canopy height data with minor errors, meeting the precision requirements for almost all forestry tasks. However, such equipment is more expensive than a spectral aerial camera system, thus there remains the challenge of obtaining the same information using low-cost methods [Matese et al., 2017]. Many works use LiDAR data as a reference and aim to find a cheaper height data source. A detailed review of the alternative approaches to LiDAR sensing is presented in [Stone et al., 2016, Lagomasino et al., 2016]. Thus, most of the current studies in the sphere of canopy height estimation use UAVs with optical aerial systems [Hartling et al., 2019, Marrs and Ni-Meister, 2019, Lin et al., 2019, Nezami et al., 2020, Nguyen et al., 2019, Swinfield et al., 2019]. Despite its advantages over field-based observations, when large regions have to be processed, the labor involved in working with vast and remote areas is problematic. Satellite data address this issue, providing a cheaper option for forest monitoring [Stone et al., 2016]. Point cloud data that is useful for estimation of the canopy height can also be derived from satellite imagery using photogrammetry approach. The comparison of such photogrammetry approach and high-density LiDAR measurements is presented in [Pearse et al., 2018], where authors showed photogrammetry method is slightly less accurate (difference in R^2 is about 0.07) compare to the LiDAR method for height measurements of the forest region in New Zealand. The important benefit of the photogrammetry method is that it could provide information for the larger scale compare to the LiDAR method, however it requires special high resolution imagery which is not always available for the particular region. The other limitation of the photogrammetric method is that it is able to characterise only the upper canopy and is not able to perform vertical characterisation of the forest such as can be done by laser scanning. The comparison of the photogremmetry obtained by unmanned aerial systems and areal laser scanning for the forest inventory in Oregon was presented in [Fankhauser et al., 2018], where authors stated that photogrammetry is slightly less accurate compare to laser scanning (difference in R^2 for height estimation is about 0.15). However, photogrammetry is easier to integrate to existing

forest monitoring methodologies.

Our work is focused on using satellite images for CHM estimation as it is more preferable data source than LiDAR derived measurements in terms of cost and spatial coverage. Neural networks allows us to conduct image processing automatically. We set up the hypothesis that neural networks can extract significant spatial features from very high-resolution RGB images of 1 m to improve performance of CHM estimation. It was expected that developing a satellite-based solution compatible with a high-resolution UAV approach would further enable the prediction of advanced forest characteristics. Thus, this study's objectives and contributions were:

1. to develop a method for vegetation height estimation utilizing deep neural networks and different configurations of input data varying spectral compound (reducing to Blue, Green, and Red), spatial resolution and by adding topography features;

2. to assess the generated height map, conducting a further investigation into the classification of dominant forest species (conifer and deciduous). For this, multispectral imagery was incorporated with height data;

3. to create the software toolchain to train a neural network to predict CHM using single satellite non-stereo imagery.

In the Chapter on forest species classification using the neural-based hierarchical approach, we showed how height measurements can adjust model performance in the forest species classification task. In this Chapter, we extend the previous findings and consider a case where LiDAR measurements are not available. The proposed method for canopy height estimation is useful for forestry and ecological applications and it also can be implemented in the aforementioned forest classification pipeline to achieve higher performance using just satellite data.

6.2 Related work

For canopy height estimation studies, spectral satellite imagery can be distinguished by the following characteristics: spatial resolution, spectral range, and availability.

The majority of works use a spatial resolution much higher than 20 m to tackle the canopy height evaluation problem. This approach is justified for particular tasks when large-scale maps are produced. In [Staben et al., 2018], they conducted a 30 m spatial resolution canopy height evaluation with Landsat imagery and showed the dynamics over 29 years in the Darwin region. In [Hansen et al., 2016], they employed Landsat 7 and 8 time-series data (30 m spatial resolution) to estimate tree heights in Africa. GLAS (Geoscience Laser Altimeter System) height measurements from the ICESat satellite were used as reference data ($60-70$ m spatial resolution). The same height data source was mentioned in [Tao et al., 2016]. In [Ghosh et al., 2020], they used Sentinel-2 images that were resampled to a 20 m pixel size to predict Mangrove forest canopy height. Other studies involving Sentinel-2 data are reported in [Verma et al., 2016, Lang et al., 2019, Puliti et al., 2020a]. In [Lee and Lee, 2018], they assessed SAR images from ALOS PALSAR, and upsampled them from 30 to 5 m as a LiDAR elevation model. The cases of very high spatial resolution (3.7 m) images from the Planet Dove implementation are presented in [Csillik et al., 2020]. However, the target height map spatial resolution for that study was 1 hectare. Very high spatial resolution (2 m) WorldView-2 satellite imagery was used in [Meddens et al., 2018], but the working spatial resolution was adjusted to 5 m.

Another important data characteristic is the spectral range and the number of channels. A wider wavelength range is available for satellites with low spatial resolutions (Landsat, Sentinel) than for some very high spatial resolution satellites. For instance, Planet (3–5 m spatial resolution) and GeoEye (2 m spatial resolution) satellites have Blue, Green, Red, and NIR bands; RapidEye (6 m spatial resolution) has Red Edge. The GeoEye panchromatic channel has a 0.4 m spatial resolution and allows RGB to be enhanced. WorldView-2 provides eight spectral bands with a spatial resolution of 2 m. An additional source of very high remote sensing data is Basemaps, with RGB bands such as those provided by Maxar one. Nevertheless, the majority of works focus on using only the wide multispectral range (more than eight bands), sacrificing the spatial resolution. From the aforementioned satellite-based studies, the minimal number of spectral bands (Blue, Green, Red, NIR) was only considered in [Csillik et al., 2020]. However, the goal of the work was the creation

of a large-scale country wide map, so the spatial resolution of the analysis was 1 hectare. Therefore, the issue of minimizing the number of required satellite bands for forest height estimation has not yet been well studied.

Satellite data are frequently accompanied with data of other sensing techniques. In [Lee et al., 2020a], they combined four Kompsat-3 multispectral bands and PALSAR-1 radar images resampled into 2.8 m to train a neural network. Few studies have implemented this into self-contained spectral satellite data [Ni et al., 2019, Lee et al., 2020b, Shah et al., 2020, Lang et al., 2019]. However, the spatial resolution of the Sentinel and Landsat images (lower than 10 m) considered in these studies is not high enough to extract small details on the surface. Thus, the satellite spatial resolution of 1-m per pixel is still beyond the scope of the majority of studies.

Data availability is also a significant aspect of implementation in practice. Sentinel and Landsat images are available free of cost, while WorldView, Planet, and RapidEye are commercial and contain a greater amount of the spatial information required in applied tasks.

After data acquisition, the obvious question of data processing arises. Computer vision algorithms enable high-quality automatic satellite imagery analysis. Such methods are usually based on key feature extraction from input spectral bands to describe some object, which can be a pixel or set of pixels. Then, the algorithm aims to ascribe a label (for classification tasks) or a value (for regression tasks) to the object. The processing methods for expansive forestry areas using satellite images are classical machine learning models, such as Random Forest [Breiman, 2001] or Support Vector Machine [Cortes and Vapnik, 1995]. Their main advantages are simplicity and straightforward interpretation in the case of linear models. Generally, spatial characteristics are not taken into consideration, and an algorithm relies on spectral values or precalculated vegetation indices. In [Staben et al., 2018], a combination of 14 vegetation indices and spectral bands were used in the Random Forest model to predict the canopy height using Landsat images. Moreover, the strong correlation between the normalized difference vegetation index (NDVI) and canopy height has been well emphasized in aerial photography [Matese et al., 2017, Lee and Lee, 2018]. Despite the importance of spectral data, other vital features can

be also processed. For instance, there is a strong correlation between forest height and canopy width, as discussed in [Verma et al., 2016], in which the canopy volume was estimated using only the crown projected area and the crown diameter combined in a particular regression equation. The deep neural network-based approach is more capable than classical machine learning methods for the following reasons: the texture and spatial features extracted by the neural networks include sufficient information about landcover; it not only handles spectral values, but also the aforementioned spatial characteristics of an object available, for instance, in UAV-based tasks [He et al., 2019a].

Tree height is correlated with tree diameter for each forest species [Özçelik et al., 2018]. In [Sharma et al., 2019], tree height was estimated from the exponential equation, including diameter at breast height value. The crown form depends on the tree species; accompanied by the crown diameter, it can provide important features for a neural network. Tree height can also be derived from spectral information only, as it depicts meaningful vegetation characteristics such as chlorophyll content [Rahimzadeh-Bajgiran et al., 2012].

6.3 Materials and methods

6.3.1 Study area

The study area is located in the Arkhangelsk region of northern European Russia with coordinates between 45°16′ and 45°89′ longitude and between 61°31′ and 61°57′ latitude (Figure 8-1). The investigated territory belongs to the middle boreal zone. The region's climate is humid, with the warmest month being July when the temperature rises to 17°C. The topography is flat, with a height difference in a range between 170 and 215 m above sea level [Aakala et al., 2011]. The main species present in the region are pine, spruce, aspen, and birch.

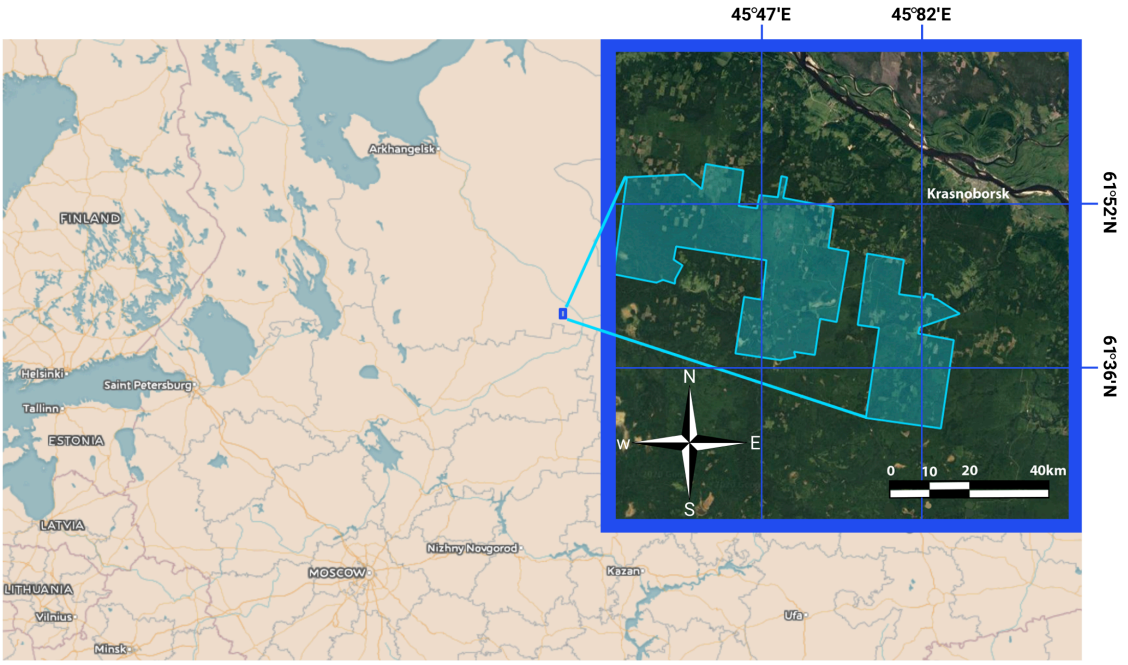


Figure 6-2: Region of interest.

6.3.2 Reference data

We used forest inventory and LiDAR-derived data covering the area of about 50 thousand hectares. LiDAR measurements were continued in the end of August of 2017 and 2018 by Leica ALS 80 HP scanner. Then the Canopy Height Model (CHM) with a 1 m spatial resolution was generated from LiDAR-derived point clouds.

The inventory data were collected in accordance with the official Russian inventory regulation in 2018 and 2019. It included such characteristics as canopy height, species percentage distribution, and age. This data was organized as a set of individual stand coordinates with appropriate characteristics based on the assumption that the crop was homogeneous. A species class markup was used in additional experiments presented as a raster map of dominant conifer and deciduous classes. The statistics of this data are shown in Table 6.3.

However, the shift in geo-referencing between the satellite data and LiDAR-derived measurements makes the target at 1 m spatial resolution less useful. As the typical shift lies between 2 and 3 m, the high spatial resolution CHM will show erroneous value for the particular point in the satellite image. This forced us to downsample the height map to 5 m to make the target value for each point represent

Table 6.1: WorldView images.

	Image ID	Date	Off-nadir angle
0	1030010056130F00	05.30.16	14
1	103001005683F200	05.30.16	14
2	1030010031934700	06.08.14	7
3	1030010032660800	06.08.14	7

Table 6.2: Sentinel images.

	Image ID	Date
0	L2A_T38VNP_A005695_20160725T082012	07.25.16
1	L2A_T38VNP_A007297_20180730T081559	07.30.18
2	L2A_T38VNP_A010986_20170730T082009	07.30.17
3	L2A_T38VNP_A013017_20190903T081606	09.03.17
4	L2A_T38VNP_A015748_20180628T082602	06.28.18
5	L2A_T38VNP_A016606_20180827T083208	08.27.18

the mean value of the area including the true location.

The distribution of the height over the study region is shown in the Figure 6-4. Although, height is usually represented as a continues value, height categories are essential for practical use in power lines services. Height classes are often required instead of continues values for decision making within protected areas [Wanik et al., 2017]. The reason is that different categories (dangerous vegetation overgrowing) have different importance and estimation in particular categories have to be more precise to reduce accidents on power lines corridors.

6.3.3 The test region selection

The training and test area was from the same satellite images, but without overlapping. The test region was manually chosen to include a diversity of height classes. The total test area was equal to 13% of the initial dataset. The spatial location is presented in the Fig 6-3. The height distribution through the test areas is presented

Table 6.3: Dataset statistics for conifer and deciduous classification.

	Training (individual stands)	Testing (individual stands)	Full dataset
Conifer	1219	534	25913 hectares
Deciduous	756	341	24397 hectares

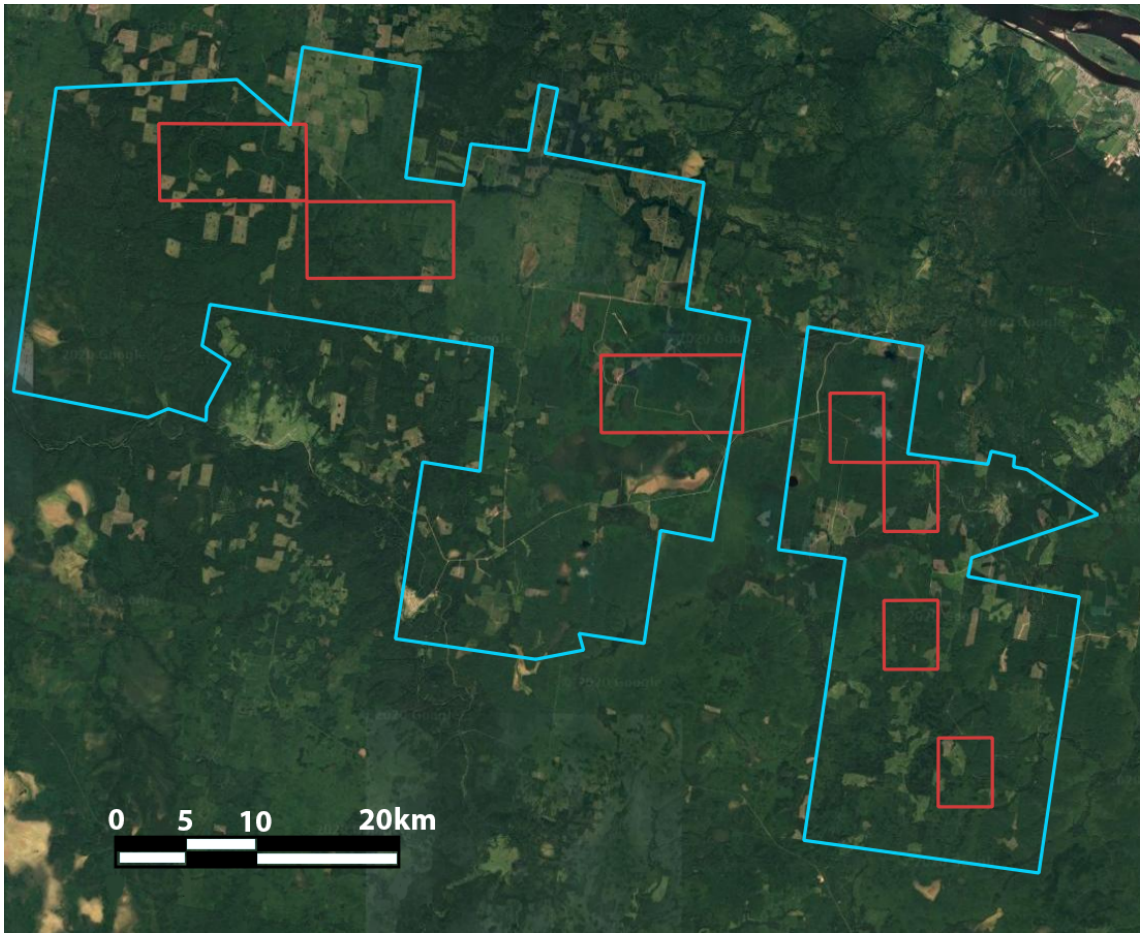


Figure 6-3: The blue lines define the study area with LiDAR measurements. The red squares are the test regions.

in the Fig 6-4.

6.3.4 Satellite data

We used Sentinel-2 and WorldView-2 satellite imagery to check the high and very high spatial resolution data sources. The boreal location of the study area resulted in a lack of cloudless images. All images were from the boreal growing season (from May to August). Image IDs and dates are presented in tables 6.1, 8.1. WorldView imagery was downloaded from GBDX. For the height estimation task, we used Red, Green, Blue, and Near-Infrared bands, while for the species classification problem, all eight bands were considered. The resolution of the WorldView images was 1, 2, or 5 m depending on the experiment setup. For CNN-based tasks, image values in the range from 0 to 1 are usually used [Vaddi and Manoharan, 2020, Debella-Gilo

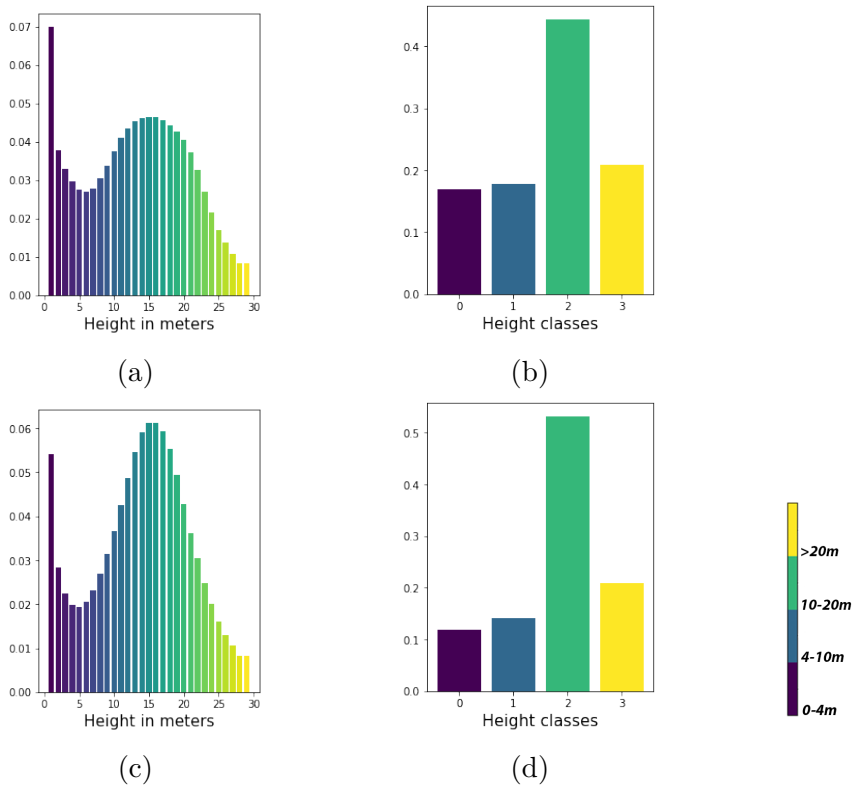


Figure 6-4: Reference LiDAR-derived height (Canopy Height Model (CHM) values) distribution for the study area. (a-b) Training dataset. (c-d) Test dataset. These height categories are the important ones for power lines services in Russia.

and Gjertsen, 2021]. Therefore, pixel values were brought into a range between 0 and 1 using Equation 8.4:

$$m = max(0, mean(I) - 3 * std(I)), \tag{6.1}$$

$$M = min(max(I), mean(I) + 3 * std(I)), \tag{6.2}$$

$$I' = (I - m)/(M - m), \tag{6.3}$$

where $mean, std$ are the mean and standard deviation of the image. In equations 8.2, 8.3, we calculate m and M (minimum and maximum of the preserved dynamic range). The standardization of the imagery according to the whole dataset statistics proves profitable for the neural network training compared to a simple scaling of the entire value range [Pal and Sudeep, 2016].

For the spatial resolution adjustment, the pansharpening procedure was implemented using a panchromatic band which was obtained in the imagery bundle with multispectral data from the data vendor. We did not consider any predefined cloud mask for WorldView. However, during training, pixels with particular properties were eliminated from consideration (see subsection 6.3.7). This allowed us to clean the dataset from erroneous labels.

For the additional analysis, freely available Sentinel-2 data were downloaded in L1C format from EarthExplorer USGS and preprocessed using Sen2Cor to an L2A format. Pixel values were brought into a range between 0 and 1 using Equation 8.4. We used the $B02, B03, B04, B05, B06, B07, B08, B11, B12$, and $B8A$ bands, which were adjusted to a 10 m resolution. 60m bands were discarded as they are more affected by atmosphere than the land surface. 20 to 10 m bands were upsampled with the nearest neighbor method to avoid initial data corruption (they can be unambiguously downsampled back to exactly initial 20m data).

Both for Sentinel and WorldView, each image covered the entire study area, and images were considered separately without any spatial averaging (the same as in [Astola et al., 2019]).

As supplementary features, we used a freely available high-resolution digital elevation model (DEM), ArcticDEM, covering boreal regions (Figure 6-5). It provides a resolution of 2 m. For some experiments, the resolution was upsampled to 1 m by interpolation (see Section 6.3.5).

Both the satellite and LiDAR data were co-registered through geo-referencing, the same as in [Meddens et al., 2018].

We used cloud-free composite orthophotomap provided by Mapbox [Mapbox, Accessed: 2020-06-17] via tile-based map service as an example of free-available high-resolution RGB data-source. This image covered the same test region and was used just for the developed model assessment. We chose this data-source, because model implementation without expensive input data demands is crucial for open-access platform that can handle a more available images. The spatial resolution was 1 m per pixel, and the preprocessing was the same as for WorldView data.

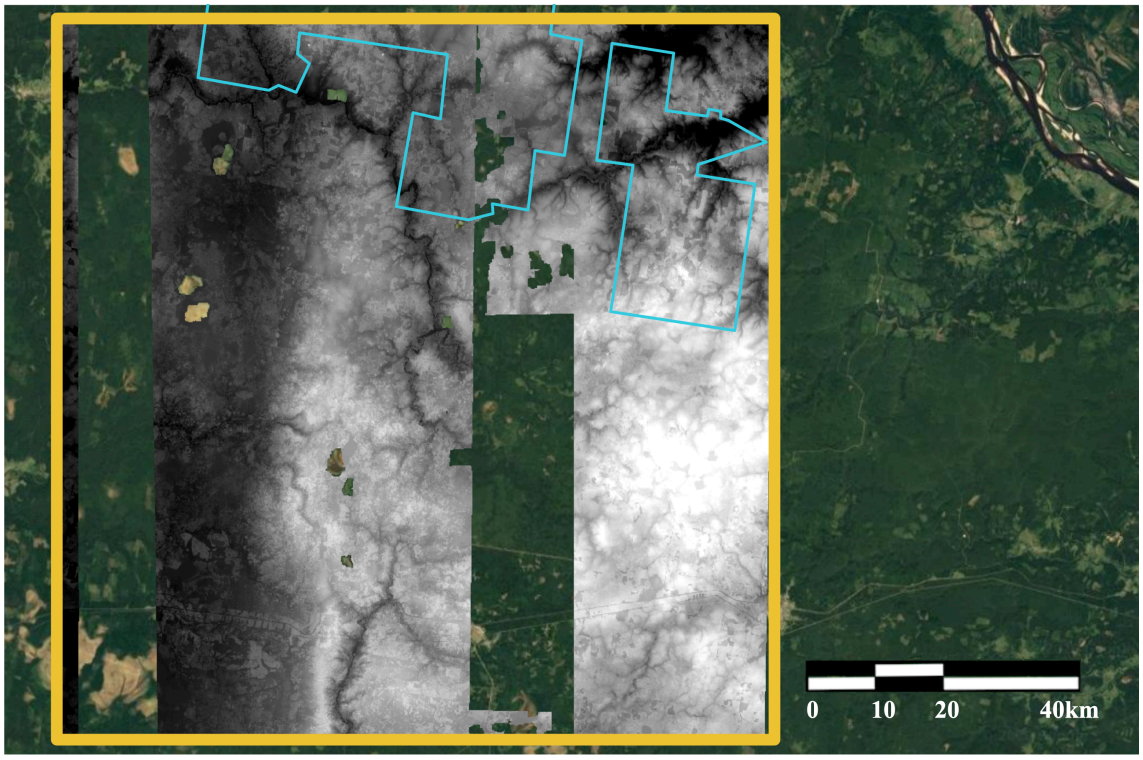


Figure 6-5: One of the ArcticDEM tiles (yellow square) with an overlay of the studied area (blue lines). Even in boreal regions, ArcticDEM layer can have some missing data.

6.3.5 Feature selection for deep neural network

Convolutional neural networks take a tensor as an input. The feature selection to create this tensor is fundamental. To find the best input data representation for the CHM estimation problem, we established a set of experiments. Firstly, we conducted a study with the WorldView bands.

The workflow of our research is shown in Figure 8-3. For each experiment, the RGB bands were used constantly. The variable part concerned the resolution changing and the supplementary features (NIR and ArcticDEM), which were combined with the RGB channel in a single input tensor for the neural network model.

We studied the original (2 m), pansharpened (1 m), and downsampled (5 m) images. For the experiments with the 1 m resolution, bands were upsampled to the target resolution by bilinear interpolation. We used bilinear interpolation for image resampling to avoid aliasing emerging in nearest neighbor and halo inherent to higher-order interpolation methods, which are more problematic for neural net-

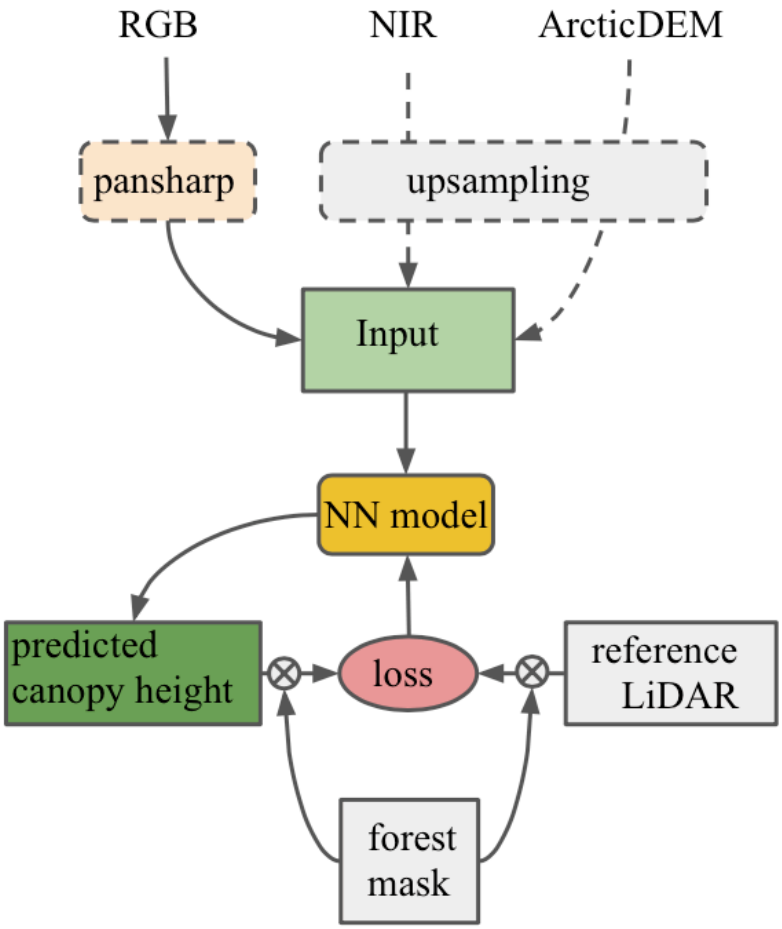


Figure 6-6: Experiment workflow for canopy height estimation with RGB WorldView bands. The dotted lines show optional steps for input tensor creation.

works than bilinear interpolation. A reference CHM was used during the training procedure to estimate the model's error. To minimize data mismatches, reference and predicted height maps were intersected with the forest cover mask before the loss function calculation stage. Therefore, we conducted the following experiments for the WorldView images:

1. RGB original resolution 2 m;
2. RGB pansharpened to 1 m;
3. RGB pansharpened to 1 m + ArcticDEM upsampled to 2 m;
4. RGB + NIR original resolution 2 m;
5. RGB + NIR original resolution 2 m + ArcticDEM upsampled to 2 m;

6. RGB pansharpened to 1 m + NIR upsampled to 1 m;

7. RGB pansharpened to 1 m + NIR upsampled to 1 m + ArcticDEM upsampled to 1 m;

8. RGB downsampled to 5 m resolution.

For experiments 1, 2, there was three-band raster; for experiments 3, 4, 6, we used four-band raster; and for experiments 5, 7, five-band raster was considered.

To assess the importance and restriction of the spatial resolution, we also checked the model's performance for the WorldView RGB bands downsampled to 5 m.

We conducted the following study to compare model's performance for high-resolution RGB images and less detailed but richer in terms of the spectral information Sentinel data with 10 bands, upsampled to 10 m. There were two experiments:

1. Multispectral bands;

2. Multispectral bands + ArcticDEM downsampled to 10 m.

6.3.6 Strategies for height prediction and evaluation metrics

Regression may naturally lead to richer (continuous) estimations for practical implementations than rigid class-based output maps. Therefore, we considered both regression and classification tasks for a comparative analysis. The regression problem statement means that we ascribe each pixel with a particular value corresponding to the height parameter. Then, the loss can be estimated as an error between real height value (CHM value) and the predicted value. The considered metrics are root mean square error (RMSE), mean absolute error (MAE), and mean bias error (MBE).

Using the same reference data we can also solve classification task. When we formalized the problem as a classification task, we divided the continuous values of height into various classes. The choice of such a division often depends on an applied task's demands. For our study, we chose intervals $0-4$, $4-10$, $10-20$, and > 20 m. We rely on the amount of classes and intervals of height that described [Peterson and Nelson, 2014]. We slightly shifted the boundaries of the height intervals, described in

[Peterson and Nelson, 2014] according to the suggestion inventory data provider from Arkhangelsk region. After splitting the continuous dataset to the aforementioned classes we can compute the portion of the wrong estimated pixel classes and use F1-score [Goutte and Gaussier, 2005] for evaluation of the trained classification models.

This refers to the area assessment, while in terms of regression, we strove to optimize each pixel value. Therefore, these two approaches can lead to a different local optimum. For example, if we split heights between 0 and 30 m into the following buckets: $0-4$, $4-10$, $10-20$, and $20-30$, then it is not important that we do not ascribe the exact values but some value from the correct bucket to some pixels. Then, it is clear that regression predictions can also be represented in terms of classification.

For the classification task, the multiclass weighted cross-entropy loss function was used to make the predictions more balanced even for classes with fewer representatives. The same approach was implemented for the regression task. We compared the simple RMSE loss (Equation 6.4) and the weighted RMSE loss (Equation 6.5):

$$\text{RMSE loss} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{N}}, \tag{6.4}$$

$$\text{Weighted RMSE loss} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 * weights(y_i)}{N}}, \tag{6.5}$$

where $\hat{y}_i$ is the predicted value of the i^{th} pixel, y_i is the target value of the i^{th} pixel, N is the number of relevant (non-masked) pixels, $weights(y_i)$ is the extra penalty depending on the target value of the i^{th} pixel.

For heights with fewer representatives, the penalty for the wrong prediction was increased by predefined weights. The weights were inversely proportional to the height distribution. There was also a threshold for the height when the weight was equal to 1 (no extra penalty). The range of weights and the threshold were chosen empirically, as shown in Figure 6-7.

We needed to manage the temporal mismatch (such as logging) between LiDAR scanning and satellite imagery. To do so, we used two heuristics. The first one was that pixels labeled as forest by the forest cover model but with a height of less than

1 m were considered to be a forest logging. The forest cover model classifies pixels covered with clouds as non-forested. Therefore, the second heuristic was that pixels not labeled as forest but with $CHM > 5$ m were considered clouds. Reference and predicted height values for these pixels were not used in the loss function calculation during the training procedure (they were treated as masked). Thus, the mask of relevant pixels was defined by the following equations:

$$logging = (height\_map < 1) * forest\_mask, \tag{6.6}$$

$$cloud = (height\_map > 5) * (forest\_mask == 0), \tag{6.7}$$

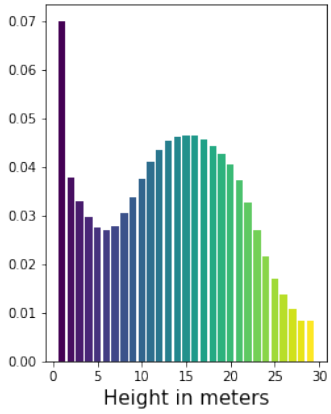
$$height\_mask = (logging == 0) * (cloud == 0), \tag{6.8}$$

where forest mask was obtained by the neural network model trained to predict forest cover with a high accuracy, especially in terms of small details using RGB bands. The model was implemented in the GeoAlert service [geoalert.io, 2019-2020].

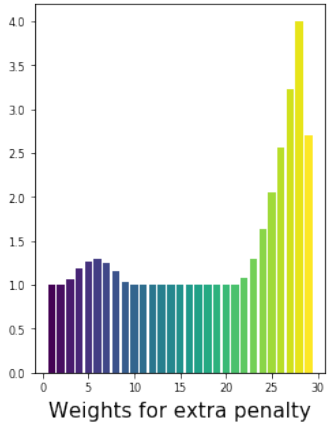
6.3.7 Experimental settings

For all the neural network models, training was performed on the Skoltech supercomputer Zhores [Zacharov et al., 2019], using Keras with a Tensorflow backend. The source code containing the implementation details is available in the aforementioned repository.

Both for the regression and classification task, U-Net [Ronneberger et al., 2015] with an Inception-ResNet-v2 [Szegedy et al., 2017] encoder was used (Figure 6-8). U-Net is a popular CNN architecture in the remote sensing domain which has shown high performance in various problems [Kattenborn et al., 2021a, Li et al., 2020b]. The upsampling layers follow the U-Net's downsampling layers. Skip connections between layers allow the convolutional neural network to manipulate vital information at large spatial scales avoiding losing local information. Skip connections also facilitate gradient flow during the training procedure that was highlighted in [Drozdzal



(a)



(b)

Figure 6-7: LiDAR-derived height distribution (a) and penalty weights for errors on corresponding height values (b). These weights are used during loss computation.

et al., 2016]. We substituted the original VGG encoder with a ResNet-based one as it has shown high results in various works [LeClerc Arrastia et al., 2021]. Residual connections in the Inception-ResNet-v2 encoder support shortcuts leading to better prediction quality [Ferreira et al., 2018] and enabling substantial simplification of the Inception blocks. We used the original U-Net decoder, where every step consists of an upsampling of the feature map followed by a 2x2 convolution. That halves the number of feature channels. The expansive path also includes concatenation with the cropped feature map from the contracting path and two 3x3 convolutions followed by a ReLU. The total number of parameters in the neural network is 62M where the encoder includes 54M. The decoder has 5 blocks, while the encoder part consist of 8 blocks. The models' implementation was based on opensource library [Yakubovskiy, 2022].

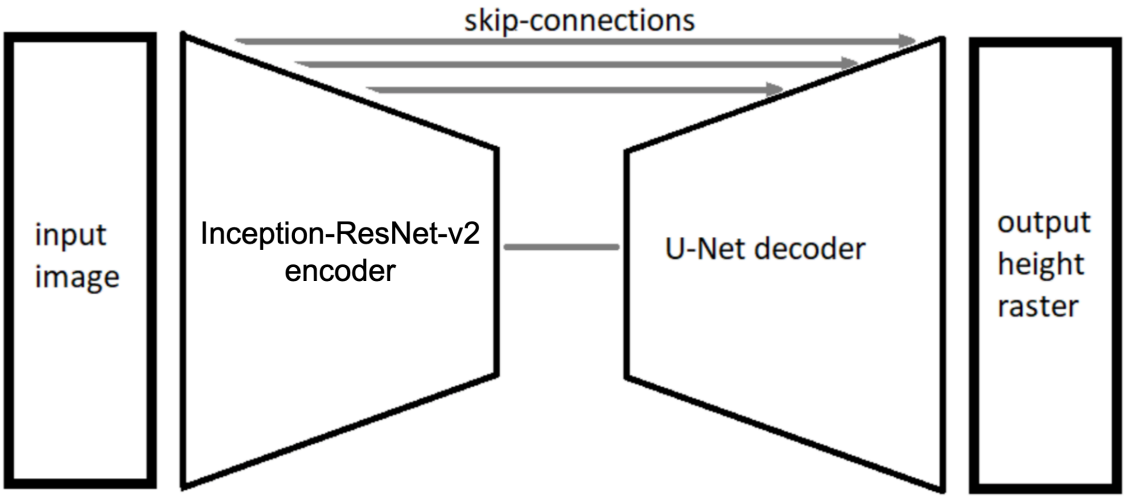


Figure 6-8: U-Net model with Inception-ResNet-v2 encoder.

Each model was trained 25 epochs for 200 training and 100 validation steps with a decreasing learning rate from 0.001 using RMSprop [Hinton and Swersky, 2012] optimizer and early stopping with patience 5 epochs. For the classification task as an activation function for the last layer, the softmax function was chosen. As an activation function for the last layer's regression model, we used linear function.

For all models, geometrical augmentations were implemented. This involves random rotations, and vertical and horizontal flipping. For models using the RGB channels only, we also implemented color transformations. For this task, the Albumentations library [Buslaev et al., 2020b] was used.

6.3.8 Classical machine learning methods

We also conducted experiments with classical machine learning methods to compare different approaches in canopy height estimation. Two approaches were considered: Random Forest (RF) [Breiman, 2001] and Gradient Boosting (GB) [Friedman, 2002]. These approaches are widely used in the remote sensing domain due to relatively high performance in various tasks. For the RF method, we implemented 300 decision trees with maximum depth equal to 8, as these parameters shown the best quality. We also compared it to decision tree numbers 100, 200, 400, 500, 600, and maximum depth values equal to 4, 5, 6, 7, 8, 9, 10. In the GB method the parameters were

200 estimators with learning rate equal to 0.1, and maximum depth equal to 7, that were also set empirically (the same grid was considered to choose number of trees and maximum depth as in the RF case). For both two methods the implementation was used from scikit-learn [Pedregosa et al., 2011a]. A proper feature space is essential for machine learning algorithms, namely in classical one. The features were selected according to the study described in [Puletti et al., 2018] as more relevant for vegetation properties estimation from Sentinel images. Therefore, the following vegetation indices were computed and accomplished initial multispectral bands resulting in Sentinel-derived features: the Normalized Difference Vegetation Index (NDVI), the Simple Ratio Index (SRI), the red-edge Normalized Difference Vegetation Index (RENDVI), and the Anthocyanin Reflectance Index 1 (ARI1). Thus, each pixel was considered as an input sample with 14 features (10 Sentinel bands and 4 vegetation indexes) for a machine learning algorithm.

The following experiments were performed:

1. RF + Sentinel-derived features (Sentinel resolution 10 m)
2. RF + Sentinel-derived features (Sentinel resolution 10 m + ArcticDEM)
3. GB + Sentinel-derived features (Sentinel resolution 10 m)
4. GB + Sentinel-derived features (Sentinel resolution 10 m + ArcticDEM)

6.3.9 Forest-type classification model

To estimate the quality of the developed models, we considered a forest-type classification problem. To train the neural network model to predict two species (conifer and deciduous), we leveraged both WorldView and Sentinel imagery. The problem was defined as the per-pixel semantic segmentation task. Forest inventory characteristics were used as reference data. Eight WorldView bands were intersected with the forest mask. Both for the Sentinel and WorldView imagery, a height map or age map was used as an additional channel. This was done to make the model more robust in terms of species diversity resulting from different forest ages. Therefore, the neural network input was formed of 10 bands.

As mentioned above, there are two familiar sources of height values: LiDAR-derivied data and forest inventory characteristics. The difference is in the data representation. Forest inventory characteristics establish height for each individual stand (small region joined according to some similar value of features such as tree species, age, density). Although real height within each stand can differ for each pixel, all pixels corresponding to a particular stand have the same height value. Thus, for this experiment we used both inventory- and LiDAR-derived height data.

We compared model predictions according to the next strategies of data leveraging:

1. just multispectral data;
2. multispectral data and CHM data;
3. multispectral data and inventory height data;
4. multispectral data and inventory age data;
5. multispectral and artificially generated CHM by the best model height.

For these experiments, we trained a smaller U-Net model with the Resnet-34 encoder [He et al., 2016]. Individual stands from the dataset were randomly split into a training and testing set shown in Table 6.3. During training, the cross-entropy loss function was computed in a per-pixel manner. For testing, the F1-score was estimated for each individual stand. The predicted class for the individual stand was defined as a dominant class among all pixels within the stand. Each forest classification model was trained 25 epochs for 200 training and 100 validation steps with a decreasing learning rate from 0.001 using RMSprop [Hinton and Swersky, 2012] optimizer and early stopping with patience 5 epochs. The activation function for the last layer was soft-max.

6.4 Results

The achieved metrics for the regression models are shown in Table 6.4. The best quality predictions, using WorldView imagery with MAE 2.47 m (Exp. 9), were

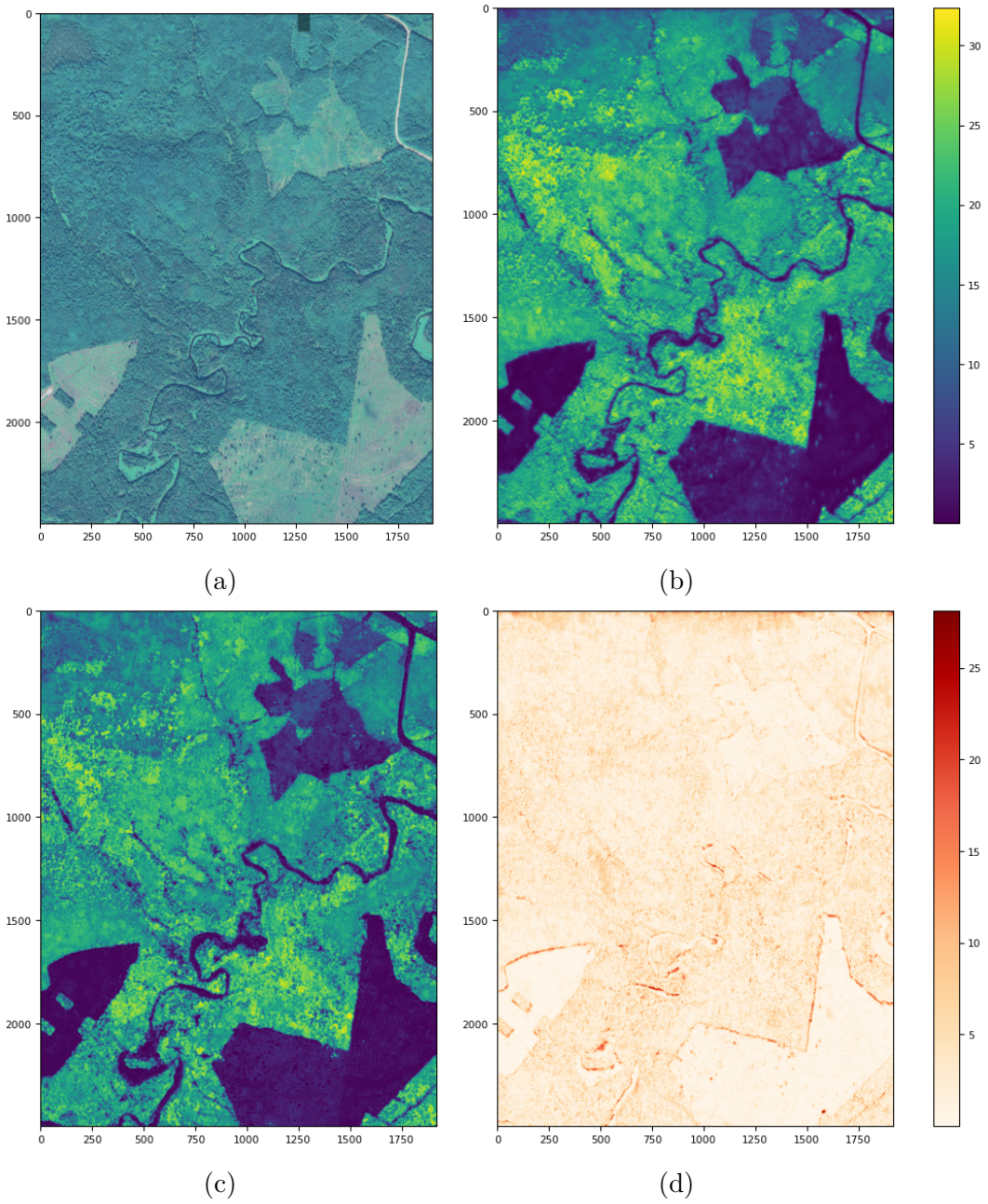


Figure 6-9: Input RGB WorldView image from test regions (a), generated CHM (b), LiDAR-derived height (c), error (d). Height measurements are in m.

achieved with a combination of Red, Green, Blue pansharpened bands, the NIR band, and the supplementary ArcticDEM raster with resolution upsampled to 1 m (Figure 6-9). The smaller region is presented in Figure 6-10. For the Sentinel imagery, only two experimental modes were considered: with ArcticDEM and without ArcticDEM. For both the Sentinel and WorldView data, ArcticDEM usage allowed us to improve the prediction results (for Sentinel, the MAE improved from 4.1 to

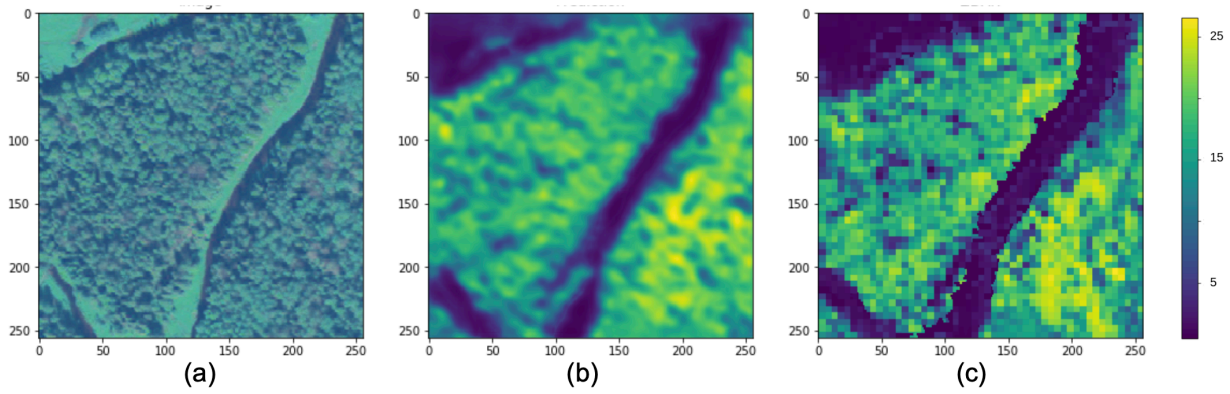


Figure 6-10: Input RGB WorldView image from test regions (pansharpened to 1 m) (a), generated height (b), LiDAR height (downsampled to 5 m) (c).

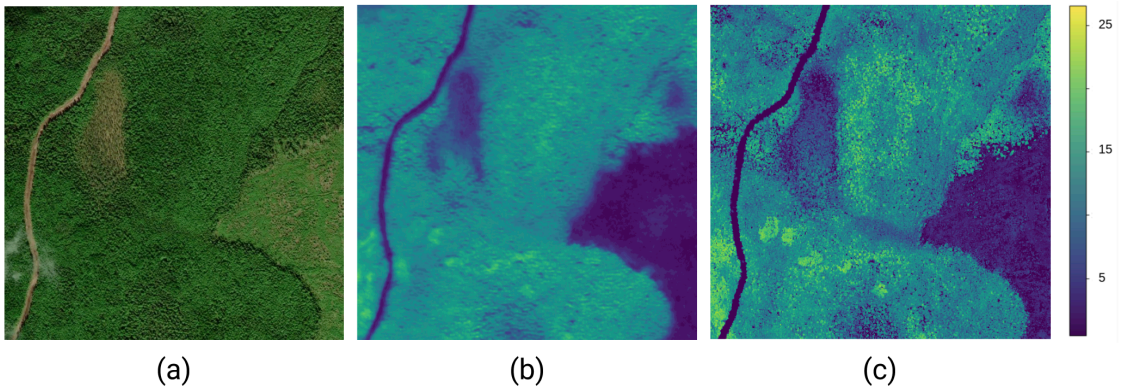


Figure 6-11: Input RGB Mapbox image from test regions (a), generated height (b), LiDAR height (c).

3.9 m, and for WorldView, the MAE improved from 2.9 to 2.58 m). The pansharpening procedure also contributed to the final result, decreasing the error from 3.3 to 3.1 m (Exp. 1 and Exp. 2) for the WorldView RGB model. The NIR band usage demonstrated an error reduction from 2.9 to 2.58 m (Exp. 3 and Exp. 7). This effect is linked to vegetation condition, which is reflected by the NIR wavelength. Additional weights during the loss computation reduced the MAE from 2.58 to 2.47 m (Exp. 7 and Exp. 9).

In Table 6.5, we can see a comparison between the regression model and the classification model (Figure 6-12). These two models were trained using the same input data. The regression model's prediction was split into four appropriate height classes and the F1-score was calculated. This confirmed the assumption that after

Table 6.4: Results for regression models with errors in meters and standard deviation for each experiment.

Exp.	Description	MAE	RMSE	MBE
1	RGB (original resolution 2 m)	3.3 ± 0.052	4.5	0.024
2	RGB (pansharpened to 1 m resolution)	3.1 ± 0.045	4.3	0.009
3	RGB (pansharpened to 1 m resolution + ArcticDEM)	2.9 ± 0.038	4.1	-0.75
4	RGB+NIR (original resolution 2 m)	2.9 ± 0.043	4.1	-0.661
5	RGB+NIR (original resolution 2 m + ArcticDEM)	2.8 ± 0.041	4	-0.132
6	RGB+NIR (RGB pansharpened to 1 m resolution)	2.9 ± 0.043	4.1	-0.8
7	RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM)	2.58 ± 0.046	3.8	-0.99
8	RGB (downsampled to 5 m resolution)	4.4 ± 0.047	5.9	0.65
9	Weighted RMSE RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM)	**2.47 ± 0.042**	**3.6**	**-0.267**
10	Multispectral (Sentinel resolution 10 m)	4.1 ± 0.046	5.7	0.79
11	Multispectral (Sentinel resolution 10 m + ArcticDEM)	3.9 ± 0.053	5.4	0.32
12	RF + Sentinel-derived features (Sentinel resolution 10 m)	4.3 ± 0.051	5.6	0.91
13	RF + Sentinel-derived features (Sentinel resolution 10 m + ArcticDEM)	4.1 ± 0.043	5.4	0.82
14	GB + Sentinel-derived features (Sentinel resolution 10 m)	4.2 ± 0.049	5.5	-0.82
15	GB + Sentinel-derived features (Sentinel resolution 10 m + ArcticDEM)	4. ± 0.039	5.4	-0.78

Table 6.5: Classification task (F1-score). Exp. 1: Weighted RMSE RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM). Exp. 2: Classification model RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM).

Exp.	0-4	4-10	10-20	> 20	Average F1-score
1	0.79	0.51	0.84	0.6	0.68 ± 0.004
2	0.79	0.49	0.78	0.62	0.67 ± 0.005

training the model to predict continuous values, the final results were not worse than the discreet ones (F1-score: 0.68 and 0.67). Moreover, the regression spectrum of values makes the model more flexible, e.g., other classes can be presented and it does not require extra training for new splitting into target classes. This approach

Table 6.6: Forest-type classification (average for all classes F1-score) for WorldView and Sentinel imagery. Generated height is derivied from the best model predictions: Exp. 9 Weighted RMSE RGB+NIR (RGB pansharpened to 1 m resolution + ArcticDEM).

Description	WorldView	Sentinel
multispectral	0.87 ± 0.002	0.88 ± 0.003
multispectral + CHM	0.9 ± 0.003	0.92 ± 0.005
multispectral + inventory height	0.9 ± 0.005	0.93 ± 0.003
multispectral + inventory age	0.93 ± 0.003	0.94 ± 0.002
multispectral + generated	0.89 ± 0.002	0.90 ± 0.004

would be of potential interest for use in other forest characteristics computations.

The recognition class that is most difficult to process is the height between 4−10 m. This is mainly caused by the spatial distribution specificity of the class, and it often occurs due to the small regions between crowns and depends dramatically on the satellite and LiDAR geo-reference data. For this study, we used LiDAR data downsampled to 5 m, while the WorldView imagery resolution was 1 or 2 m. This allowed us to save high-resolution spatial surface characteristics.

To assess the importance of texture information, we experimented with RGB bands downsampled to 5 m (Table 6.4). The MAE for this case was 4.4 m. This result is lower than that of the Sentinel images (4.1 m) and confirms that when we reduced the spectral information, we faced stricter demands for spatial resolution.

We checked the generated height in the forestry task of species classification. The results are presented in Table 6.6. The first objective of the experiment was to show how supplementary features can enhance the quality of applied tasks. Both LiDAR and inventory data helped to improve classification in comparison with simple multispectral data. The second goal was to show that the generated height is of sufficient quality to beat the base model using just satellite data. We did not intend to conduct a comparison between WorldView and Sentinel sources. For this reason, in both experiments, observation dates were not equal in the data used. The superior results for the Sentinel imagery, as compared with the WorldView data, were partially due to the wider dataset.

We also evaluated the regression model trained using RGB WorldView (pan-

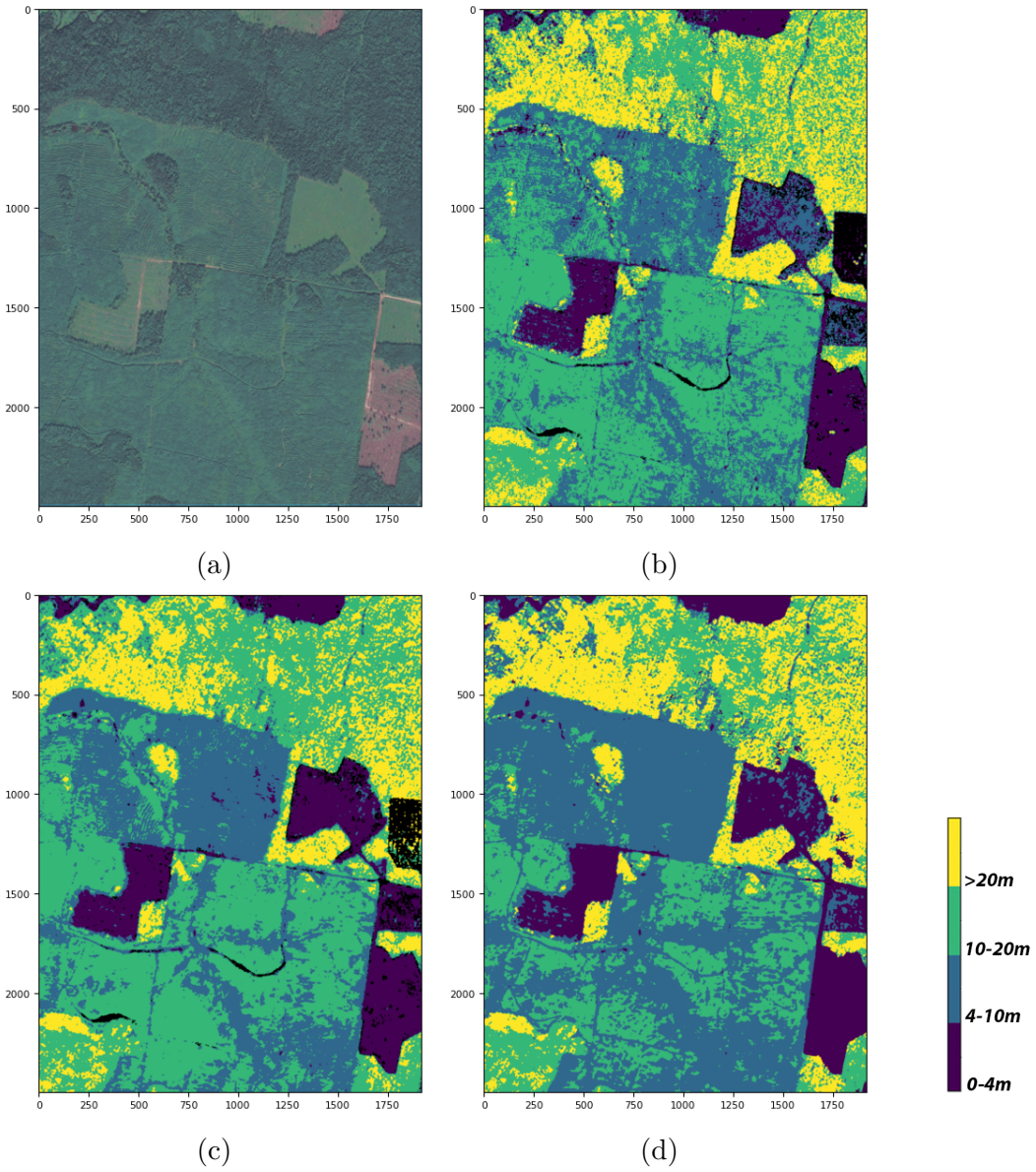


Figure 6-12: Input RGB WorldView image from test regions (a), original height classes (b), generated height classes in regression (c) and classification (d) problem statement.

sharpened to 1 m resolution) image on a cloud-free composite orthophotomap provided by Mapbox [Mapbox, Accessed: 2020-06-17] and covering the same test area. For this experiment, the MAE was equal to 3.5, and the RMSE was 4.6. Prediction example is shown in Figure 6-11. This promising result allows cheaper CHM estimation for large areas using only high-resolution free-available satellite RGB data.

We conducted experiments with classical machine learning algorithms using Sentinel-derived features to compare this approach to the proposed one, namely

the CNN-based with high-resolution data. The best results were achieved for the GB algorithm and combination of Sentinel-derived features with ArcticDEM, where MAE was equal to 4 and RMSE was equal to 5.4 (Figure 6.4).

6.5 Discussion

It is challenging to perform a fair comparison between the majority of studies related to height estimation for various reasons. The main reason is the difference in height distribution. For example, in [Meddens et al., 2018], the predicted height was limited by 30 m, the spatial resolution was 5 m, and the final RMSE was 2.2 m. However, according to the presented plots, the mean value was less than 10 m, while in our study, it was about 15 m. In [Staben et al., 2018], the validation pixels range was defined as being from 0 to 25 m, with a mean value of 7 m. The model's spatial resolution was 30 m. For this height distribution, an RMSE from 2.3 to 4.1 m was achieved. In [Ghosh et al., 2020], they studied the ranges between 0 to 18 m and 3 to 15 m, by leveraging satellite (both spectral and radar) data with a 20 m resolution. In contrast to our work, field-based observations with a sampling frequency of the 10 largest trees per inventory plot were used as reference material. Therefore, the achieved result (an RMSE of 1.48 m) cannot be compared with our model's performance. Other obstacles impeding a fair comparison are the species diversity and regional conditions.

It is worth mentioning that although ArcticDEM provides a stable improvement in canopy height estimation (see Table 6.4, Exp. 6 and Exp. 7), it does not cover central or southern regions. For these areas, more powerful base models need to be implemented, leveraging just satellite imagery.

We showed that high-resolution WorldView 3-bands images provided more significant features than low resolution Sentinel with 10 spectral bands (see Table 6.4, Exp. 2 and Exp. 10). However, resolution adjustment from 2 m to 5 m for the same WorldView dataset leads to a loss of important information, in particular texture information (see Table 6.4, Exp. 2 and Exp. 8). The aforementioned experiments, which was performed on the same dataset and using the same NNs with only one

difference - the adjusted spatial resolution, showed that neural networks can extract additional spatial features from very high-resolution optical images of 1 m. Thus we experimentally confirmed the initial hypothesis that by using high resolution data it is possible to make CHM estimation more accurate.

Creating the model with only high-resolution RGB channels allows it to be implemented in more available satellite images, such as RGB mosaic basemaps (google, yandex, and Mapbox). Therefore, an opportunity to replace WorldView data with satellite images derived from other sources, making the provided model more universal. We made a prediction for cloud-free composite orthophotomap provided by Mapbox [Mapbox, Accessed: 2020-06-17] using the CNN model trained on RGB 1 m bands. The achieved quality (MAE = 3.5) confirms the opportunity for further model application for basemaps analysis.

There are the following directions for future research. The first involves improving the co-registration between LiDAR and satellite data. Now the developed RGB-based model shows the ability to reconstruct the main patterns corresponding to the CHM (Figure 6-10); large individual trees and spots within forest are detected successfully. However, satellite data has a slight shift in comparison with LiDAR data. Improving co-registration would allow the model's performance to be assessed more accurately for resolutions of less or equal to 1 m and also could probably improve the poor performance for the class of 4–10 m.

The ability of the model to be transferred to new regions is another essential question. As we did not have data from other regions, it is impossible to judge the model robustness for new areas. Moreover, for some regions, the ArcticDEM layer is not available; therefore, additional training for new areas might improve prediction quality. However, the neural network approach has proven to be powerful enough to extract the necessary spatial information and adapt to changing natural conditions. Augmentation and image diversity are often applied to overcome this weakness in real-life applications.

Another possible objective for future research is a canopy height estimation for areas with complex topography. Neural network models rely on landcover's spectral and texture characteristics, making the initial approach promising even when

topography is not flat. However, shadows on slopes pose additional challenges to the multispectral satellite image analysis. LiDAR data additional preprocessing is also considered for study areas with complex topography [Liu et al., 2017].

In this study, we used all available images both for training and testing (splitting them into training and testing regions) as it is a common choice in the remote sensing domain [Saralioglu and Gungor, 2020]. However, in the future work, image-based cross-validation techniques can be used and robustness for new environmental conditions can be considered [Illarionova et al., 2021a].

6.6 Conclusions

Overall, in this study we confirm the hypothesis that neural networks can extract significant spatial features from very high-resolution RGB images, which can be used for more precise canopy height estimation. We also checked whether it is possible to get an accuracy of canopy height estimation by using of satellite-based solutions compatible with measurements obtained by UAV approach. For checking our assumptions, we analysed the potential of very high-resolution images with limited spectral information in the task of canopy height model estimation. We created a software toolchain based on a state-of-the-art neural network architecture that enable us to extract spatial features from very high-resolution images. The proposed approach led to a reduction in the mean absolute error to 2.4 m, while leveraging just four spectral bands and the supplementary features from ArcticDEM. However, in southern regions where ArcticDEM is not available and without other sufficiently accurate DEM, the model achieved an MAE of 2.9 m. We also examined how generated height can be successfully used in the forest classification task. Our canopy height model estimation results using RGB bands indicated the prospect of replacing expensive LiDAR sensing data with easily attainable satellite data. Depending on the region of study, our technique allows a customer to promptly collect all the necessary relevant forestry inventory information without ground-based observations.

Chapter 7

Generation of the NIR Spectral Band for Satellite Images with Convolutional Neural Networks

7.1 Introduction

Machine learning techniques allow researchers to achieve high performance in a wide range of remote sensing tasks by leveraging spectral bands of different wavelengths [Maxwell et al., 2018]. One essential spectrum interval for the remote sensing image analysis is represented by the near-infrared (NIR) channel. The classical approaches in landcover classification tasks often use NIR-based spectral indices such as the Normalized Difference Vegetation Index (NDVI) or the Enhanced Vegetation Index (EVI) to assess the vegetation state [Huete et al., 1999]. This spectral band is widely used in many applications, including forestry [Li et al., 2019a, Illarionova et al., 2020], agriculture [Kussul et al., 2017, Navarro et al., 2016], and general landcover classification [Scott et al., 2017, Fan et al., 2017]. However, there are still cases when the NIR band is not presented in the available data [Flood et al., 2019, Alias et al., 2018]. Thus, the researchers rely only on RGB. For example, the Maxar Open Data Program [Maxar, Accessed: 2020] provides only RGB images. Many aerial imaging systems are also limited to visible wavelength ranges.

The NIR band cannot be extracted from RGB bands. A simple example is

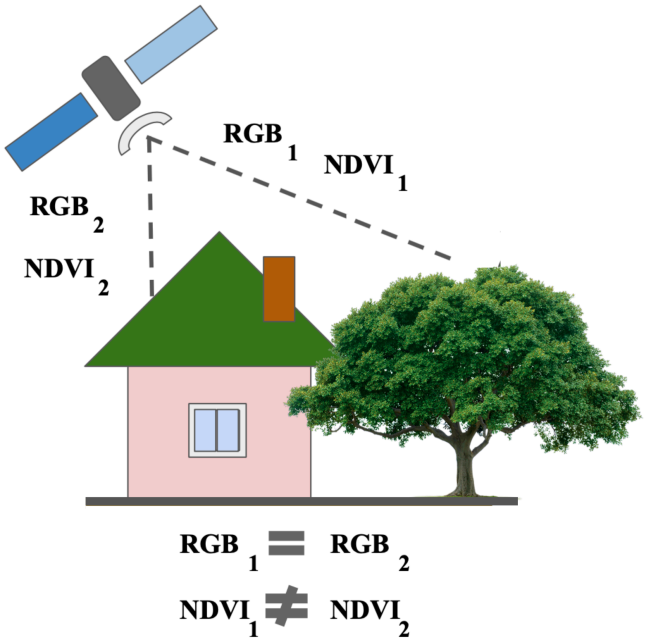


Figure 7-1: Objects with the same spectral values in the RGB range can belong to significantly different classes. For these objects, spectral values beyond the visible range differ. These differences can be illustrated using vegetation indices such as the NDVI in the case of an artificial object and a plant during the vegetation period.

provided in Figure 7-1. For both the green tree and the green roof, the RGB values are the same. However, the values differ drastically in the NIR spectral range, as the metal roof does not have the vegetation properties that affect the NIR. On the other hand, indirect features can be used to evaluate the NIR value. In general, all roofs have a lower NIR values than any healthy tree during the vegetation period. Therefore, it is possible to make assumptions about the NIR value based on the object's shape and texture. This study investigates how neural networks can be applied to solve the NIR generation task by learning the statistical distribution of a large unlabeled dataset of satellite images.

In [de Lima et al., 2019], a similar problem of generating the NIR channel from RGB was described. The proposed solution was based on the K-Nearest Neighbor classification algorithm and was focused on the agricultural domain. The researchers show in [de Lima et al., 2019] a high demand for the generated NIR data, which can solve particular problems. However, the neural network approach was beyond the scope of the present study for image generation. In [Gravey et al., 2019], they

generated synthetic spectral bands for archive satellite images using Landsat data. Synthetic satellite imagery generation from Sentinel-2 (with the spatial resolution more than 10 meters per pixel) was considered in [Abady et al., 2020, Mohandoss et al., 2020]. However, in our work, we were focused on high-resolution satellite images as they provide valuable texture information.

Generative adversarial networks (GANs) have achieved great results in recent years [Alqahtani et al., 2019]. The basis of this approach consists of two neural network models that are trained to beat each other. The first network (generator) aims to create instances as realistically as possible, and the second network (discriminator) learns to verify whether the instance is fake or real. Conditional GANs (cGAN) have proven to be a promising approach in various fields using additional conditions in the generation process. Conditional GANs were implemented to solve different tasks such as image colorization [Nazeri et al., 2018], including infrared input [Suárez et al., 2017] and remote sensing data [Wu et al., 2019, Li et al., 2018a, Tang et al., 2020, Singh and Komodakis, 2018], and style transfer [Zhu et al., 2017a, Isola et al., 2017].

Pix2pix GAN, as described in [Isola et al., 2017], proposes an image-to-image translation approach. Previous studies have shown a lack of generalization for other problems. Authors of [Isola et al., 2017] aimed to develop an efficient framework that can be successfully implemented to solve a wide variety of tasks, such as image colorization, synthesizing images from a labeled map, generating land-cover maps from remote sensing images, changing the style, etc. Pix2pix GAN uses a U-Net-based architecture as a generator and a convolutional PatchGAN as a discriminator. The model was trained to estimate image originality separately for each small region. The authors used the following objective function $G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G)$ to train the model. The Pix2pix approach enhancements were provided by the authors of [Qu et al., 2019].

One prevalent computer vision task is image colorization, which is required to obtain color images from grayscale ones [Wang and Liu, 2021]. One of the earliest works using texture information for this task is [Welsh et al., 2002]. In recent years, GANs (in particular cGANs) have become a popular approach for such a

challenge, in particular, in the remote sensing domain [Li et al., 2018a, Wu et al., 2020b]. In the image colorization task, cGANs take a condition that should be utilized for new image generation. The results for such a task can be evaluated visually. This challenge share similarities with the NIR generation problem. As an input, grayscale images are received, and as an output, an RGB image is created. In contrast, for NIR, we strive to obtain one channel from three channels. Unlike mapping grayscale to RGB, NIR does not include a mixture of RGB; NIR even lies in a distant wavelength region from RGB. It makes the task more challenging. Moreover, in the colorization problem, the choice of color sometimes depends on the statistical distribution in the training set (for example, the color of the car might depend on the number of cars for each color). Such mismatches in colorization might not be treated as a severe mistake, and it does not corrupt the sense of the natural source of objects or phenomena. In contrast, for NIR in vegetation tasks, there is a strong connection between chlorophyll content and the intensity of the channel value [Yang et al., 2020]. A neural network can extract structure features such as shape and texture characteristics. We attempt to combine them with RGB values to generate the NIR band artificially and save the physical sense of this channel as much as possible.

In the remote sensing domain, the opportunity to work with multiple satellite data simultaneously is essential in various cases [Vandal et al., 2021]. In [Kwan et al., 2018], the authors consider WorldView and Planet imagery. WorldView has a higher spatial resolution, while Planet has a higher temporal resolution. Therefore, by combining these data, researchers can solve particular problems rapidly and with better quality. In [Zhang et al., 2014], Modis and Landsat images fusion was considered in the flood mapping case. In [Sedano et al., 2021], they combine images from WorldView2, Rapid Eye, and PlanetScope platforms to solve the forest degradation problem. When images from several sensors were available, the highest spatial resolution images were always preferred. Therefore, in the remote sensing domain, acquisition dates can vary for different satellites, and for monitoring, it is crucial to work with all available data sources. However, when a computer vision model uses data from different distributions, it can decrease prediction quality. One

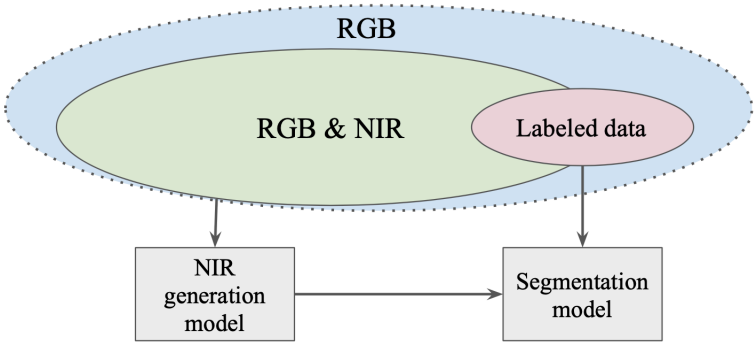


Figure 7-2: A large amount of RGB & NIR data without markup that can be further leveraged in semantic segmentation tasks when NIR is not available in some particular cases.

of the objectives of our study was to examine the importance of the NIR band for cross-domain stability.

In our study, we examine whether the cGAN image generation approach can produce sufficient results for image segmentation purposes. Multiscale contextual features and spatial details are highly important in the remote sensing domain [He et al., 2021]. Therefore, we aim to apply the NIR generation as a feature-engineering method, creating a new feature (NIR reflectance) that is not present in the original feature space (RGB reflectance). We also study original and artificially generated NIR in the cross-domain stability problem, as convolutional neural network (CNN) robustness for various data is vital in the remote sensing domain [Illarionova et al., 2021a]. We aim to use a vast amount of RGB & NIR data without markup that can be further leveraged in semantic segmentation tasks when NIR is not always available Figure 7-2.

We propose and validate an efficient approach to produce an artificial NIR band from the RGB satellite image. A state-of-the-art Pix2pix GAN technique is implemented for this task and compared with a common CNN-based approach for the regression task. WorldView-2 high-resolution data are leveraged to conduct image translation from RGB to NIR with further verification on PlanetScope and Spot-5 RGB images. We also investigate how original and artificially generated NIR bands affect both CNN and Random Forest (RF) [Pal, 2005] predictions in forest segmentation tasks compared to only RGB data. The experiments involve two significant practical cases: two data source combinations (PlanetScope and Spot-5) and differ-

ent amount of labeled training data (the total dataset size for the segmentation task is 500.000 hectares). The contribution of the presented work is as follows:

- We propose the approach for feature-engineering based on the NIR channel generation via cGANs.
- We investigate the impact of artificially generated and real NIR data on the model performance in the satellite image segmentation task. We also examine the NIR channel contribution in reducing labeled dataset size with minimum quality loss. The NIR channel for satellite cross-domain stability is considered.

In the Chapter on the canopy height estimation, we discussed the role of the near-infrared spectral band. In this Chapter, we show how we can generate this band artificially and apply it to forestry tasks.

7.2 Materials and Methods

7.2.1 Dataset

We leveraged WorldView-2 satellite imagery downloaded from GBDX [GBDX, Accessed: 2020] to train the generative models. For forest segmentation experiments, we used the satellite data provided by the SPOT-5 satellite and the PlanetScope satellite group. The imagery has a high spatial resolution of 2–3 meters per pixel in four spectral channels (red, green, blue, near-infrared). All images were georeferenced and had values equal to the surface reflectance.

Overall, two datasets were used in this work:

The first dataset used in this work was for cGAN model training. The dataset consists of RGB and NIR channels from the same satellite (WorldView-2). It covers different regions of Russia and Kazakhstan with approximately the same climate and ecological conditions. The total territory is about 900,000 ha. The datasets consist of varying land cover classes such as crops, forests, non-cultivated fields, and human-made objects. Images with dates from May to September were chosen to represent the high-vegetation period.

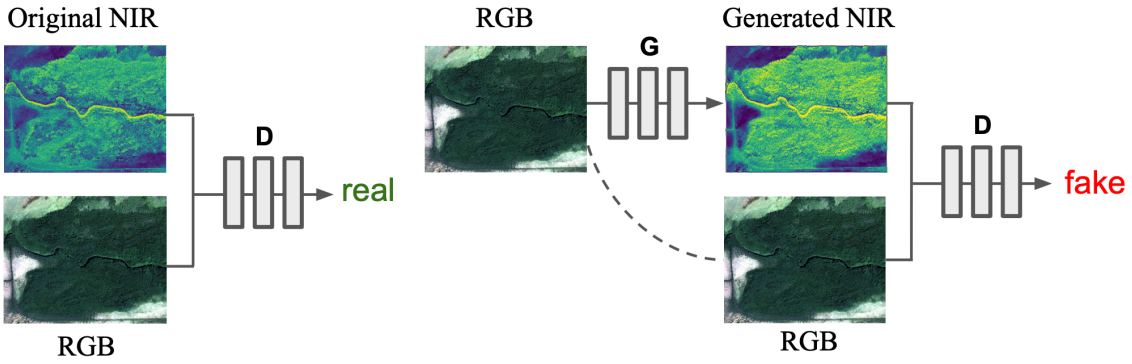


Figure 7-3: Training procedure for GAN using the RGB image as an input and the NIR band as a condition.

The second dataset was used to test the real and artificial NIR channel's influence compared to the bare RGB image. This dataset includes PlanetScope and Spot-5 imagery. The resolution of images ranges between 2 and 3 meters, depending on the view angle. The markup for the study region consists of the binary masks of the forested areas and other classes in equal proportion, covering 500,000 ha. The labeled markup was used for the binary image segmentation problem. The region was split into test and train parts in the proportion of 0.25:0.75.

7.2.2 Artificial NIR Channel Generation

To generate the NIR band from RGB, we used cGAN. We chose the Pix2pix approach for this task because it performs quite well for image translation problems [Salehi and Chalechale, 2020, Ren et al., 2019]. For the generator, we used the U-Net [Ronneberger et al., 2015] architecture with the Resnet-34 [Szegedy et al., 2017] encoder. For the discriminator, the PatchGAN as described in [Isola et al., 2017] with various receptive field sizes was used. The training procedure is shown in Figure 7-3. There were two models: the generator and the discriminator. The generator was trained to create artificial NIR images, using the RGB image as a conditional input. The discriminator received an RGB image in combination with the alleged NIR image. Then, there were few possible scenarios: (1) the NIR was original, and the discriminator succeeded in ascertaining it; (2) the NIR was fake, but the discriminator failed by treating it as original; (3) the NIR was original, but the discriminator mistook for fake; (4) the NIR was fake, and the discriminator

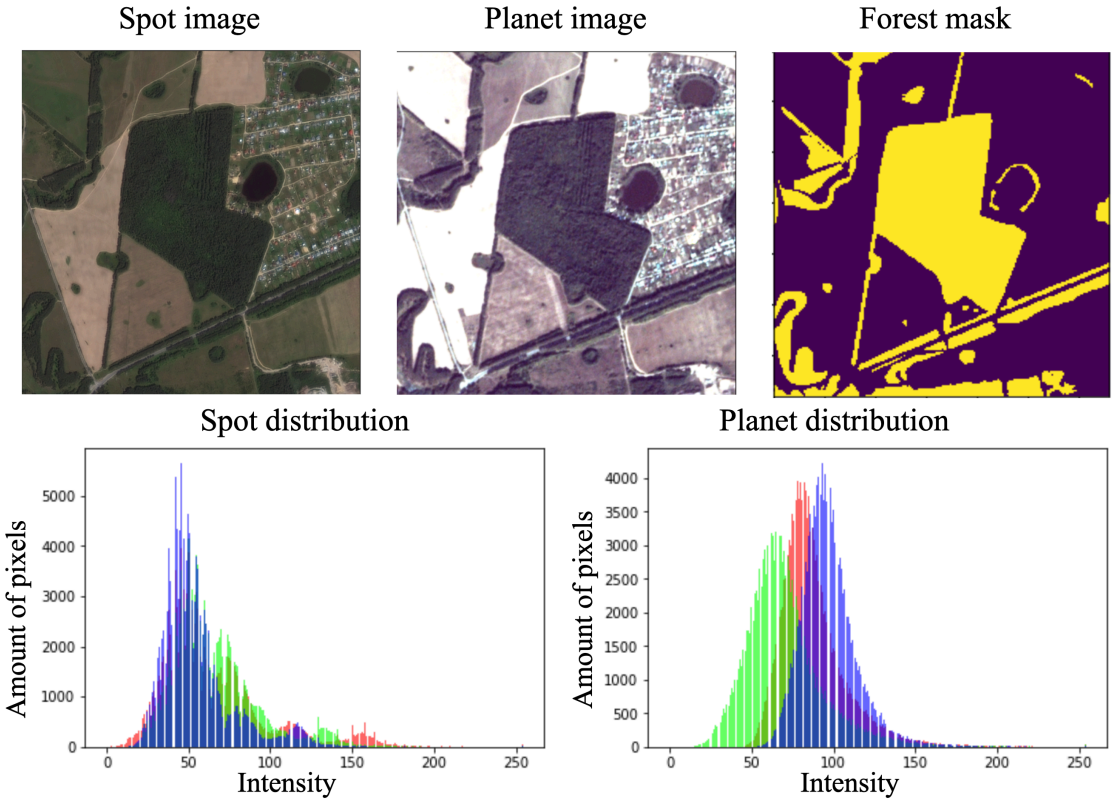


Figure 7-4: Original SPOT and Planet images (without any enhancements) and their RGB spectral values distribution. The histograms were computed within the forest area. Although the presented images are from the summer period, their spectral values differ drastically, as the histogram shows.

exposed it. Although this model was trained simultaneously, we ultimately strove to receive a high performing generative model, to solve the objective of the study. For further analysis, only the generator was considered. Unlike classical machine learning techniques, which usually work only with one particular point (see [de Lima et al., 2019]), the U-Net generator processes a particular neighborhood and learns how to summarize 3-dimensional information.

We compared the cGAN-based approach with the simple CNN-based one where U-Net with Resnet-34 encoder was trained to solve the regression problem.

We considered the root mean square error (RMSE), mean absolute error (MAE), mean bias error (MBE), (PSNR) for the model performance evaluation.

7.2.3 Forest Segmentation Task

To empirically evaluate the usefulness of the original and artificially generated NIR channel to solve real image segmentation problems, we considered the forest segmentation task with high-resolution satellite imagery. In this task, a CNN model was trained to ascribe each pixel with the forest content label.

We used the common solution for the image semantic segmentation: U-Net [Ronneberger et al., 2015] with the ResNet-34 [He et al., 2016] encoder. The chosen architecture is widely implemented in the remote sensing domain [Kattenborn et al., 2021b]. We conducted experiments with different input channels: only RGB, RGB & original NIR, and RGB & generated NIR. The model output was a binary mask of the forest landcover, which was evaluated against the ground truth with an F1-score. We also assessed the original and artificially generated NIR in the same task using classical machine learning approach. We trained a Random Forest (RF) classifier [Pal, 2005]. The RF implementation was from [Pedregosa et al., 2011a] with the default parameters the same as in [Pal, 2005]. Each pixel was considered as an object for the classification.

7.2.4 NIR Channel Usage

We conducted an experiment that estimated the dependency of the segmentation quality on the training dataset size in both RGB and RGB & NIR cases. We randomly split and chose 50% and 30% of the initial training dataset (test data were the same for these random splits). The same experiment was repeated both for the SPOT and Planet imagery but separately for each data source.

In the second study, we considered data from different sources (both PlanetScope and SPOT data) simultaneously. Even if we have two images of the same date, region, and resolution but from various providers, sensors systems and image preprocessing can make them radically different from each other. The intensity distribution for images from Spot and Planet are shown in Figure 7-4. Such differences can be crucial for machine vision algorithms and lead to a reduction in prediction quality. Therefore, it can be treated as a case of a more complex multi-domain satellite

segmentation task. To estimate the importance of the original and artificial NIR channels for different satellite data, we conducted the following experiment. The CNN model was trained using the Planet and SPOT data simultaneously. To evaluate the model's performance, three test sets were considered: only the Planet test images, only the SPOT test images, and both the Planet and SPOT images. The images for Planet and Spot covered the same territory.

7.2.5 Training Setup

The training of all neural network models was performed on a PC with GTX-1080Ti GPUs, using Keras [Keras, Accessed: 20 November 2021] with a Tensorflow [TensorFlow, Accessed: 20 November 2021] backend. For the simple regression model, the following training parameters were set. An optimizer RMSprop was chosen with a learning rate of 0.001, which was reduced with patience 5. There were 20 epochs with 100 steps per epoch. The batch size was specified to be 30 with an image size of 256×256 pixels [Isola et al., 2017]. A model based on GAN training parameters was constructed as follows. The loss functions were chosen to be binary cross-entropy and MAE. The optimizer was Adam. The batch size and image size were the same as for the simple model. The models were trained for 600 epochs, 100 steps per epoch, and a batch size of 30. For the Planet data, we also conducted a fine-tuning procedure of the pretrained generative model using a small area without the necessity of markup. For the SPOT data, there was no additional training.

7.3 Results and Discussion

The results for NIR generation by cGAN are presented in Table 7.1 for the WorldView, SPOT, and Planet satellite data. All values for real and generated NIR were in the range $[0, 1]$. For PSNR evaluation, we consider images in the range $[0, 255]$ as a more common representation. The simple CNN regression approach showed significantly poor results (the MAE was 0.21 for WorldView). Therefore, we did not select this approach for future study. The principal difference between cGANs and the regression CNN model is the type of loss function. As our experiments show,

both MAE and MSE loss in the regression CNN model led to the local optimum, which was far from the global one. The loss function can be affected by the distribution of RGB values. Compared to the regression CNN model, the results of cGAN were significantly closer to the real NIR values.

Another approach to evaluate the generated NIR band involves the forest segmentation task. The segmentation model was trained on the original NIR channels to predict the forest segmentation mask using RGB & generated NIR. The results are presented in Table 7.2, which shows that the additional NIR channel improved the cross-domain stability of the model. The example of segmentation prediction is shown in Figure 7-5. The model using the generated NIR provided more accurate results than the model trained only on RGB bands. The original NIR usage obtained an F1-score of 0.953, the generated NIR obtained an F1-score of 0.947, and the model using only RGB bands obtained an F1-score of 0.914. The predicted NIR channel is shown in Figure 7-6, which confirms a high level of similarity between generated and original bands. Therefore, this approach allows more efficient CNN model usage in practical cases when data from different Basemaps are processed and cross-domain tasks occur.

We also assessed the generated and original NIR bands using classical machine learning approach. Results are presented in Table 7.2. For RF, the NIR band usage improves the classification quality from 0.841 to 0.877. The F1-score for the generated NIR is 0.874. This experiment shows that for the classification approach without spatial information the generated band is also provide significant information.

The results for different dataset sizes are presented in Table 7.3 and show that leveraging the NIR channel was beneficial in the case of smaller dataset sizes, whereas its effect decreased with the growing amount of the training data.

GANs aim to learn dataset distribution. It is conducted by minimizing the overall distance between the real and the generated distribution. We made an assumption that the dataset size is enough to approximate the distribution. Thus, we train the generator to sample according to the target distribution. The trained generator allowed a high-realistic image-to-image translation (G : $\{RGB\}$ NIR) such that the

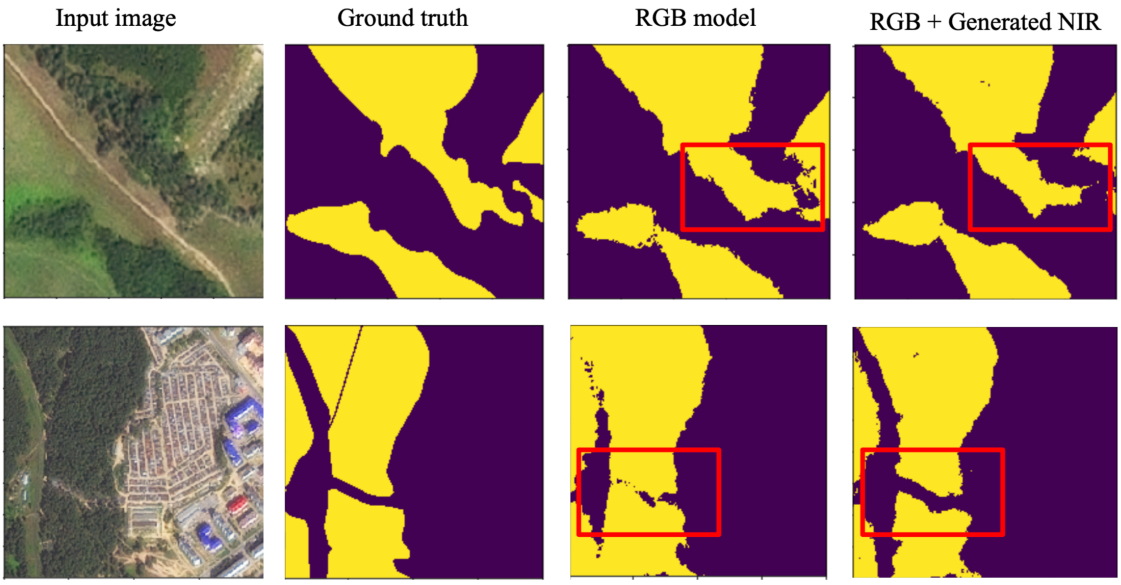


Figure 7-5: Forest segmentation predictions on the test regions (SPOT). One model was trained just on RGB images; another model used RGB and generated NIR.

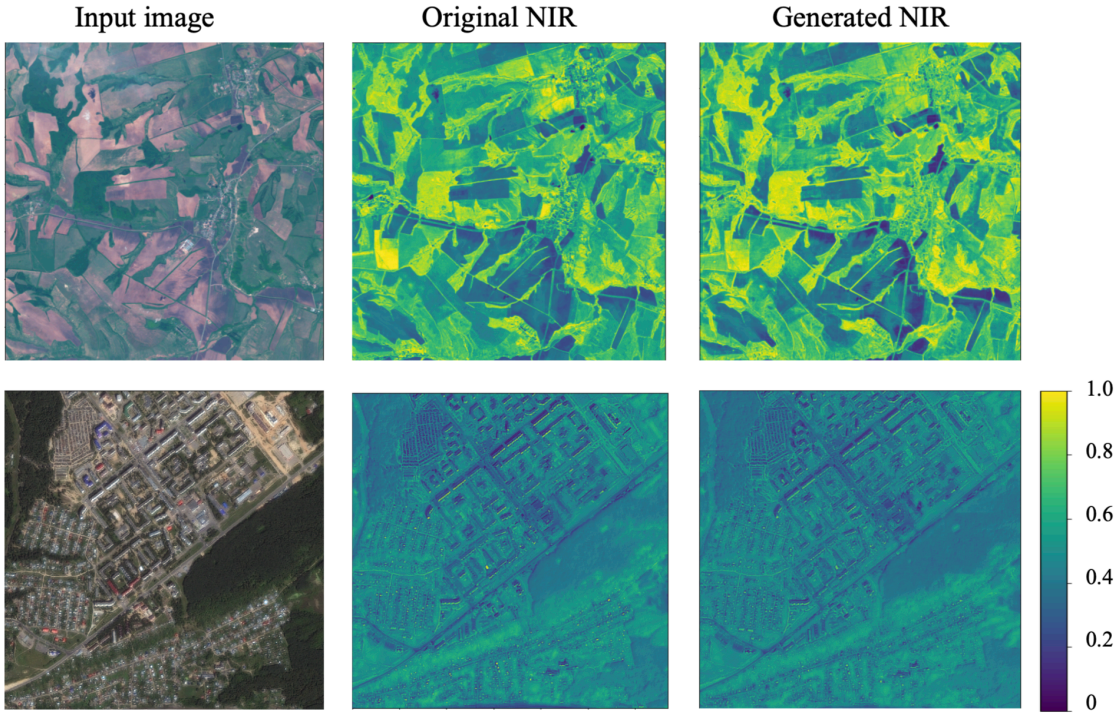


Figure 7-6: Example of generated NIR on the test set. The first row presents the SPOT image; the second row is the WorldView image.

Table 7.1: Error of the artificial NIR band for the test WorldView, SPOT, and Planet imagery. Standard deviation is computed for PSNR values.

	MAE	**RMSE**	**Mean Bias**	**PSNR**
WorldView	0.09	0.31	0.058	27.62 ± 0.205
SPOT	0.037	0.194	−0.0029	28.53 ± 0.261
Planet	0.16	0.41	0.088	26.14 ± 0.217

Table 7.2: The results of the forest segmentation experiments with different data sources. Both the RGB model and the RGB and NIR model were trained on Planet and Spot images simultaneously. The F1-score was computed on the test set individually for Planet and Spot and for the joined Planet and Spot test set. Standard deviation is computed for each experiment.

Test images	RGB	RGB and NIR	RGB and artificial NIR
		U-Net	
SPOT	0.954	0.961	0.96
Planet	0.857	0.939	0.936
SPOT + Planet	0.932	0.96	0.945
Average	0.914 ± 0.001	0.953 ± 0.003 (+0.039)	0.947 ± 0.002 (+0.033)
		RF	
SPOT	0.874	0.892	0.889
Planet	0.815	0.863	0.861
SPOT + Planet	0.836	0.876	0.872
Average	0.841 ± 0.002	0.877 ± 0.002 (+0.036)	0.874 ± 0.001 (+0.033)

Table 7.3: The results for the forest segmentation experiments with different dataset sizes. The F1-score for SPOT and Planet on the test set. The entire dataset size was 500,000 ha.

	Bands	**All Data**	**1/2**	**1/3**
SPOT	RGB	0.97 ± 0.003	0.956 ± 0.003	0.942 ± 0.002
	RGB and NIR	0.97 ± 0.002	0.963 ± 0.004	0.961 ± 0.002
Planet	RGB	0.939 ± 0.002	0.933 ± 0.001	0.874 ± 0.001
	RGB and NIR	0.95 ± 0.001	0.942 ± 0.002	0.927 ± 0.001

obtained NIR band is similar to those belonging to the target domain.

Example of a green roof is presented in Figure 7-7. Although, the color of the object is green, NIR value is low. It shows that the model had a sufficient amount of the training samples to learn such cases.

The experiments indicate that the generated NIR provides additional information to the segmentation model. We assume that the generative model incorporates the hidden statistical connections between the spectral channels that can be learned from the significant amount of real RGB and NIR data. As opposed to the segmentation or classification approach, the channel generation does not require the manual ground truth markup to significantly increase the dataset size. Therefore, this approach can be used as a feature-engineering tool to create a new feature similar to the NIR band of multispectral remote sensing imagery.

We set the goal to predict exactly the NIR band instead of vegetation indexes such as NDVI or EVI. These indexes use the NIR band in combination with the Red band. The NIR band generation allows further computation of other indexes without a requirement for extra model training. The future study can be extended by implementing different vegetation indexes. Moreover, in the case of using neural networks with the generated NIR band, it is enough to provide input NIR and Red bands separately (not in the form of the computed indexes) because a neural network can approximate nonlinear functions such as vegetation indexes.

One example of a failure case is a green lake (Figure 7-8) that might be mistaken for a green lawn. The reason is insufficient representation in the training dataset. Another possible example is an artificial turf such as an open-air stadium. The model can erroneously treat it as a landcover with high NIR value. On the other hand, if we add a significant amount of such samples, it is possible that the model learns such a distribution both.

Pix2pix architecture includes 54M parameters in the generator part and 6M parameters in the discriminator part. The future study can be focused on trainable parameters reduction. Light-weighted neural network models are studied currently and show promising results in the remote sensing domain [He et al., 2021] (just 4M parameters are used).

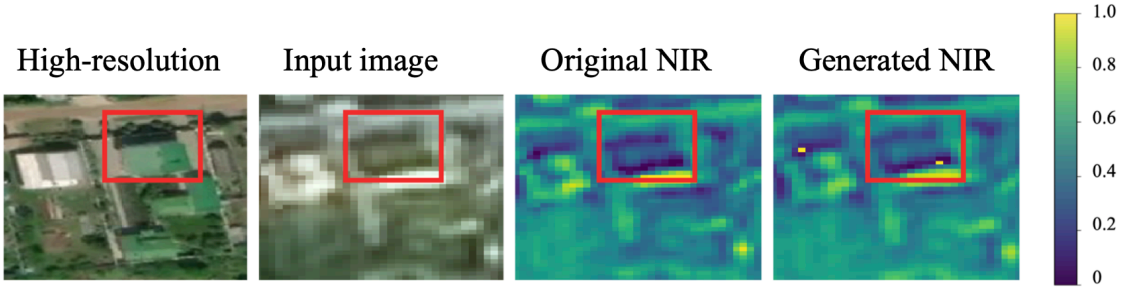


Figure 7-7: Example of a case with a green roof (SPOT image). The green roof has low NIR values both for original and generated NIR bands.

Training models independently for each data source often leads to better results. However, it is a more expensive approach. In this study, we considered the case when we minimize the cost. In future research, separate models training for each datasource should be studied and analyzed.

Data providers aim to minimize time and other costs while providing imagery to customers. For this purpose, online services for data acquisition are created [Securewatch, OneAtlas] that allows one to analyze data "on the fly". The most spread and cheap format for such platforms is RGB images, even when original imagery includes more spectral channels. The proposed NIR generation approach can be implemented for such products as "basemaps". That requires further study.

In the future, we seek to implement this feature-engineering approach to other remote sensing tasks, such as agriculture classification and land-cover semantic segmentation. In addition, the proposed approach holds potential to solve challenges when only drones RGB channels are available. Another direction is to combine this feature-engineering approach with different augmentation techniques for remote sensing tasks [Yu et al., 2017, Illarionova et al., 2021b].

It is promising to investigate the application of NIR generation methods beyond remote sensing problems in future works. Since NIR provides valuable auxiliary data in plant phenotyping tasks, NIR generation can be extended for greenhouses where high precision is vital [Nesteruk et al., 2021].

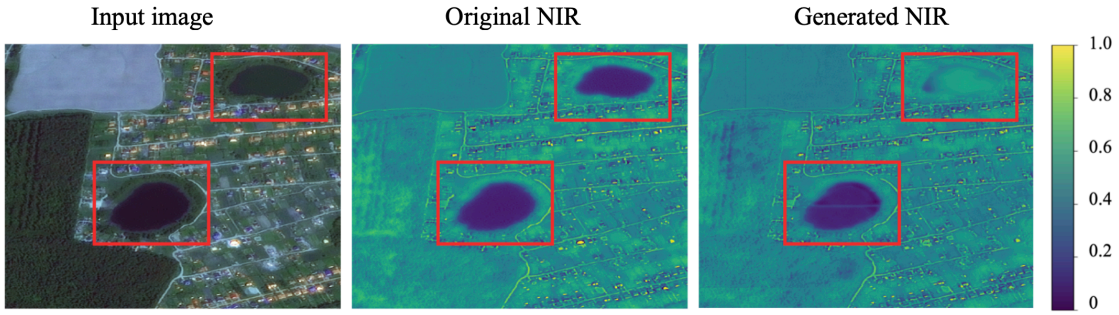


Figure 7-8: Example of a failure case (SPOT image). Green lake is erroneously treated as a surface with high NIR value.

7.4 Conclusions

The NIR band contains essential properties for landcover tasks. However, in particular cases, this band is not available. This study investigated Pix2pix cGAN implementation for image-to-image translation from RGB space imagery to the NIR band. We proposed an efficient feature-engineering approach based on an artificial NIR band generation. We conducted forest segmentation experiments to assess the importance of the NIR band in cases of small datasets and different satellite data sources. The proposed approach improved the model's robustness to data source diversity and reduced the requirement to mark the dataset size, which is crucial for machine learning challenges. We assume that this data generation strategy can be implemented in practical tasks that require the NIR channel. This method can be extended to other spectral channels and remote sensing data sources.

Chapter 8

MixChannel: Advanced Augmentation for Multispectral Satellite Images

8.1 Introduction

Freely available remote sensing images with medium spatial resolution allow solving various environmental tasks using advanced computer vision tools such as convolutional neural networks (CNN) [Jia et al., 2021]. In comparison with ordinary RGB images, satellite data usually consists of multispectral bands. Larger feature dimensionality ensures solving more complicated tasks [Setiyoko et al., 2017] that would not be possible to solve just by using the RGB spectrum in case of medium spatial resolution (such as 10 m per pixel) [Wicaksono et al., 2019]. Therefore, the lack of texture information can be efficiently compensated by a wide spectral range. However, larger feature space poses extra complexity to features connection that describes target objects. Changes in this relationship can lead to a severe CNN model deterioration for new images.

In most works for relatively small remote sensing datasets, model robustness to new territories and images is still beyond the study's scope. Splitting into training and testing objects is conducted within the same images, and only objects' locations vary. For instance, in [Wicaksono et al., 2019], they used just a single image from

WorldView-2 for tropical seagrass classification. In [Saralioglu and Gungor, 2020], they also used a single WorldView-2 image both for training and validation in the task of land cover semantic segmentation. The same imagery limitations were faced in [Erinjery et al., 2018] (two Sentinel-2 images were considered). It can lead to particular challenges trying to implement the trained models on new data. For instance, when the target territory for prediction does not have cloud-free images for the exact dates used during model training. One of the approaches to overcome this problem is discussed in [Zhou et al., 2020] where authors developed the spatiotemporal image fusion approach based on pixels replacement for cloudy image reconstruction. However, computer vision (CV) model generalization in such cases is usually not studied.

In remote sensing tasks, more than one image covering the same area for different dates is usually available. Therefore, we provide a brief overview of this topic. Additional satellite images complement the spectral information, and a multi-temporal dataset increases a model's predictive power [Persson et al., 2018b]. Combining multi-year imagery observed from a single sensor during different parts of the growing season allows one to evaluate a complete vegetation growth trajectory. However, in practice, time series can be boisterous due to the incomplete recording of the vegetation life cycle [Zeng et al., 2020]. Therefore, the main approaches for multi-temporal data leveraging are: find optimal observation dates for a particular study case and available images [Skriver, 2011]; aggregate images for different dates by averaging [Viskovic et al., 2019].

In [Viskovic et al., 2019], they proposed a method for agricultural field classification that relies on multi-temporal properties of Sentinel-2A and Sentinel-2B satellite images. A sequence of images during the year was collected and aggregated by averaging pixel values with the exact location for each band. Then, standard vegetation indices were computed to train classification models. The specificity of the study region, namely California, is a vast amount of cloudless images per year (24 to 37 images, depending on a geographical area) that would not be available for boreal territories. Thus, the described approach should be verified in the case of minimal satellite observations. In [Watkins and van Niekerk, 2019], they used seven

cloud-free Sentinel-2 images for agriculture field boundary delineation. The edge detection algorithm was implemented for red, blue, green, and near infrared (NIR) bands and resulted in an individual edge layer for each band. Then, the same as in [Viskovic et al., 2019], multi-temporal properties were used, combining edge images for different dates into one composite.

To overcome the limitation in the number of available training images, it is common to use image augmentation. It adds variability to the data and therefore makes a model more robust [Buslaev et al., 2020b]. Among popular image augmentations, there exist basic geometrical transformations and color transformations that applied to the original image. Another approach is to generate new training samples with generative adversarial networks (GANs) [Yi et al., 2017]. All of the listed approaches are successfully applied for RGB images in various fields, including remote sensing [Li et al., 2021]. However, they should be additionally studied for multispectral data for the following reasons. Geometrical transformations do not provide enough variability for satellite images with medium spatial resolution (such as 10 meters per pixel). It is complicated to apply color transformations for such multispectral data in the environmental domain, where dependencies between channels are more crucial than in general CV tasks with high-resolution RGB data. No works successfully use GANs for multispectral satellite image augmentation to the best of our knowledge. This work presents an augmentation approach that targets multispectral images and does not require training auxiliary models to generate samples.

In this study, we explore the efficiency of CNNs to learn spectral characteristics in the case study of conifer and deciduous boreal forests classification using Sentinel-2 [Drusch et al., 2012] images. A straightforward approach for training a CNN classification model is to take a set of available satellite images for a given territory during a period of active vegetation. The training set is constructed by taking a random patch of a large image, see Section 8.2.3 for details. However, if we test the obtained model for the image, taken on the date that was not included in the training set, the accuracy can drop dramatically. This situation gets even worse when the model is tested on new territory. It is supposed that the accuracy drop mentioned above happened due to changes in the characteristics of the

distribution (see Section 8.2.2 for examples).

In this Chapter, we propose a novel MixChannel augmentation method aiming to address robustness for multispectral satellite (Sentinel) images. We enlarge the training dataset generating new samples artificially with the following procedure. The method is based on substituting bands from original images with the same bands from images of another date covering the same area. While all available images are used during training, only a single image is required for inference time. For this study, only summer images of the active vegetation period are used for conifer and deciduous species classification. We trained CNN models with different architectures to compare the proposed method with the standard augmentation techniques. The result of our MixChannel augmentation consistently outperforms commonly used normalization and augmentation strategies.

The main contributions of this research are:

- We showcase the problem of poor generalization of CNNs for multispectral satellite images of middle resolution.
- We propose a simple and efficient augmentation scheme that improves CNN model generalization for multispectral satellite images.
- We test the proposed method on conifer and deciduous forest types classification and show that our approach outperforms state-of-the-art solutions.
- We show that the MixChannel approach can be efficiently combined with other methods to achieve the synergy effect.

This Chapter finalizes the research conducted as a part of the PhD study. It comprises an important question of model robustness through different territories and observation dates. The presented Chapter shares ideas and findings on augmentation approaches for multispectral satellite imagery. It supports the previous Chapters with additional useful tools for computer vision methods in Remote Sensing of Environment.

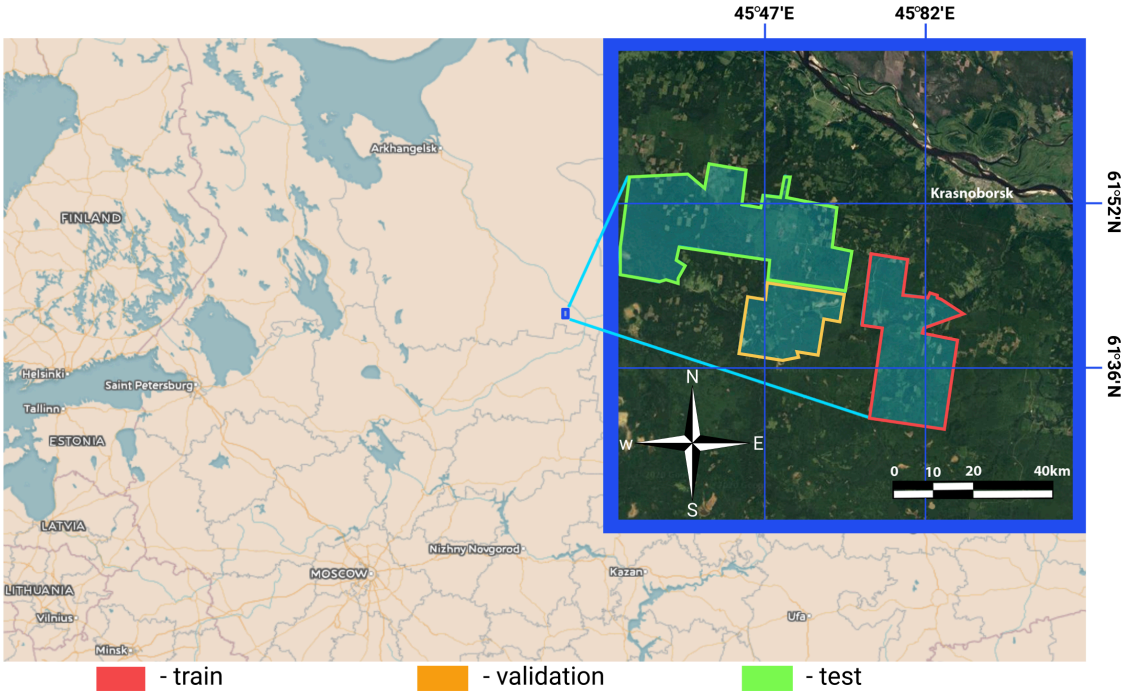


Figure 8-1: Investigated region. Selected train, validation, and test sub-areas with available ground truth labels used for image data samples creation.

8.2 Materials and Methods

8.2.1 Study Area and Dataset

The study area is located in the Arkhangelsk region of northern European Russia with coordinates between 45°16′ and 45°89′ longitude and between 61°31′ and 61°57′ latitude that belongs to the middle boreal zone (Figure 8-1). The total area is about

For the study, we used forest inventory data collected according to the official Russian inventory regulation. This data was organized as a set of individual stands with appropriate characteristics based on the assumption that the stand was homogeneous. We used such a characteristic as dominant species and canopy height for an additional experiment. Thus, inventory data was converted in a raster map of dominant conifer and deciduous classes and a raster with height values. The assumption on homogeneous means that for particular stands defined as conifer or deciduous dominant types, these individual stands can contain another class representative (but less than 50%). We excluded from the study non-forest areas and areas with the equivalent conifer and deciduous composition.

Table 8.1: Sentinel images used in this study. Date format is: month, day, year.

	Image ID	Date
0	L2A_T38VNP_A016606_20180827T083208	08.27.18
1	L2A_T38VNP_A010986_20170730T082009	07.30.17
2	L2A_T38VNP_A005695_20160725T082012	07.25.16
3	L2A_T38VNP_A007297_20180730T081559	07.30.18
4	L2A_T38VNP_A015748_20180628T082602	06.28.18
5	L2A_T38VNP_A013017_20190903T081606	09.03.19

8.2.2 Satellite Data

The data source used in this study is Sentinel-2 satellite multispectral images. Sentinel-2 satellite is a part of the Sentinel program with a mission focusing on high-resolution land cover monitoring. It was launched in 2015. Sentinel includes 13 spectral bands with a spatial resolution of 10, 20, and 60 m.

For the forest classification task, we selected images over the vegetation period between the years 2016 and 2019 close to the date of taxation. The study region is boreal forests with high cloud coverage during a year; therefore, the number of appropriate imagery was severely limited. The available image IDs selected for the study are presented in Table 8.1.

We downloaded Sentinel data in L1C format from EarthExplorer USGS [USGS, Accessed: 2020] and preprocessed them using Sen2Cor [Sen2Cor, Accessed: 2020]. Sen2Cor is a semi-empirical algorithm that removes atmospheric effects from Sentinel-2 images and creates a level L2A Bottom of Atmosphere (BoA) reflectance product. This atmospheric correction processor is based on a set of Look-Up tables created by libRadtran model [Martins et al., 2017]. Preprocessed data are more suitable for further analysis than level L1C product. The pixel values were in the range $[0, 10,000]$. We used 10 bands with the following central wavelengths [Drusch et al., 2012]: Band 2: Blue, 492.4 nm; Band 3: Green, 559.8 nm; Band 4: Red, 664.6 nm; Band 5: Red-edge I (R-edge I), 704.1 nm; Band 6: Red-edge II (R-edge II), 740.5 nm; Band 7: Red-edge III (R-edge III), 782.8 nm; Band 8: Near infrared (NIR), 832.8 nm; Band 8A: Narrow Near infrared (NNIR), 864.7 nm; Band 11: Shortwave infrared-1 (SWIR1), 1613.7 nm; Band 12: Shortwave infrared-2 (SWIR2),

2202.4 nm. The bands at 20 m resolution were adjusted to 10 m resolution before classification using the same procedure discussed in [Erinjery et al., 2018].

The average values for each channel and each image within forested areas are presented in Figure 8-2. Here, in the plot for the entire study area, it is shown that the distribution of the mean values for images changes drastically. Even images of the same day but one year apart (images with IDs 1 and 2 for the 30 July 2017, and 2018 respectively) have markedly different mean spectral values. Moreover, for each band, changes are not equivalent. Figure 8-2 also presents three random crops 200×200 pixels each. It is shown that depending on a particular area, the mean values for each band change. Therefore, it is impossible to bring auxiliary training data within the same image distribution using linear transformations or noise.

For classification tasks using CNN, image values are often brought to the interval from 0 to 1 [Vaddi and Manoharan, 2020, Debella-Gilo and Gjertsen, 2021]. It can be done using different approaches. The first approach is to divided by the maximum value such as in [Illarionova et al., 2020]. In our case, this values is 10,000 (the maximum physical surface reflectance value for Sentinel-2 in level L2A):

$$I' = I/10{,}000. \tag{8.1}$$

Another way is to normalize data by the min–max normalization technique. In satellite remote sensing domain, it was used in [Prathap and Afanasyev, 2018] and aims to reduce noise of each channel:

$$m = max(0, mean(I) - 2 * std(I)), \tag{8.2}$$

$$M = min(max(I), mean(I) + 2 * std(I)), \tag{8.3}$$

$$I' = (I - m)/(M - m), \tag{8.4}$$

where $mean, std$ are the mean and standard deviation of the image. In Equations (8.2) and (8.3), we calculate m and M (minimum and maximum of the preserved dynamic range). In Equation (8.4), values are scaled to 0 and 1 linearly.

We used both normalization techniques for evaluating our proposed approach

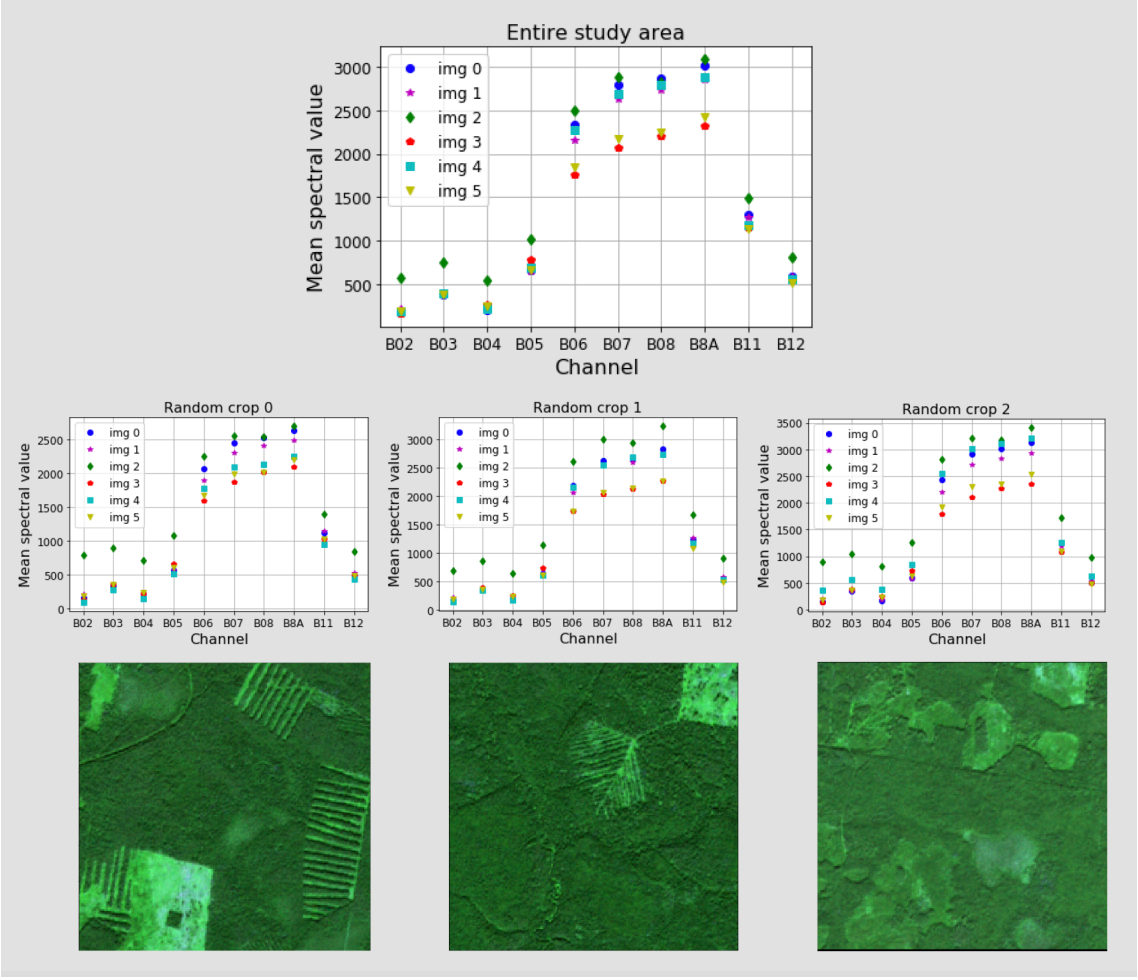


Figure 8-2: Example for mean values for each channel for entire study area and for random image crops (the crop size is 200 × 200 pixels). The mean values are calculated from the extracted spectral information in the forested areas.

(see Section 8.2.4).

8.2.3 Baseline Description

We solve the image semantic segmentation task where a CNN model is trained to create an output map with target classes for each pixel by processing a multispectral input image. Therefore, the output consists of pixels for which forest types are assigned. The batch for model training is formed as follows. For each patch in a batch, one image is chosen from the image set, and a patch of predefined size is cropped randomly. The batch and the patch sizes are presented in Section 8.2.6. A patch consists of 10 multispectral normalized bands, and it is used as a ten-layer input for a CNN model instead of the usually used three-layer input tensor. For model

training, namely model loss function computing, masks with target values are given for each patch. The CNN architecture for the baseline model is U-Net [Ronneberger et al., 2015].

8.2.4 MixChannel Augmentation

The proposed MixChannel augmentation algorithm operates by substituting some channels of the original image by channels from the other images that cover the same territory (Algorithm 1). MixChannel takes the set of images of the exact location, chooses one as an anchor image, and with the predefined probability substitutes some channels of the anchor image with the matching channels from non-anchor images from the same set. The workflow of the developed augmentation algorithm, in particular, the creation of the new data sample, is schematically presented in Figure 8-3.

Input: S, P
Output: I
$I \subseteq S$, $\#I = 1$
$\acute{S} = S \setminus I$
for $c \in \{0, 1, \ldots, C-1\}$ **do**
  if $P_C > R$ **then**
    $\acute{I} \subseteq \acute{S}, \#\acute{I} = 1$
    $I_C = \acute{I}_C$
  end
end

Algorithm 1: MixChannel $\mathcal{T}(S, \acute{P})$

$\mathcal{T}()$ is the MixChannel algorithm; $S, \#S \geq 1$ is the set of images covering the same area; $P = \{p_0, p_1, \ldots, p_{C-1}\}, p \sim \mathcal{U}([0, 1])$ is the set of probabilities to substitute each channel; $I, I \in S$ is the anchor image; C—is the number of channels in images; $R \sim \mathcal{U}([0, 1])$ is a random variable from the uniform distribution; I_C is the c-th channel of the image I; letters with the stroke sign denote temporal variables.

The probability choice of channel substituting is an essential parameter of the algorithm to be studied. Therefore, we considered different probabilities with the step of 0.1. The range was set from 0 to 0.7 where 0 probability is equal to the absence of the MixChannel augmentation and defined as a baseline. To compare the proposed

augmentation with other approaches, we conducted the following experiments (see the short summary of experiments in the Table 8.2):

- Average-channel. This experiment is based on the approach proposed for multispectral Sentinel data in [Viskovic et al., 2019]. The idea of the method is described in Section 8.1. For each pixel of the particular band, the corresponding value is averaged within all images that cover the same territory.

- Channel-dropout. In this experiment, we used augmentation described in [Tompson et al., 2015] where it was proposed for RGB images. It aims to prevent a CNN model from overfitting for particular data. Our study implemented this approach by substituting each channel with the predefined probability by zero values. We investigated different probabilities in the range from 0 to 0.5 with the step of 0.1.

- Color jittering. Color jittering [Taylor and Nitschke, 2018] is commonly used for RGB image augmentation. In the color jittering experiment, we multiply values in each band by the random value (fixed within each band) in the range of 0.8–1.2. The approach aims to add variability to the initial data.

- Patching. As an additional experiment, we implemented MixChannel augmentation for patch parts independently. The patch was divided into four equal parts; for each part, channels can be substituted by bands from different images.

- Optimization. In this experiment, we search for the optimal probabilities for band substitution using a greedy optimization approach. The detailed description of the MixChannel optimization procedure is presented in Section 8.2.8.

- Height adding. In this experiment, we complemented the spectral data with height data and used them both as input data for CNNs. Experiments MixChannel augmentation for data that include height and Baseline + height are described in details in Section 8.2.5;

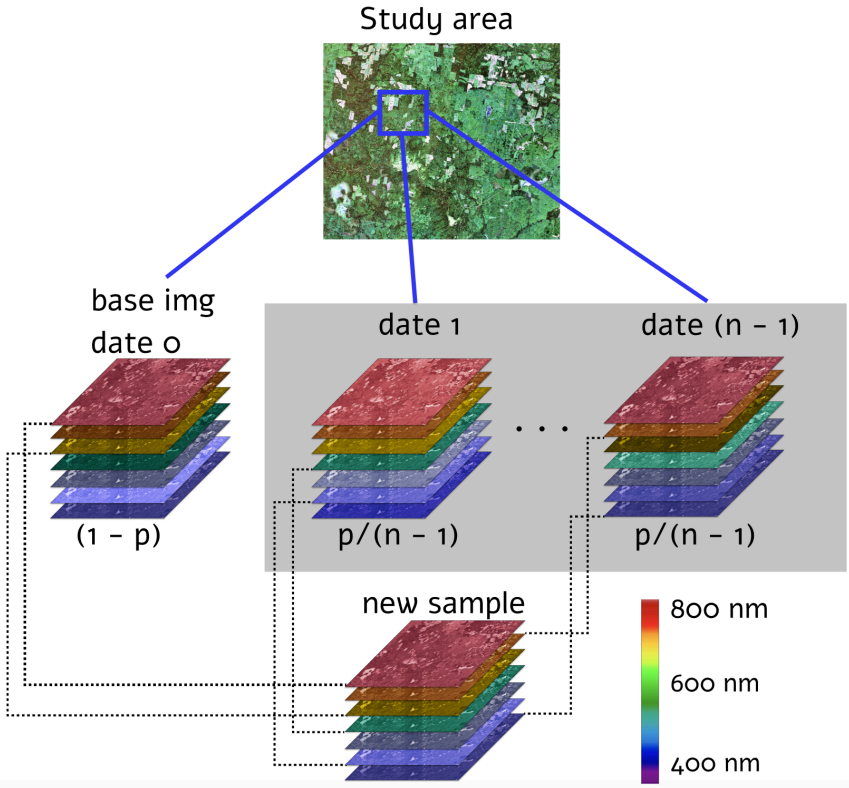


Figure 8-3: MixChannel algorithm. Schematic workflow of new image sample creation using spectral channels from other images in the investigated region with certain probabilities.

For all experiments except channel-normalization, data were normalized using the Equation (8.1) described in Section 8.2.2. In the Channel-normalization experiment, we used Equation (8.4) for data preprocessing. In all experiments, geometrical transformations such as rotation and random flip were applied.

8.2.5 Height Data for Stronger Robustness

As was previously shown in [Illarionova et al., 2020], additional height data can significantly improve model performance in the forest species classification task. Therefore, we conduct further experiments to evaluate extra height data importance for model robustness in new images and territory. We also check the assumption that MixChannel can be efficiently combined with other techniques to achieve the so-called synergistic effect.

For this experiment, height measurements from inventory data were converted into raster by assigning the same height value to each pixel within an individual

Table 8.2: Experiments description.

No.	Method	Description
1	Baseline	Without any data transformations or aggregations (except geometrical).
2	Baseline + height	Add extra input layer with height values.
3	Channel normalization	Use normalization defined in Equation (8.4).
4	Average-channel	Aggregate images for various dates by averaging.
5	Channel-dropout	Substitute random channels with zero values.
6	Color jittering	Multiply each channel by a random value.
7	MixChannel	Our approach.
8	MixChannel + height	Add extra input layer with height values.

stand. This layer was normalized by dividing by 100 and clipping into $[0, 1]$ range to have the same range as multispectral input data for a CNN model. The obtained layer was stacked to initial input layers to add additional information to our model.

8.2.6 Neural Networks Models and Training Details

To evaluate the MixChannel approach on different CNN architectures, we considered U-Net [Ronneberger et al., 2015], U-Net++ [Zhou et al., 2018], and DeeplLab [Chen et al., 2017b]. For all mentioned architectures, we use ResNet-34 [He et al., 2016] encoder. As a base architecture, we choose U-Net. The models' architecture implementation was based on opensource library [Yakubovskiy, 2022] and used PyTorch framework.

For each model, we set the following training parameters. There were 50 epochs with 32 training steps per epoch and the same for validation. An Adam optimizer [Kingma and Ba, 2014] with a learning rate of 0.001, which was reduced after 25 epochs. Early stopping was chosen with the patience of 10. The best model according to the validation score was considered. The batch size was specified to be 16 with a patch size of 256×256 pixels. These sizes were chosen to meet memory restrictions for computing using one GPU. For each model, the activation function for the last layer was Softmax [Gao and Pavel, 2017]. As a loss function, categorical cross entropy (3.1) was used.

The training of all the neural network models was performed at Zhores [Zacharov

et al., 2019] supercomputer with 16Gb Tesla V100-SXM2 GPUs.

8.2.7 Evaluation

Cross-validation is an effective technique for machine learning model assessment [Roberts et al., 2017]. It makes model evaluation more reliable. However, in most works for relatively small datasets (where the study area can be covered by a single satellite tile), splitting for testing and training samples is performed only within the same images. Moreover, the cross-validation technique is not so popular for CNN tasks because it requires extra computational resources. In cases of CNN, fixed splitting into testing and training areas is often used [Nesteruk et al., 2021]. This study implements an image-based cross-validation approach to evaluate CNN model robustness both for new images and territory for a relatively small dataset.

Splitting into folds for cross-validation was organized as follows (Figure 8-4). Test, train, and validation territories are shown in Figure 8-1. Six images were used (see Table 8.1). For each fold, one image was set aside for testing, while the other five images were leveraged to train a model in only the training territory (see Figure 8-1). Validation was conducted using the same five images but for the validation territory. Thus, the reported result is reliable because it was obtained on unseen images and territories and aggregated across five cross-validation folds.

The model outputs masks of two target classes, which are compared with the ground truth by pixel-wise F1-score. For each experiment, a model was trained three times with different random seeds for averaging model performance on different initialization of trained parameters.

8.2.8 Optimization

The MixChannel algorithm supports changing the probabilities to substitute image channels (see Algorithm 1). Different values of probability have various effects on the final accuracy and robustness of the trained model. Thus, a task of channel substitution probabilities optimization appears. Optimization of these probabilities leads to better results and will be shown in Section 8.3. However, it should be noted

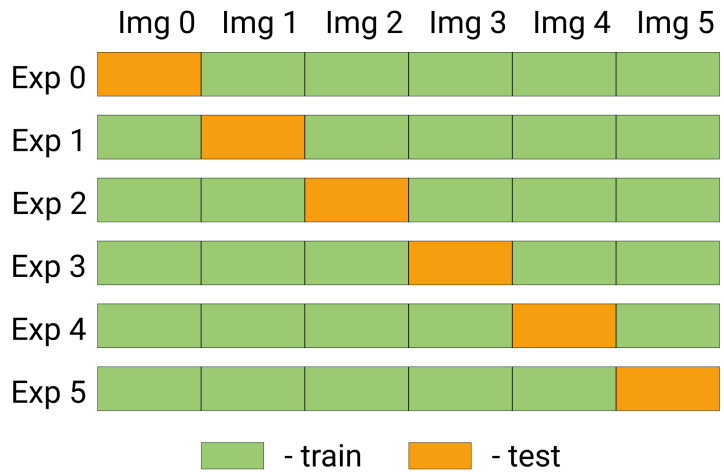


Figure 8-4: Cross-validation scheme. Each experiment (Exp) in the cross-validation procedure iteratively uses one image (Img) that represents the whole study area at the certain time as the test (only test sub-area according to Figure 8-1). Training data for CNNs is generated from the train sub-area (see Figure 8-1) of the rest images.

that performance evaluation using each selected probability set requires a full model training cycle. Therefore, it is very computation-intensive to iterate over all possible options. More precisely, it would have exponential complexity with respect to the number of channels.

When computational resources are minimal, the baseline approach assumes that the optimal values for all channels are the same. Then, it is possible to iterate over several probability values and set a single global substitution probability to each channel. The advantage of this approach is that it has constant complexity with respect to the number of image channels because it iterates only over substitution probabilities and does not explore interactions between channels. It allows finding suboptimal probabilities but does not consider that optimal probability may vary severely for some channels. This section proposes a greedy optimization scheme that aims at finding optimal channel substitution probabilities.

Let $\mathcal{J} : H \rightarrow \mathcal{R}$ be the objective function. $\mathcal{J}$ maps hyperparameters H that include model, MixChannel parameters and dataset to the resulting F_1-score value. Then, the optimization problem formulates as $P^\star = \underset{P}{argmax}\mathcal{J}(\theta^\star|\mathcal{T}(S, P))$.

The greedy optimization algorithm for MixChannel probabilities tuning operates by iteratively searching for the optimal substitution probability for each channel with

other channels' probabilities fixed to sub-optimal values (Algorithm 2).

Input: S, q, n, p_{max}
Output: $\theta^\star, P, r$
$P = \{0, 0, \ldots, 0\}, \#P = C$
$r = 0$
for $iter \in \{0, 1, \ldots, n-1\}$ **do**
  for $c \in \{0, 1, \ldots, C-1\}$ **do**
    for $p \in \{0, p_{max}/q, 2p_{max}/q, \ldots, p_{max}\}$ **do**
      $\acute{P} = P$
      $\acute{P}_C \leftarrow P$
      $\acute{r} = \mathcal{J}(\theta^\star | \mathcal{T}(S, \acute{P}))$
      if $\acute{r} > r$ **then**
        $P = \acute{P}$
      end
      else
        $P = P$
      end
      $r = max(r, \acute{r})$
    end
  end
end

Algorithm 2: Greedy MixChannel Optimization

$\theta^\star$—optimal model weights found via the gradient descent algorithm for the defined hyperparameters; q is the the number of probability quantization levels; n is the number of iterations; $p_{max} \leq 1$ is the is the highest considered value of probability; r is the the F_1-score of the trained model with the considered hyperparameters; v is the the number of images in the dataset covering the same area.

The described optimization algorithm considers the effect of each channel on every other channel. It can be efficiently applied because it has linear complexity with respect to the number of image channels.

8.3 Results

This section describes the results of the experiments with MixChannel and compares them with other approaches.

MixChannel Augmentation

Table 8.3 presents details of MixChannel performance. Considering the small number of available training samples, Table 8.3 shows cross-validation results to increase the reliability of the score. Each model is trained on five training images and is validated on the remaining one image. Columns represent a single global substitution probability, set to each channel. Zero probability means that the MixChannel algorithm is not applied. For a more straightforward interpretation, results for each model aggregated to show average and standard deviation. Bold font highlights the best result for each model. It should be noted that a better model must have a higher F1-score but a lower standard deviation.

The baseline model shows poor performance for particular images (Figure 8-5). It leads to a low average score (0.696) and a high standard deviation (0.17) (see Table 8.3). The model with the same CNN architecture, namely U-Net, but trained using the proposed MixChannel augmentation, beats the baseline approach confidently. For the best substituting probability, it achieves an F1-score of 0.77 for U-Net architecture. Moreover, the model performance for each test image became more stable. One of the outstanding results is that, for some cases, by using MixChannel augmentation we were able to double the scores. For example, an image with ID 5 was complex for the baseline approach (F1-score 0.381) and after application of MixChannel augmentation the F1-score doubled and reached 0.775. The drop of the average standard deviation from 0.17 to 0.069 proves that MixChannel enables better model generalization. We compared different probabilities for channel substituting. For the U-Net model, the best one is 0.6. However, it is clear that the proposed approach leads to higher results even with not the optimal substitution probability.

To evaluate the MixChannel augmentation for different CNN architectures, we conducted experiments with U-Net (as the base model), DeepLab, and U-Net++. Our approach confirms to be preferable for each architecture choice than the baseline approach trained for the same architecture. Moreover, as shown in Table 8.3 the best score for each architecture is approximately equals to 0.77. However, the best probability for channel substituting differs: for U-Net, it is 0.6, for U-Net++ 0.1, and for DeepLab, it is 0.3. Unfortunately, we cannot expect the optimal substi-

Table 8.3: MixChannel predictions with different channels replacing probabilities (F1-score). Bold text in each row indicates the best result for the model.

Model	**Probabilities**	**0 (Baseline)**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**
U-Net	Test image 0	0.8	0.762	0.79	0.8	0.77	0.813	0.815
	Test image 1	0.607	0.606	0.59	0.605	0.58	0.611	0.625
	Test image 2	0.86	0.829	0.83	0.81	0.835	0.84	0.826
	Test image 3	0.849	0.814	0.825	0.82	0.815	0.83	0.825
	Test image 4	0.675	0.733	0.76	0.745	0.725	0.771	0.775
	Test image 5	0.381	0.72	0.71	0.685	0.685	0.77	0.775
	Average	0.696	0.744	0.75	0.744	0.735	**0.77**	**0.77**
	Standard deviation	0.17	0.073	0.082	0.076	0.086	0.077	**0.069**
Deeplab	Test image 0	0.804	0.793	0.784	0.803	0.817	0.805	0.806
	Test image 1	0.614	0.631	0.615	0.633	0.633	0.636	0.615
	Test image 2	0.855	0.811	0.824	0.829	0.832	0.833	0.829
	Test image 3	0.851	0.834	0.82	0.824	0.812	0.809	0.821
	Test image 4	0.697	0.76	0.761	0.789	0.774	0.771	0.777
	Test image 5	0.38	0.664	0.758	0.784	0.722	0.742	0.759
	Average	0.7	0.749	0.76	**0.777**	0.765	0.766	0.768
	Standard deviation	0.167	0.076	0.069	**0.066**	0.069	**0.066**	0.0725
U-Net++	Test image 0	0.79	0.803	0.819	0.824	0.825	0.814	0.817
	Test image 1	0.49	0.639	0.61	0.618	0.64	0.648	0.605
	Test image 2	0.861	0.837	0.832	0.811	0.837	0.834	0.837
	Test image 3	0.851	0.826	0.823	0.822	0.809	0.83	0.828
	Test image 4	0.6	0.795	0.795	0.765	0.739	0.789	0.775
	Test image 5	0.38	0.761	0.64	0.735	0.768	0.719	0.774
	Average	0.66	**0.777**	0.753	0.762	0.769	0.772	0.773
	Standard deviation	0.185	**0.066**	0.091	0.072	0.067	0.069	0.079

tution probability to be the same for each model because it is a hyperparameter, and therefore should be tuned for each new case. Every model represents features differently, and augmentation affects these representations differently.

We compared MixChannel performance with the popular solutions for multispectral data. The first experiment was focused on the standard normalization techniques implemented to enhance image spectral properties. As presented in Table 8.4, image normalization did not lead to F1-score improvement (0.678) compared to the baseline (0.696) where spectral values were dividing by the max possible value. Another considered approach for multispectral augmentation was Channel-dropout. As shown in Table 8.5, it outperforms the baseline model with the best F1-score of 0.753. However, it still does not achieve MixChannel's results. We also compared our approach with channel averaging. As presented in Table 8.4, it did not improve the baseline model results achieved 0.672 F1-score. Color jittering also did not outperform MixChannel (F1-score 0.685).

Experiments with additional height data are presented in Table 8.4. Both for the baseline and MixChannel approaches, it leads to the higher results. For MixChannel F1-score improves from 0.77 to 0.81, while for the baseline, F1-score increases from 0.696 to 0.74.

Table 8.4: MixChannel comparison with other approaches. Predictions for U-Net models (F1-score). Results of MixChannel application are in blue. Bold text indicates the best result that was obtained by application of MixChannel with height. Avg is average value, Std is Standard deviation.

Img#	**Baseline**	**Normalization**	**Average Channel**	**Color Jittering**	**Channel Dropout**	**MixChannel**	**Baseline + Height**	**MixChannel + Height**
0	0.8	0.839	0.786	0.806	0.809	0.813	0.812	0.845
1	0.607	0.408	0.495	0.551	0.56	0.611	0.605	0.66
2	0.86	0.79	0.844	0.865	0.806	0.84	0.872	0.85
3	0.849	0.859	0.855	0.853	0.816	0.83	0.879	0.865
4	0.675	0.487	0.67	0.579	0.752	0.771	0.73	0.835
5	0.381	0.685	0.38	0.457	0.778	0.77	0.567	0.8
Avg	0.696	0.678 (−1.8%)	0.672 (−2.4%)	0.685 (−1%)	0.753 (+5.7%)	0.77 (+7.5%)	0.74 (+4.5%)	**0.81** (+11%)
Std	0.17	0.175 (+0.005)	0.179 (+0.01)	0.162 (−0.01)	0.089 (−0.08)	0.077 (−0.1)	0.12 (−0.05)	**0.069** (−0.1)

Table 8.5: Channel-dropout predictions for U-Net with different channels replacing probabilities (F1-score). Bold text indicates the best result that was obtained by application of Channel-dropout.

Probabilities	**0 (Baseline)**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**
Test image 0	0.8	0.802	0.802	0.809	0.794	0.761
Test image 1	0.607	0.57	0.576	0.56	0.504	0.55
Test image 2	0.86	0.814	0.81	0.806	0.791	0.775
Test image 3	0.849	0.804	0.803	0.816	0.791	0.624
Test image 4	0.675	0.752	0.753	0.752	0.756	0.737
Test image 5	0.381	0.689	0.739	0.778	0.766	0.733
Average	0.696	0.738	0.747	**0.753**	0.733	0.696
Standard deviation	0.17	0.086	**0.081**	0.089	0.1	0.0816

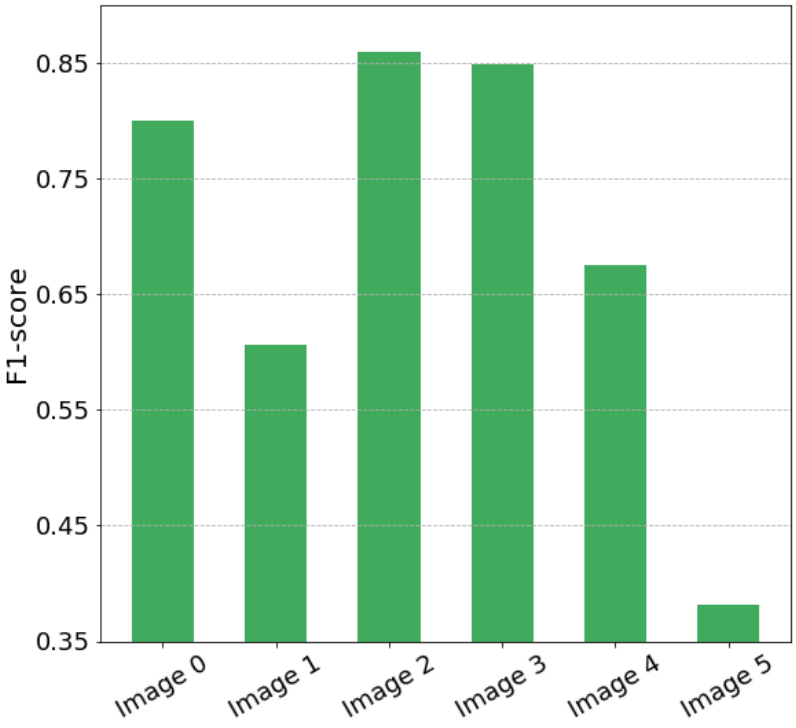


Figure 8-5: Baseline prediction.

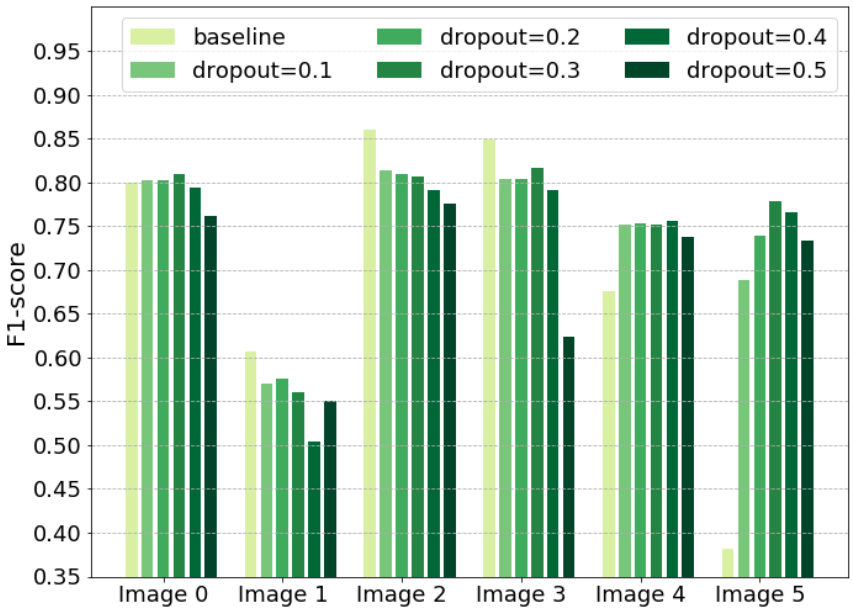


Figure 8-6: Channel-dropout predictions.

Table 8.6: MixChannel for four crop parts (F1-score). Bold text indicates the best result that was obtained by application of MixChannel for four crop parts.

Probabilities	**Baseline**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**
Test image 0	0.8	0.798	0.81	0.8	0.798	0.806	0.77
Test image 1	0.607	0.595	0.594	0.624	0.585	0.61	0.616
Test image 2	0.86	0.833	0.83	0.833	0.819	0.835	0.835
Test image 3	0.849	0.828	0.823	0.823	0.82	0.815	0.823
Test image 4	0.675	0.782	0.739	0.77	0.754	0.768	0.77
Test image 5	0.381	0.597	0.674	0.615	0.758	0.71	0.72
Average	0.696	0.738	0.745	0.744	0.756	**0.757**	0.755
Standard deviation	0.17	0.1	0.0869	0.09	0.08	0.077	**0.073**

In this section above, we showed that the MixChannel algorithm consistently improves model accuracy even with default substitution probabilities. Then, we showed that it is possible to obtain better results tuning a single global probability for each channel (Table 8.3). Our further experiments show that Algorithm 2 allows finding optimal substitution probabilities separately for each channel. Our optimization setup is as follows. We used the U-Net model; two algorithm iterations n; initial probability values $P = \{0.5, 0.5, ..., 0, 5\}$; the highest probability value $p_{max} = 0.7$; the number of considered probability values $v = 8$. It gives us 160 model training loops in total and increased the previous best result by 1% from 0.777% to 0.791. It is a minor improvement, but it shows that MixChannel can be tuned further. However, for the practical application, we suggest using global probability tuning because it can noticeably increase model accuracy in a few iterations and can be performed in a parallel fashion.

In this Chapter we use standard deviation to assess prediction quality through different images. Considering standard deviation in a more common sense, we can evaluate statistical significance of the achieved results for described approaches. For each experiment, standard deviation through different algorithm running did not exceed the value of 0.005. It confirms the advantage of the proposed approaches comparing with the baseline method.

8.4 Discussion

Usually, in the remote sensing domain, we do not have a sufficient amount of well-labeled training data for solving particular tasks. The main limitation in getting more data for boreal regions is cloud coverage. Obtaining new labeled data is a time-consuming and costly process because it is often necessary to conduct field-based measurements. Therefore, it is practically reasonable to find techniques that will allow us to enhance the existing image datasets in order to obtain better results in CV models with minimal additional enforces. One of the commonly-used approaches for enhancing the dataset characteristics is image augmentation. However, as is shown above, the standard augmentation techniques are not able to principally improve

the scores of trained on multispectral data models. Thus, it is natural to use the distinctive feature of multispectral image data, namely different spectral channels. We showed that generic image augmentations that include color jittering and changing brightness do not ensure robustness for new multispectral images (Table 8.4). Randomly changing color values in different channels pushes the augmented image out of the distribution of initial images. It may lead to better model robustness against noise but does not ensure better model generalization. As shown in [Hataya et al., 2020], image augmentations that better suit the distribution of the original dataset provide better model performance than augmentations that push images out of distribution. However, it is challenging to preserve the same data distribution with multispectral images because the high number of dimensions makes it difficult to reveal the dependencies between bands.

The MixChannel augmentation algorithm proposed in this study, in contrast, tries to preserve the distribution of the original dataset. It cannot save the joint distributions across all bands, but it saves every separate bands' distribution. MixChannel substitutes some channels of the anchor image with channels from other images of the same location. The enormous number of possible channel mixing combinations ensures the increase of the number of useful training data images while preserving the distribution characteristics of the dataset. Our experiments show that MixChannel reduced both bias and variance error of all the considered models. The results of the comparison of the predictions for testing and validation areas obtained by baseline models and by using proposed augmentation are presented in Figure 8-7. From Figure 8-7 we can visually notice that the proposed approach works better and the prediction results are closer to ground truth. The MixChannel algorithm gains in model performance utilizing the availability of multiple images of the exact location. Therefore, the apparent limitation of the method is the need for more than one image at the same spot. It is suitable in such cases as remote monitoring and continuous stationary imaging. In our investigations, we mainly focus on some image channels substitution with channels from other images. More flexible schemes are also can be considered. It is possible to substitute only some parts (Table 8.6) or patches in a channel by mask instead of the entire channel.

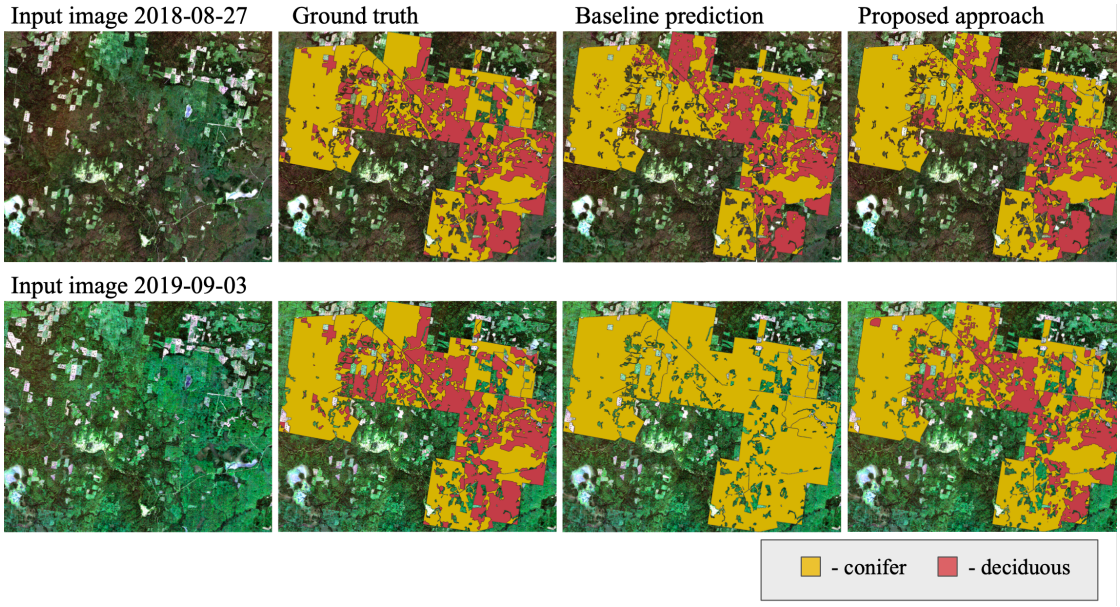


Figure 8-7: Predictions for testing and validation areas obtained by baseline models and by using proposed augmentation. F1-score for the image with date 2018-08-27 (image0) is 0.8 for the Baseline and 0.813 for MixChannel approach. F1-score for the image with date 2019-09-03 (image 5) is 0.38 for the Baseline and 0.77 for MixChannel approach (with the same U-Net architecture).

Substitution masks can be either based on segmentation masks or random.

We test the MixChannel algorithm using the images with ten channels as an input to CNN models for training them in order to distinguish two classes. In further studies, we will examine the dependency between the number of image channels and the gain of the MixChannel augmentation. It seems promising to test it with three-channel RGB images. The other possible future extension of the current research is to try out more forest species and other classification pipelines (such as a hierarchical approach for species classification described in [Illarionova et al., 2020]). Other target classes of vegetation can be studied (such as [Wicaksono et al., 2019]). For instance, it can be applied for solving some tasks in precision agriculture such as crop boundaries delineation [Watkins and van Niekerk, 2019]. Such augmentation techniques can be applied for hyperspectral data which is widely used for environmental tasks. The MixChannel algorithm allows for picking different channel substitution probabilities. Our experiments show that the optimal values are not the same for different models. Moreover, the optimal values vary from channel to channel. In practical tasks, we suggest starting with channel substitution probabilities equal

to 0.3 for all channels. Then, depending on the available computational resources, an optimization algorithm can be applied to tune the probabilities if needed.

In addition to MixChannel, we show other promising ways to achieve more robust results for CNN model predictions. Channel-dropout demonstrates significantly higher performance than Baseline approach (Table 8.5, Figure 8-6). Although Channel-dropout does not outperform MixChannel, it can be applied in cases when just a single multispectral image is available. Both MixChannel and Channel-dropout approaches prevent the model from overfitting on training images and allows extracting relevant information for better predictions. The combination of these augmentations should be studied further. Additional height data is also a powerful way to increase the model robustness (Table 8.4). It makes the model less sensitive to shifting in spectral distribution. However, height data are not often available.

The design of the MixChannel algorithm uses the variability of the spectrum from image to image. It brings new information when channel values may differ for the target object within the same part of the year. Therefore, this approach is practical for the environmental domain where vegetation characteristics correlate in diverse locations and different years but do not match exactly. In contrast, artificial objects such as buildings remain the same distribution over time and will not benefit in the same way from the MixChannel algorithm. Another limitation arises from the assumption that the objects of interest have no significant changes across the image set. For instance, any crop will differ too much before and after harvesting. Consequently, it is not recommended to apply MixChannel when images for the location are spread across the year, and a CNN model is not supposed to handle such massively different data.

8.5 Conclusions

This work examines the problem of inconsistency of convolutional neural network generalization in the remote sensing domain. The problem occurs when the training set and the test set of images are from different locations or times of the year. Image

exploration shows that even the exact locations at similar dates, but different years, can vary dramatically. It leads to model overfitting on the training set and a drop in performance dramatically on the test set. This problem is crucial when the size of the training set is small. This study proposes and evaluates a novel image augmentation approach called MixChannel. MixChannel uses multiple multispectral images of the exact location at various dates of the vegetation period to augment the training set. MixChannel was applied to the task of forest types classification in the Northern regions of Russia. This approach shows a noticeable increase in performance with all the tested convolutional neural networks, namely U-Net, Deeplab, and U-Net++. In comparison with other augmentation and preprocessing techniques popular for multispectral images, MixChannel provides better generalization. It is superior in both prediction bias and variance on the unseen test images. The average gain over the baseline solution is 7.5% from 0.696 F1-score to 0.77, while the average variance drops more than twice from 0.17 to 0.077. Further improvement was achieved by adding auxiliary heights data, giving the overall accuracy of 0.81. It proves that the proposed approach can be combined with other techniques to get the synergy effect. Our study shows that MixChannel is a promising approach that enables training more precise models for remote sensing in the environmental domain.

Chapter 9

Conclusion

The Thesis is focused on the estimation of vegetation characteristics using artificial intelligence algorithms on satellite data. The most appropriate remote sensing data was estimated to be used in the subsequent calculation of the tree species, canopy height, and forest mask. Novel approaches were proposed to address the tasks in the remote sensing domain using more available data sources covering larger areas. The developed approaches will be helpful to solve related tasks. More precise vegetation variables will allow improving environmental studies, in particular, connected with global climate changes. Overall, the main contribution of the present work is the following:

- We propose a novel neural network based approach for a high-detailed forest mask creation. It involves a novel advanced object-based augmentation approach that outperforms standard color and geometrical image transformations in particular remote sensing tasks. The presented method combines target objects from georeferenced satellite images with new backgrounds to produce more diverse realistic training samples. We implement an object-based augmentation technique for a minimum amount of labeled high-detailed data. Using this augmented data we fine-tune the models, trained on a large forest dataset with less precise labeled masks. The provided algorithm is tested for multiple territories in Russia. The developed model is available in an SAAS platform through the link [Mapflow.ai, Accessed: 10 Febuary 2022]. It allows one to easily create the detailed and precise forest mask and then use it for

solving various applied problems;

- We represent the multi-class forest classification problem as a hierarchical set of binary classification tasks, which allows us to reach better results with both high- and medium-resolution satellite imagery. We also examine supplementary data such as tree height to improve the species classification results for wider tree age diversity. The proposed approach is tested on sample territories in Leningrad Oblast of Russia, for which the field-based observations were acquired and made publicly available as a single dataset. The proposed approach shows significantly better results than a conventional multi-class classification;

- We enhance tree species classification based on a neural network approach providing automatic markup adjustment and improving sampling technique. For forest species markup adjustment, we propose using a weakly supervised learning approach based on the knowledge of dominant species content within each stand. We also propose substituting the commonly used CNN sampling approach with the object-wise one to reduce the effect of the spatial distribution of forest stands. We consider four species commonly found in Russian boreal forests: birch, aspen, pine, and spruce. We use imagery from the Sentinel-2 satellite, which has multiple bands (in the visible and infrared spectra) and a spatial resolution of up to 10 meters. A data set of images for Leningrad Oblast of Russia is used to assess the methods. This approach is promising for future studies to obtain more specific information about stands composition even using incomplete data;

- Leveraging typical data from airplane-based LiDAR (Light Detection and Ranging), we train a deep neural network to predict the vegetation height. The provided approach is less expensive than the commonly used drone measurements, and the predictions have a higher spatial resolution (less than 5 m) than the vast majority of studies using satellite data (usually more than 30 m). The experiments, which were conducted in Russian boreal forests, demonstrated a strong correlation between the prediction and LiDAR-derived measurements. Moreover, we tested the generated CHM as a supplementary

feature in the species classification task. Among different input data combinations and training approaches, we achieved the mean absolute error equal to 2.4 m using U-Net with Inception-ResNet-v2 encoder, high-resolution RGB image, near-infrared band, and ArcticDEM. The obtained results show promising opportunities for advanced forestry analysis and management. We also developed the easy-to-use open-access solution for solving these tasks based on the approaches discussed in the study cloud-free composite orthophotomap provided by mapbox via tile-based map service;

- Modern achievements in image processing via deep neural networks make it possible to generate artificial spectral information, for example, to solve the image colorization problem. We investigate whether this approach can produce not only visually similar images but also an artificial spectral band that can improve the performance of computer vision algorithms for solving remote sensing tasks. We study the use of a generative adversarial network (GAN) approach in the task of the NIR band generation using only RGB channels of high-resolution satellite imagery. We evaluate the impact of a generated channel on the model performance to solve the forest segmentation task. The presented study shows the advantages of generating the extra band such as the opportunity to reduce the required amount of labeled data;

- We examine the problem of inconsistency of convolutional neural network generalization in the remote sensing domain. The problem occurs when the training set and the test set of images are from different locations or times of the year. Image exploration shows that even the exact locations at similar dates, but different years, can vary dramatically. It leads to model overfitting on the training set and a drop in performance dramatically on the test set. This problem is crucial when the size of the training set is small. We propose and evaluate a novel image augmentation approach called MixChannel. MixChannel uses multiple multispectral images of the exact location at various dates of the vegetation period to augment the training set. Our study shows that MixChannel is a promising approach that enables training more precise

models for remote sensing in the environmental domain.

We propose different approaches for environmental studies and show their applications in forestry tasks. However, the same ideas can be modified and applied in other similar specific domains that suffer from incorrect labels and a lack of high-quality training data. Data usage from various sensors is promising for model quality adjustment not only in the Remote sensing domain. The same as augmentation techniques implementation or auxiliary data generation is beneficial in a wide range of tasks to extend existing training datasets. However, method transfer to new tasks requires knowledge about data specificity and practical limitations.

Bibliography

Order of the federal forestry agency (rosleskhoz) of december 12, 2011 n 516 moscow "on approval of the forest inventory instruction"] "prikaz federal'nogo agentstva lesnogo hozyajstva (rosleskhoz) ot 12 dekabrya 2011 g. n 516 g. moskva "ob utverzhdenii lesoustroitel'noj instrukcii", 2012.

Terra aqua moderate resolution imaging spectroradiometer (modis). https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/modis/, Accessed: 20 November 2021).

Tuomas Aakala, Timo Kuuluvainen, Tuomo Wallenius, and Heikki Kauhanen. Tree mortality episodes in the intact picea abies-dominated taiga in the arkhangelsk region of northern european russia. *Journal of Vegetation Science*, 22(2):322–333, 2011.

Lydia Abady, Mauro Barni, Andrea Garzelli, and Benedetta Tondi. Gan generation of synthetic multispectral satellite images. In *Image and Signal Processing for Remote Sensing XXVI*, volume 11533, page 115330L. International Society for Optics and Photonics, 2020.

Moloud Abdar, Farhad Pourpanah, Sadiq Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U Rajendra Acharya, et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297, 2021.

Azadeh Abdollahnejad, Dimitrios Panagiotidis, Shaban Shataee Joybari, and Peter Surovỳ. Prediction of dominant forest tree species using quickbird and environmental data. *Forests*, 8(2):42, 2017.

Khaled Abutaleb, Solomon W Newete, Shelter Mangwanya, Elhadi Adam, and Marcus J Byrne. Mapping eucalypts trees using high resolution multispectral images: A study comparing worldview 2 vs. spot 7. *The Egyptian Journal of Remote Sensing and Space Science*, 24(3):333–342, 2021.

Italian Space Agency. URL https://www.asi.it/en/earth-science/prisma/. Accessed: 11.11.2022.

Nouman Ahmed, Sudipan Saha, Muhammad Shahzad, Muhammad Moazam Fraz, and Xiao Xiang Zhu. Progressive unsupervised deep transfer learning for forest mapping in satellite image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 752–761, 2021.

Oumer S Ahmed, Adam Shemrock, Dominique Chabot, Chris Dillon, Griffin Williams, Rachel Wasson, and Steven E Franklin. Hierarchical land cover and vegetation classification using multispectral data acquired from an unmanned aerial vehicle. *International journal of remote sensing*, 38(8-10):2037–2052, 2017.

Jiwoon Ahn, Sunghyun Cho, and Suha Kwak. Weakly supervised learning of instance segmentation with inter-pixel relations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019a.

Jiwoon Ahn, Sunghyun Cho, and Suha Kwak. Weakly supervised learning of instance segmentation with inter-pixel relations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2209–2218, 2019b.

Airbus. URL https://earth.esa.int/eogateway/catalog/spot-6-and-7-esa-archive. Accessed: 2022-05-2022.

B. Alias, R. Karthika, and L. Parameswaran. Classification of high resolution remote sensing images using deep learning techniques. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1196–1202, 2018.

Hamed Alqahtani, Manolya Kavakli-Thorne, and Gulshan Kumar. Applications of generative adversarial networks (gans): An updated review. *Archives of Computational Methods in Engineering*, pages 1–28, 2019.

Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

William RL Anderegg, Anna T Trugman, Grayson Badgley, Christa M Anderson, Ann Bartuska, Philippe Ciais, Danny Cullenward, Christopher B Field, Jeremy Freeman, Scott J Goetz, et al. Climate-driven risks to the climate mitigation potential of forests. *Science*, 368(6497):eaaz7005, 2020.

Maria Giuseppa Angelini, Domenica Costantino, and Attilio Di Nisio. Aster image for environmental monitoring change detection and thermal map. In *2017 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE, 2017.

Chiara Aquino, Edward Mitchard, Iain McNicol, Harry Carstairs, Andrew Burt, Beisit Luz Puma Vilca, and Mathias Disney. Using experimental sites in tropical forests to test the ability of optical remote sensing to detect forest degradation at 0.3-30 m resolutions. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 677–680. IEEE, 2021.

Heikki Astola, Tuomas Häme, Laura Sirro, Matthieu Molinier, and Jorma Kilpi. Comparison of sentinel-2 and landsat 8 imagery for forest variable prediction in boreal region. *Remote Sensing of Environment*, 223:257–273, 2019. ISSN 0034-4257. doi:https://doi.org/10.1016/j.rse.2019.01.019. URL https://www.sciencedirect.com/science/article/pii/S0034425719300252.

Heikki Astola, Lauri Seitsonen, Eelis Halme, Matthieu Molinier, and Anne Lönnqvist. Deep neural networks with transfer learning for forest variable estimation using sentinel-2 imagery in boreal forest. *Remote Sensing*, 13(12):2392, 2021.

Elias Ayrey and Daniel J Hayes. The use of three-dimensional convolutional neural networks to interpret lidar for forest inventory. *Remote Sensing*, 10(4):649, 2018.

Andras Balazs, Eero Liski, Sakari Tuominen, and Annika Kangas. Comparison of neural networks and k-nearest neighbors methods in forest stand variable estimation using airborne laser data. *ISPRS Open Journal of Photogrammetry and Remote Sensing*, 4:100012, 2022.

Laurel Ballanti, Kristin B Byrd, Isa Woo, and Christopher Ellings. Remote sensing for wetland mapping and historical change detection at the nisqually river delta. *Sustainability*, 9(11):1919, 2017.

Sangeeta Bansal, Deeksha Katyal, Ridhi Saluja, Monojit Chakraborty, and Jai K Garg. Remotely sensed modis wetland components for assessing the variability of methane emissions in indian tropical/subtropical wetlands. *International journal of applied earth observation and geoinformation*, 64:156–170, 2018.

Frank Barrett, Ronald E McRoberts, Erkki Tomppo, Emil Cienciala, and Lars T Waser. A questionnaire-based review of the operational use of remotely sensed data by national forest inventories. *Remote Sensing of Environment*, 174:279–289, 2016.

Bjorn Barz and Joachim Denzler. Deep learning on small datasets without pre-training using cosine loss. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1371–1380, 2020.

Yakoub Bazi, Laila Bashmal, Mohamad M Al Rahhal, Reham Al Dayil, and Naif Al Ajlan. Vision transformers for remote sensing image classification. *Remote Sensing*, 13(3):516, 2021.

Mariana Belgiu and Lucian Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114:24–31, 2016.

Thomas Blaschke. Object based image analysis for remote sensing. *ISPRS journal of photogrammetry and remote sensing*, 65(1):2–16, 2010.

Leo Gallus Bont, Andreas Hill, Lars T Waser, Anton Bürgi, Christian Ginzler, and Clemens Blattert. Airborne-laser-scanning-derived auxiliary information discriminating between broadleaf and conifer trees improves the accuracy of models for predicting timber volume in mixed and heterogeneously structured forests. *Forest Ecology and Management*, 459:117856, 2020.

Abdelmalek Bouguettaya, Hafed Zarzour, Amine Mohammed Taberkit, and Ahmed Kechida. A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. *Signal Processing*, 190:108309, 2022.

Clément Bourgoin, Lilian Blanc, Jean-Stéphane Bailly, Guillaume Cornu, Erika Berenguer, Johan Oszwald, Isabelle Tritsch, François Laurent, Ali F Hasan, Plinio Sist, et al. The potential of multisource remote sensing for mapping the biomass of a degraded amazonian forest. *Forests*, 9(6):303, 2018.

L Bragagnolo, Roberto Valmir da Silva, and José Mario Vicensi Grzybowski. Amazon forest cover change mapping based on semantic segmentation by u-nets. *Ecological Informatics*, 62:101279, 2021.

Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Sebastian Briechle, Peter Krzystek, and George Vosselman. Silvi-net–a dual-cnn approach for combined classification of tree species and standing dead trees from remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 98:102292, 2021.

Alexander R Brown, George P Petropoulos, and Konstantinos P Ferentinos. Appraisal of the sentinel-1 & 2 use in a large-scale wildfire assessment: A case study from portugal's fires of 2017. *Applied geography*, 100:78–89, 2018.

Wolfgang Buermann, Sassan Saatchi, Thomas B Smith, Brian R Zutta, Jaime A Chaves, Borja Milá, and Catherine H Graham. Predicting species distributions across the amazonian and andean regions using remote sensing data. *Journal of Biogeography*, 35(7):1160–1176, 2008.

Leszek Bujoczek, Małgorzata Bujoczek, and Stanisław Zięba. How much, why and where? deadwood in forest ecosystems: The case of poland. *Ecological Indicators*, 121:107027, 2021.

Eric L Bullock, Sean P Healey, Zhiqiang Yang, Rasmus Houborg, Noel Gorelick, Xiaojing Tang, and Carole Andrianirina. Timeliness in forest change monitoring: A new assessment framework demonstrated using sentinel-1 and a continuous change detection algorithm. *Remote Sensing of Environment*, 276:113043, 2022.

Joseph Bullock, Carolina Cuesta-Lázaro, and Arnau Quera-Bofarull. Xnet: A convolutional neural network (cnn) implementation for medical x-ray image segmentation suitable for small datasets. In *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10953, page 109531Z. International Society for Optics and Photonics, 2019.

Alexander Buslaev, Vladimir I Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A Kalinin. Albumentations: fast and flexible image augmentations. *Information*, 11(2):125, 2020a.

Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020b. ISSN 2078-2489. doi:10.3390/info11020125. URL https://www.mdpi.com/2078-2489/11/2/125.

Kim Calders, Jennifer Adams, John Armston, Harm Bartholomeus, Sebastien Bauwens, Lisa Patrick Bentley, Jerome Chave, F Mark Danson, Miro Demol, Mathias Disney, et al. Terrestrial laser scanning in forest ecology: Expanding the horizon. *Remote Sensing of Environment*, 251:112102, 2020.

Manuel Campos-Taberner, Francisco Javier García-Haro, Beatriz Martínez, Emma Izquierdo-Verdiguier, Clement Atzberger, Gustau Camps-Valls, and María Amparo Gilabert. Understanding deep learning in land use classification based on sentinel-2 time series. *Scientific reports*, 10(1):1–12, 2020.

Emmanuelle Cano, Jean-Philippe Denux, Mar Bisquert, Laurence Hubert-Moy, and Véronique Chéret. Improved forest-cover mapping based on modis time series and landscape stratification. *International Journal of Remote Sensing*, 38(7): 1865–1888, 2017.

Jingjing Cao, Wanchun Leng, Kai Liu, Lin Liu, Zhi He, and Yuanhui Zhu. Object-based mangrove species classification using unmanned aerial vehicle hyperspectral images and digital surface models. *Remote Sensing*, 10(1):89, 2018.

Kaili Cao and Xiaoli Zhang. An improved res-unet model for tree species classification using airborne high-resolution images. *Remote Sensing*, 12(7):1128, 2020.

Abhishek Chaurasia and Eugenio Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE Visual Communications and Image Processing (VCIP)*, pages 1–4. IEEE, 2017.

Bangqian Chen, Xiangming Xiao, Xiangping Li, Lianghao Pan, Russell Doughty, Jun Ma, Jinwei Dong, Yuanwei Qin, Bin Zhao, Zhixiang Wu, et al. A mangrove forest map of china in 2015: Analysis of time series landsat 7/8 and sentinel-1a imagery in google earth engine cloud computing platform. *ISPRS Journal of Photogrammetry and Remote Sensing*, 131:104–120, 2017a.

Hao Chen, Zipeng Qi, and Zhenwei Shi. Remote sensing image change detection with transformers. *IEEE Transactions on Geoscience and Remote Sensing*, 2021a.

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017b.

Na Chen, Nandin-Erdene Tsendbazar, Eliakim Hamunyela, Jan Verbesselt, and Martin Herold. Sub-annual tropical forest disturbance monitoring using harmonized landsat and sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation*, 102:102386, 2021b.

Shuxiao Chen, Edgar Dobriban, and Jane H Lee. Invariance reduces variance: Understanding data augmentation in deep learning and beyond. *arXiv preprint arXiv:1907.10905*, 2019a.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA, 2016a. ACM. ISBN 978-1-4503-4232-2. doi:10.1145/2939672.2939785. URL http://doi.acm.org/10.1145/2939672.2939785.

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016b.

Yuehong Chen, Yanan Zhou, Yong Ge, Ru An, and Yu Chen. Enhancing land cover mapping through integration of pixel-based and object-based classifications from remotely sensed imagery. *Remote Sensing*, 10(1):77, 2018.

Yun Chen, Juan P Guerschman, Zhibo Cheng, and Longzhu Guo. Remote sensing for vegetation monitoring in carbon capture storage regions: A review. *Applied energy*, 240:312–326, 2019b.

Xun Cheng and Jianbo Yu. Retinanet with difference channel attention and adaptively spatial feature fusion for steel surface defect detection. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11, 2020.

Hannah V Cooper, Christopher H Vane, Stephanie Evers, Paul Aplin, Nicholas T Girkin, and Sofie Sjögersten. From peat swamp forest to oil palm plantations: The stability of tropical peatland carbon. *Geoderma*, 342:109–117, 2019.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

Jérémy Cours, Laurent Larrieu, Carlos Lopez-Vaamonde, Jörg Müller, Guillem Parmain, Simon Thorn, and Christophe Bouget. Contrasting responses of habitat conditions and insect biodiversity to pest-or climate-induced dieback in coniferous mountain forests. *Forest Ecology and Management*, 482:118811, 2021.

Luís Guilherme Teixeira Crusiol, Marcos Rafael Nanni, Renato Herrig Furlanetto, Everson Cezar, and Guilherme Fernando Capristo Silva. Reflectance calibration of uav-based visible and near-infrared digital images acquired under variant altitude and illumination conditions. *Remote Sensing Applications: Society and Environment*, 18:100312, 2020.

Ovidiu Csillik, Pramukta Kumar, and Gregory P Asner. Challenges in estimating tropical forest canopy height from planet dove imagery. *Remote Sensing*, 12(7): 1160, 2020.

Andrew M Cunliffe, Jakob J Assmann, Gergana N Daskalova, Jeffrey T Kerby, and Isla H Myers-Smith. Aboveground biomass corresponds strongly with drone-derived canopy height but weakly with greenness (ndvi) in a shrub tundra landscape. *Environmental Research Letters*, 15(12):125004, 2020.

Luciana Borges da Costa, Osmar Luiz Ferreira de Carvalho, Anesmar Olino de Albuquerque, Roberto Arnaldo Trancoso Gomes, Renato Fontes Guimarães, and Osmar Abílio de Carvalho Júnior. Deep semantic segmentation for detecting eucalyptus planted forests in the brazilian territory using sentinel-2 imagery. *Geocarto International*, pages 1–13, 2021.

Anca Dabija, Marcin Kluczek, Bogdan Zagajewski, Edwin Raczko, Marlena Kycko, Ahmed H Al-Sulttani, Anna Tardà, Lydia Pineda, and Jordi Corbera. Comparison of support vector machines and random forests for corine land cover mapping. *Remote Sensing*, 13(4):777, 2021.

Michele Dalponte, Lorenzo Bruzzone, and Damiano Gianelle. Tree species classification in the southern alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and lidar data. *Remote sensing of environment*, 123:258–270, 2012.

An Thi Ngoc Dang, Subrata Nandy, Ritika Srinet, Nguyen Viet Luong, Surajit Ghosh, and A Senthil Kumar. Forest aboveground biomass estimation using machine learning regression algorithm in yok don national park, vietnam. *Ecological Informatics*, 50:24–32, 2019.

Daniel Caio de Lima, Diego Saqui, Steve Ataky, Lúcio A de C Jorge, Ednaldo José Ferreira, and José Hiroki Saito. Estimating agriculture nir images from aerial rgb data. In *International Conference on Computational Science*, pages 562–574. Springer, 2019.

Misganu Debella-Gilo and Arnt Kristian Gjertsen. Mapping seasonal agricultural land use types using deep learning on sentinel-2 image time series. *Remote Sensing*, 13(2):289, 2021.

Wolfgang Deigele, Melanie Brandmeier, and Christoph Straub. A hierarchical deep-learning approach for rapid windthrow detection on planetscope and high-resolution aerial image data. *Remote Sensing*, 12(13):2121, 2020.

Evan R DeLancey, John F Simms, Masoud Mahdianpari, Brian Brisco, Craig Mahoney, and Jahan Kariyeva. Comparing deep learning and shallow learning for large-scale wetland classification in alberta, canada. *Remote Sensing*, 12(1):2, 2019.

Evan R DeLancey, John F Simms, Masoud Mahdianpari, Brian Brisco, Craig Mahoney, and Jahan Kariyeva. Comparing deep learning and shallow learning for large-scale wetland classification in alberta, canada. *Remote Sensing*, 12(1):2, 2020.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

Yago Diez, Sarah Kentsch, Motohisa Fukuda, Maximo Larry Lopez Caceres, Koma Moritake, and Mariano Cabezas. Deep learning in forestry using uav-acquired rgb data: A practical review. *Remote Sensing*, 13(14):2837, 2021.

Ivica Dimitrovski, Dragi Kocev, Suzana Loskovska, and Sašo Džeroski. Hierarchical classification of diatom images using ensembles of predictive clustering trees. *Ecological Informatics*, 7(1):19–29, 2012.

Luofan Dong, Huaqiang Du, Fangjie Mao, Ning Han, Xuejian Li, Guomo Zhou, Junlong Zheng, Meng Zhang, Luqi Xing, Tengyan Liu, et al. Very high resolution remote sensing imagery classification using a fusion of random forest and deep learning techniquesubtropical area for example. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:113–128, 2019.

Luofan Dong, Huaqiang Du, Ning Han, Xuejian Li, Dien Zhu, Fangjie Mao, Meng Zhang, Junlong Zheng, Hua Liu, Zihao Huang, et al. Application of convolutional neural network on lei bamboo above-ground-biomass (agb) estimation using worldview-2. *Remote Sensing*, 12(6):958, 2020.

Iryna Dronova, Sophie Taddeo, Kyle S Hemes, Sara H Knox, Alex Valach, Patricia Y Oikawa, Kuno Kasak, and Dennis D Baldocchi. Remotely sensed phenological heterogeneity of restored wetlands: linking vegetation structure and function. *Agricultural and Forest Meteorology*, 296:108215, 2021.

Michal Drozdzal, Eugene Vorontsov, Gabriel Chartrand, Samuel Kadoury, and Chris Pal. The importance of skip connections in biomedical image segmentation. In *Deep learning and data labeling for medical applications*, pages 179–187. Springer, 2016.

Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. Sentinel-2: Esa's optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.

Ralph Dubayah, James Bryan Blair, Scott Goetz, Lola Fatoyinbo, Matthew Hansen, Sean Healey, Michelle Hofton, George Hurtt, James Kellner, Scott Luthcke, et al. The global ecosystem dynamics investigation: High-resolution laser ranging of the earths forests and topography. *Science of remote sensing*, 1:100002, 2020.

Timothy Dube, Tawanda W Gara, Onisimo Mutanga, Mbulisi Sibanda, Cletah Shoko, Amon Murwira, Mhosisi Masocha, Henry Ndaimani, and Chipo M Hatendi. Estimating forest standing biomass in savanna woodlands as an indicator of forest productivity using the new generation worldview-2 sensor. *Geocarto International*, 33(2):178–188, 2018.

Giovanni DAmico, Elia Vangi, Saverio Francini, Francesca Giannetti, Antonino Nicolaci, Davide Travaglini, Lorenzo Massai, Yamuna Giambastiani, Carlo Terranova, and Gherardo Chirici. Are we ready for a national forest information system? state of the art of forest maps and airborne laser scanning data availability in italy. *iForest-Biogeosciences and Forestry*, 14(2):144, 2021.

Joseph J. Erinjery, Mewa Singh, and Rafi Kent. Mapping and assessment of vegetation types in the tropical rainforests of the western ghats using multispectral sentinel-2 and sar sentinel-1 satellite imagery. *Remote Sensing of Environment*, 216:345–354, 2018. ISSN 0034-4257. doi:https://doi.org/10.1016/j.rse.2018.07.006. URL https://www.sciencedirect.com/science/article/pii/S003442571830333X.

Carlos Esse, Alfonso Condal, Patricio de Los Ríos-Escalante, Francisco Correa-Araneda, Roberto Moreno-García, and Roderick Jara-Falcón. Evaluation of classification techniques in very-high-resolution (vhr) imagery: A case study of the identification of deadwood in the chilean central-patagonian forests. *Ecological Informatics*, page 101685, 2022.

European and the Space Agency. URL https://sentinel.esa.int/web/sentinel/user-guides. Accessed: 2022-05-2022.

Timothy J Fahey, Peter B Woodbury, John J Battles, Christine L Goodale, Steven P Hamburg, Scott V Ollinger, and Christopher W Woodall. Forest carbon storage: ecology, management, and policy. *Frontiers in Ecology and the Environment*, 8 (5):245–252, 2010.

Jiayuan Fan, Tao Chen, and Shijian Lu. Unsupervised feature learning for land-use scene recognition. *IEEE Transactions on Geoscience and Remote Sensing*, PP: 1–12, 01 2017. doi:10.1109/TGRS.2016.2640186.

Kathryn E Fankhauser, Nikolay S Strigul, and Demetrios Gatziolis. Augmentation of traditional forest inventory and airborne laser scanning with unmanned aerial systems and photogrammetry for forest monitoring. *Remote Sensing*, 10(10):1562, 2018.

A Fernandez-Carrillo, D de la Fuente, FW Rivas-Gonzalez, and A Franco-Nieto. A sentinel-2 unsupervised forest mask for european sites. In *Earth Resources and Environmental Remote Sensing/GIS Applications X*, volume 11156, page 111560Y. International Society for Optics and Photonics, 2019.

Angel Fernandez-Carrillo, Zdeněk Patočka, Lumír Dobrovolnỳ, Antonio Franco-Nieto, and Beatriz Revilla-Romero. Monitoring bark beetle forest damage in central europe. a remote sensing approach validated with field data. *Remote Sensing*, 12(21):3634, 2020.

Carlos A Ferreira, Tânia Melo, Patrick Sousa, Maria Inês Meyer, Elham Shakibapour, Pedro Costa, and Aurélio Campilho. Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2. In *International conference image analysis and recognition*, pages 763–770. Springer, 2018.

Matheus Pinheiro Ferreira, Fabien Hubert Wagner, Luiz EOC Aragão, Yosio Edemir Shimabukuro, and Carlos Roberto de Souza Filho. Tree species classification in tropical forests using visible to shortwave infrared worldview-3 images and texture

analysis. *ISPRS journal of photogrammetry and remote sensing*, 149:119–131, 2019.

Neil Flood, Fiona Watson, and Lisa Collett. Using a u-net convolutional neural network to map woody vegetation extent from high resolution satellite imagery across queensland, australia. *International Journal of Applied Earth Observation and Geoinformation*, 82:101897, 2019.

José Juan Flores-Martínez, Anuar Martínez-Pacheco, Eduardo Rendón-Salinas, Jorge Rickards, Sahotra Sarkar, and Víctor Sánchez-Cordero. Recent forest cover loss in the core zones of the monarch butterfly biosphere reserve in mexico. *Frontiers in Environmental Science*, 7:167, 2019.

Giles M Foody. Status of land cover classification accuracy assessment. *Remote sensing of environment*, 80(1):185–201, 2002.

Andreas Forstmaier, Ankit Shekhar, and Jia Chen. Mapping of eucalyptus in natura 2000 areas using sentinel 2 imagery and artificial neural networks. *Remote Sensing*, 12(14):2176, 2020.

Janet Franklin, Riley Andrade, Mark L Daniels, Patrick Fairbairn, Maria C Fandino, Thomas W Gillespie, Grizelle González, Otto Gonzalez, Daniel Imbert, Valerie Kapos, et al. Geographical ecology of dry forest tree communities in the west indies. *Journal of Biogeography*, 45(5):1168–1181, 2018.

Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

Anmin Fu, Guoqing Sun, Zhifeng Guo, and Dianzhong Wang. Forest cover classification with modis images in northeastern asia. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 3(2):178–189, 2010.

Yuanyuan Fu, Hong S He, Todd J Hawbaker, Paul D Henne, Zhiliang Zhu, and David R Larsen. Evaluating k-nearest neighbor (k nn) imputation models for species-level aboveground forest biomass mapping in northeast china. *Remote sensing*, 11(17):2005, 2019.

Selina Ganz, Petra Adler, and Gerald Kändler. Forest cover mapping based on a combination of aerial images and sentinel-2 satellite data compared to national forest inventory data. *Forests*, 11(12):1322, 2020.

Bolin Gao and Lacra Pavel. On the properties of the softmax function with application in game theory and reinforcement learning. *arXiv preprint arXiv:1704.00805*, 2017.

Xiaoxia Gao, Shikui Dong, Shuai Li, Yudan Xu, Shiliang Liu, Haidi Zhao, Jane Yeomans, Yu Li, Hao Shen, Shengnan Wu, et al. Using the random forest model and validated modis with the field spectrometer measurement promote the accuracy of estimating aboveground biomass and coverage of alpine grasslands on the qinghai-tibetan plateau. *Ecological Indicators*, 112:106114, 2020a.

Yan Gao, Margaret Skutsch, Jaime Paneque-Gálvez, and Adrian Ghilardi. Remote sensing of forest degradation: a review. *Environmental Research Letters*, 15(10): 103001, 2020b.

GBDX. GBDX. https://gbdxdocs.digitalglobe.com/, Accessed: 2020.

geoalert.io. Geoalert analytics platform. https://www.geoalert.io/en-US/, 2019-2020.

Cynthia Gerlein-Safdi, A Anthony Bloom, Genevieve Plant, Eric A Kort, and Christopher S Ruf. Improving representation of tropical wetland methane emissions with cygnss inundation maps. *Global Biogeochemical Cycles*, 35(12): e2020GB006890, 2021.

G Gerylo, RJ Hall, SE Franklin, A Roberts, and EJ Milton. Hierarchical image classification and extraction of forest species composition and crown closure from airborne multispectral images. *Canadian Journal of Remote Sensing*, 24(3):219–232, 1998.

Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D Cubuk, Quoc V Le, and Barret Zoph. Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2918–2928, 2021.

Sujit Madhab Ghosh, Mukunda Dev Behera, and Somnath Paramanik. Canopy height estimation using sentinel series images through machine learning models in a mangrove forest. *Remote Sensing*, 12(9):1519, 2020.

Francesca Giannetti, Matteo Pecchi, Davide Travaglini, Saverio Francini, Giovanni DAmico, Elia Vangi, Claudia Cocozza, and Gherardo Chirici. Estimating vaia windstorm damaged forest area in italy using time series sentinel-2 imagery and continuous change detection algorithms. *Forests*, 12(6):680, 2021.

Colin J Gleason and Jungho Im. Forest biomass estimation from airborne lidar data using machine learning approaches. *Remote Sensing of Environment*, 125:80–91, 2012.

Cyril Goutte and Eric Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European conference on information retrieval*, pages 345–359. Springer, 2005.

Ewa Grabska, David Frantz, and Katarzyna Ostapowicz. Evaluation of machine learning algorithms for forest stand species mapping using sentinel-2 imagery and environmental data in the polish carpathians. *Remote Sensing of Environment*, 251:112103, 2020.

Mathieu Gravey, Luiz Gustavo Rasera, and Gregoire Mariethoz. Analogue-based colorization of remote sensing images using textural information. *ISPRS Journal of Photogrammetry and Remote Sensing*, 147:242–254, 2019. ISSN 0924-2716. doi:https://doi.org/10.1016/j.isprsjprs.2018.11.003. URL https://www.sciencedirect.com/science/article/pii/S0924271618302995.

Thomas Gschwantner, Iciar Alberdi, Sébastien Bauwens, Susann Bender, Dragan Borota, Michal Bosela, Olivier Bouriaud, Johannes Breidenbach, Jānis Donis, Christoph Fischer, et al. Growing stock monitoring by european national forest inventories: Historical origins, current methods and harmonisation. *Forest Ecology and Management*, 505:119868, 2022.

David Gudex-Cross, Jennifer Pontius, and Alison Adams. Enhanced forest cover mapping using spectral unmixing and object-based classification of multi-temporal landsat imagery. *Remote sensing of Environment*, 196:193–204, 2017.

Sercan Gülci, Abdullah E Akay, Neşe Gülci, and İnanç Taş. An assessment of conventional and drone-based measurements for tree attributes in timber volume estimation: A case study on stone pine plantation. *Ecological Informatics*, 63: 101303, 2021.

Alkan Günlü, İlker Ercanlı, Muammer Şenyurt, and Sedat Keleş. Estimation of some stand parameters from textural features from worldview-2 satellite image using the artificial neural network and multiple regression methods: a case study from turkey. *Geocarto International*, 36(8):918–935, 2021.

Sheng Guo, Weilin Huang, Haozhi Zhang, Chenfan Zhuang, Dengke Dong, Matthew R. Scott, and Dinglong Huang. Curriculumnet: Weakly supervised learning from large-scale web images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018a.

Yiqing Guo, Xiuping Jia, and David Paull. Effective sequential classifier training for svm-based multitemporal remote sensing image classification. *IEEE Transactions on Image Processing*, 27(6):3036–3048, 2018b.

Radoslaw Guzinski, Hector Nieto, Inge Sandholt, and Georgios Karamitilios. Modelling high-resolution actual evapotranspiration through sentinel-2 and sentinel-3 data fusion. *Remote Sensing*, 12(9):1433, 2020.

David Gwenzi, Eileen H Helmer, Xiaolin Zhu, Michael A Lefsky, and Humfredo Marcano-Vega. Predictions of tropical forest biomass and biomass growth based on stand height or canopy area are improved by landsat-scale phenology across puerto rico and the us virgin islands. *Remote Sensing*, 9(2):123, 2017.

Nam Thang Ha, Merilyn Manley-Harris, Tien Dat Pham, and Ian Hawes. A comparative assessment of ensemble-based machine learning and maximum likelihood methods for mapping seagrass using sentinel-2 imagery in tauranga harbor, new zealand. *Remote Sensing*, 12(3):355, 2020.

Helena Haakana et al. Multi-source forest inventory data for forest production and utilization analyses at different levels. *Finnish Society of Forest Science, Helsinki, Finland*, 2017.

Alireza Hamedianfar and Mohamed Barakat A. Gibril. Large-scale urban mapping using integrated geographic object-based image analysis and artificial bee colony

optimization from worldview-3 data. *International Journal of Remote Sensing*, 40(17):6796–6821, 2019.

Hamid Hamraz. Automated tree-level forest quantification using airborne lidar. 2018.

Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *arXiv preprint arXiv:1804.06872*, 2018.

Matthew C Hansen, Yosio E Shimabukuro, Peter Potapov, and Kyle Pittman. Comparing annual modis and prodes forest cover change data for advancing monitoring of brazilian forest cover. *Remote Sensing of Environment*, 112(10):3784–3793, 2008.

Matthew C Hansen, Peter V Potapov, Scott J Goetz, Svetlana Turubanova, Alexandra Tyukavina, Alexander Krylov, Anil Kommareddy, and Alexey Egorov. Mapping tree height distributions in sub-saharan africa using landsat 7 and 8 data. *Remote Sensing of Environment*, 185:221–232, 2016.

Nancy L Harris, David A Gibbs, Alessandro Baccini, Richard A Birdsey, Sytze De Bruin, Mary Farina, Lola Fatoyinbo, Matthew C Hansen, Martin Herold, Richard A Houghton, et al. Global maps of twenty-first century forest carbon fluxes. *Nature Climate Change*, 11(3):234–240, 2021.

Sean Hartling, Vasit Sagan, Paheding Sidike, Maitiniyazi Maimaitijiang, and Joshua Carron. Urban tree species classification using a worldview-2/3 and lidar data fusion approach and deep learning. *Sensors*, 19(6):1284, 2019.

Ryuichiro Hataya, Jan Zdenek, Kazuki Yoshizoe, and Hideki Nakayama. Faster autoaugment: Learning augmentation strategies using backpropagation. In *European Conference on Computer Vision*, pages 1–16. Springer, 2020.

Paweł Hawryło and Piotr Wężyk. Predicting growing stock volume of scots pine stands using sentinel-2 satellite imagery and airborne image-derived point clouds. *Forests*, 9(5):274, 2018.

Haiqing He, Yeli Yan, Ting Chen, and Penggen Cheng. Tree height estimation of forest plantation in mountainous terrain from bare-earth points using a dog-coupled radial basis function neural network. *Remote sensing*, 11(11):1271, 2019a.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Qibin He, Xian Sun, Zhiyuan Yan, and Kun Fu. Dabnet: Deformable contextual and boundary-weighted network for cloud detection in remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–16, 2021. doi:10.1109/TGRS.2020.3045474.

Yuhong He, Jian Yang, John Caspersen, and Trevor Jones. An operational workflow of deciduous-dominated forest species classification: Crown delineation, gap elimination, and object-based classification. *Remote Sensing*, 11(18):2078, 2019b.

Ana Hernando, Luis Puerto, Blas Mola-Yudego, José Antonio Manzanera, Antonio García-Abril, Matti Maltamo, and Rubén Valbuena. Estimation of forest biomass components using airborne lidar and multispectral sensors. *iForest-Biogeosciences and Forestry*, 12(2):207, 2019.

Andreas Hill, Henning Buddenbaum, and Daniel Mandallaz. Combining canopy height and tree species map information for large-scale timber volume estimations under strong heterogeneity of auxiliary data and variable sample plot sizes. *European Journal of Forest Research*, 137(4):489–505, 2018.

Srivastava N. Hinton, G. and K. Swersky. Lecture 6d - a separate, adaptive learning rate for each connection. slides of lecture neural networks for machine learning, 2012.

Manuela Hirschmugl, Janik Deutscher, Carina Sobe, Alexandre Bouvet, Stéphane Mermoz, and Mathias Schardt. Use of sar and optical time series for tropical forest disturbance mapping. *Remote Sensing*, 12(4):727, 2020.

Samuel Hislop, Simon Jones, Mariela Soto-Berelov, Andrew Skidmore, Andrew Haywood, and Trung H Nguyen. Using landsat spectral indices in time-series to assess wildfire disturbance and recovery. *Remote sensing*, 10(3):460, 2018.

Yang Hu, Xuelei Xu, Fayun Wu, Zhongqiu Sun, Haoming Xia, Qingmin Meng, Wenli Huang, Hua Zhou, Jinping Gao, Weitao Li, et al. Estimating forest stock volume in hunan province, china, by integrating in situ plot data, sentinel-2 images, and linear and machine learning regression models. *Remote Sensing*, 12(1):186, 2020.

Chih-Sheng Huang, Chun-Ling Lin, Li-Wei Ko, Sheng-Yi Liu, Tung-Ping Sua, and Chin-Teng Lin. A hierarchical classification system for sleep stage scoring via forehead eeg signals. In *2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, pages 1–5. IEEE, 2013.

Cho-ying Huang, William RL Anderegg, and Gregory P Asner. Remote sensing of forest die-off in the anthropocene: From plant ecophysiology to canopy structure. *Remote Sensing of Environment*, 231:111233, 2019a.

Hai Huang, Hao Zhou, Xu Yang, Lu Zhang, Lu Qi, and Ai-Yun Zang. Faster r-cnn for marine organisms detection and recognition using data augmentation. *Neurocomputing*, 337:372–384, 2019b. ISSN 0925-2312. doi:https://doi.org/10.1016/j.neucom.2019.01.084. URL https://www.sciencedirect.com/science/article/pii/S0925231219301274.

Huabing Huang, Caixia Liu, Xiaoyi Wang, Gregory S Biging, Yanlei Chen, Jun Yang, and Peng Gong. Mapping vegetation heights in china using slope correction icesat data, srtm, modis-derived and climate data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 129:189–199, 2017.

Sha Huang, Lina Tang, Joseph P Hupy, Yang Wang, and Guofan Shao. A commentary review on the use of normalized difference vegetation index (ndvi) in the era of popular remote sensing. *Journal of Forestry Research*, 32(1):1–6, 2021.

Xiaoman Huang and Mark A Friedl. Distance metric-based forest cover change detection using modis time series. *International Journal of Applied Earth Observation and Geoinformation*, 29:78–92, 2014.

Alfredo Huete, Chris Justice, and Wim Van Leeuwen. Modis vegetation index (mod13). *Algorithm theoretical basis document*, 3(213), 1999.

Eric Hyyppä, Juha Hyyppä, Teemu Hakala, Antero Kukko, Michael A Wulder, Joanne C White, Jiri Pyörälä, Xiaowei Yu, Yunsheng Wang, Juho-Pekka Virtanen, et al. Under-canopy uav laser scanning for accurate forest field measurements. *ISPRS Journal of Photogrammetry and Remote Sensing*, 164:41–60, 2020.

Kotaro Iizuka, Yuichi S Hayakawa, Takuro Ogura, Yasutaka Nakata, Yoshiko Kosugi, and Taichiro Yonehara. Integration of multi-sensor data to estimate plot-level stem volume using machine learning algorithms–case study of evergreen conifer planted forests in japan. *Remote Sensing*, 12(10):1649, 2020.

Svetlana Illarionova. Satellite object augmentation. https://github.com/LanaLana/satellite_object_augmentation, 2021.

Svetlana Illarionova, Alexey Trekin, Vladimir Ignatiev, and Ivan Oseledets. Neural-based hierarchical approach for detailed dominant forest species classification by multispectral satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:1810–1820, 2020.

Svetlana Illarionova, Sergey Nesteruk, Dmitrii Shadrin, Vladimir Ignatiev, Maria Pukalchik, and Ivan Oseledets. Mixchannel: Advanced augmentation for multispectral satellite images. *Remote Sensing*, 13(11):2181, 2021a.

Svetlana Illarionova, Sergey Nesteruk, Dmitrii Shadrin, Vladimir Ignatiev, Mariia Pukalchik, and Ivan Oseledets. Object-based augmentation for building semantic segmentation: Ventura and santa rosa case study. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 1659–1668. IEEE, 2021b.

Svetlana Illarionova, Dmitrii Shadrin, Alexey Trekin, Vladimir Ignatiev, and Ivan Oseledets. Generation of the nir spectral band for satellite images with convolutional neural networks. *Sensors*, 21(16), 2021c. ISSN 1424-8220. URL https://www.mdpi.com/1424-8220/21/16/5646.

Svetlana Illarionova, Alexey Trekin, Vladimir Ignatiev, and Ivan Oseledets. Tree species mapping on sentinel-2 satellite imagery with weakly supervised classification and object-wise sampling. *Forests*, 12(10):1413, 2021d.

Markus Immitzer, Clement Atzberger, and Tatjana Koukal. Tree species classification with random forest using very high spatial resolution 8-band worldview-2 satellite data. *Remote Sensing*, 4(9):2661–2693, 2012.

Markus Immitzer, Francesco Vuolo, and Clement Atzberger. First experience with sentinel-2 data for crop and tree species classifications in central europe. *Remote Sensing*, 8(3):166, 2016.

Markus Immitzer, Martin Neuwirth, Sebastian Böck, Harald Brenner, Francesco Vuolo, and Clement Atzberger. Optimal input features for tree species classification in central europe based on multi-temporal sentinel-2 data. *Remote Sensing*, 11(22):2599, 2019.

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

A. Shilova A. Katrutsa D. Bershatsky X. Zhao L. Eyraud-Dubois O. Shlyazhko D. Dimitrov I. Oseledets J. Gusak, D. Cherniuk and O. Beaumont. Survey on efficient training of large neural networks. In *Thirty-First International Joint Conference on Artificial Intelligence IJCAI-22*, pages 5494–5501, 2022. doi:https://doi.org/10.24963/ijcai.2022/769.

Piyush Jain, Sean CP Coogan, Sriram Ganapathi Subramanian, Mark Crowley, Steve Taylor, and Mike D Flannigan. A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28(4):478–505, 2020.

G Janssens-Maenhout, B Pinty, M Dowell, H Zunker, E Andersson, G Balsamo, J-L Bézy, T Brunhes, H Bösch, B Bojkov, et al. Toward an operational anthropogenic co 2 emissions monitoring and verification support capacity. *Bulletin of the American Meteorological Society*, 101(8):E1439–E1451, 2020.

Farzaneh Dadrass Javan, Farhad Samadzadegan, Soroosh Mehravar, Ahmad Toosi, Reza Khatami, and Alfred Stein. A review of image fusion techniques for pan-sharpening of high-resolution satellite imagery. *ISPRS journal of photogrammetry and remote sensing*, 171:101–117, 2021.

JAXA. URL `https://www.eorc.jaxa.jp/ALOS/en/alos-2/a2_sensor_e.htm`. Accessed: 2022-05-2022.

T Jayalakshmi and A Santhakumaran. Statistical normalization and back propagation for classification. *International Journal of Computer Theory and Engineering*, 3(1):1793–8201, 2011.

Sadeepa Jayathunga, Toshiaki Owari, and Satoshi Tsuyuki. Digital aerial photogrammetry for uneven-aged forest management: Assessing the potential to reconstruct canopy structure and estimate living biomass. *Remote Sensing*, 11(3): 338, 2019.

Kejin Jia. Agricultural image denoising, compression and enhancement based on wavelet transform. *Agronomia*, 36(2), 2019.

Sen Jia, Shuguo Jiang, Zhijie Lin, Nanying Li, Meng Xu, and Shiqi Yu. A survey: Deep learning for hyperspectral image classification with

few labeled samples. *Neurocomputing*, 448:179–204, 2021. ISSN 0925-2312. doi:https://doi.org/10.1016/j.neucom.2021.03.035. URL https://www.sciencedirect.com/science/article/pii/S0925231221004033.

Yufeng Jiang, Li Zhang, Min Yan, Jianguo Qi, Tianmeng Fu, Shunxiang Fan, and Bowei Chen. High-resolution mangrove forests classification with machine learning using worldview and uav hyperspectral data. *Remote Sensing*, 13(8):1529, 2021.

David John and Ce Zhang. An attention-based u-net for detecting deforestation within satellite sensor imagery. *International Journal of Applied Earth Observation and Geoinformation*, 107:102685, 2022.

Lilli Kaarakka, Meredith Cornett, Grant Domke, Todd Ontl, and Laura E Dee. Improved forest management as a natural climate solution: A review. *Ecological Solutions and Evidence*, 2(3):e12090, 2021.

Annika Kangas, Rasmus Astrup, Johannes Breidenbach, Jonas Fridman, Terje Gobakken, Kari T Korhonen, Matti Maltamo, Mats Nilsson, Thomas Nord-Larsen, Erik Næsset, et al. Remote sensing and forest inventories in nordic countries–roadmap for the future. *Scandinavian Journal of Forest Research*, 33(4):397–412, 2018.

DV Karelin, DG Zamolodchikov, and AS Isaev. Unconsidered sporadic sources of carbon dioxide emission from soils in taiga forests. In *Doklady biological sciences*, volume 475, pages 165–168. Springer, 2017.

Martin Karlson, Heather Reese, and Madelene Ostwald. Tree crown mapping in managed woodlands (parklands) of semi-arid west africa using worldview-2 imagery and geographic object based image analysis. *Sensors*, 14(12):22643–22669, 2014.

Teja Kattenborn, Jens Leitloff, Felix Schiefer, and Stefan Hinz. Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:24–49, 2021a. ISSN 0924-2716.

Teja Kattenborn, Jens Leitloff, Felix Schiefer, and Stefan Hinz. Review on convolutional neural networks (cnn) in vegetation remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 173:24–49, 2021b.

Yinghai Ke and Lindi J Quackenbush. Forest species classification and tree crown delineation using quickbird imagery. In *Proceedings of the ASPRS 2007 Annual Conference*, pages 7–11, 2007.

Keras. Keras. 2020–2021. https://keras.io/, Accessed: 20 November 2021).

Nour Eldeen Khalifa, Mohamed Loey, and Seyedali Mirjalili. A comprehensive survey of recent trends in deep learning for digital images augmentation. *Artificial Intelligence Review*, pages 1–27, 2021.

Salman Khan, Muzammal Naseer, Munawar Hayat, Syed Waqas Zamir, Fahad Shahbaz Khan, and Mubarak Shah. Transformers in vision: A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41, 2022.

T Khovratovich, S Bartalev, A Kashnitskii, I Balashov, and A Ivanova. Forest change detection based on sub-pixel tree cover estimates using landsat-oli and sentinel 2 data. In *IOP Conference Series: Earth and Environmental Science*, volume 507, page 012011. IOP Publishing, 2020.

Jinki Kim, Duk-Byeong Park, and Jung Il Seo. Exploring the relationship between forest structure and health. *Forests*, 11(12), 2020. ISSN 1999-4907. doi:10.3390/f11121264. URL https://www.mdpi.com/1999-4907/11/12/1264.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Alexander V Kirdyanov, Matthias Saurer, Rolf Siegwolf, Anastasia A Knorre, Anatoly S Prokushkin, Olga V Churakova, Marina V Fonti, and Ulf Büntgen. Long-term ecological consequences of forest fires in the continuous permafrost zone of siberia. *Environmental Research Letters*, 15(3):034061, 2020.

Uwe Knauer, Cornelius Styp von Rekowski, Marianne Stecklina, Tilman Krokotsch, Tuan Pham Minh, Viola Hauffe, David Kilias, Ina Ehrhardt, Herbert Sagischewski, Sergej Chmara, et al. Tree species classification based on hybrid ensembles of a convolutional neural network (cnn) and random forest classifiers. *Remote Sensing*, 11(23):2788, 2019.

Michael J Koontz, Andrew M Latimer, Leif A Mortenson, Christopher J Fettig, and Malcolm P North. Cross-scale interaction of host tree size and climatic water deficit governs bark beetle-induced tree mortality. *Nature communications*, 12(1): 1–13, 2021.

Kirill A Korznikov, Dmitry E Kislov, Jan Altman, Jiří Doležal, Anna S Vozmishcheva, and Pavel V Krestov. Using u-net-like deep convolutional neural networks for precise tree recognition in very high resolution rgb (red, green, blue) satellite images. *Forests*, 12(1):66, 2021.

VV Kozoderov and EV Dmitriev. Models of pattern recognition and forest state estimation based on hyperspectral remote sensing data. *Izvestiya, Atmospheric and Oceanic Physics*, 54(9):1291–1302, 2018.

VV Kozoderov, TV Kondranin, and EV Dmitriev. Hyperspectral remote sensing imagery processing: an overview. *Climate&Nature*, (1):2–18, 2017.

Nataliia Kussul, Andrii Shelestov, Mykola Lavreniuk, Igor Butko, and Sergii Skakun. Deep learning approach for large scale land cover mapping based on remote sensing data fusion. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 198–201. IEEE, 2016.

Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrii Shelestov. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5):778–782, 2017.

Chiman Kwan, Xiaolin Zhu, Feng Gao, Bryan Chou, Daniel Perez, Jiang Li, Yuzhong Shen, Krzysztof Koperski, and Giovanni Marchisio. Assessment of spatiotemporal fusion algorithms for planet and worldview images. *Sensors*, 18(4), 2018. ISSN 1424-8220. doi:10.3390/s18041051. URL https://www.mdpi.com/1424-8220/18/4/1051.

Sookyung KWON, Eunhee KIM, Joongbin LIM, and A-Ram YANG. The analysis of changes in forest status and deforestation of north korea's dmz using rapideye satellite imagery and google earth. *Journal of the Korean Association of Geographic Information Studies*, 24(4):113–126, 2021.

David Lagomasino, Temilola Fatoyinbo, SeungKuk Lee, Emanuelle Feliciano, Carl Trettin, and Marc Simard. A comparison of mangrove canopy height using multiple independent measurements from land, air, and space. *Remote sensing*, 8(4): 327, 2016.

Petro Lakyda, Anatoly Shvidenko, Andrii Bilous, Viktor Myroniuk, Maksym Matsala, Sergiy Zibtsev, Dmitry Schepaschenko, Dmytrii Holiaka, Roman Vasylyshyn, Ivan Lakyda, et al. Impact of disturbances on the carbon cycle of forest ecosystems in ukrainian polissya. *Forests*, 10(4):337, 2019.

Nico Lang, Konrad Schindler, and Jan Dirk Wegner. Country-wide high-resolution vegetation height mapping with sentinel-2. *Remote Sensing of Environment*, 233: 111347, 2019.

Pan-European High Resolution Layers. URL https://land.copernicus.eu/pan-european/high-resolution-layers. Accessed: 2022-10-2022.

Alex M Lechner, Giles M Foody, and Doreen S Boyd. Applications in remote sensing to forest ecology and management. *One Earth*, 2(5):405–412, 2020.

Won-Jin Lee and Chang-Wook Lee. Forest canopy height estimation using multi-platform remote sensing dataset. *Journal of Sensors*, 2018, 2018.

Yong-Suk Lee, Sunmin Lee, Won-Kyung Baek, Hyung-Sup Jung, Sung-Hwan Park, and Moung-Jin Lee. Mapping forest vertical structure in jeju island from optical and radar satellite images using artificial neural network. *Remote Sensing*, 12(5): 797, 2020a.

Yong-Suk Lee, Sunmin Lee, and Hyung-Sup Jung. Mapping forest vertical structure in gong-ju, korea using sentinel-2 satellite images and artificial neural networks. *Applied Sciences*, 10(5):1666, 2020b.

Yang Lei, Paul Siqueira, Diya Chowdhury, and Nathan Torbick. Generation of large-scale forest height mosaic and forest disturbance map through the combination of

spaceborne repeat-pass insar coherence and airborne lidar. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 5342–5345. IEEE, 2016.

Jean LeClerc Arrastia, Nick Heilenkötter, Daniel Otero Baguer, Lena Hauberg-Lotte, Tobias Boskamp, Sonja Hetzer, Nicole Duschner, Jörg Schaller, and Peter Maass. Deeply supervised unet for semantic segmentation to assist dermatopathological assessment of basal cell carcinoma. *Journal of Imaging*, 7(4):71, 2021.

Feimo Li, Lei Ma, and Jian Cai. Multi-discriminator generative adversarial network for high resolution gray-scale satellite image colorization. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 3489–3492. IEEE, 2018a.

Hui Li, Baoxin Hu, Qian Li, and Linhai Jing. Cnn-based tree species classification using airborne lidar data and high-resolution satellite image. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium*, pages 2679–2682. IEEE, 2020a.

Ke Li, Gang Wan, Gong Cheng, Liqiu Meng, and Junwei Han. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159:296–307, 2020b. ISSN 0924-2716. doi:https://doi.org/10.1016/j.isprsjprs.2019.11.023. URL https://www.sciencedirect.com/science/article/pii/S0924271619302825.

Weijia Li, Runmin Dong, Haohuan Fu, and Le Yu. Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks. *Remote Sensing*, 11(1):11, 2019a.

Yansheng Li, Wei Chen, Yongjun Zhang, Chao Tao, Rui Xiao, and Yihua Tan. Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning. *Remote Sensing of Environment*, 250:112045, 2020c. ISSN 0034-4257. doi:https://doi.org/10.1016/j.rse.2020.112045. URL https://www.sciencedirect.com/science/article/pii/S0034425720304156.

Yansheng Li, Te Shi, Yongjun Zhang, Wei Chen, Zhibin Wang, and Hao Li. Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 175:20–33, 2021. ISSN 0924-2716. doi:https://doi.org/10.1016/j.isprsjprs.2021.02.009. URL https://www.sciencedirect.com/science/article/pii/S0924271621000423.

Ying Li, Haokui Zhang, Xizhe Xue, Yenan Jiang, and Qiang Shen. Deep learning for remote sensing image classification: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(6):e1264, 2018b.

Zhen Li, Qijie Zan, Qiong Yang, Dehuang Zhu, Youjun Chen, and Shixiao Yu. Remote estimation of mangrove aboveground carbon stock at the species level using a low-cost unmanned aerial vehicle system. *Remote Sensing*, 11(9):1018, 2019b.

Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

Zhiwei Lin, Qilu Ding, Jiahang Huang, Weihao Tu, Dian Hu, and Jinfu Liu. Study on tree species classification of uav optical image based on densenet. *Remote Sensing Technology and Application*, 34(4):704–711, 2019.

David B Lindenmayer, Chris R Margules, and Daniel B Botkin. Indicators of biodiversity for ecologically sustainable forest management. *Conservation biology*, 14 (4):941–950, 2000.

Jing Liu, Andrew K Skidmore, Marco Heurich, and Tiejun Wang. Significant effect of topographic normalization of airborne lidar data on the retrieval of plant area index profile in mountainous forests. *ISPRS journal of photogrammetry and remote sensing*, 132:77–87, 2017.

Yanan Liu, Weishu Gong, Xiangyun Hu, and Jianya Gong. Forest type identification with random forest using sentinel-1a, sentinel-2a, multi-temporal landsat-8 and dem data. *Remote Sensing*, 10(6):946, 2018.

Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152:166–177, 2019. ISSN 0924-2716. doi:https://doi.org/10.1016/j.isprsjprs.2019.04.015. URL `https://www.sciencedirect.com/science/article/pii/S0924271619301108`.

Masoud Mahdianpari, Bahram Salehi, Mohammad Rezaee, Fariba Mohammadimanesh, and Yun Zhang. Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery. *Remote Sensing*, 10(7):1119, 2018.

Masoud Mahdianpari, Bahram Salehi, Fariba Mohammadimanesh, Brian Brisco, Saeid Homayouni, Eric Gill, Evan R DeLancey, and Laura Bourgeau-Chavez. Big data for a big country: the first generation of canadian wetland inventory map at a spatial resolution of 10-m using sentinel-1 and sentinel-2 data on the google earth engine cloud computing platform. *Canadian Journal of Remote Sensing*, 46 (1):15–33, 2020.

Titta Majasalmi, Stephanie Eisner, Rasmus Andreas Astrup, Jonas Fridman, and Ryan M Bright. An enhanced forest classification scheme for modeling vegetation–climate interactions based on national forest inventory data. 2018.

Ramandeep Kaur M Malhi, Akash Anand, Prashant K Srivastava, Sumit K Chaudhary, Manish K Pandey, Mukund Dev Behera, Amit Kumar, Prachi Singh, and G Sandhya Kiran. Synergistic evaluation of sentinel 1 and 2 for biomass estimation in a tropical forest of india. *Advances in Space Research*, 69(4):1752–1767, 2022.

Radek Malinowski, Stanisław Lewiński, Marcin Rybicki, Ewa Gromny, Małgorzata Jenerowicz, Michał Krupiński, Artur Nowakowski, Cezary Wojtkowski, Marcin Krupiński, Elke Krätzschmar, et al. Automated production of a land cover/use map of europe based on sentinel-2 imagery. *Remote Sensing*, 12(21):3523, 2020.

Eylül Malkoç, Marius Rüetschi, Christian Ginzler, and Lars T Waser. Country-wide mapping of trees outside forests based on remote sensing data in switzerland. *International Journal of Applied Earth Observation and Geoinformation*, 100:102336, 2021.

Mapbox. Mapbox service. https://www.mapbox.com/maps, Accessed: 2020-06-17.

Mapflow.ai. Mapflow.ai. https://docs.mapflow.ai/userguides/pipelines.html, Accessed: 10 Febuary 2022).

Julia Marrs and Wenge Ni-Meister. Machine learning techniques for tree species classification using co-registered lidar and hyperspectral data. *Remote Sensing*, 11(7):819, 2019.

Vitor Souza Martins, Claudio Clemente Faria Barbosa, Lino Augusto Sander De Carvalho, Daniel Schaffer Ferreira Jorge, Felipe de Lucia Lobo, and Evlyn Márcia Leão de Moraes Novo. Assessment of atmospheric correction methods for sentinel-2 msi images applied to amazon floodplain lakes. *Remote Sensing*, 9(4): 322, 2017.

Alexander Marx and Birgit Kleinschmit. Sensitivity analysis of rapideye spectral bands and derived vegetation indices for insect defoliation detection in pure scots pine stands. *iForest-Biogeosciences and Forestry*, 10(4):659, 2017.

Alexander Marx and Gideon Okpoti Tetteh. A forest vitality and change monitoring tool based on rapideye imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(6):801–805, 2017.

Axel H Masquelin, Nicholas Cheney, C Matthew Kinsey, and Jason HT Bates. Wavelet decomposition facilitates training on small datasets for medical image classification by deep learning. *Histochemistry and Cell Biology*, pages 1–9, 2021.

Alessandro Matese, Salvatore Filippo Di Gennaro, and Andrea Berton. Assessment of a canopy height model (chm) in a vineyard using uav-based multispectral imaging. *International Journal of Remote Sensing*, 38(8-10):2150–2160, 2017.

MAXAR. URL https://earth.esa.int/eogateway/missions/worldview-2. Accessed: 2022-05-2022.

Maxar. Maxar basemaps. https://www.maxar.com/products/imagery-basemaps, Accessed: 2020.

Aaron E Maxwell, Timothy A Warner, and Fang Fang. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9):2784–2817, 2018.

Arjan JH Meddens, Lee A Vierling, Jan UH Eitel, Jyoti S Jennewein, Joanne C White, and Michael A Wulder. Developing 5 m resolution canopy height and digital terrain models from worldview and arcticdem data. *Remote Sensing of Environment*, 218:174–188, 2018.

Jukka Miettinen, Hans-Jürgen Stibig, and Frédéric Achard. Remote sensing of forest degradation in southeast asiaaiming for a regional view through 5–30 m satellite data. *Global Ecology and Conservation*, 2:24–36, 2014.

Gourav Misra, Fiona Cawkwell, and Astrid Wingler. Status of phenological research using sentinel-2 data: A review. *Remote Sensing*, 12(17):2760, 2020.

Mthembeni Mngadi, John Odindi, Kabir Peerbhay, and Onisimo Mutanga. Examining the effectiveness of sentinel-1 and 2 imagery for commercial forest species mapping. *Geocarto International*, pages 1–12, 2019.

Mthembeni Mngadi, John Odindi, Kabir Peerbhay, and Onisimo Mutanga. Examining the effectiveness of sentinel-1 and 2 imagery for commercial forest species mapping. *Geocarto International*, 36(1):1–12, 2021.

Jahangir Mohammadi, Shaban Shataee, and Manoocher Babanezhad. Estimation of forest stand volume, tree density and biodiversity using landsat etm+ data, comparison of linear and regression tree analyses. *Procedia Environmental Sciences*, 7:299–304, 2011.

Midhun Mohan, Carlos Alberto Silva, Carine Klauberg, Prahlad Jat, Glenn Catts, Adrián Cardil, Andrew Thomas Hudak, and Mahendra Dia. Individual tree detection from unmanned aerial vehicle (uav) derived canopy height model in an open canopy mixed conifer forest. *Forests*, 8(9):340, 2017.

Tharun Mohandoss, Aditya Kulkarni, Daniel Northrup, Ernest Mwebaze, and Hamed Alemohammad. Generating synthetic multispectral satellite imagery from sentinel-2. *arXiv preprint arXiv:2012.03108*, 2020.

Pinki Mondal, Sonali Shukla McDermid, and Abdul Qadir. A reporting framework for sustainable development goal 15: Multi-scale monitoring of forest degradation using modis, landsat and sentinel data. *Remote Sensing of Environment*, 237: 111592, 2020.

Douglas C Morton, Ruth S DeFries, Yosio E Shimabukuro, Liana O Anderson, Fernando Del Bon Espírito-Santo, Matthew Hansen, and Mark Carroll. Rapid assessment of annual deforestation in the brazilian amazon using modis data. *Earth Interactions*, 9(8):1–22, 2005.

Lichao Mou and Xiao Xiang Zhu. Im2height: Height estimation from single monocular imagery via fully residual convolutional-deconvolutional network. *arXiv preprint arXiv:1802.10249*, 2018.

ME Muhd-Ekhzarizal, I Mohd-Hasmadi, O Hamdan, MK Mohamad-Roslan, and S Noor-Shaila. Estimation of aboveground biomass in mangrove forests using vegetation indices from spot-5 image. *Journal of Tropical Forest Science*, 30(2): 224–233, 2018.

Matteo Mura, Francesca Bottalico, Francesca Giannetti, Remo Bertani, Raffaello Giannini, Marco Mancini, Simone Orlandini, Davide Travaglini, and Gherardo Chirici. Exploiting the capabilities of the sentinel-2 multi spectral instrument for predicting growing stock volume in forest ecosystems. *International Journal of Applied Earth Observation and Geoinformation*, 66:126–134, 2018.

L Naidoo, Moses A Cho, R Mathieu, and G Asner. Classification of savanna tree species, in the greater kruger national park region, by integrating hyperspectral and lidar data in a random forest data mining environment. *ISPRS journal of Photogrammetry and Remote Sensing*, 69:167–179, 2012.

NASA. URL https://modis.gsfc.nasa.gov/about/specifications.php. Accessed: 2022-05-2022.

NASA and the U.S. Geological Survey. URL https://landsat.gsfc.nasa.gov/satellites/landsat-8/. Accessed: 2022-05-2022.

Alexey Natekin and Alois Knoll. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7:21, 2013.

UNITED NATIONS. *Handbook and National Accounting: Integrated Environmental and Economic Accounting*. 1992.

José Antonio Navarro, Nur Algeet, Alfredo Fernández-Landa, Jessica Esteban, Pablo Rodríguez-Noriega, and María Luz Guillén-Climent. Integration of uav, sentinel-1, and sentinel-2 data for mangrove plantation aboveground biomass monitoring in senegal. *Remote Sensing*, 11(1):77, 2019.

Pedro J Navarro, Fernando Pérez, Julia Weiss, and Marcos Egea-Cortines. Machine learning and computer vision system for phenotype data acquisition and analysis in plants. *Sensors*, 16(5):641, 2016.

Kamyar Nazeri, Eric Ng, and Mehran Ebrahimi. Image colorization using generative adversarial networks. In *International conference on articulated motion and deformable objects*, pages 85–94. Springer, 2018.

S. Nesteruk, D. Shadrin, M. Pukalchik, A. Somov, C. Zeidler, P. Zabel, and D. Schubert. Image compression and plants classification using machine learning in controlled-environment agriculture: Antarctic station use case. *IEEE Sensors Journal*, 2021. doi:10.1109/JSEN.2021.3050084.

Sergey Nesteruk, Svetlana Illarionova, Timur Akhtyamov, Dmitrii Shadrin, Andrey Somov, Mariia Pukalchik, and Ivan Oseledets. Xtremeaugment: Getting more from your data through combination of image collection and image augmentation. *IEEE Access*, 10:24010–24028, 2022.

Petteri Nevavuori, Nathaniel Narra, and Tarmo Lipping. Crop yield prediction with deep convolutional neural networks. *Computers and electronics in agriculture*, 163:104859, 2019.

Somayeh Nezami, Ehsan Khoramshahi, Olli Nevalainen, Ilkka Pölönen, and Eija Honkavaara. Tree species classification of drone hyperspectral and rgb imagery with deep learning convolutional neural networks. *Remote Sensing*, 12(7):1070, 2020.

Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pages 443–449, 2015.

Hoang Minh Nguyen, Begüm Demir, and Michele Dalponte. Weighted support vector machines for tree species classification using lidar data. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 6740–6743. IEEE, 2019.

Trung H Nguyen, Simon D Jones, Mariela Soto-Berelov, Andrew Haywood, and Samuel Hislop. A spatial and temporal analysis of forest dynamics using landsat time-series. *Remote sensing of environment*, 217:461–475, 2018.

Xiliang Ni, Min Xu, Chunxiang Cao, Wei Chen, Bin Yang, and Bo Xie. Forest height estimation and change monitoring based on artificial neural network using geoscience laser altimeter system and landsat data. *Journal of Applied Remote Sensing*, 14(2):022207, 2019.

Sascha Nink, Joachim Hill, Henning Buddenbaum, Johannes Stoffels, Thomas Sachtleber, and Joachim Langshausen. Assessing the suitability of future multi- and hyperspectral satellite systems for mapping the spatial distribution of norway spruce timber volume. *Remote Sensing*, 7(9):12009–12040, 2015.

Davide Notti, Daniele Giordan, Fabiana Caló, Antonio Pepe, Francesco Zucca, and Jorge Pedro Galve. Potential and limitations of open satellite data for flood mapping. *Remote sensing*, 10(11):1673, 2018.

German Novikov, Alexey Trekin, Georgy Potapov, Vladimir Ignatiev, and Evgeny Burnaev. Satellite imagery analysis for operational damage assessment in emergency situations. In *International Conference on Business Information Systems*, pages 347–358. Springer, 2018.

Nuri Erkin Ocer, Gordana Kaplan, Firat Erdem, Dilek Kucuk Matci, and Ugur Avdan. Tree extraction from multi-scale uav images using mask r-cnn with fpn. *Remote sensing letters*, 11(9):847–856, 2020.

OneAtlas. OneAtlas. https://www.intelligence-airbusds.com/imagery/oneatlas/. Accessed: 2021-08-17.

Masanori Onishi and Takeshi Ise. Explainable identification and mapping of trees using uav rgb image and deep learning. *Scientific reports*, 11(1):1–15, 2021.

Todd A Ontl, Maria K Janowiak, Christopher W Swanston, Jad Daley, Stephen Handler, Meredith Cornett, Steve Hagenbuch, Cathy Handrick, Liza McCarthy, and Nancy Patch. Forest management for carbon sequestration and climate adaptation. *Journal of Forestry*, 118(1):86–101, 2020.

Optuna. URL https://optuna.org/. Accessed: 2022-10-2022.

Viviana Otero, Ruben Van De Kerchove, Behara Satyanarayana, Columba Martínez-Espinosa, Muhammad Amir Bin Fisol, Mohd Rodila Bin Ibrahim, Ibrahim Sulong, Husain Mohd-Lokman, Richard Lucas, and Farid Dahdouh-Guebas. Managing mangrove forests from the sky: Forest inventory using field data and unmanned aerial vehicle (uav) imagery in the matang mangrove forest reserve, peninsular malaysia. *Forest ecology and management*, 411:35–45, 2018.

MA Othman, ZH AshAari, AZ Aris, and MF Ramli. Tropical deforestation monitoring using ndvi from modis satellite: a case study in pahang, malaysia. In *IOP Conference Series: Earth and Environmental Science*, volume 169, page 012047. IOP Publishing, 2018.

Ramazan Özçelik, Quang V Cao, Guillermo Trincado, and Nilsun Göçer. Predicting tree height from tree diameter and dominant height using mixed-effects and quantile regression models for two species in turkey. *Forest Ecology and Management*, 419:240–248, 2018.

Ana María Pacheco-Pascagaza, Yaqing Gou, Valentin Louis, John F Roberts, Pedro Rodríguez-Veiga, Polyanna da Conceição Bispo, Fernando DB Espírito-Santo, Ciaran Robb, Caroline Upton, Gustavo Galindo, et al. Near real-time change detection system using sentinel-2 and machine learning: a test for mexican and colombian forests. *Remote Sensing*, 14(3):707, 2022.

K. K. Pal and K. S. Sudeep. Preprocessing for image classification by convolutional neural networks. In *2016 IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, pages 1778–1781, 2016. doi:10.1109/RTEICT.2016.7808140.

Mahesh Pal. Random forest classifier for remote sensing classification. *International journal of remote sensing*, 26(1):217–222, 2005.

Klaudia Weronika Pałaś and Jarosław Zawadzki. Sentinel-2 imagery processing for tree logging observations on the białowieża forest world heritage site. *Forests*, 11 (8):857, 2020.

Prem Chandra Pandey, Akash Anand, and Prashant K Srivastava. Spatial distribution of mangrove forest species and biomass assessment using field inventory

and earth observation hyperspectral data. *Biodiversity and Conservation*, 28(8-9): 2143–2162, 2019.

Valerie J Pasquarella, Christopher E Holden, and Curtis E Woodcock. Improved mapping of forest type using spectral-temporal landsat features. *Remote Sensing of Environment*, 210:193–207, 2018.

N. Paton. Automating data preparation: Can we? should we? must we? *In Proceedings of the 21st International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data*, 2019.

Grant D Pearse, Jonathan P Dash, Henrik J Persson, and Michael S Watt. Comparison of high-density lidar and satellite photogrammetry for forest inventory. *ISPRS journal of photogrammetry and remote sensing*, 142:257–267, 2018.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011a.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011b.

Henrik J Persson, Håkan Olsson, Maciej J Soja, Lars MH Ulander, and Johan ES Fransson. Experiences from large-scale forest mapping of sweden using tandem-x data. *Remote Sensing*, 9(12):1253, 2017.

Magnus Persson, Eva Lindberg, and Heather Reese. Tree species classification with multi-temporal sentinel-2 data. *Remote Sensing*, 10(11):1794, 2018a.

Magnus Persson, Eva Lindberg, and Heather Reese. Tree species classification with multi-temporal sentinel-2 data. *Remote Sensing*, 10(11), 2018b. ISSN 2072-4292. URL https://www.mdpi.com/2072-4292/10/11/1794.

Glen P Peters. Beyond carbon budgets. *Nature Geoscience*, 11(6):378–380, 2018.

Birgit Peterson and Kurtis J Nelson. Mapping forest height in alaska using glas, landsat composites, and airborne lidar. *Remote Sensing*, 6(12):12409–12426, 2014.

Dirk Pflugmacher, Andreas Rabe, Mathias Peters, and Patrick Hostert. Mapping pan-european land cover using landsat spectral-temporal metrics and the european lucas survey. *Remote sensing of environment*, 221:583–595, 2019.

Tien Dat Pham, Dieu Tien Bui, Kunihiko Yoshino, and Nga Nhu Le. Optimized rule-based logistic model tree algorithm for mapping mangrove species using alos palsar imagery and gis in the tropical region. *Environmental earth sciences*, 77 (5):1–13, 2018.

Kim Pingoud, Tommi Ekholm, Risto Sievänen, Saija Huuskonen, and Jari Hynynen. Trade-offs between forest carbon stocks and harvests in a steady state–a multi-criteria analysis. *Journal of environmental management*, 210:96–103, 2018.

Planet. URL https://www.planet.com/products/planet-imagery/. Accessed: 2022-05-2022.

Sorin C Popescu, Randolph H Wynne, and Ross F Nelson. Measuring individual tree crown diameter with lidar and assessing its influence on estimating forest volume and biomass. *Canadian journal of remote sensing*, 29(5):564–577, 2003.

Geesara Prathap and Ilya Afanasyev. Deep learning approach for building detection in satellite multispectral imagery. In *2018 International Conference on Intelligent Systems (IS)*, pages 461–465, 2018. doi:10.1109/IS.2018.8710471.

Nicola Puletti, Francesco Chianucci, and Cristiano Castaldi. Use of sentinel-2 for forest classification in mediterranean environments. *Ann. Silvic. Res*, 42(1):32–38, 2018.

S Puliti, M Hauglin, J Breidenbach, P Montesano, CSR Neigh, J Rahlf, S Solberg, TF Klingenberg, and R Astrup. Modelling above-ground biomass stock over norway using national forest inventory data with arcticdem and sentinel-2 data. *Remote Sensing of Environment*, 236:111501, 2020a.

Stefano Puliti, Svetlana Saarela, Terje Gobakken, Göran Ståhl, and Erik Næsset. Combining uav and sentinel-2 auxiliary data for forest growing stock volume estimation through hierarchical model-based inference. *Remote sensing of environment*, 204:485–497, 2018.

Stefano Puliti, Johannes Breidenbach, and Rasmus Astrup. Estimation of forest growing stock volume with uav laser scanning data: can it be done without field data? *Remote Sensing*, 12(8):1245, 2020b.

Tao Qi, Haowei Zhu, Junguo Zhang, Zihe Yang, Lei Chai, and Jiangjian Xie. Patch-u-net: tree species classification method based on u-net with class-balanced jigsaw resampling. *International Journal of Remote Sensing*, 43(2):532–548, 2022.

Rui Qiao, Ali Ghodsi, Honggan Wu, Yuanfei Chang, and Chengbo Wang. Simple weakly supervised deep learning pipeline for detecting individual red-attacked trees in vhr remote sensing images. *Remote Sensing Letters*, 11(7):650–658, 2020. doi:10.1080/2150704X.2020.1752410. URL https://doi.org/10.1080/2150704X.2020.1752410.

Yuanwei Qin, Xiangming Xiao, Jinwei Dong, Geli Zhang, Masanobu Shimada, Jiyuan Liu, Chungan Li, Weili Kou, and Berrien Moore III. Forest cover maps of china in 2010 from multiple approaches and data sources: Palsar, landsat, modis, fra, and nfi. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109:1–16, 2015.

Zixuan Qiu, Zhong-Ke Feng, Mingming Wang, Zhenru Li, and Chao Lu. Application of uav photogrammetric system for monitoring ancient tree communities in beijing. *Forests*, 9(12):735, 2018.

Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8160–8168, 2019.

Parinaz Rahimzadeh-Bajgiran, Masashi Munehiro, and Kenji Omasa. Relationships between the photochemical reflectance index (pri) and chlorophyll fluorescence parameters and plant pigment indices at different leaf growth stages. *Photosynthesis research*, 113(1-3):261–271, 2012.

W Gareth Rees, Jack Tomaney, Olga Tutubalina, Vasily Zharko, and Sergey Bartalev. Estimation of boreal forest growing stock volume in russia from sentinel-2 msi and land cover classification. *Remote Sensing*, 13(21):4483, 2021.

Johannes Reiche, Eliakim Hamunyela, Jan Verbesselt, Dirk Hoekman, and Martin Herold. Improving near-real time deforestation monitoring in tropical dry forests by combining dense sentinel-1 time series with landsat and alos-2 palsar-2. *Remote Sensing of Environment*, 204:147–161, 2018.

Florian Reiner, Martin Brandt, Xiaoye Tong, Ankit Kariryaa, Compton Tucker, and Rasmus Fensholt. Mapping continental african tree cover at individual tree level with planet nanosatellites. In *AGU Fall Meeting Abstracts*, volume 2021, pages B55E–1257, 2021.

Hui Ren, Jia Li, and Nan Gao. Two-stage sketch colorization with color parsing. *IEEE Access*, 8:44599–44610, 2019.

G Reyes-Palomeque, JM Dupuy, CA Portillo-Quintero, JL Andrade, FJ Tun-Dzul, and JL Hernández-Stefanoni. Mapping forest age and characterizing vegetation structure and species composition in tropical dry forests. *Ecological Indicators*, 120:106955, 2021.

Mohammad Rezaee, Masoud Mahdianpari, Yun Zhang, and Bahram Salehi. Deep convolutional neural network for complex wetland classification using optical remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(9):3030–3039, 2018.

Clarence W Richardson. Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water resources research*, 17(1):182–190, 1981.

Herbert Robbins and Sutton Monro. A stochastic approximation method. *The annals of mathematical statistics*, pages 400–407, 1951.

David R Roberts, Volker Bahn, Simone Ciuti, Mark S Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, 2017.

Pedro Rodríguez-Veiga, James Wheeler, Valentin Louis, Kevin Tansey, and Heiko Balzter. Quantifying forest biomass carbon stocks from space. *Current Forestry Reports*, 3(1):1–18, 2017.

Rebecca Roelofs, Vaishaal Shankar, Benjamin Recht, Sara Fridovich-Keil, Moritz Hardt, John Miller, and Ludwig Schmidt. A meta-analysis of overfitting in machine learning. *Advances in Neural Information Processing Systems*, 32, 2019.

Brendan M Rogers, Kylen Solvik, Edward H Hogg, Junchang Ju, Jeffrey G Masek, Michael Michaelian, Logan T Berner, and Scott J Goetz. Detecting early warning signals of tree mortality in boreal north america using multiscale satellite data. *Global change biology*, 24(6):2284–2304, 2018.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Amirhossein Rostami, Reza Shah-Hosseini, Shabnam Asgari, Arastou Zarei, Mohammad Aghdami-Nia, and Saeid Homayouni. Active fire detection from landsat-8 imagery using deep multiple kernel learning. *Remote Sensing*, 14(4):992, 2022.

Yaser Sadeghi, Benoît St-Onge, Brigitte Leblon, Jean-François Prieur, and Marc Simard. Mapping boreal forest biomass from a srtm and tandem-x based on canopy height model and landsat spectral indices. *International journal of applied earth observation and geoinformation*, 68:202–213, 2018.

Anastasiia Safonova, Siham Tabik, Domingo Alcaraz-Segura, Alexey Rubtsov, Yuriy Maglinets, and Francisco Herrera. Detection of fir trees (abies sibirica) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote sensing*, 11(6):643, 2019.

Sancho Salcedo-Sanz, Pedram Ghamisi, María Piles, M Werner, Lucas Cuadra, A Moreno-Martínez, Emma Izquierdo-Verdiguier, Jordi Muñoz-Marí, Amirhosein Mosavi, and Gustau Camps-Valls. Machine learning information fusion in earth observation: A comprehensive review of methods, applications and data sources. *Information Fusion*, 63:256–272, 2020.

Pegah Salehi and Abdolah Chalechale. Pix2pix-based stain-to-stain translation: A solution for robust stain normalization in histopathology images analysis. In *2020 International Conference on Machine Vision and Image Processing (MVIP)*, pages 1–7. IEEE, 2020.

Maurizio Santoro, Oliver Cartus, Nuno Carvalhais, Danaë Rozendaal, Valerio Avitabile, Arnan Araza, Sytze De Bruin, Martin Herold, Shaun Quegan, Pedro Rodríguez-Veiga, et al. The global forest above-ground biomass pool for 2010 estimated from high-resolution satellite observations. *Earth System Science Data*, 13(8):3927–3950, 2021.

Ekrem Saralioglu and Oguz Gungor. Semantic segmentation of land cover from high resolution multispectral satellite images by spectral-spatial convolutional neural network. *Geocarto International*, pages 1–21, 2020.

Takeshi Sasaki, Junichi Imanishi, Keiko Ioki, Yukihiro Morimoto, and Katsunori Kitada. Object-based classification of land cover and tree species by integrating airborne lidar and high spatial resolution imagery data. *Landscape and Ecological Engineering*, 8(2):157–171, 2012.

Dmitry Schepaschenko, Elena Moltchanova, Stanislav Fedorov, Victor Karminov, Petr Ontikov, Maurizio Santoro, Linda See, Vladimir Kositsyn, Anatoly Shvidenko, Anna Romanovskaya, et al. Russian forest sequesters substantially more carbon than previously reported. *Scientific reports*, 11(1):1–7, 2021.

Felix Schiefer, Teja Kattenborn, Annett Frick, Julian Frey, Peter Schall, Barbara Koch, and Sebastian Schmidtlein. Mapping forest tree species in high resolution uav-based rgb-imagery by means of convolutional neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing*, 170:205–215, 2020.

Michael Schmitt, Jonathan Prexl, Patrick Ebel, Lukas Liebel, and Xiao Xiang Zhu. Weakly supervised semantic segmentation of satellite images for land cover mapping–challenges and opportunities. *arXiv preprint arXiv:2002.08254*, 2020.

Sequential model-based optimization in Python scikit optimize. URL https://scikit-optimize.github.io/stable/. Accessed: 2022-10-2022.

Grant J Scott, Matthew R England, William A Starms, Richard A Marcum, and Curt H Davis. Training deep convolutional neural networks for land–cover classification of high-resolution imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(4):549–553, 2017.

Securewatch. Securewatch. https://www.maxar.com/products/securewatch. Accessed: 2021-08-17.

F Sedano, S N Lisboa, R Sahajpal, L Duncanson, N Ribeiro, A Sitoe, G Hurtt, and C J Tucker. The connection between forest degradation and urban energy demand in sub-saharan africa: a characterization based on high-resolution remote sensing data. *Environmental Research Letters*, 16(6):064020, may 2021. doi:10.1088/1748-9326/abfc05. URL https://doi.org/10.1088/1748-9326/abfc05.

Nathalie Seddon, Alexandre Chausson, Pam Berry, Cécile AJ Girardin, Alison Smith, and Beth Turner. Understanding the value and limits of nature-based solutions to climate change and other global challenges. *Philosophical Transactions of the Royal Society B*, 375(1794):20190120, 2020.

Sebastian Seibold, Werner Rammer, Torsten Hothorn, Rupert Seidl, Michael D Ulyshen, Janina Lorz, Marc W Cadotte, David B Lindenmayer, Yagya P Adhikari, Roxana Aragón, et al. The contribution of insects to global forest deadwood decomposition. *Nature*, 597(7874):77–81, 2021.

Sen2Cor. Sen2cor. https://step.esa.int/main/third-party-plugins-2/sen2cor/, Accessed: 2020.

Sentinel-Hub. Sentinel Hub. https://www.sentinel-hub.com/explore/eobrowser/. Accessed: 2020-08-17.

Sentinel-2 L2A SentinelHub. URL https://docs.sentinel-hub.com/api/latest/data/sentinel-2-l2a/. Accessed: 2022-10-2022.

Andie Setiyoko, I Gede Wahyu Surya Dharma, and Toto Haryanto. Recent development of feature extraction and classification multispectral/hyperspectral images: a systematic literature review. In *Journal of Physics: Conference Series*, volume 801, page 012045. IOP Publishing, 2017.

Seyd Teymoor Seydi, Mahdi Hasanlou, and Jocelyn Chanussot. Burnt-net: Wildfire burned area mapping with single post-fire sentinel-2 data and deep learning morphological neural network. *Ecological Indicators*, 140:108999, 2022.

D. Shadrin, A. Menshchikov, A. Somov, G. Bornemann, J. Hauslage, and M. Fedorov. Enabling precision agriculture through embedded sensing with artificial intelligence. *IEEE Transactions on Instrumentation and Measurement*, 69(7): 4103–4113, 2020.

Syed Aamir Ali Shah, Muhammad Asif Manzoor, and Abdul Bais. Canopy height estimation at landsat resolution using convolutional neural networks. *Machine Learning and Knowledge Extraction*, 2(1):23–36, 2020.

Riyaaz Uddien Shaik, Lorenzo Fusilli, and Laneve Giovanni. New approach of sample generation and classification for wildfire fuel mapping on hyperspectral (prisma) image. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 5417–5420. IEEE, 2021.

Yunxiao Shan, Xiangtong Yao, Hongquan Lin, Xuesong Zou, and Kai Huang. Lidar-based stable navigable region detection for unmanned surface vehicles. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2021.

Zhenfeng Shao and Linjing Zhang. Estimating forest aboveground biomass by combining optical and sar data: a case study in genhe, inner mongolia, china. *Sensors*, 16(6):834, 2016.

Ram P Sharma, Zdeněk Vacek, Stanislav Vacek, and Miloš Kučera. Modelling individual tree height–diameter relationships for multi-layered and multi-species forests in central europe. *Trees*, 33(1):103–119, 2019.

David Sheeren, Mathieu Fauvel, Veliborka Josipović, Maïlys Lopes, Carole Planque, Jérôme Willm, and Jean-François Dejoux. Tree species classification in temperate forests using formosat-2 satellite image time series. *Remote Sensing*, 8(9):734, 2016a.

David Sheeren, Mathieu Fauvel, Veliborka Josipovic, Mailys Lopes, Carole Planque, Jérôme Willm, and Jean-François Dejoux. Tree species classification in temperate forests using formosat-2 satellite image time series. *Remote Sensing*, 8:734, 09 2016b. doi:10.3390/rs8090734.

Shiwen Shen, Simon X Han, Denise R Aberle, Alex A Bui, and William Hsu. An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification. *Expert systems with applications*, 128:84–95, 2019.

Emily T Shinzato, Yosio E Shimabukuro, Nicholas C Coops, Piotr Tompalski, and Esthevan AG Gasparoto. Integrating area-based and individual tree detection approaches for estimating tree volume in plantation inventory using aerial image and airborne laser scanning data. *iForest-Biogeosciences and Forestry*, 10(1):296, 2016.

ET Shinzato, YE Shimabukuro, NC Coops, P Tompalski, and EA Gasparoto. Integrating area-based and individual tree detection approaches for estimating tree volume in plantation inventory using aerial image and airborne laser scanning data. *iForest - Biogeosciences and Forestry*, (1):296–302, 2017. doi:10.3832/ifor1880-009. URL `https://iforest.sisef.org/contents/?id=ifor1880-009`.

Connor Shorten and T. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6:1–48, 2019.

Arunima Singh and Sunni Kanta Prasad Kushwaha. Forest degradation assessment using uav optical photogrammetry and sar data. *Journal of the Indian Society of Remote Sensing*, 49(3):559–567, 2021.

Praveer Singh and Nikos Komodakis. Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 1772–1775. IEEE, 2018.

David L Skole, Jay H Samek, Cheikh Mbow, Michael Chirwa, Dan Ndalowa, Tangu Tumeo, Daud Kachamba, Judith Kamoto, Alfred Chioza, and Francis Kamangadazi. Direct measurement of forest degradation rates in malawi: Toward a national forest monitoring system to support redd+. *Forests*, 12(4):426, 2021.

Henning Skriver. Crop classification by multitemporal c-and l-band single-and dual-polarization and fully polarimetric sar. *IEEE Transactions on Geoscience and Remote Sensing*, 50(6):2138–2149, 2011.

Jia Song, Shaohua Gao, Yunqiang Zhu, and Chenyan Ma. A survey of remote sensing image classification based on cnns. *Big earth data*, 3(3):232–254, 2019.

Camile Sothe, Michele Dalponte, Cláudia Maria de Almeida, Marcos Benedito Schimalski, Carla Luciane Lima, Veraldo Liesenberg, Gabriela Takahashi Miyoshi,

and Antonio Maria Garcia Tommaselli. Tree species classification in a highly diverse subtropical forest integrating uav-based photogrammetric point cloud and hyperspectral data. *Remote Sensing*, 11(11):1338, 2019.

Billie F Spencer Jr, Vedhus Hoskere, and Yasutaka Narazaki. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, 5(2): 199–222, 2019.

Ben Spracklen and Dominick V Spracklen. Synergistic use of sentinel-1 and sentinel-2 to map natural forest and acacia plantation and stand ages in north-central vietnam. *Remote Sensing*, 13(2):185, 2021.

Ritika Srinet, Subrata Nandy, Hitendra Padalia, Surajit Ghosh, Taibanganba Watham, NR Patel, and Prakash Chauhan. Mapping plant functional types in northwest himalayan foothills of india using random forest algorithm in google earth engine. *International Journal of Remote Sensing*, 41(18):7296–7309, 2020.

Grant Staben, Arko Lucieer, and Peter Scarth. Modelling lidar derived tree canopy height from landsat tm, etm+ and oli satellite imagerya machine learning approach. *International journal of applied earth observation and geoinformation*, 73:666–681, 2018.

Andrei Stoian, Vincent Poulain, Jordi Inglada, Victor Poughon, and Dawa Derksen. Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems. *Remote Sensing*, 11(17):1986, 2019.

C Stone, M Webster, J Osborn, and I Iqbal. Alternatives to lidar-derived canopy height models for softwood plantations: a review and example using photogrammetry. *Australian Forestry*, 79(4):271–282, 2016.

Christoph Straub, Jiaojiao Tian, Rudolf Seitz, and Peter Reinartz. Assessment of cartosat-1 and worldview-2 stereo imagery in combination with a lidar-dtm for timber volume estimation in a highly structured forest in germany. *Forestry*, 86 (4):463–473, 2013.

Premysl Stych, Barbora Jerabkova, Josef Lastovicka, Martin Riedl, and Daniel Paluba. A comparison of worldview-2 and landsat 8 images for the classification of forests affected by bark beetle outbreaks using a support vector machine and a neural network: A case study in the sumava mountains. *Geosciences*, 9(9): 396, 2019.

Patricia L Suárez, Angel D Sappa, and Boris X Vintimilla. Infrared image colorization based on a triplet dcgan architecture. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 18–23, 2017.

Damien Sulla-Menashe, Josh M Gray, S Parker Abercrombie, and Mark A Friedl. Hierarchical mapping of annual global land cover 2001 to present: The modis collection 6 land cover product. *Remote Sensing of Environment*, 222:183–194, 2019.

Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

Xian Sun, Bing Wang, Zhirui Wang, Hao Li, Hengchao Li, and Kun Fu. Research progress on few-shot learning for remote sensing image interpretation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14: 2387–2402, 2021.

Ying Sun, Jianfeng Huang, Zurui Ao, Dazhao Lao, and Qinchuan Xin. Deep learning approaches for the mapping of tree species diversity in a tropical wetland using airborne lidar and high-spatial-resolution remote sensing images. *Forests*, 10(11): 1047, 2019.

Tom Swinfield, Jeremy A Lindsell, Jonathan V Williams, Rhett D Harrison, Elva Gemita, Carola B Schönlieb, David A Coomes, et al. Accurate measurement of tropical forest canopy heights and aboveground carbon using structure from motion. *Remote Sensing*, 11(8):928, 2019.

Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-first AAAI conference on artificial intelligence*, 2017.

Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

Chengliang Tang, María Uriarte, Helen Jin, Douglas C Morton, and Tian Zheng. Large-scale, image-based tree species mapping in a tropical forest using artificial perceptual learning. *Methods in Ecology and Evolution*, 12(4):608–618, 2021.

Rongxin Tang, Hualin Liu, and Jingbo Wei. Visualizing near infrared hyperspectral images with generative adversarial networks. *Remote Sensing*, 12(23):3848, 2020.

Xiaojing Tang, Eric L Bullock, Pontus Olofsson, Stephan Estel, and Curtis E Woodcock. Near real-time monitoring of tropical forest disturbance: New algorithms and assessment framework. *Remote Sensing of Environment*, 224:202–218, 2019.

Shengli Tao, Qinghua Guo, Chao Li, Zhiheng Wang, and Jingyun Fang. Global patterns and determinants of forest canopy height. *Ecology*, 97(12):3265–3270, 2016.

Luke Taylor and Geoff Nitschke. Improving deep learning with generic data augmentation. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1542–1547. IEEE, 2018.

TensorFlow. Tensorflow. 2020–2021. https://github.com/tensorflow/tensorflow, Accessed: 20 November 2021).

Andualem Aklilu Tesfaye and Berhan Gessesse Awoke. Evaluation of the saturation property of vegetation indices derived from sentinel-2 in mixed crop-forest ecosystem. *Spat. Inf. Res*, 29:109–121, 2021.

Thejna Tharammal, Govindasamy Bala, Narayanappa Devaraju, and Ramakrishna Nemani. A review of the major drivers of the terrestrial carbon uptake: model-based assessments, consensus, and uncertainties. *Environmental Research Letters*, 14(9):093005, 2019.

Nathan Thomas, Priscilla Baltezar, David Lagomasino, Atticus Stovall, Zaheer Iqbal, and Lola Fatoyinbo. Trees outside forests are an underestimated resource in a country with low forest cover. *Scientific reports*, 11(1):1–13, 2021.

Nathan Marc Thomas, Priscilla Baltezar, David Lagomasino, Seung-Kuk Lee, Temilola Fatoyinbo, Justin Green, and Mizanur Rahman. Extent and canopy height maps of trees outside forest (tof) for bangladesh. *AGUFM*, 2018:B31I–2599, 2018.

Jonathan Tompson, Ross Goroshin, Arjun Jain, Yann LeCun, and Christoph Bregler. Efficient object localization using convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 648–656, 2015.

Bang Nguyen Tran, Mihai A Tanase, Lauren T Bennett, and Cristina Aponte. Evaluation of spectral indices for assessing fire severity in australian temperate forests. *Remote sensing*, 10(11):1680, 2018.

Claire C Treat, Maija E Marushchak, Carolina Voigt, Yu Zhang, Zeli Tan, Qianlai Zhuang, Tarmo A Virtanen, Aleksi Räsänen, Christina Biasi, Gustaf Hugelius, et al. Tundra landscape heterogeneity, not interannual variability, controls the decadal regional carbon balance in the western russian arctic. *Global Change Biology*, 24(11):5188–5204, 2018.

Alexey Trekin, Vladimir Ignatiev, and Pavel Yakubovskii. Deep neural networks for determining the parameters of buildings from single-shot satellite imagery. *Journal of Computer and Systems Sciences International*, 59:755–767, 2020.

Øivind Due Trier, Arnt-Børre Salberg, Jörg Haarpaintner, Dagrun Aarsten, Terje Gobakken, and Erik Næsset. Multi-sensor forest vegetation height mapping methods for tanzania. *European Journal of Remote Sensing*, 51(1):587–606, 2018.

Bambang H Trisasongko, Dyah R Panuju, David J Paull, Xiuping Jia, and Amy L Griffin. Comparing six pixel-wise classifiers for tropical rural land cover mapping using four forms of fully polarimetric sar data. *International Journal of Remote Sensing*, 38(11):3274–3293, 2017.

Bangira Tsitsi. Remote sensing of aboveground forest biomass: A review. *Tropical Ecology*, 57(2):125–132, 2016.

Sakari Tuominen, Andras Balazs, Eija Honkavaara, Ilkka Pölönen, Heikki Saari, Teemu Hakala, and Niko Viljanen. Hyperspectral uav-imagery and photogrammetric canopy height model in estimating forest stand variables. *Silva Fennica*, 51, 2017.

Md Palash Uddin, Md Al Mamun, and Md Ali Hossain. Pca-based feature reduction for hyperspectral remote sensing image classification. *IETE Technical Review*, 38 (4):377–396, 2021.

EarthExplorer USGS. Earthexplorer usgs. https://earthexplorer.usgs.gov/, Accessed: 2020.

Radhesyam Vaddi and Prabukumar Manoharan. Hyperspectral image classification using cnn with spectral and spatial features integration. *Infrared Physics Technology*, 107:103296, 2020. ISSN 1350-4495. doi:https://doi.org/10.1016/j.infrared.2020.103296. URL https://www.sciencedirect.com/science/article/pii/S1350449520300876.

Thomas J Vandal, Daniel McDuff, Weile Wang, Kate Duffy, Andrew Michaelis, and Ramakrishna R Nemani. Spectral synthesis for geostationary satellite-to-satellite translation. *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

Elia Vangi, Giovanni DAmico, Saverio Francini, Francesca Giannetti, Bruno Lasserre, Marco Marchetti, and Gherardo Chirici. The new hyperspectral satellite prisma: Imagery for forest types discrimination. *Sensors*, 21(4):1182, 2021.

Mikko Vastaranta, Xiaowei Yu, Ville Luoma, Mika Karjalainen, Ninni Saarinen, Michael A Wulder, Joanne C White, Henrik J Persson, Markus Hollaus, Tuomas Yrttimaa, et al. Aboveground forest biomass derived using multiple dates of worldview-2 stereo-imagery: quantifying the improvement in estimation accuracy. *International Journal of Remote Sensing*, 39(23):8766–8783, 2018.

Luis Alberto Vega Isuhuaylas, Yasumasa Hirata, Lenin Cruyff Ventura Santos, and Noemi Serrudo Torobeo. Natural forest mapping in the andes (peru): A comparison of the performance of machine-learning algorithms. *Remote Sensing*, 10(5): 782, 2018.

Martin D Venturas, Henry N Todd, Anna T Trugman, and William RL Anderegg. Understanding and predicting forest mortality in the western united states using long-term forest inventory data and modeled hydraulic damage. *New Phytologist*, 230(5):1896–1910, 2021.

Niva Kiran Verma, David W Lamb, Nick Reid, and Brian Wilson. Comparison of canopy volume measurements of scattered eucalypt farm trees derived from high spatial resolution imagery and lidar. *Remote Sensing*, 8(5):388, 2016.

Lucija Viskovic, Ivana Nizetic Kosovic, and Toni Mastelic. Crop classification using multi-spectral and multitemporal satellite imagery with machine learning. In *2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, pages 1–5, 2019. doi:10.23919/SOFTCOM.2019.8903738.

Nicolena vonHedemann, Zachary Wurtzebach, Thomas J Timberlake, Emily Sinkular, and Courtney A Schultz. Forest policy and management approaches for carbon dioxide removal. *Interface focus*, 10(5):20200001, 2020.

Fabien H Wagner, Alber Sanchez, Marcos PM Aidar, Andre LC Rochelle, Yuliya Tarabalka, Marisa G Fonseca, Oliver L Phillips, Emanuel Gloor, and Luiz EOC Aragao. Mapping atlantic rainforest degradation and regeneration history with indicator species using convolutional network. *PloS one*, 15(2):e0229448, 2020.

Fabien Hubert Wagner, Matheus Pinheiro Ferreira, Alber Sanchez, Mayumi CM Hirye, Maciel Zortea, Emanuel Gloor, Oliver L Phillips, Carlos Roberto de Souza Filho, Yosio Edemir Shimabukuro, and Luiz EOC Aragão. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. *ISPRS journal of photogrammetry and remote sensing*, 145: 362–377, 2018.

Kiri Joy Wallace and Bruce D Clarkson. Urban forest restoration ecology: a review from hamilton, new zealand. *Journal of the Royal Society of New Zealand*, 49(3): 347–369, 2019.

Hao Wang and Xuedong Liu. Overview of image colorization and its applications. In *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, volume 5, pages 1561–1565. IEEE, 2021.

Kepu Wang, Tiejun Wang, and Xuehua Liu. A review: Individual tree species classification using integrated airborne lidar and optical imagery with a focus on the urban environment. *Forests*, 10(1):1, 2018.

Sherrie Wang, William Chen, Sang Michael Xie, George Azzari, and David B. Lobell. Weakly supervised deep learning for segmentation of remote sensing imagery. *Remote Sensing*, 12(2), 2020. ISSN 2072-4292. doi:10.3390/rs12020207. URL https://www.mdpi.com/2072-4292/12/2/207.

DW Wanik, JR Parent, EN Anagnostou, and BM Hartman. Using vegetation management and lidar-derived tree height data to improve outage predictions for electric utilities. *Electric Power Systems Research*, 146:236–245, 2017.

RH Waring, NC Coops, W Fan, and JM Nightingale. Modis enhanced vegetation index predicts tree species richness across forested ecoregions in the contiguous usa. *Remote Sensing of Environment*, 103(2):218–226, 2006.

Barry Watkins and Adriaan van Niekerk. A comparison of object-based image analysis approaches for field boundary delineation using multi-temporal sentinel-2 imagery. *Computers and Electronics in Agriculture*, 158:294–302, 2019. ISSN 0168-1699. doi:https://doi.org/10.1016/j.compag.2019.02.009. URL https://www.sciencedirect.com/science/article/pii/S0168169918317733.

WeatherSpark. Weather spark. https://weatherspark.com/, 2020.

Tomihisa Welsh, Michael Ashikhmin, and Klaus Mueller. Transferring color to greyscale images. *ACM Trans. Graph.*, 21(3):277280, July 2002. ISSN 0730-0301. doi:10.1145/566654.566576. URL https://doi.org/10.1145/566654.566576.

Mathias Wessel, Melanie Brandmeier, and Dirk Tiede. Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on sentinel-2 data. *Remote Sensing*, 10(9):1419, 2018.

Pramaditya Wicaksono, Muhammad Afif Fauzan, Ignatius Salivian Wisnu Kumara, Rifka Noviaris Yogyantoro, Wahyu Lazuardi, and Zhafirah Zhafarina. Analysis of reflectance spectra of tropical seagrass species and their value for mapping using multispectral satellite images. *International Journal of Remote Sensing*, 40(23): 8955–8978, 2019.

Phil Wilkes, Mathias Disney, Matheus Boni Vicari, Kim Calders, and Andrew Burt. Estimating urban above ground biomass with multi-scale lidar. *Carbon balance and management*, 13(1):1–20, 2018.

Bichen Wu, Chenfeng Xu, Xiaoliang Dai, Alvin Wan, Peizhao Zhang, Zhicheng Yan, Masayoshi Tomizuka, Joseph Gonzalez, Kurt Keutzer, and Peter Vajda. Visual transformers: Token-based image representation and processing for computer vision. *arXiv preprint arXiv:2006.03677*, 2020a.

Bingfang Wu, Yuan Zeng, and Dan Zhao. Land cover mapping and above ground biomass estimation in china. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3535–3536. IEEE, 2016.

Min Wu, Xin Jin, Qian Jiang, Shin-Jye Lee, Lin Guo, Yide Di, Shanshan Huang, and Jinfang Huang. Remote sensing image colorization based on multiscale senet gan. In *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–6. IEEE, 2019.

Min Wu, Xin Jin, Qian Jiang, Shin-jye Lee, Wentao Liang, Guo Lin, and Shaowen Yao. Remote sensing image colorization using symmetrical multi-scale dcgan in yuv color space. *The Visual Computer*, pages 1–23, 2020b.

Gui-Song Xia, Xiang Bai, Jian Ding, Zhen Zhu, Serge Belongie, Jiebo Luo, Mihai Datcu, Marcello Pelillo, and Liangpei Zhang. Dota: A large-scale dataset for object detection in aerial images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3974–3983, 2018a.

Junshi Xia, Naoto Yokoya, and Tien Dat Pham. Probabilistic mangrove species mapping with multiple-source remote-sensing datasets using label distribution learning in xuan thuy national park, vietnam. *Remote Sensing*, 12(22):3834, 2020.

Qing Xia, Cheng-Zhi Qin, He Li, Chong Huang, and Fen-Zhen Su. Mapping mangrove forests based on multi-tidal high-resolution satellite imagery. *Remote Sensing*, 10(9):1343, 2018b.

Jingfeng Xiao, Frederic Chevallier, Cecile Gomez, Luis Guanter, Jeffrey A Hicke, Alfredo R Huete, Kazuhito Ichii, Wenjian Ni, Yong Pang, Abdullah F Rahman, et al. Remote sensing of the terrestrial carbon cycle: A review of advances over 50 years. *Remote Sensing of Environment*, 233:111383, 2019.

Gang Xu, Zhigang Song, Zhuo Sun, Calvin Ku, Zhe Yang, Cancheng Liu, Shuhao Wang, Jianpeng Ma, and Wei Xu. Camel: A weakly supervised learning framework for histopathology image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

Guang Xu, Xuan Zhu, and Nigel Tapper. Using convolutional neural networks incorporating hierarchical active learning for target-searching in large-scale remote sensing images. *International Journal of Remote Sensing*, 41(11):4057–4079, 2020.

Jinru Xue and Baofeng Su. Significant remote sensing vegetation indices: A review of developments and applications. *Journal of sensors*, 2017, 2017.

Pavel Yakubovskiy. Segmentation models. https://github.com/qubvel/segmentation_models, 2022.

Guang Yang, Yaolong Zhao, Baoxin Li, Yuntao Ma, Ruren Li, Jiangbo Jing, and Yuanyong Dian. Tree species classification by employing multiple features acquired from integrated sensors. *Journal of Sensors*, 2019, 2019.

Peiqi Yang, Christiaan van der Tol, Petya KE Campbell, and Elizabeth M Middleton. Fluorescence correction vegetation index (fcvi): A physically based reflectance index to separate physiological and non-physiological information in far-red sun-induced chlorophyll fluorescence. *Remote sensing of environment*, 240: 111676, 2020.

Junho Yeom, Youkyung Han, Taeheon Kim, and Yongmin Kim. Forest fire damage assessment using uav images: A case study on goseong-sokcho forest fire in 2019. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, 37(5):351–357, 2019.

Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

Tuomas Yrttimaa, Ninni Saarinen, Ville Kankare, Niko Viljanen, Jari Hynynen, Saija Huuskonen, Markus Holopainen, Juha Hyyppä, Eija Honkavaara, and Mikko Vastaranta. Multisensorial close-range sensing generates benefits for characterization of managed scots pine (pinus sylvestris l.) stands. *ISPRS International Journal of Geo-Information*, 9(5):309, 2020.

Jinxiang Yu, Xiyuan Peng, Shaoli Li, Yibo Lu, and Wenjia Ma. A lightweight ship detection method in optical remote sensing image under cloud interference. In *2021 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, pages 1–6. IEEE, 2021.

Xingrui Yu, Xiaomin Wu, Chunbo Luo, and Peng Ren. Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. *GIScience & Remote Sensing*, 54(5):741–758, 2017.

Igor Zacharov, Rinat Arslanov, Maksim Gunin, Daniil Stefonishin, Andrey Bykov, Sergey Pavlov, Oleg Panarin, Anton Maliutin, Sergey Rykovanov, and Maxim Fedorov. zhorespetaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in skolkovo institute of science and technology. *Open Engineering*, 9(1):512–520, 2019.

George N Zaimes, Dimitrios Gounaridis, and Elias Symenonakis. Assessing the impact of dams on riparian and deltaic vegetation using remotely-sensed vegetation indices and random forests modelling. *Ecological indicators*, 103:630–641, 2019.

Linglin Zeng, Brian D. Wardlow, Daxiang Xiang, Shun Hu, and Deren Li. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. *Remote Sensing of Environment*, 237:111511, 2020. ISSN 0034-4257. doi:https://doi.org/10.1016/j.rse.2019.111511. URL https://www.sciencedirect.com/science/article/pii/S0034425719305309.

Yelu Zeng, Dalei Hao, Alfredo Huete, Benjamin Dechant, Joe Berry, Jing M Chen, Joanna Joiner, Christian Frankenberg, Ben Bond-Lamberty, Youngryel Ryu, et al. Optical vegetation indices for monitoring terrestrial ecosystems globally. *Nature Reviews Earth & Environment*, pages 1–17, 2022.

Bin Zhang, Lin Zhao, and Xiaoli Zhang. Three-dimensional convolutional neural network model for tree species classification using airborne hyperspectral images. *Remote Sensing of Environment*, 247:111938, 2020.

Fang Zhang, Xiaolin Zhu, and Desheng Liu. Blending modis and landsat images for urban flood mapping. *International Journal of Remote Sensing*, 35(9): 3237–3253, 2014. doi:10.1080/01431161.2014.903351. URL https://doi.org/10.1080/01431161.2014.903351.

Fanyi Zhang, Xin Tian, Haibo Zhang, and Mi Jiang. Estimation of aboveground carbon density of forests using deep learning and multisource remote sensing. *Remote Sensing*, 14(13):3022, 2022a.

Guanwen Zhang, Jien Kato, Yu Wang, and Kenji Mase. How to initialize the cnn for small datasets: Extracting discriminative filters from pre-trained model. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 479–483. IEEE, 2015.

Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*, 2017.

Huifang Zhang, Yi Sun, Li Chang, Yu Qin, Jianjun Chen, Yan Qin, Jiaxing Du, Shuhua Yi, and Yingli Wang. Estimation of grassland canopy height and aboveground biomass at the quadrat scale using unmanned aerial vehicle. *Remote sensing*, 10(6):851, 2018.

Jianrong Zhang, Hongwei Zhao, and Jiao Li. Trs: Transformers for remote sensing scene classification. *Remote Sensing*, 13(20):4143, 2021a.

Rong Zhang, Mingming Jia, Zongming Wang, Yaming Zhou, Dehua Mao, Chunying Ren, Chuanpeng Zhao, and Xianzhao Liu. Tracking annual dynamics of mangrove forests in mangrove national nature reserves of china based on time series sentinel-2 imagery during 2016–2020. *International Journal of Applied Earth Observation and Geoinformation*, 112:102918, 2022b.

Wei Zhang, Ping Tang, and Lijun Zhao. Remote sensing image scene classification using cnn-capsnet. *Remote Sensing*, 11(5):494, 2019a.

Xin Zhang, Ling Du, Shen Tan, Fangming Wu, Liang Zhu, Yuan Zeng, and Bingfang Wu. Land use and land cover mapping using rapideye imagery based on a novel band attention deep learning method in the three gorges reservoir area. *Remote Sensing*, 13(6):1225, 2021b.

Yihang Zhang, Feng Ling, Giles M Foody, Yong Ge, Doreen S Boyd, Xiaodong Li, Yun Du, and Peter M Atkinson. Mapping annual forest cover by fusing palsar/palsar-2 and modis ndvi during 2007–2016. *Remote Sensing of Environment*, 224:74–91, 2019b.

Yuzhen Zhang, Shunlin Liang, and Lu Yang. A review of regional and global gridded forest biomass datasets. *Remote Sensing*, 11(23):2744, 2019c.

Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

Fuqun Zhou, Detang Zhong, and Rihana Peiman. Reconstruction of cloud-free sentinel-2 image time-series using an extended spatiotemporal image fusion approach. *Remote Sensing*, 12(16):2595, 2020.

Yingwei Zhou. Slot based image augmentation system for object detection. *arXiv preprint arXiv:1907.12900*, 2019.

Zhi-Hua Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5(1):44–53, 08 2017. ISSN 2095-5138. doi:10.1093/nsr/nwx106. URL https://doi.org/10.1093/nsr/nwx106.

Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017a.

Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui-Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, 5 (4):8–36, 2017b. doi:10.1109/MGRS.2017.2762307.

Katarzyna Zielewska-Büttner, Petra Adler, Sven Kolbe, Ruben Beck, Lisa Maria Ganter, Barbara Koch, and Veronika Braunisch. Detection of standing deadwood from aerial imagery products: two methods for addressing the bare ground misclassification issue. *Forests*, 11(8):801, 2020.

Barret Zoph, Ekin D Cubuk, Golnaz Ghiasi, Tsung-Yi Lin, Jonathon Shlens, and Quoc V Le. Learning data augmentation strategies for object detection. In *European Conference on Computer Vision*, pages 566–583. Springer, 2020.