Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Daniil Kononenko

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Learnable warping-based approach to image re-synthesis with application to gaze redirection.

Supervisor: Professor Victor Lempitsky

Chair of PhD defense Jury: Professor Maxim Fedorov **Email:** m.fedorov@skoltech.ru

Date of Thesis Defense: October 30, 2017

The thesis document includes the following changes in answer to the external review process

Reviewer Comment 1: The structure of the chapter 1 is slightly complicated and could benefit by splitting the discussion of related work into further subsections or paragraphs.

Author: Agreed. I allocated sub-sections about VAEs and GANs and also split the section about related work on gaze redirection.

Reviewer Comment 2: The caption of Figure 1.14 was slightly confusion and could be rephrased.

Author: The caption is rephrased.

Reviewer Comment 3: It would have been interesting to have additional discussion about effect of camera position and a possibility to utilize eye tracker in data collection.

Author: Using the eye tracker could benefit the data collection. It can validate that user actually followed the point on the screen, and, otherwise, the bad shots could be removed from the training data. Eye tracker was not used in this work, therefore the process of removing outliers was more complicated, including manually looking through images with high training error. The different camera positions could potentially enlarge the training database and be useful for test-time scenarios with unusual camera position, such as camera on the side of the monitor.

Reviewer Comment 4: All videos seem to be recorded with same distance from the camera. The effect of distance could have been discussed and perhaps evaluated further.

Author: The bounding box is defined to be covariant to the scale of the eye (see Section 2.3). Therefore, it is sufficient to use the same distance to the camera for all training images.

Reviewer Comment 5: It was also unclear if the dataset was collected as a part of this thesis work. In such case, it could have been clearly listed as a contribution in Chapter 1.

Author: The data collection protocol was suggested by the thesis author. However, the collection of the dataset required substantial efforts of several members of our research group and I don't claim it as an exclusively personal contribution.

Reviewer Comment 6: It would have been interesting to have further discussion about the temporal stability of the method, how face registration affects the practical performance of the method, and a short description how the numerical parameters were obtained.

Author: Considering short description how the numerical parameters were obtained – added to the thesis text. The temporal stability of the method could not be better than the temporal stability of the face alignment method used because it affects both the bounding box and the input features. Qualitative testing showed that with particular face alignment method used the temporal stability of the method was satisfactory for use in practice. Face registration slightly decreases both training and validation error on a Columbia dataset, where images are worse registered. However, this effect was not noticed on the Skoltech dataset, so face registration is not applied in the final variant of the method.

Reviewer Comment 7: The description of the interpolation kernel and related backpropagation on page 63 could be slightly elaborated.

Author: The description was detailed.

Reviewer Comment 8: Description of the models in Section 4.6.1.1 could have been more comprehensive to help the reader to understand the modifications done to DeepWarp architecture.

Author: The details on how the DeepWarp model was simplified were added to the text.

Reviewer Comment 9, 10: The main challenge is to understand how the differences in the MSE errors reflect the practical performance differences of the methods. In my opinion, the work could have included slightly more comprehensive discussion about practical validity of the quantitative evaluation criteria.

Perhaps some discussion about the limitation of the user study approach could have been included. For instance, it only evaluates only the photorealism of the results and as far as I understand a trivial method that produces zero warping field would get high score.

Author: The Section 2.4.1, describing comparative advantages and drawbacks of assessment methods, used in the work, was added to the thesis text.

Reviewer Comment 11: There could have been discussion about the possibility to include similar lighting correction module also to random forest approach.

Author: Yes, it could be included for example as the second forest, predicting the single value of the lightness map.

Reviewer Comment 12: Nice illustrations of the resulting warping field are presented in Figure 5.1., similar could have been produced also in previous chapters.

Author: More illustrations were added.

Reviewer Comment 13: A table summarizing the previous work on gaze redirection with their main properties could be useful as well.

Author: The table was added to the Section 1.2.

Reviewer Comment 14: It would have been interesting to know if gaze direction estimation network could have been utilized to evaluate the analogy property.

Author: Agreed, thanks for the valuable suggestion. Proposed experiment would be conducted in the future work.

Reviewer Comment 15: It would have been interesting to have some examples of the actual warping fields in the document or a discussion of their inherent dimensionality.

Author: More examples of warping fields were added. To determine inherent functionality, an experiment with PCA learned on predicted warping fields was added to the Section 5.3.3.

Reviewer Comment 16: The authors claim generality of the method to image re-synthesis problems, however there are no examples or discussion of any other domains where the methods would be applicable. The gaze redirection problem has very strong priors on the appearance of the re-synthesised image, so it's not obvious which other problems have such strong priors.

Author: Agreed, the evaluation experiments presented in the thesis do not go beyond gaze redirection task. However, a review in the Section 1.1.2 contains some recent works on image re-synthesis via warping, which were published, when the work on this thesis was underway. They applied the same idea to such tasks as novel view synthesis, manipulating facial expressions, video interpolation.

Reviewer Comment 17: In this case it would be preferred a passive voice to be used instead of an active voice.

Author: Agreed, during proofreading I aimed to change to passive voice.

Reviewer Comment 18: An attempt to verify the result using the prediction of the real angle is done, but this neural network (denoted model E on page 80) gives quite large variance.

Author: Such variance corresponds to quite low validation error of the evaluation network. Therefore, I don't agree that the variance is large, I would call it normal. However, I agree that since evaluation network is not perfect, it can't be used directly to evaluate each testing example because this very example could be an outlier for the evaluation network. This is why in the thesis text the distributions overall validation set are compared with the distribution for the ground truth.

Reviewer Comment 19: A user study has been done, but the statistical significance of the results of this study (and the question if the selection of users was representative) is a separate story.

Author: Applying statistical tests to analyze results is difficult, because of the high variance in users' results: some people are much attentive than other. Therefore, for example, increasing the number of people does not solve the problem, because there is still a significant amount of very good and very bad results. Instead, we should increase the number of tests for each user, but the user study, which takes several hours per user, is more complicated (and expensive) to organize.

Reviewer Comment 20: How the methods from this work can be extended to other similar problems?

Author: Depending on the problem statement, the pipeline for image re-synthesis could be slightly changed. For example, for the novel view synthesis, it could be impossible to generate a novel view because of dis-occlusions. In this case, several input views of the object could be used, and several warping fields predicted, along with the map predicting the occlusions, which determines the particular input image for each pixel.

Reviewer Comment 21: How these methods can be developed further?

Author: Possible direction of the developing is combining different losses. For example, for deep learning based methods, an attribute loss could be useful. Basically, it is the same evaluation network, which is used for assessment of the redirection angle. But it could be used as an additional loss during the training and possibly trained jointly with the warping field predictor.

Reviewer Comment 22: Overall, what is the fundamental impact of the work (in terms of making impact on the specific areas of research = image re-synthesis)?

Author: Some papers on image re-synthesis using the same idea of learnable warping fields came out when the work on the thesis was approaching the finish. Some of these papers are citing our works. Therefore, I would say, that, apart from the contributions of the work itself, stated in the text of the thesis, the work has the fundamental impact of encouraging other research on learnable warping field approach to image re-synthesis.

Reviewer Comment 23: The authors claim generality of the method to image re-synthesis problems, however there are no examples or discussion of any other domains where the methods would be applicable. The gaze redirection problem has very strong priors on the appearance of the re-synthesised image, so it’s not obvious which other problems have such strong priors.

Author: Agreed. The clarifications are added to the text of thesis. The approach is applicable to image re-synthesis problems, where the transformations can be well approximated by warping, in particular where dis-occlusion and global color changes are minimal.