Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Daniil Kononenko

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Gaze redirection in Images Using Maching Learning.

Supervisor: Professor Victor Lempitsky

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

Date of Thesis Defense: October , 2017

Name of Reviewer: Dimitris Samaras

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** **Date:** ***24-10-2017***

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to forward a completed copy of this report to the Chair of the Jury at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.
If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

The thesis presents an approach to the gaze redirection task, where the goal is to re-synthesize an image of a person in such a way that gaze is redirected by a certain angle. To overcome the blurriness that plagues standard generative models, a warping based approach is utilized, where the warping field is predicted from the input image and the predictor is learned from a dataset of input and output images.

The first chapter presents the image re-synthesis, gaze redirection and gaze correction problems. A comprehensive literature review of generative models and gaze redirection, includes recent work, contemporaneous to the thesis. The chapter also describes the author's contributions. Those are the suggested methods for learning a warping field predictor for the image re-synthesis task. The suggested methods are useful for gaze correction in videoconferencing and for image and video post-processing. The results of the work have beennpublished in CVPR and ECCV articles, and in a TPAMI paper.

The warping approach to image re-synthesis is described in the second chapter. In the case of gaze redirection, only the eye regions are warped. The chapter also describes the collection of the specific Skoltech dataset for the gaze redirection problem. The literature review of face alignment methods is given. Finally, the chapter presents general approaches to image re-synthesis and in particular, to gaze redirection, given the learned warping field predictor.

The main contributions of the thesis are the next four chapters proposing methods for learning a warping filed predictor from a dataset of images. Chapter 3 presents an approach, based on a weakly-supervised random forest. The tree contains a structured output in its leaves, predicting the warping error distribution for a particular pixel in the output image. The prediction sums up these distributions from different trees, by minimizing the overall error. The chapter starts with an introduction of the random forest approach in machine learning and its applications in computer vision. Further, the weakly-supervised random forest for image re-synthesis is described in detail and its experimental evaluation is given. The method produces photorealistic results and works in real time on a single SPU core. However, the trained model is restricted to a redirection by a fixed angle.

Chapter 4 presents a DeepWarp approach for image re-synthesis, based on a deep learning. The neural network predicts an output for an input image and an arbitrary redirection angle. Except for the warping field, the network also predicts a special lightness correction map to increase the photorealism of the result. The chapter presents related work in deep learning and an experimental comparison with a forest-based method. The DeepWarp model outperforms the forest-based one but is slower.

Chapter 5 presents a fully-supervised method, which is a hybrid of the first two methods. It is based on the same forest architecture but it is trained using the predicted warping field of a neural network instead of ground-truth images. The chapter starts with a review of teacher-student architectures. All three methods described previously are evaluated qualitatively and quantitatively, using the MSE error, a user study and an additional network trained to assure, that the gaze was redirected on the desired angle. The hybrid method produces quite good results and is fast due to the forest-based architecture.

Chapter 6 introduces a semi-supervised method for gaze redirection, suitable when there is not enough labeled data. Related work on image analogies is discussed. The suggested architecture is trained in an unsupervised way and utilizes a small part of labeled data in testing time. The method shows a substantial advantage over fully supervised methods when only a small part of the training dataset has been labeled.

Chapter 7 concludes the thesis and summarizes the presented methods. In the latter part of the chapter, several limitations of the approach are discussed and comparative advantages and drawbacks of the methods for learning a warping field predictor are summarized.

Overall the thesis is well written and well argued. The sequence of the proposed methods has internal logic and follows the arc of recent computer vision into deep learning. This thesis proposes a solution to a practical program, using state-of-the-art machine learning methods. The evaluation is well thought out, with a diverse set of experiments. The psycho-physical experiment is interesting even though not formally designed by a trained psychologist. The literature review is thorough, although a bit fragmented. There is a number of language and style mistakes, that somewhat impede the flow of the text. Readability would improve, if the thesis was proofread by a native English speaker. 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. It would have been interesting to have some examples of the actual warping fields in the document or a discussion of their inherent dimensionality. These comments can be easily addressed.

Provisional Recommendation

X☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*

☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*

☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*