Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Dmitry Ulyanov

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Image Generation with Convolutional Neural Networks

Supervisor: Prof. Victor Lempitsky

Name of the Reviewer: Associate professor Evgeny Burnaev

I confirm the absence of any conflict of interest	**Signature:**  **Date: 11-11-2019**

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 submit signed copy of the report 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 addresses the problems related to image synthesis with convolutional neural networks in different contexts such as style transfer, image super-resolution, etc.

The authors considered several important challenges, namely: how to make image style transfer and texture synthesis methods scalable? how to define a texture synthesis loss for fast and diverse texture synthesis? how to define adversarial generator-encoder network capable to perform a backward mapping (to a latent space)? how to construct perceptual discriminators? how to define an image prior for single image and how to construct a method capable to perform the main image restoration tasks? how to construct neural networks capable of full-body rendering of a person for varying body pose and camera position.

The introduction gives the description of the scope, topic, tasks and the context of the thesis. The above-mentioned results are presented in the six corresponding chapters of the thesis and structured in an appropriate manner. Although these chapters are provided in the form of the separate papers, they are based on several solid mathematical/algorithmic ideas and have common grounds; in particular, the author proposed the reparameterization idea to represent a feed-forward approach for style transfer in replace of a standard optimization-based approach. Similar idea was used to define the deep image prior model, which turned out to be a very efficient image prior, suited for all main image restoration tasks. The neural avatar application then benefits from using these results/architectures as building blocks for its own development.

In the second chapter the author considers the reparameterization approach to approximate a local minimum of some energy instead of running local minima search every time we need to generate an image when doing synthesis of textures and stylization of images.

In the third chapter the author considers the same problem statement as in the previous case. The main added value is a new mathematical formulation of the loss function that allows unbiased sampling of textures. Besides that, the author introduces an instance normalization module to improve performance.

In the fourth chapter the adversarial generator-encoder network is considered. The proposed objective compares the divergences of each of the real and the generated data distributions with the prior distribution in the latent space. This helps to avoid using and training any external mappings during the process of learning.

In the fifth chapter the author suggests basing the GAN discriminator on the perceptual statistics computed by the reference network on the input image. The motivation is that a discriminator that uses perceptual features has a better chance to learn good statistics than a discriminator initialized to a random network.

In the six chapter the author developed a seminal Deep Image Prior approach, which already has several hundred citations. The idea is to use the reparameterization approach from the second chapter, but for defining image prior. The authors showed superior performance for different image restoration tasks.

The application – 2.5D framework for realistic rendering of human avatars that combines neural rendering with classical rendering techniques – is considered in the seventh chapter.

The application uses as basic building blocks some of the developments, described in the previous chapters.

The presented results are demonstrated for many publicly available datasets, as well as private ones, and show the suggested methods to be relevant to the considered tasks. The scientific significance is evidenced by the quality of the corresponding publications (A* – level conferences).

Thus, the overall structure of the dissertation is logical; separate sections of the dissertations correspond to different aspects of image synthesis with convolutional neural networks in different contexts such as style transfer, image super-resolution, etc.

The topic of the dissertation is fully relevant to its actual content. The methods are novel, and results of experiments support their efficiency. The obtained results are significant as they are based on solid mathematical ideas and provide grounds for the development of efficient image analysis methods; competitive performance with SOTA methods as well as papers in top venues supports compliance of the results with the international level and current state of the art. The dissertation contains all necessary experimental evaluation, which this proves the relevance of the obtained results to applications. The neural avatar model is novel and innovative.

The thesis is accurately written. There were some technical misprints, which had been operatively corrected by the author.

During the defense it could be interesting to discuss the following issues:

- There exist different functions used as losses when doing synthesis of textures and stylization of images. Which of those functions turned out to be the best when combined with the reparameterization approach?
- The reparameterization approach is actually some method to find an optimum of the function; here we parameterized that optimum by a convolutional neural network. Is this method applicable to other optimization tasks?
- The entropy estimate, used in the third chapter, is not efficient in multidimensional space. In our case as objects we consider images, which have a very high-dimension. Could the author comment somehow on why the regularization based on the entropy estimate still works?
- Fitting the deep image prior is a computationally intensive task. Thus, the method is not appropriate to be used in real-time applications. How can this be fixed?

In any case, the work is solid, it fulfils all requirements to PhD thesis. Actually, I do not expect detailed answers to questions above (as they require a separate line of research), rather I would like to listen to the author's opinion.

Yes: *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*