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

Co-advisor: Prof. Andrea Vedaldi, University of Oxford

Date of Thesis Defense: 11 December 2019

Name of the Reviewer: Ivan Laptev

I confirm the absence of any conflict of interest	**Signature:**  **Date: 30-11-2019**

Reviewer's Report
The thesis of Dmitry Ulyanov presents a comprehensive and outstanding work including multiple original ideas, impressive results and new findings that have already influenced the field. The thesis is focused on image generation and advances the recently successful paradigm of generative neural networks. Starting with efficient methods for texture generation and style transfer, the work develops a set of new conditional generators addressing drawbacks of previous methods and advancing applications. Besides improved understanding of generative neural networks and their training, this thesis contributes to solutions of several practical problems including image denoising, super-resolution and inpainting. The desire and art to create images goes back to prehistoric drawings of Chauvet Cave and alike. The advent of computers and computer graphics has made a significant impact on the society by revolutionizing movie and gaming industries. While such industries still heavily rely of manual work for content creation, the most recent progress in convolutional neural networks presents a promise of a new revolution for automatic image generation and editing. In this context, the current thesis address a timely topic with large potential. The work of the thesis builds on the successful backbones of convolutional neural networks (CNNs) and Generative Adversarial Networks (GANs). The developed methods improve the training and the inference of previous methods, combine advantages of alternative approaches and contribute to the general understanding of generative networks. The impact of this work has been validated through publications

in top conferences on computer vision and machine learning, many of which have already obtained high citation counts, e.g., ICML 2016 (408 citations), CVPR 2017 (170 citations) and CVPR 2018 (328 citations).

The PhD thesis is organized in eight chapters where Chapter 1 presents a short motivation followed by an overview of contributions. Chapters 2-7 present main contributions organized in a chronological order as a collection of papers by the author. The thesis concludes with a summary in Chapter 8. While the thesis is logically organized, and contains minimal introduction and conclusions, I would expect to have more discussion of the context. Given the breadth of proposed contributions, a more extensive review of prior work in the field as well as the discussion of application areas would have been an advantage. A chapter presenting details of CNN, GAN, VAE and other related methods would be appropriate. Finally, the thesis would benefit from an outlook and discussion of remaining open challenges.

Chapter 2 proposes Texture Networks (TNs) – a new fast method for generating image textures and stylized images. The method builds on the previous work by Gatys et al 2015 producing high-quality images by optimizing a perceptual loss for each new image. While such an iterative optimization procedure takes several seconds per image, the innovation of TNs is to use the perceptual loss to pre-train a feedforward neural network. Such networks can then be used to generate images with a single inference pass, hence, providing a speed up of two orders of magnitude compared to the original method. Chapter 3 extends TNs by introducing instance normalization and a diversity term that further improves the quality of generated images. The presented results visually look appealing. Presenting results on more images and emphasizing failure cases would have been an advantage.

Chapter 4 presents Adversarial Generator-Encoder networks (AGE). AGE aims to combine advantages of Variational Auto Encoders (VAEs) and Generative Adversarial Networks (GANs). While VAEs enable both image generation and inference by mapping images to a low-dimensional latent subspace, GANs generate better images. The idea of AGE is to define an adversarial game in the latent space. Given an encoder e(X) and a generator g(Z), this is done by (i) optimizing e to maximize the divergence between e(X) and g(e(X)), while at the same time (ii) optimizing g to minimize the same divergence. This is an interesting idea that has been developed to a working solution. Compared to the most related method ALI [Dumoulin et al., 2017], AGE does not require training a discriminator and provides a more convenient training procedure. The comparison of ALI and AGE in practice demonstrates the better performance of AGE especially for limited number of training iterations.

Chapters 5 and 6 present the work on Image Manipulation with Perceptual Discriminators and Textured Neural Avatars. The first work introduces perceptual losses for the training of GAN-like architectures. The second work is focused on image generation for the particular but important case of generating human avatars. The main contributions in this work was done by other co-authors, while Dmitry Ulyanov was helping with experiments and baselines.

Chapter 7 is the last technical chapter presenting the work on Deep Image Prior. CNNs are known to perform very well for various imaging tasks, yet the understanding of this success is still largely lacking. In contrast to the common belief that the success of CNNs comes from the training on a large amount of data, the work on Deep Image Prior investigates an alternative hypothesis. Interestingly and rather surprisingly, it is shown that common tasks such as image denoising, super-resolution and inpainting can be addressed by fitting a randomly initialized CNN to a single test image. The study in this chapter suggests that CNNs are well-adapted to imaging tasks already by their structure without training parameters on a large corpus of images. As mentioned in the thesis, such a behavior is related to the previous work on dictionary learning and self-similarity. It would be interesting to investigate this relation further both

theoretically and practically. The chapter would also benefit from ablation studies, for example, by measuring PSNR as a function of the number of optimization steps.

Overall, the present thesis contains original and innovative ideas to image generation. The material of the thesis is clearly presented. The work of the author has been established in the community which is confirmed by the high number of citations and publications at top conferences of the field. Based on the above, **I highly recommend Dmitry Ulyanov for a PhD degree** at Skolkovo Institute of Science and Technology.

Provisional Recommendation

☒ *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*