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: Dr. Radu Timofte

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:** **Date: 03-12-2019**

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

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications

The summary of issues to be addressed before/during the thesis defense

The thesis written by Dmitry Ulyanov is based on a collection of several papers. It describes algorithmic contributions and results of convolutional networks (ConvNets) with application to image generation, restoration and manipulation. The core chapters of the thesis are based on published peer-reviewed papers at top computer vision and machine learning venues and follow a logical order. The chapters use extensively 6 peer-reviewed published papers while other published papers of the author were not included in the thesis despite being related to its overall topic. It is a bit unfortunate that it is not clear from the thesis why these related papers of the PhD candidate remained unused in the final PhD thesis. On the other hand, what remains is still sufficient by far.

The abstract succinctly provides an overview of the research topics approached in the PhD thesis. The main topics and the corresponding contributions follow the chapter organization of the thesis: i) a fast style transfer method; ii) an improved style transfer method in several aspects; iii) a novel formulation of a GAN game allowing mapping from image space to latent space; iv) a perceptual discriminator for GANs; v) a GAN-based image generation system to synthesize humans guided by pose; vi) a study of ConvNets as deep priors, by decoupling the impact of the architecture from learning on a dataset.
The chapter continues by listing the publications used in this PhD thesis, making a clear distinction between four papers with Dmitry Ulyanov as first author and two papers without, but with listed significant contributions made by Dmitry Ulyanov.

The introductory chapter states the motivation of the conducted research, 'how to generate and manipulate images with convolutional neural networks?'. Then it provides summaries of each of the following chapters based on the six main contributions: fast style transfer, improved texture networks, an Adversarial Generator-Encoder network, perceptual discriminators for image manipulation, textured neural avatars and deep image priors.

The first contribution on fast style transfer is described in the second chapter. It follows the structure of the published ICML2015 conference paper it builds upon: problem formulation, related work, proposed method description, experimental setup and results analysis. The novel efficient texture network is based on the approach of Gatys et al. (2015), a seminal method from the style transfer field. If Gatys et al. require thousands of forward-backward iterations in the network for each processed input image, the novel feed-forward texture network avoids this and greatly improves runtime and memory requirements. At the same time the proposed solution matches the texture capability of Gatys et al. (2015).

The second contribution of the thesis, on improved texture networks, is introduced in the third chapter. The chapter is based on a CVPR 2017 published paper, and introduces the problem and related work to then describe the improved texture networks, the experiments and the results. It addresses the shortcomings of the previous proposed feed-forward texture networks. It is shown that better quality and diversity in the generated outputs can be achieved by replacing batch normalization with a proposed instance normalization module and by using a novel formulation to encourage unbiased sampling from the Julesz texture ensemble by the generators. The nets are much faster but achieve similar quality as the generation-via-optimization methods.

The third contribution of the thesis is in hybrid autoencoders and adversarial networks (based on an AAAI2018 paper). In the proposed Adversarial Generator-Encoder Networks (AGE) the adversarial game is set between the encoder and the generator (decoder) with no external mappings trained in the learning process. The divergences of real and fake generated data distributions are compared in the game objective with the prior distribution in the latent space. The proposed AGE is capable of producing results with a quality comparable to recently-introduced more complex solutions.

The fourth major contribution is described in the fifth chapter of the thesis -- perceptual discriminators for image manipulation (an ECCV2018 paper). The perceptual loss and the adversarial discriminator are combined in a principled and non-additive manner for unpaired image translation tasks. The proposed perceptual discriminator embeds parts of a classification pretrained network within the discriminator network. Quantitative and qualitative results support the design in comparison with state-of-the-art frameworks.

The fifth contribution to learning full-body neural avatars is the system from chapter six. This solution was described in a CVPR 2019 paper and builds upon the classical graphics pipeline and the recent generation of human images based on image to image translation deep learning approaches. The key element is the explicit 2D texture representation of the model surface without explicit 3D shape modeling. A fully-convolutional net learns to map directly configurations of body feature points w.r.t. the camera to the 2D texture coordinates in the image frame. This design is shown to generalize better than other systems using direct image-to-image translation.

The sixth contribution from chapter seven was published at CVPR 2018. The authors show that the structure of a generator network captures a large amount of low-level image statistics without applying any learning. A randomly-initialized network can be used for inverse problems as a handcrafted prior and the results are surprisingly good. The same prior can also be used for inverting deep neural representations or to restore images. This work bridges the gap between learning-free methods based on handcrafted priors such as self-similarity and deep learning-based methods.

The PhD thesis concludes with chapter eight, a brief summary of the contributions.

The PhD thesis has a modular structure and is easy to follow. It is a collection of 6 (peer reviewed) published papers in top vision venues. Each contributed paper/chapter can be seen as a standalone problem formulation, solution description and evaluation. This also is my main criticism: the proposed contributions are rarely studied together. Another criticism is due to the use of research paper style guidelines for the chapters. Concepts and techniques from the (recent) literature - while referenced properly - are seldom sufficiently described and/or reintroduced to facilitate their understanding by the less knowledgeable reader. These shortcomings are mainly due to the collection format adopted by the author for the PhD thesis. The research questions are stated clearly, as well as the contributions and the novelty. The author relates the contributions in the thesis to the relevant works from the research literature and most of the necessary background, technical and theoretical details are provided. Moreover, the author always experimentally validates the proposed solutions against state-of-the-art methods on public benchmarks to achieve comparable or significant improvements. The publicly released codes (https://dmitryulyanov.github.io/) prove the adherence to open research of the PhD candidate and the strong reproducibility of his work.

I, Dr. Radu Timofte, consider that the PhD thesis contains important contributions with significant impact in the research community.

I recommend acceptance of the PhD thesis of the PhD candidate Dmitry Ulyanov.

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*