Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Evgeniya Ustinova

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Image-based human re-identification and recognition using deep learning methods

Supervisor: Prof. Victor Lempitsky

Name of the Reviewer: Andrzej Cichocki

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **31-10-2019** **Date: DD-MM-YYYY**

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

Provisional Recommendation

The topic of the PhD thesis of Ms. Evgeniya USTINOVA is devoted, in general, to human recognition using deep neural networks. The Author of the thesis investigated two challenging and important sub-problems: Person Re-identification and Face Recognition. By person re-identification the author understands matching of full-body pedestrian images across possibly non-overlapping camera views and face recognition task is to match face images. Both sub-problems can be considered as retrieval problems, where a key task is to estimate the similarity of set of pairs of images in the presence of different sources of variation, e.g., various poses, illuminations. The key issue was to estimate the similarity between images pairwise in such a way that semantically related images have a relative high similarity value and semantically unrelated have a relative low similarity value. This is an important and hot topic of research. The author investigated several modern deep learning techniques for person re-identification, surveillance face recognition and deep embedding learning in general. Particularly, the training of Siamese neural networks was investigated in detail. Three important aspects of training Siamese neural network for retrieval and human recognition are investigated by Ms. Ustinova in details: Architecture design, objective (cost, loss) functions design, and cross-domain training. The Author of the thesis mostly investigated feed-forward architectures, layers design, objective function design, as well as applications of recent deep learning methods, especially image-to-image translation and domain-adversarial training. In all cases, standard backpropagation techniques based on state of the arts gradient-based optimization algorithms (Stochastic Gradient Descent with momentum, ADAM) were exploited.

Deep learning has already approach*ed* human performance in face recognition. However, for person re-identification, the results achieved by deep neural networks were rather modest and did not clearly outperform of some "non-deep" methods. Therefore, Ms. Ustinova investigated more carefully deep learning methods for person re-identification and suggested several interesting and innovative improvements in comparison to the existing methods (on the moment of her publications), especially aspects related to neural network architectures and construction of objective functions. Additionally, the Author investigated several cross-domain scenarios related to human recognition. Cross-domain training, which in my opinion, was a very important direction of research, since in many cases we observed a big performance drop, when the system trained on data which is quite different from the test data. For the cross-domain topic the Authors did not suggest completely new methods but rather combines several existing approaches in order to find good practical strategies of domain adaptation.

In very interesting and well written Chapter 3, a novel loss (cost) function for similarity learning was suggested, that does not require any kind of tunable hyper parameters (like e.g., thresholds). As baselines, the Author considers several widely used objective functions: Binomial deviance [Yi et al., 2014], Triplet [Schroff *et al*., 2015] and LSSS [Song *et al*., 2016] losses. The LSSS loss was shown to outperform popular contrastive and triplet losses for a number of retrieval datasets earlier, therefore Ms. Ustinova compares her results with state of the arts approaches. In Chapter 4, the author of the thesis attempts to improve performance over existing methods exploiting deep learning for person re-identification. The closest to her approaches are works of [Yi *et al*., 2014, Cheng et al., 2016] as they used separate streams to process horizontal stripes of the pedestrian image, both works consider 2-depth convolutional streams. Several other approaches [Li *et al*., 2014, Ahmed *et al*., 2015, Wu *et al.,* 2016] also train deep neural networks for person re-identification. However, they do not build a mapping to a descriptor space, but map a pair of images directly to a similarity score. Such approaches may allow modeling the image difference more explicitly but are less scalable than that of building a mapping to a descriptor space. Chapter 5 is devoted to cross-domain person re-identification. For cross-domain person re-identification, on the moment of submission, the non-deep method of Ma *et al*.

[2015] demonstrated the best results in cross-domain person re-identification, where the authors mimic the negative pairs in target domain by random sampling. It is worth emphasizing that results presented in this chapter for cross-domain person re-identification were slightly lower. However, the Author validated the domain-adversarial training to non-classification tasks (being the part of her original paper on this method). It should be mentioned, that later successful methods for cross-domain person re-identification are based on image-level adaptation that the author of the thesis applied for a different task (surveillance face recognition) in Chapter 6. In Chapter 6, the Author demonstrated that domain-specific data augmentation is essential for training face recognition systems. Moreover, an extensive evaluation and validation of image-level adaptation were performed for surveillance face recognition. Additionally, the feature-level domain adaptation was also included for the comparison purpose in Chapter 6. All the methods and algorithms described in Chapters 3, 4, 5 were implemented and tested using Caffe-package (C++ and Python) package for deep learning. Furthermore, the code for Chapter 6 was developed using Pytorch (Python) and utilized the existing CycleGAN code. These software is potentially useful and helpful for many researchers working in these areas.

In my opinion, the most important and original contributions were presented in Chapter 3.

The methods and results presented in this thesis were already published in high quality and influential highly cited papers published in prestigious conferences NIPS, AVSS and Journal of Machine Learning :

• Chapter 5 is closely related to journal paper Yaroslav Ganin, **Evgeniya Ustinova**, Hana Ajakan, Pascal Germain, Hugo Larochelle, Francois Laviolette, Mario Marchand and Victor Lempitsky. Domain-adversarial training of neural networks. The Journal of Machine Learning Research, JMLR, 17(1), pages 2096-2030, 2016. **(cited already over 1100 times).**

• Chapter 3, which in my opinion, it is the best one, is related to excellent publication of Ms Ustinova as leading couthor in prestigious conference: Advances in Neural Information Processing Systems, NIPS, 2016 **(cited over 140 times)** and Russian patent **Evgeniya Ustinova** and Victor Lempitsky. Learning deep embeddings with histogram loss The Russian Federation patent 2641447 has also been issued for this work.

• Chapter 4 is based on following publication **Evgeniya Ustinova**, Yaroslav Ganin and Victor Lempitsky, Multiregion bilinear convolutional neural networks for person re-identification published in Advanced Video and Signal based Surveillance, (AVSS), 2017.

Although this is very high quality PhD thesis, I should mention some minor weakness: Although, some new results have been submitted to arXiv 2018 and 2019, however, no any significant papers hav been published or accepted in the last two years. Also in Bibliography, I could not find any references to very recent related works published in 2019 and quite few form 2018. Since the area of research is extremely competitive and hot/ popular and develops very fast, so now there exist several competitive works showing much better results than those demonstrated in this thesis. For example, the methods that explicitly utilize pose prediction or segmentation and more modern general-purpose architectures improve the results [Saquib Sarfraz et al., 2018, Suh et al., 2018, Kalayeh et al., 2018]. Therefore, in my opinion, the results presented in Chapter 4 become rather more historical and academic than practically useful today. However, this Chapter is an important contribution, since it can be considered as a part of an interesting research direction dedicated to using multiplicative feature interactions in neural networks. However, these are minor remarks, and in general, I am very impressed by quality and contents of the PhD thesis of Ms. Evgeniya Ustinova. In conclusion, the PhD thesis of Evgeiya Ustinova

consists many important, interesting and innovative results. 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*