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

Date of Thesis Defense: 11 December 2019

Name of the Reviewer: Associate professor Ondrej Chum

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **Date: 04-12-2019**

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 addresses methods for building deep learning system and their application to person re-identification and face recognition. Three directions are followed: design of a histogram loss function, regional bi-linear pooling, and domain adaptation. The thesis is divided into seven chapters.

The first chapter introduces the problem, datasets and neural-network architectures used in the research area, the contributions are also listed. Related work relevant to the thesis is discussed in Chapter 2. The core contributions are presented in Chapters 3 – 6. Chapter 7 concludes the thesis.

Chapter 3 introduces a novel loss function for embedding learning through deep neural networks. The histogram loss is designed to separate the distributions of distances between pairs of positive and pairs of negative training examples. Unlike previously used loss functions, such as contrastive of triplet loss, the proposed method does not require any sensitive parameters such as thresholds or margins. The only parameter used in the method is the bin size of the histogram. The method is shown to perform well on a number of tasks.

Chapter 4 is dedicated to architecture design beneficial for person re-identification. The area of interest is divided into three horizontal stripes. Each stripe is described using a standard bi-linear pooling method within each spatial region (stripe). The stripe descriptors are then combined by a fully connected layer. The idea behind such a procedure is to model significant geometric changes (coming from articulated objects) within the regions and to maintain the geometric information in the vertical direction. The experimental results show superior performance of the proposed architecture. It would be interesting to show, whether combining features from different regions (e.g. neighbouring) would be beneficial or not. Is it important to share parameters of the convolutions between the parts, or would part-specific convolutions improve the accuracy?

Chapter 5 addresses a general problem of domain adaptation, where labelled data are available for one, source, domain and only unlabelled data are available for the target domain. The proposed approach of domain-adversarial neural networks combines two losses. One loss is evaluated on labelled source domain data and tries to minimize the true objective of the task. The other loss is evaluated on both source and target domain data on an additional domain-discriminator branch. This branch takes in the feature embeddings and tries to determin the domain of origin. The goal is to train embedding so that even a good discriminator cannot discriminate between the domains. This is achieved by placing gradient reversal between the last embedding layer and the domain discriminator in standard backpropagation. It was shown that the proposed method improves results for domain adaptation in person re-identification, i.e., training on source camera system with labelled data and a different target camera system with unlabelled data (simulated by two different benchmarks).

Chapter 6 tackles image-based domain adaptation techniques for face recognition. As a source domain, annotated Internet faces are used, the target domain being faces from surveillance cameras. The mapping between the domains is learned by CycleGAN. It is shown, that classifier trained on union of the initial and degraded images performs the best, outperforming also the domain-adversarial approach described in Chapter 5.

Overall, the thesis is well organized and written in good English. It contains a number of contributions that are timely and of high quality. Individual contributions were published in a top machine learning journal JMLR 2016 (1265 citations according to Google Scholar), at a top-quality (CORE A*) machine learning conference NIPS 2016 (156 citations), and at an international conference (CORE B) AVSS 2017 (137 citations). While at the conference publications, the candidate was the first author, she is second author on the journal publication. It is a good practice to also report personal contributions for the

publications listed as thesis related. The impact of the work of the candidate can be seen from the high number of citations to this work.

Minor issues. There are some problems with references that could be easily removed. For example, in the beginning of Chapter 3 and in Chapter 6, the formatting of references differs from the rest of the thesis (only the year is in the brackets). Repeated incorrect reference to Section 1.3 as describing architectures, while the section introduces datasets. Three letter references that do not correspond to any paper, e.g., [Zhe] on page 49, [Kin] on page 52, etc. Reference to Chapter ?? on page 64. In the caption of Fig. 6.2.: “last two rows” should be columns.

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*