ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

REVIEWER'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	**Deep neural network for satellite image classification using OpenStreetMap**
Author's name:	**Vladimír Kunc**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Computer Science
Thesis reviewer:	Radoslav Škoviera
Reviewer's department:	Robotics and Machine Perception, Czech Institute of Informatics, Robotics, and Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment — **challenging**

The difficulty of the assignment was above average as the student had to familiarize himself deeply with two research topics and several frameworks and tools. However, the most difficult parts, from my point of view, were the implementation and the analysis of the related work. If evaluated alone, I would rate the difficulty of the requirements on own research as average at most.

Satisfaction of assignment — **fulfilled**

The assignment was fulfilled completely and thoroughly.

Method of conception — **correct**

The student used state of the art convolutional neural networks which are themselves a state of the art method in the area of computer vision. In addition, the student did a thorough evaluation to find the most suitable configuration of the used CNNs.

The methodology used for dataset creation is adequate. However, as it is also pointed out in the thesis itself, the dataset contains several problems (e.g. an image of a village with ground truth set to power_hydro, indicating presence of a hydroelectric power plant; maybe because there is a brook with some homemade hydroelectric generator, see p. 119). Such problems could negatively affect the overall accuracy of the classifier, depending on the number of the problematic images. While the problems come mainly from the used data sources, if the student focused more on the research of the dataset creation techniques, at least some of the problems could have been avoided. Admittedly, in combination with the classification part of the assignment, it would probably extend the complexity of the thesis beyond the scope of a master thesis. Also, the created dataset is still better in many aspects than other available datasets.

What is missing from the thesis is any attempt of addressing the problem with presence of irrelevant objects in the images. For example, some sort of mask (provided with the training dataset and used during training of the CNN) that would cover parts of an image that do not contain the annotated object. This technique was already successfully applied in conjunction with CNNs in a different application. While this is not possible for objects denoted as nodes in the OSM, it would be interesting to create such dataset from objects denoted as areas (i.e. closed ways).

Technical level — **A - excellent.**

In order to fulfill the assignment, the student had to study a lot of expert literature as the required knowledge was beyond the scope of the knowledge gained by the study at the university. The selection and use of data sources was also excellent.

Formal and language level, scope of thesis — **B - very good.**

The formal and language level of thesis is mostly good. It contains several grammar mistakes and only a very few formal ones. The use of formal notation was good and the whole thesis was explained clearly.

Selection of sources, citation correctness — **A - excellent.**

The thesis uses an overwhelming 209 sources. Many of them are repositories of programming language libraries and their "readme" files. However, I still rate the overall work with literature as excellent as the list of used literature still contains

sufficient amount of scientific papers and other expert literature. In addition, proper citation of the resources and tools developed by others that were used during the implementation shows good ethics of the author.

Additional commentary and evaluation

All of the goals of the thesis were fulfilled and the results were very good. Although, as stated above, not a lot of own research and innovation was, in my opinion, required. This is, however, not the student's fault but rather a shortcoming of the assignment. Nevertheless, the successful completion of the thesis required a lot of study of literature and a great deal of implementation.

My suggestion regarding the future work section of the thesis:

> The first point suggests further testing of different, newer CNN architectures. It is true that better CNN architectures and configurations are developed every year. It is to be expected that if a new CNN outperforms other CNNs on many databases, it will probably outperform the older CNNs on the satellite imagery as well. After all, this thesis proves it, by showing that the state of the art architecture outperformed other architectures. While for a specific classification task it is important to find the best classifier, the strong point of this thesis is the creation of a very good dataset. A well-made dataset can improve performance of all classifiers and their configurations (past or future). It is therefore much greater contribution to the scientific community and the industry than showing that some well-known classifier with certain parameters performs well on a dataset. I would therefore defer from further CNN architecture testing and focus solely on the improvement of the data gathering and annotation.

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

The theoretical part of the thesis contains and excellent explanation of neural networks and deep network architectures, including history of their development. In my opinion, the text could even be used as an introductory text for courses teaching about neural networks. The rest of the used concepts and methodology is also described very clearly.

The main contribution of the work from my point of view is the creation of automatically annotated dataset which is something missing from the area of the classification of the imagery of Earth. More research should be done on the methodology of the dataset creation but the thesis did a good job of paving the road for such research.

The results of classification of the images are less exciting, as recent years already showed that CNNs with sufficient amount of training data can provide great, even above-human classification accuracy.

Questions:

1. It is stated in the thesis that the existing satellite imagery datasets are not suitable for training of the CNNs, yet there are papers stated in the related work that used CNNs for classification of satellite imagery. Can you, please, comment on this?
2. According to the thesis, dataset with a constant scale of the images is better than a dataset with varying scale. Wouldn't it be better to have a dataset with varying image scale in order to gain classifier robust to varying scale of the objects? Also, what about types of objects (or instances of objects that have smaller size) that cannot be meaningfully annotated or detected at certain scales? (e.g. small parking lots in an image containing large city area)
3. Images were taken mostly from Europe. What do you think would happen to the classification performance of the trained network(s) if it is presented with images from tropical or desert areas or areas covered in snow (e.g. recognition of buildings with snow coverage)?

4. Judging by the graphs of per epoch validation classification accuracy in the transfer learning, the validation in many cases improved only slightly in the first few epochs. Afterwards, it only oscillated around certain value for the rest of the training. What is your interpretation of this result (since this was not mentioned in the thesis; only the decreasing performance was explained – correctly – by overfitting)? Also, why didn't you use some stopping criterion for the training that would take the non-increasing or decreasing performance on the validation dataset into account (e.g. some weighted average of testing + validation performance)? In my opinion, stopping the training before the validation performance dropped drastically, could have resulted in greater testing performance.

I evaluate handed thesis with classification grade **A - excellent.**

Date: **7.6.2017** Signature: