

## I. IDENTIFICATION DATA

<b>Thesis title:</b>	<b>Deep neural network for city mapping using Google Street View data</b>
<b>Author's name:</b>	<b>Varun Burde</b>
<b>Type of thesis :</b>	master
<b>Faculty/Institute:</b>	Faculty of Electrical Engineering (FEE)
<b>Department:</b>	Department of Cybernetics
<b>Thesis reviewer:</b>	Ing. Michal Reinštein, Ph.D.
<b>Reviewer's department:</b>	Department of Cybernetics

## II. EVALUATION OF INDIVIDUAL CRITERIA

<b>Assignment</b>	<b>challenging</b>
<i>How demanding was the assigned project?</i>	
The aim of the thesis was to design, implement and experimentally evaluate a deep neural network based solution for city mapping using Google Street View images. The proposed software solution should allow the user to request Google Street View imagery for any location, perform analysis and feature extraction using deep neural network(s) (DNN) and output vectorized description projected and visualized over an underlying map. This thesis aimed for the integration of Google APIs and cloud services with state-of-the-art DNN based technology to create an application allowing user to extract insights from Google Street View imagery. The topic was demanding since it expected gaining knowledge in different fields as well as to understand deep learning.	

<b>Fulfilment of assignment</b>	<b>fulfilled with major objections</b>
<i>How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.</i>	
The thesis assignment defined 5 milestones to be met, 3 were fulfilled as requested, 2 were fulfilled with major objections. First objection: the comparison of the achieved results with the state-of-the-art is not sufficient. Second objection: the user interface based on Google Colab is badly designed – current solution lacks any ability to configure the proposed pipeline since it consists only of hardcoded calls to predefined scripts. The interface also lacks any documentation or comments to guide the user – the user cannot use it without proper knowledge of the code repository. The visualisation of the results as demonstrated in the thesis were sufficient, however, they were not included into the user interface as requested. On the other hand, the user interface is not the essential part of the thesis. The proposed pipeline connecting the Google APIs and the DNN based model is the important part and it was demonstrated by the student successfully.	

<b>Activity and independence when creating final thesis</b>	<b>D - satisfactory.</b>
<i>Assess whether the student had a positive approach, whether the time limits were met, whether the conception was regularly consulted and whether the student was well prepared for the consultations. Assess the student's ability to work independently.</i>	
The student attended regular consultations that were happening weekly or bi-weekly mostly prepared. Detailed guidance was necessary otherwise the progress was very slow. The student was not meeting internal deadlines repeatedly and in the final he heavily underestimated the effort necessary to write the master thesis. This was the reason the thesis was not submitted by the original deadline and required one semester extension. Even during the extension, the majority of the work on the document was done at last minute, leaving several issues unresolved. The student adhered to rather simple solutions without any inclination to dive deeper into the problem and to seek innovative ways how to approach the topic.	

<b>Technical level</b>	<b>C - good.</b>
<i>Is the thesis technically sound? How well did the student employ expertise in his/her field of study? Does the student explain clearly what he/she has done?</i>	
The given task required a combination of various software related technical skills: working with APIs, handling deep neural network based models in TensorFlow, understanding evaluation of machine learning algorithms, understanding geospatial data, and integrating geospatial analytics with a computer vision solution. Starting without any apriori knowledge of the topic, the student was able to tackle all of these skills up to a good level that was sufficient to design and implement basic	

solution to the given problem. The topic was open for innovation in many different ways, however, only the most straightforward solution was explored and implemented. The proposed solution does not scale to larger data volumes. The implementation is not sufficiently robust and does not meet common standards of writing a clean code. On the other hand, the student invested a lot of effort to learn and improve in many different fields of research that were new to him and this is something I highly appreciate. The student studied large amount of scientific papers and related literature, however, the text of the thesis does not fully demonstrate this understanding.

## Formal level and language level, scope of thesis

**E - sufficient.**

*Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?*

Although the student proposed, implemented and evaluated a good solution to the problem, more effort should have been invested into the text of master thesis. The structure of the thesis is good, but the chapters 5 and 6 should be restructured due to overlapping content. The extent of the thesis meets the requirements; the thesis contains sufficient amount of figures that help the reader understand the proposed solution, however, the figure descriptions are lacking. However, the language of the thesis is the weakest part. Some parts of the thesis contain incomplete or confusing sentences that do not make sense. The narrative of the thesis is weak, majority of the explanations lack technical details and the thesis suffers from a numerous grammar issues. The thesis advisor strongly and repeatedly highlighted theses issues and although the thesis improved a lot during the iterations between different versions, the result is unfortunately only satisfactory.

## Selection of sources, citation correctness

**D - satisfactory.**

*Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?*

The master thesis contains a large number of references (81) to public sources, white papers and scientific literature. All the listed sources are well chosen and related to the work presented. Unfortunately, correctness of citations and correctness of paraphrasing was one of the issues that repeatedly came up during the process of creation of this document. Previous versions of the thesis repeatedly contained portions of text copied from other sources without a reference to the original. We discussed this serious issue during regular meetings and all the identified cases were corrected, however, I cannot claim that all cases were actually identified due to last minute submission. The student worked hard and invested a lot of effort into learning how to work with scientific literature and made a great progress in this regard.

## Additional commentary and evaluation (optional)

*Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.*

The final grade reflects the final state of the thesis as well as the quality of the technical solution with respect to the original assignment and given milestones. Although the grade is low, I appreciate very much the amount of effort the student invested into the thesis and all the knowledge he gained while he worked on the thesis.

## III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

*Summarize your opinion on the thesis and explain your final grading.*

Based on the above criteria and considering the above named issues I consider the thesis to be satisfactory.

The grade that I award for the thesis is **D - satisfactory**.

Date: **26.1.2020**

Signature:

20 January, 2020  
Québec

Ing. Vladimír Kubelka, Ph.D.

## Master's thesis review / Posudek oponenta

Author of the reviewed thesis: **Varun Burde**

Thesis title: **Deep neural network for city mapping using Google Street View data**

The aim of the master's thesis of Varun Burde was to design and test a new software pipeline that would exploit state-of-the-art image classifiers to analyze Google Street View images from a given area selected by the user. The detected objects found in the images would be later inserted into a digital map as additional points of interest. The motivation for this work is the automatic enhancement of existing maps by the means of machine learning algorithms.

I appreciate the effort put into re-implementing various components necessary for the whole pipeline, especially the state-of-the-art deep neural networks (NN) utilized for the image analysis. The thesis contains a brief overview of available NN architectures both for object detection and classification and for the depth reconstruction. The author selected the best-performing ones. These main building blocks were integrated with the Street View and other map components to fulfill the thesis goal. The user interface was implemented using the Google Colab environment.

However, regarding the technical part, I see several issues that need clarification (these are also my questions). The author reports the problem of obtaining detections of a single object from multiple views. That is actually expected, the difficulty is to match these together into a single point added to the map in the end. Clustering using the k-means algorithm was not satisfactory and it was not used in the final implementation. How was the problem resolved then? Are all the detections just added to the map including the duplicates? There is a possible resolution proposed in the future work which would utilize a tool from Google to be able to download images in sequence along a street that would allow searching for the related detections. Is that necessary? The downloaded images contain information about their location, why it is not possible just to sample the area as proposed in the thesis and then process the images in sequence according to their coordinates? Since the depth estimation is not always perfect, this would also allow better localization instead of using fixed thresholds.

My other comment is about the way the location (longitude and latitude) of each detection is computed from the pixel coordinates. The proposed formula (6.2) is very approximative, the left-right displacement w.r.t. the camera depends on the distance, not only the pixel

coordinates. How does the author justify this approximation? Is the allowed distance so constrained that the dependency does not matter in the formula (6.2)? There was a proposal in Sec. 6.14 to resolve this problem by finding an explicit mapping between the pixel coordinates and the world coordinates, but this was later rejected as unsatisfactory. How large the position errors actually are? I am missing a clear comparison or an experiment.

The presentation of the work done is, unfortunately, the weaker part of the thesis. The structure of the thesis is fine up to the not-so-logical split of the pipeline description between chapters 5 and 6, these two parts could have been just one chapter, perhaps with a dedicated section for discussion of the encountered problems. The text is however difficult to read, with whole sentences that do not make sense. For example, the abstract contains statements as: "A bar graph to visualize the number of detection per class." That is not a sentence. Sentences "*The author projected a general framework for classifying the practicality of individual buildings.*" or "*Network Architecture: Mask R-CNN have multiple architectures) Convolution backbone architecture used for feature extraction over an entire image) network head for bounding box recognition (classification and regression).*"(sic) also do not help the reader to understand the message. Several figures lack an explanation of the units used (Figs. 3.8, 3.9: what is the x-axis?, Fig. 6.4: what is the indicated depth, meters?). Abbreviations should be explained when used for the first time in the text, e.g. "CNN" and "GSV" in Sec. 1.1. What is "Overpass API"? It was used in the work but not explained properly. There are many more issues like these and they degrade the thesis. I do acknowledge that the thesis was written in English but I propose to the author to use more proofreading next time to avoid this unnecessary problem.

In conclusion, the goal of the thesis was generally satisfied except for the last point from the assignment guidelines (the comparison of the whole pipeline to the state-of-the-art approaches is missing). The pipeline accepts the input from the user, downloads and analyses the related Google Street View images, and finally puts obtained detections into a digital map. The presented solution would benefit from a better way of dealing with multiple detections of a single object and with a failing depth estimation. The formal part of the thesis is the weakest point. As the final mark for the thesis, I propose "**GOOD**" (C).

In Québec, 20 January 2020

Vladimír Kubelka