REVIEWER'S OPINION OF FINAL THESIS

I. IDENTIFICATION DATA

Thesis name:	**Representation learning for trademark search with limited supervision**
Author's name:	**Šuma Pavel**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Computer Science
Thesis reviewer:	Oleksandr Shekhovtsov
Reviewer's department:	Department of Computer Science

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **challenging**

Evaluation of thesis difficulty of assignment.

The assignment is defined as gradual steps, where the first steps are easy and the student is to familiarize himself with the data and simple baseline solutions of the problem. The latter steps 3,4 of the assignment offer a space for some creativity and exploration in setting up the semi-supervised learning problem and evaluation and exploring semi-supervised learning methods. Semi-supervised learning problem is generally a challenging topic. It is considered here for a specific application of finding similar trader logos and the challenge in my understanding is to explore these specifics to aid semi-supervised learning.

Satisfaction of assignment **fulfilled**

Assess that handed thesis meets assignment. Present points of assignment that fell short or were extended. Try to assess importance, impact or cause of each shortcoming.

I think the assignment is fulfilled and exceeded in few points. For example, a new dataset is collected to improve possibilities to evaluate and select the learning techniques, several methods explored and compared for each step of the assignment. One bit that I was missing (perhaps due to a misunderstanding from my side) is a demonstration of test time improvement when applying semi-supervised method with all available labeled and unlabelled data.

Method of conception **correct**

Assess that student has chosen correct approach or solution methods.

The approach taken follows the assignment guidelines, addressing the problem with methods of increased complexity, always comparing to the preceding simpler baselines. The evaluation methods are correct from the point of view of machine learning theory. The approach to formalize semi-supervised learning and the explored approaches to improve its performance are heuristic in nature. I therefore cannot rate them as conceptually outstanding. However the approaches are evaluated and experimentally analyzed in a sound way, which is positive.

Technical level **A - excellent.**

Assess level of thesis specialty, use of knowledge gained by study and by expert literature, use of sources and data gained by experience.

The student gained experience with modern machine learning tools: development environments, libraries, learning methods, data analysis and visualization tools, applied math skills to pose the problems and evaluate complicated metrics, deep metric learning and semi-supervised learning methods. The thesis appropriately uses the expert literature to build baseline models and as supportive evidence for the arguments regarding design choices.

One aspect left out of scope is the connection between cosine similarity (and its definition indeed) of normalized vectors and the Euclidean distance. Perhaps this considered obvious. Another related notion, "maximum inner product search", could be linked as well. It is somewhat unclear from the beginning, whether each trademark has one representative image of a logo or could have many possible drawings with some variations of their logo. In either case the set of all trademarks and their respective clusters seems to be finite and the concept of hard negatives appears to be well defined (unlike e.g. learning a metric for multi-view matching).

REVIEWER'S OPINION OF FINAL THESIS

Formal and language level, scope of thesis	**A - excellent.**
Assess correctness of usage of formal notation. Assess typographical and language arrangement of thesis.	
The thesis is well-organized, I think all important aspects are sufficiently discussed; the explanations are clear and sufficiently detailed. The mathematical notation is used where necessary, it is correct and all relevant concepts are properly defined. I have no complains. I believe that the writing, technical and typographical qualities are on a high level. There are really very few places whether I noticed that the language could be improved (perhaps fewer than in this review).	

Selection of sources, citation correctness	**B - very good.**
Present your opinion to student's activity when obtaining and using study materials for thesis creation. Characterize selection of sources. Assess that student used all relevant sources. Verify that all used elements are correctly distinguished from own results and thoughts. Assess that citation ethics has not been breached and that all bibliographic citations are complete and in accordance with citation convention and standards.	
As far as I can see, all used elements are correctly distinguished from own results and thoughts and used as building steps for the assignment. It is very positive that the work builds a series of baselines, to measure the improvements in each stage, and includes ablation studies of which components are more essential for the performance. I would still remark the following two directions where I think the connections to the prior work can be enhanced: 1) a discussion of semi-supervised learning approaches beyond pseudo-labeling. From my perspective, I know a number of works for Hamming distance retrieval. Two examples for image and text retrieval are: Kang et al. 2019 "Maximum-Margin Hamming Hashing" Nanculef et al. 2020 "Self-Supervised Bernoulli Autoencoders for Semi-Supervised Hashing". These methods are more advanced than pseudo-labeling in that they try to employ generative models to learn a better representation space of codes, which is advantageous in the unsupervised setting (unlabeled data) or semi-supervised setting. Although these are more complicated methods, and not guarantee better results in practice, they may be considered as more principled to address the semi-supervised learning problem. 2) comparison of the results achieved In the thesis with baselines from the literature. I am not familiar with the exact SOTA on the trademarks retrieval and perhaps I missed something. It would be desirable (if technically possible) to include published performance results from existing methods to the final comparison.	

Additional commentary and evaluation
Present your opinion to achieved primary goals of thesis, e.g. level of theoretical results, level and functionality of technical or software conception, publication performance, experimental dexterity etc.
Some additional comments on the work are as follows. I like how the wok is build gradually from simpler baseline methods to more complex ones and that the evaluation methodology is devoted a major place in the work and is discussed in detail. I was not completely convinced the evaluation methodology is appropriated for the specific task of the trademark similarity search. Suppose a new company wants to register an original trademark, which is distinct from the existing ones. Then the mAP score should be zero. It is not clear whether such cases were excluded from the metric at all. The metric also does not distinguish the cases when some query has very similar analogues and other query has somewhat similar analogues but nevertheless distinct enough to allow its legal registration. Along with metrics universally used for retrieval, also the training objectives universally used for retrieval are used. If the system is to be deployed to work with an existing dataset of trademarks, it may be reasonable to assume that at the test time, the dataset itself (not the test queries) stays the same, which I think is not utilized in the thesis. From this perspective I had impression that the specifics of the trademark retrieval problem are somewhat underexplored. Although, the features characteristic to trademarks are indeed explored. The work could also elaborate more on the degree how the new trademark can vary in style and drawing and how a sufficient dissimilarity is decided legally, or at least some official guidelines. The final evaluation shows that the semi-supervised method gives an improvement over the supervised method with few training labels (2-3 similar trademarks of each similarity cluster). At the same time it would be desirable that the semi-supervised method would also lead to an improvement (even if a small one) in the case when all of the annotated training data is used (which is still a very small subset of the entire image data collected). The thesis does not include such evaluations, which indicates that either it does not work towards improvement with a larger supervision set or the author

for some reason is not interested in building the best performing trademark retrieval system.

Nevertheless, I would like to remark that chapter 5 (semi-supervised learning) investigates quite complex combinations of self-labeling strategies, proposes intermediate evaluation and ablation studies to investigate their effect and I found it to be done to some depth and with high standards.

The problem posed is challenging from a practical perspective and does not admit a concise mathematical formulation. The heretic methods investigated experimentally are therefore rather appropriate here. Respectively, the theoretical advancement of the problem or obtaining theoretical results was not amongst posed goals.

III. OVERALL EVALUATION, QUESTIONS FOR DEFENSE, CLASSIFICATION SUGGESTION

Summarize thesis aspects that swayed your final evaluation. Please present apt questions which student should answer during defense.

Summary:

I think this work included many technical challenges in working with software, organizing and collecting datasets, running multiple evaluations with different performance metrics and ablation studies. I like that the work is build gradually, considering several alternatives of similar complexity at each stage and incrementally building stronger baselines. The final semi-supervised methods I think are not very innovative, rather a selection amongst several heuristics is made along with defining adaptive thresholds and selection of good feature combinations and post-processing steps. However I am sure the whole work is done by the student individually, and I find it to be done to a high quality standards. This is in my understanding fully appropriate for a master thesis.

Questions:

1. When introducing adaptive thresholds for self-labeling methods in chapter 5, they depend on the distances within currently discovered clusters, such as average distance between all positive examples in the cluster. However they do not seem to depend on the distance to some known negative examples. If the method is run for many iterations, without stopping by the validation set performance, would such adaptive thresholds lead to clusters of "similar" trademarks ever grow, absorbing all examples or to stop at some point?

2. If at the test time, we query with a new logo, that has no similar analogues in the dataset, what is the expected output of the method and what is the contribution of this query to the performance metric?

3. I understand that the shortest path search is not a computation bottleneck in the method, nevertheless, the following question. When searching for the shortest path in the graph from a vertex to a subset of vertices (a given cluster of examples) with the allowed path length / number of edges in the path bounded by a small number, would it not be more efficient to use breadth-first Dijkstra search instead of depth-first A* search (that has to be repeated for all possible targets)? Is the Euclidean distance-based heuristic in A* efficient for high-dimensional spaces (2560)?

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

Date: **20.1.2021** Signature: