EVALUATION OF THE DOCTORAL THESIS

Author: **ING. MARTIN KOLÁŘ, PH.D.**

Title: **COMPUTER VISION WITH ACTIVE LEARNING**

Study direction: Computer Technology and Informatics

Supervisor: prof. Dr. Ing. Pavel Zemčík

Affiliation: Brno University of Technology, Faculty of Information Technology, Department of Computer Graphics and Multimedia

The doctoral thesis of Ing. Martin Kolář, Ph.D. is focussed on the field of active learning in improving the accuracy of models in the machine vision field mainly via the optimizing of expert labelling. The benefit of using the active learning was experimentally validated in various scenarios. It is a promising research field, which is currently being studied very intensively. **The topic of the PhD. thesis clearly falls into the study direction – computer technology and informatics.** It can be also said that the subject of the thesis – Active Learning and also Transform Learning are **interesting and relevant to current needs of the scientific community in the given field.**

The author has formulated the following hypothesis of his doctoral thesis:

- Human and algorithmic expert annotation using Active Learning improves the accuracy of contemporary Computer Vision methods.

This hypothesis was validated in several experiments:

- Active Learning (AL) has been shown to increase the quality of Generative Adversarial Networks (GAN) for Font Capture.
- AL has enhanced the applicability of GANs for generating faces.
- AL has reduced time to manually annotate tags on images.
- In the case of weak supervision, AL improve the classification accuracy of images classifier.

The PhD. work contains original parts and I see the original contributions of the work in following.

1. On the scientific level, the main contribution is validation of the hypothesis above:
    - via increasing the efficiency of manual labour for annotation,
    - via using Transfer Learning by applying pre-trained neural network models.
2. On the application level:
    - a series of experimental demonstrations of the hypothesis (above),
    - specific model improvements in Font Capture,
    - One-shot-learning for image classification,
    - the tagging GUI for more comfortable annotation process.

Publication activity of the author is above average. He has 7 publications, of which 4 are publications in journals (1 in current content database, 2 Springer and 1 Elsiever with IF=2.991), 2 publications in conference proceedings and 1 technical report. On the other hand,

2 publication are still under review. I think, **the core of the PhD. thesis was published at the required level.** The doctoral thesis satisfies conditions of a creative scientific work and **the list of scientific activities confirms the author's scientific erudition**.

The doctoral thesis is well written, the style of the work is erudite, well organised. It proves author's broad overview of the domain (122 references). The work is intuitively well understandable, but sometimes too brief. Related works are described very shortly. The work could have a larger scope and some issues could be better explained. For example, the exact understanding of formula notations on page 6 and 33 is effectively prevented by missing definitions of some variables, indexes, or symbols.

I have several notes, some of them deserve discussion:

1. What is "Superhumen accuracy" (page 18)?
2. How can be SVM used for finding similar images (page 26)?
3. What is the range of values of the score function for human and random annotation? What is the purpose of the random annotation (page 26)?
4. Within Figure 3.3, it would be useful to specify which annotations belongs to which retrieved images. Some annotations start with a lowercase letter, so they seem to be a continuation of the previous ones.
5. The images in Figure 3.4 are fake or real in what sense?
6. According to which formulas the success rate and the adjust success rate in Table 5.1 were computing?
7. What is the interpretation of the concept "believe of label" (b or B in pages 50 and 51)?
8. For the accuracy computing (Figure 5.14), the numbers of TP, FP, TN, and FN are essential. What is relation of these numbers and "correct images" in Figure 5.13. The results in Figure 5.14 are referenced by Table 5.14 in the first sentence on page 54. There is not Table 5.14 in the text of the work.
9. Figure 3.2 refers to the reference [LGRN09] but the same figure is mentioned in relation to [YCN+15] in the text (page 23). What is the relationship between the two references?
10. The unigrams "track" and "tracks" in Table3.3 represent one word. A reduction to the stem form of words would be beneficial.
11. The reference [JVK+20] is not in bibliography.
12. To provide a uniform form of references, [CZ] should be replaced by [CZ05]
13. By substituting equations (4.2) into each other, we would imply that M is an argument of M.

These comments do not reduce the very good level of the work. Finally, I can state, that the author of the thesis Ing. Martin Kolář, Ph.D. proved to have an ability to perform research and to achieve scientific results. **In my opinion, the thesis meets the generally accepted requirements for obtaining the academic degree of Ph.D.**

Košice, March 1, 2021

assoc. prof. Ing. Kristína Machová, Ph.D.
Department of Cybernetics and Artificial Intelligence
Technical University of Košice
Letná 9, 042 00 Košice, Slovakia