# Supervisor's review of master thesis

| | |
|---|---|
| Author of the review: | RNDr. Pavel Pecina Ph.D. |
| Author of the thesis: | Nguyen Tien Dat |
| Title of the thesis: | Toward concept visualization through image generation |

The work presented in the thesis of Nguyen Tien Dat belongs to the area of deep learning applied to textual and visual data. The author exploits well-known methods to represent words and images in vector spaces (embeddings) and a method to map between these spaces. The mapping is trained on a set of images provided with textual descriptions and allows to "translate" between word embeddings and image representations. Such a system allows, for example, to build an image annotation pipeline which for a given image, generates its textual description (labels). In that setting, the system generates an image representation which is then mapped to the space of word embeddings and the closest neighbours are used to generate the output labels. The presented work, however, attempts to build the inverse pipeline where, for a textual input, its image counterpart is produced. For a given word, the pipeline generates its word embedding which is mapped to image representation and this vector is projected back to a pixel representation using an inversion algorithm. The entire idea is very challenging and has become of interest only recently.

The thesis is structured into five sections plus a lists of figures, tables and an attachment providing more detailed results of some of the experiments. The main text spans 50 pages. After the introduction in Section 1, the author presents related work in Section 2, and details of the components of the pipeline in Section 3. Section 4 then describes the experiments conducted within this work. The first part focus on parameter optimization, the second part then presents results of five evaluation tests. Section 5 then concludes the work.

The text is well structured, written in a good English. The introductory and theoretical parts of the thesis are all right and show that the author studied a large portion of related work (the references contains almost 70 papers). The experimental part, however, could have been improved, especially due to the fact that the goal of the thesis was not accomplished and the thesis mostly presents negative results (the author did not deliver a pipeline which can produce meaningful results). There are two main issues. First, the description of the experiments is often very brief and shallow and the results are not analysed thoroughly. Second, the design of the experiments is questionable. For example, in Section 4.1 and 4.2, the author describes experiments to select the optimal methods for generating image representations for mapping between word and image embeddings. This is done by presenting several image examples generated by different methods for a given concept (label) to a group of people who voted which picture represented the given concept best. Although the results were tested for statistical significance, I am not convinced that there is any significant difference among the options at all. The author did not show that using alternative options leads to different (worse) results in the subsequent steps. The remaining part of Section 4 presents the main experiments, which basically say that people are not able to correctly distinguish between results of the presented pipeline and random pictures. The author claims that in some cases, the method helps to distinguish between e.g. man-made and organic objects,

but this is mostly decided based on colour and texture only. Such features can be extracted from images using much simpler descriptors (e.g. SIFT, MPEG-7). These experiments miss details and often it is not clear how exactly they were conducted (what was presented to the participants and how; what they were instructed to do; how the results were analysed, etc.)

From the implementation point of view, most components of the pipeline were adopted and the main work was to integrate everything together and conduct the experiments. No code and data (experimental results) were attached to the text of the thesis.

Despite all these issues, the author showed a good knowledge of the field and performed a significant amount of work in the experimental part and I recommend the thesis to be defended.

Prague 26.1.2016