Thesis advisor review of a bachelor thesis

Author: Ing. arch. Alena Moravová
Title: Text recognition in images using recurrent neural networks
Thesis advisor: Mgr. Jan Šochman, PhD
Department: Department of Cybernetics, FEE, CTU in Prague

Traditionally, the text recognition problem is solved using convolutional neural networks (CNN) only. However, the text is naturally a sequence of characters and sequences are usually well processed by recurrent neural networks (RNN). The thesis takes as its starting point one very recent approach, CRNN [1], which combines the power of both CNN and RNN architectures. This combination makes the approach more intuitive and appealing for possible modifications.

The goal of the thesis was to analyse the performance of the algorithm in various situations and on the most recent benchmarks, as only limited evaluation is presented in the original paper. The thesis was also supposed to build on the original algorithm and propose and test some useful extensions and modifications.

The thesis presents several experiments on the latest ICDAR'15 dataset as well as on training for case sensitive text recognition. These results are valuable for comparison with other state-of-the-art methods used in the literature.

The thesis proposes two modifications of the original algorithm: (i) padding the images instead of scaling them to a fixed size, and (ii) imposing blanks in the final layer text decoding. As expected, the first one is shown to bring significant improvement on short words compared to the original scaling approach. The second one, does not seem to bring any significant improvements.

Finally, the thesis presents an analysis of the training and test data which reveals several flaws in the publicly available datasets and training protocols.

The achieved results are interesting and sufficient for a bachelor thesis. Judging from my experience working with Alena Moravová, I also believe we could go even further with the experiments haven't there be technical problems with running the algorithm on the Metacentrum cluster which lasted nearly a month. In spite of that, I would very much like to express my satisfaction with Alena's work on the thesis. She is very quick in understanding even complex theoretical topics and also in implementing the proposed solutions. She works independently and comes often with her own ideas for experiments. She very quickly grasped the idea of a scientific research and prepared the experiments in a systematic way. I hope that our cooperation will continue in the future.

Regarding the text of the thesis itself, some of the formulations are still too vague for me. But overall I am happy with the structure of the thesis, presentation style and the language used. The feedback given was again quickly understood and integrated into the text.

I recommend the thesis for defense and suggest the mark A - excellent.

In Prague, 9 June 2014, Jan Šochman

References

[1] Baoguang Shi, Xiang Bai, and Cong Yao. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *CoRR*, abs/1507.05717, 2015.