Posudek diplomové práce
Matematicko-fyzikální fakulta Univerzity Karlovy

Autor práce       Bc. Jakub Hajič
Název práce       Visual Question Answering
Rok odevzdání     2017
Studijní program  Informatika
Studijní obor     Umělá inteligence

Autor posudku    RNDr. Jakub Lokoč, Ph.D.
Pracoviště       KSI MFF UK
Role              Oponent

Text posudku:
The author of the thesis focuses on so-called Visual Question Answering, which is a recently introduced multimodal machine learning task. The input to the task is an image with an associated natural language question. The output is the answer to the question. Few years ago, such task was considered to be very complex and difficult for machines. However, with the reign of deep learning, significant breakthroughs have been achieved. The thesis builds on the very recent state-of-the-art in the VQA area, combining in a single end-to-end model deep convolutional neural networks, long short-term memory units, three types of attention mechanisms and multimodal compact bilinear pooling. The thesis describes several extensions and presents results of several proposed trained architectures. The modifications exhibit similar performance as the original state-of-the-art model and the results are discussed. The thesis is written in English on a good level (except just few typos in the second part of the thesis), the text is clear and well structured in eight sections. Just a minor remark - Section 4 could contain more details about multimodal compact bilinear model. For example, how are the gradients propagated through this multimodal combination unit? What hash functions were used? The trained model shows very interesting results, corresponding to the accuracy presented in the experimental section. Given an image of a white bus on a road, the model answered correctly questions like "What is on the picture?", "What is the color of the bus?". The same good performance was for an image of a bench. For yes/no questions and the second image, the model answered yes for questions "Is there a bench?" or "Is there a cat?", while the answer was no for question "Is there an airplane?". Overall, the thesis investigates an important topic and presents interesting findings for the machine learning community. The performance of the implemented model provides promising results for future investigation.

Práci doporučuji k obhajobě.

Práci nenavrhuji na zvláštní ocenění.

Pokud práci navrhujete na zvláštní ocenění (cena děkana apod.), prosím uveďte zde stručné zdůvodnění (vzniklé publikace, významnost tématu, inovativnost práce apod.).

Datum       17. August 2017
Podpis