Supervisor's review of doctoral thesis

Charles University, Faculty of Mathematics and Physics
Study programme: Computer Science
Study branch: Mathematical Linguistics

Student: Mgr. Jindřich Libovický
Title: Multimodality in Machine Translation

Supervisor: doc. RNDr. Pavel Pecina, Ph.D.
Department: Institute of Formal and applied Linguistics MFF UK

The thesis of Jindřich Libovický belongs to the broad research area combining automatic processing of language and vision – two modalities that naturally interact in various tasks. The presented work focuses on mutlimodal machine translation and specifically the task of translation of image captions, where the source data is presented in the form of an image plus its textual description, then the goal is to generate a translation of the caption into a target language given the visual information in the picture. The thesis also tackles a more general problem – multi-source sequence-to-sequence learning which can be applied not only to multimodal machine translation but other tasks too.

This field is relatively new and undergoing a rapid developmend due to the recent advances in research of deep artificial neural networks. Before the deep learning era, the state of the art in image representation was based on global descriptors which were not really able to represent "semantic content" of an image. The current methods are able to generate representations which are much closer to "semantics than before and they also allow generating joint representations based on multiple inputs of potentially different modalities. Deep learning also have boosted research in the traditional text-to-text machine translation and its quality is getting close to the quality of human-produced translations. This situation creates a very strong baseline for the research of multimodal translation which aims to improve this *text-to-text* baseline by exploiting visual information from the picture in *text+image-to-text* setting.

The thesis of Jndřich Libovický is very well structured. It includes 7 chapters and several appendices including a list of Jindřich's own publications related to the thesis. Chapter 1 provides an introduction to the thesis and motivation for the work. In Chapter 2, Jindřich provides a very well presented, comprehensive and coherent overview of deep learning methods relevant in the area of language and vision. In Chapter 3, Jindřich discuses how these two modalities can be combined and where this combination can be exploited. Chater 4 introduces the task of multimodal machine translation, the datasets the authors worked with including the Multi30k dataset which is a practical outcome of the thesis. Chapters 5 and 6 contains the original contribution of the author, such as the innovation of deep learning architectures for sequence-to-sequence learning combining multiple sources and generating singel output sequence, series of experiments, their analysis and discussion of the results. Chapter 7 concludes the thesis and summarize the main contributions of the work. The text is very well written with very rare errors in English. The text is readable, experiments are well described, results well analyzed and discussed.

I'd like to emphasize that Jindřich started his research before the beginning of the (above mentioned) deep-learning era which caused a paradigm shift in many areas including natural language processing and computer vision. Jindřich had very quickly adopted this research direction and became able to conduct experiments and design new methods. His research results are evident from his bibliography. Jindřich is an author of about 20 peer-reviewed papers published at major international venues including ACL where his 2017 paper was selected as an Outstanding Paper. Many of Jindřich's publications are joint works with researchers from other research labs (national as well as international). So far Jindřich's publications have collected more than 100 citations (in peer-reviewed papers, excluding self-citations). These figures are quite exceptional for a PhD student not only at our department but world-wide and confirm the strong impact of Jindřich's work to the research community.

I fully recommend the doctoral thesis of Jindřich Libovický to be defended.

doc. RNDr. Pavel Pecina, Ph.D.   30.5.2019, Prague