Gated Hierarchical Attention for Image Captioning

SPEAKER **Mr WANG Qingzhong**

PhD Student
Department of Computer Science
City University of Hong Kong
Hong Kong

DATE 29 November 2018 (Thursday)
TIME 10:00 am - 11:00 am
VENUE B4701, 4th Floor, Blue Zone
Yeung Kin Man Academic Building
City University of Hong Kong
83 Tat Chee Avenue
Kowloon Tong

ABSTRACT

Attention modules connecting encoder and decoders have been widely applied in the field of object recognition, image captioning, visual question answering and neural machine translation, and significantly improves the performance. In this paper, we propose a bottom-up gated hierarchical attention (GHA) mechanism for image captioning. Our proposed model employs a CNN as the decoder which is able to learn different concepts at different layers, and apparently, different concepts correspond to different areas of an image. Therefore, we develop the GHA in which low-level concepts are merged into high-level concepts and simultaneously low-level attended features pass to the top to make predictions. Our GHA significantly improves the performance of the model that only applies one level attention, e.g. the CIDEr score increases from 0.923 to 0.999, which is comparable to the state-of-the-art models that employ attributes boosting and reinforcement learning (RL). We also conduct extensive experiments to analyze the CNN decoder and our proposed GHA, and we find that deeper decoders cannot obtain better performance, and when the convolutional decoder becomes deeper the model is likely to collapse during training.

This paper will be presented in 14th Asian Conference on Computer Vision (ACCV), Dec. 2-6, 2018, Perth, Australia.

Supervisor: Dr CHAN Antoni Bert

Research Interests: Computer vision, Deep learning and Natural language generation

All are welcome!