Summary about VQA project

Visual Question Answering is a task that it requires the model to give an answer to a text question based on a given graph. It should be able to infer specific information from the elements in the graph as well as some basic knowledge. It is a cross domain of computer vision and natural language processing. Generally, VQA models use computer vision model, like convolutional neural network, to detect elements in the graph, and use natural language processing model, like long short-term memory, to find what the question needs before combining these information supply and demand and giving the answer.

Visual Question Answering is widely used in many fields, such as graphic reading, blind visual care, AI custom service, cross-modal search and so on. For example, some online shops use AI custom service to answer the questions from consumers. The AI can read some related information from the poster of shop so that they can answer many performance questions of consumers, which reduces labor costs.

Here are some related papers which I have read, or I plan to read about VQA.

[1] Gao H, Mao J, Zhou J, et al. Are you talking to a machine? dataset and methods for multilingual image question[C]//Advances in neural information processing systems. 2015: 2296-2304.

[2] Wu Qi, Teney Damien, Wang Peng, Shen Chunhua, Dick Anthony, and van den Hengel Anton. 2017. Visual question answering: A survey of methods and datasets. Comput. Vis. Image Underst. 163 (2017), 21–40.

[3] Gupta Akshay Kumar. 2017. Survey of visual question answering: Datasets and techniques. arXiv preprint arXiv:1705.03865 (2017).

[4] A. Das, H. Agrawal, C.L. Zitnick, D. Parikh, D. Batra, Human attention in visual question answering: Do humans and deep networks look at the same regions? Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP) (2016), pp. 932-937

[5] Zhou Yu, Jun Yu, Yuhao Cui, Dacheng Tao, Qi Tian. Deep Modular Co-Attention Networks for Visual Question Answering. In CVPR, 2019.

[6] W. Zheng, L. Yin, X. Chen, Z. Ma, S. Liu, B. Yang Knowledge base graph embedding module design for visual question answering model Pattern Recognit., 120 (2021), p. 108153