

EC601Project1

Xin Wang

September 2022

1 Problem Statement

The topic is Visual Question Answering (VQA). VQA means open-ended questioning about images, and receiving answers. As "VQA: Visual Question Answering" published on International Conference on Computer Vision (ICCV) in 2015 introduced that VQA takes an image and a free-formed, open-ended, nature-language question about the image and produce a nature-language answer as the output. For example, "What color is the shirt?", or "What is the man doing?".

It is important because VQA is the next generation of AI algorithm, which is a multi-discipline task and has a well-defined quantitative evaluation metric to track progress. The goal of the research is to help visual-impaired people or intelligence analysts actively elicit visual information.[1]

2 List of resource

[1]Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, Devi Parikh (2015). VQA: Visual Question Answering. In International Conference on Computer Vision (ICCV).

[2]Yash Goyal, Tejas Khot, Douglas Summers-Stay, Dhruv Batra, Devi Parikh (2017). Making the V in VQA Matter: Elevating the Role of Image Understanding in Visual Question Answering. In Conference on Computer Vision and Pattern Recognition (CVPR).

[3]Peng Zhang, Yash Goyal, Douglas Summers-Stay, Dhruv Batra, Devi Parikh (2016). Yin and Yang: Balancing and Answering Binary Visual Questions. In Conference on Computer Vision and Pattern Recognition (CVPR).

[4]Shih, K., Singh, S., Hoiem, D. (2016). Where to Look: Focus Regions for Visual Question Answering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[5]Qi Wu, Damien Teney, Peng Wang, Chunhua Shen, Anthony Dick, Anton van den Hengel (2017). Visual question answering: A survey of methods and datasets. Computer Vision and Image Understanding, 163, 21-40.

3 Citation

[1] Stanislaw Antol, Aishwarya Agrawal, Jiasen Lu, Margaret Mitchell, Dhruv Batra, C. Lawrence Zitnick, Devi Parikh (2015). VQA: Visual Question Answering. In International Conference on Computer Vision (ICCV).