Problem Statement

Definition

In our daily life, people can describe what they have seen from the picture, more than just recognizing the object. Visual Question Answering is a research area that we can build a computer system to answer questions presented in an image through natural language. Like the computer system can not only says that it is a car but also says that the car is moving.

Why is it useful

VQA can combine the advantages of both computer vision and natural language processing to help computers achieve human-like judgement functions of vision and the ability to express natural language output. It is a dynamic solution to problems in the medical and disability support fields.

Application

New scenarios are opening up to support clinical decision making by medical staff and to enhance diagnosis through computer-based "second opinions" and help visually impaired people to live without visual impairment. It can help blind users to communicate through pictures, attract customer of online shopping sites by giving "semantically" satisfying results for their search queries, allow learners engaged in educational services to interact with images, help the analysts in surveillance data analysis to summarize the available visual data.

Open source links

https://towardsdatascience.com/deep-learning-and-visual-question-answering-c8c8093941bc

https://www.sciencedirect.com/science/article/abs/pii/S0167865521003147

https://openaccess.thecvf.com/content_iccv_2015/papers/Antol_VQA_Visual_Question_ICCV_2015_paper.pdf

https://nlp.stanford.edu/seminar/details/aagrawal.pdf

https://openaccess.thecvf.com/content_ICCV_2017/papers/Hu_Learning_to_Reason_ICCV_2017_paper.pdf

https://www.sciencedirect.com/science/article/pii/S1077314217301170?casa_token=0GfiCESTTKIAAAAA:tAzCpivq1iSZELb-kbbbF6q0q38WmCyV2PFRr-AVv3kkXP29HRMMADFm2r_vSKOHg6bEsC7qEQ