Visual Question Answering (VQA) is a cross-field task that combines both Computer Vision (CV) and Question Answering (QA) fields. In a VOA system, the input should be pictures and questions and the output should be the answers related to the questions based on the pictures. An ideal VQA system can give feedback correctly and accurately and the user does not need to consider how to ask to make sure that the system can understand. This requires the system to understand the pictures and questions completely, which is harder than most of the other CV tasks.

CV covers many sub-problems, such as Object recognition (What is …), Object detection (Are there …), Attribute classification (What color …), Sence classification, Counting, and so on. There are also some more complex sub-problems. For example, what is between object A and object B (Spatial relationship), why is he crying (Common sense reasoning questions) [3]. Most of the approaches are based on convolutional neural networks (CNN), and the development of the hardware also plays important role in the development of CV [2].

QA is also a complex task that is related to natural language. It is an old research area in computer science. And thanks to forerunners of computer science, there are relatively mature question answering databases available now. But it is no doubt that the database still cannot cover all situations, so it is still a challenge to all researchers [1].

The development of VQA is important to that of Artificial Intelligence (AI). Maybe we can recognize it as a subfield of common Artificial Intelligence. Apart from the importance in the technique area, the development of VQA is also important to all of us because the applications of VQA can be a benefit to everyone.

VQA can be widely used in many areas. VQA can improve human-computer interaction. The user can get familiar with their device faster and use it easier. For example, VQA can help people with poor eyesight to make their lives more convenient [3]. Besides, the existing AI can be improved with an embedded VQA system. For example, with an embedded VQA system, AlphaGo may be a good Go teacher. So, in short, the research to VQA can contribute to not only individuals, but also companies, communities, and nations. It can provide individuals convenience, give companies income and reputation, improve techniques in communities and strengthen the nations.

Reference

[1] Diefenbach, D., Lopez, V., Singh, K. *et al.* Core techniques of question answering systems over knowledge bases: a survey. *Knowl Inf Syst* **55,**529–569 (2018). <https://doi.org/10.1007/s10115-017-1100-y>

[2] Feng, X., Jiang, Y., Yang, X., Du, M., & Li, X. (2019). Computer vision algorithms and hardware implementations: A survey. *Integration*, *69*, 309–320. https://doi.org/10.1016/j.vlsi.2019.07.005

[3] PatrickZH. (n.d.). *Algorithm\_Interview\_Notes-Chinese/D-视觉问答-1\_综述.MD at master · patrickzh/algorithm\_interview\_notes-chinese*. GitHub. Retrieved September 18, 2021, from https://github.com/PatrickZH/Algorithm\_Interview\_Notes-Chinese/blob/master/B-%E8%87%AA%E7%84%B6%E8%AF%AD%E8%A8%80%E5%A4%84%E7%90%86/D-%E8%A7%86%E8%A7%89%E9%97%AE%E7%AD%94-1\_%E7%BB%BC%E8%BF%B0.md.