Summary

Renyu Jiang, BU ID: U18427594

**Problem Statement:**

Visual Question Answering (VQA) is a recently emerging field of machine learning, and it is a mixed fields requiring visual recognition and natural language processing. Visual question answering usually takes image and question in natural language as input and gives out answer in natural language according to human’s question. However, it is a very challenging task due to its openness, both in question and answer.

**Why is it useful? What are its applications?**

With the rising labor costs, people are relying on AI technologies more and more, visual question answering is such a technology that can help people recognize in a cost-efficient way. In some occasions, to identify an object, people need to search database for images and compare by themselves, vqa technology can accomplish these tasks faster and more accurate.

A typical application of this technology is to help blind people recognize things nearby, what the user needs to do is asking and the machine would give answer, when compared with human guide, this technology is much cheaper. Aside from helping blind people, helping scientists analyze things in a fast speed is also an application, for example, analyze the properties of organic matter based on molecular structure diagrams.

**List of papers or open-source links:**

1. Antol, Stanislaw, et al. "Vqa: Visual question answering." Proceedings of the IEEE international conference on computer vision. 2015.
2. <https://blog.paperspace.com/introduction-to-visual-question-answering/>
3. <https://tryolabs.com/blog/2018/03/01/introduction-to-visual-question-answering>
4. Teney, Damien, et al. "Tips and tricks for visual question answering: Learnings from the 2017 challenge." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
5. Kafle, Kushal, and Christopher Kanan. "An analysis of visual question answering algorithms." *Proceedings of the IEEE international conference on computer vision*. 2017.