

Z.H.C.E.T., Aligarh Muslim University

B. Tech Project Part - I

(COC 4980)

Visual QnA System

Week Ending 20th Sep

Under the Supervision of **Prof. Mohammad Sarosh Umar**

Ravi Sahni
17 COB 085
GJ 7718

Tanuj Singhal
17 COB 033
GI 1585

Objective:

We intend to implement the task of free-form and open-ended Visual Question Answering.

Abstract:

Given an image and a natural language question about the image, the task is to provide an accurate natural language answer. Starting from basic images-less features like small shapes and moving on to real life images with huge feature set.

Progress:

This week we continued our study on NLP in VQA through resources available via the Internet while also going through the research papers we sorted last week. We also spent some time trying to find ways to reduce the size of the dataset & we are still working on the issue.

Currently, we've been discussing about meeting up sometime next week to move ahead with the implementation sides of the thing, to discuss various tools and platforms we can use to collaborate remotely.

The references we are going through right now are –

- [1] Agrawal, A., Lu, J., Antol, S., Mitchell, M., Zitnick, C., Parikh, D. and Batra, D., 2016. VQA: Visual Question Answering. International Journal of Computer Vision, [online] 123(1), pp.4-31. Available: <<https://arxiv.org/pdf/1505.00468.pdf>>.

[2] Goyal, Y., Khot, T., Agrawal, A., Summers-Stay, D., Batra, D. and Parikh, D., 2018. Making the V in VQA Matter: Elevating the Role of Image Understanding in Visual Question Answering. International Journal of Computer Vision, [online] 127(4), pp.398414. Available: <<https://arxiv.org/pdf/1612.00837.pdf>>.
