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

B. Tech Project Part - I

(COC 4980)

Visual QnA System

Week Ending 27th 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:

Up till now we had found ourselves stuck in handling the large dataset of VQA and not wanting to sacrifice by training a weak model in future due to using other poor-quality datasets, now we have finally found a way to go around this problem.

We have decided to take the dataset from Georgia Tech website visualQa.org.

- Description of Dataset –

VQA Input Images

COCO

Training images: 82,783 images

Validation images: 40,504 images

Testing images: 81,434 images

VQA Input Questions

Training questions 2017 v2.0: 443,757 questions

Validation questions 2017 v2.0: 214,354 questions

Testing questions 2017 v2.0: 447,793 questions

VQA Annotations

Training annotations 2017 v2.0: 4,437,570 answers

Validation annotations 2017 v2.0: 2,143,540 answers

The main problem here is the sheer size of the dataset (around 40+ GB) which is mostly due to the images. So, we have decided that rather than taking images in raw form as input, we will rather take a processed form of these images i.e. only features extracted from these images by a VGG model. We found these features after an extensive search online, the size of these features is manageable but we will have to only work on cloud for this project as much the size and the processing needed for the code base is out of scope for our local systems and we will try to manage this using Google Colab.

Now we are working on automating the pre-processing of this dataset and removing any corrupt tuples along with going through natural language processing tutorials for the same.

****