

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

B. Tech Project Part - I

(COC 4980)

Visual QnA System

Synopsis

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.

Introduction:

Vision and language problems such as image captioning and visual question answering (VQA) have gained popularity in recent years as computer vision research is progressing rapidly.

In a general way we can define a VQA system as *an algorithm that takes as input an image and a natural language question about the image and generates a natural language answer as the output*. This is by nature a multi-discipline research problem. It combines Computer Vision (CV), Natural Language Processing (NLP), and Knowledge Representation & Reasoning (KR).

Applications:

In the list of applications of VQA, probably the most direct application would be to help blind and visually-impaired users. A VQA system could provide information about an image on the Web or any social media. Another application is to integrate VQA into image retrieval systems. This could have a huge impact on social media or e-commerce. VQA can also be used for educational or recreational purposes.

VQA is directly applicable to a variety of applications of high societal impact that involve humans eliciting situationally-relevant information from visual data; where humans and machines must collaborate to extract information from pictures.
