CS4089 PROJECT ABSTRACT

Title

Answering questions based on image using deep learning.

Group Members

K A SIVA VARDHAN REDDY	B160333CS
P SATYANARAYANA	B160340CS
B MAHESWARARAO	B160349CS
R VAMSI KRISHNA	B160109CS

Guided By

Lijiya.A

Assistant Professor.

Abstract

This project deals with the problem of Visual Question Answering (VQA). We develop neural network-based models to answer open-ended questions that are grounded in images. Our model makes use of two popular neural network architecture: Convolutional Neural Nets (CNN) and Long Short Term Memory Networks (LSTM).

In order to do this, our model would need to understand several things - let's break them down into sub-tasks:

1. Identifying the various objects in the image (the train, traffic signals, tracks, pavement, person, etc).
2. Processing the text of the question itself, which can be processed as a 'sequence' of words.

3. Mapping the appropriate sections of the image to the given input question.
4. Generating natural language text in the form of an answer with an acceptable certainty.

Dataset

1. Visual Question Answering (VQA) dataset:

   The computer system needs to address issues, such as, a binary classification problem (Is the umbrella upside down?), a counting problem (How many children are on the bed?) or an open-ended question (Who is wearing glasses? Where is the child sitting?).
2. FigureQA Dataset:

   The computer system needs to answer questions presented by bar charts, pie charts, or line plots.(Is Aqua the maximum? Is midnight blue greater than aqua?)
3. DAQUAR - Dataset for Question Answering on Real-world images:

   This was one of the earliest datasets on image question and answering.