Subject: Bigdata Analytics and Applications
Document Scope: Project Proposal
Team #: 5

CS5542 - BIGDATA ANALYTICS AND APPLICATIONS
PROJECT PROPOSAL

1. Group Information:

Team #: 5

Team members:

1. Koushik Katakam – 10

2. Saitejaswi Koppuravuri – 13

3.Venkata Lakshmi Korrapati – 14

4. Pavankumar Manchala – 16

5. Zakari Abdulmuhaymin Ahmad H - 29

2. Project Goals and Objective:

2.1 Motivation:

Have you ever thought of generating a caption for an image? Yes, these days caption generation for an image has become an important task in the area of research of machine learning and Artificial Intelligence. No only captioning is a primary goal but predicting the objects in an image and express their relation in a process of natural language processing. The process of image captioning has been made little advanced on the advancement of neural networks.

2.2 Significance/Uniqueness:

As this arena is emerging these days, there are quite a known number of applications which provide the image captioning for an image. Similar applications include Microsoft Seeing AI, Envision AI and couple more. Recently google has come up with an idea called Google Lookout especially for the disabled. But these applications have lacked a little accuracy and facing problems especially in particular lighting conditions. Especially, these applications in particular are developed for IOS.

2.3 Objectives:

Our main motto is to make a visual world into an audible one. The main objective is to create a caption generator model for to understand images and text. Deep learning models are used in order to verify the best. Our major area of interest is on “FOOD”. Various images of food are collected from the COCO Data set and train the model with these images. Finally, generating a perfect caption for an image.

2.4 System Features:

- The main feature is to generate a perfect caption for the image.
- Providing user, the captions with highest point of accuracy which can be done by training the implementing it with different models and choosing the correct optimizers when needed.
- Food is the major area of focus on which we are trying to collect images and train the model using Show, Attend and Tell methodology.

Subject: Bigdata Analytics and Applications
Document Scope: Project Proposal
Team #: 5

2.5 Related Work:

These are all the existing apps at present.

Envision AI:

Features: The main feature highlighted in the Envision AI is describing a scene. For example, when a scene is scanned it determines what is the image.

Pros: Object detection works well.

Cons: Fails to work in dark lighting conditions.

Google Lens:

Features: The main interesting feature is handwriting detector.
Pros: Barcode Scanner works very accurate.
Cons: Object Detection has an issue.

Microsoft Seeing AI

Features:
The features in the Microsoft seeing AI is that Barcode scanner, person detection, document scanner, handwriting detection, light detection, scene detection, color recognition, currency detection.
Pros: Each feature is working very accurate.
Cons: Some letters are not identified correctly.

Aipoly Vision

Features:
It includes object detection, location recognition, color recognition.
Pros: Speech feature is appropriate.
Cons: When group of objects are present the object cannot be identified.

Google Lookout: This is an application recently launched by the Google for capturing the images and generating a caption and sends a voice message for the disabled people.

Show Attend and Tell Model: Using this methodology we are trying to implement and check whether the model gives us the pretty quiet accuracy.

Paper Link: http://proceedings.mlr.press/v37/xuc15.pdf

Bibliography:

http://proceedings.mlr.press/v37/xuc15.pdf

https://www.analyticsvidhya.com/blog/2018/04/solving-an-image-captioning-task-using-deep-learning/