MINOR PROJECT REPORT

ON

"Image Caption Generator Using Deep Learning"

Submitted in partial fulfillment of the requirementsfor the award of the degree of

Bachelor of technology In

Computer Science and Engineering

Under the guidance of :
Ms. Jyotsna

Submitted By:
Gurman Singh(111/CSE2/2019)
Harneet Singh(103/CSE2/2019)
Kuldeep Singh(117/CSE2/2019)
Sahibjot Singh(065/CSE2/2019)

GURU TEGH BAHADUR INSITUTE OF TECHNOLOGY

RAJOURI GARDEN, DELHI -110027

*Affiliated To*

## GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY

Introduction

- Generating accurate captions for an image has remained as one of the major challenges in Artificial Intelligence with plenty of applications ranging from robotic vision to helping the visually impaired.

- "Image caption generator": the name itself suggests that we aim to build an optimal system which can generate semantically and grammatically accurate captions for an image.

- Researchers have been involved in finding an efficient way to make better predictions, therefore we have discussed a few methods to achieve good results.

- There are two phases: feature extraction from the image using Convolutional Neural Networks (CNN) and generating sentences in natural language based on the image using LSTM (Long Short Term Memory) for framing our sentences from the input images given.

Objective

- Image captioning has various applications such as recommendations in editing applications, usage in virtual assistants, for image indexing.

- Probably, will be useful in cases/fields where text is most used and with the use of this, you can infer/generate text from images.

- There are many NLP applications right now, which extract insights/summary from a given text data or an essay etc.

- The same benefits can be obtained by people who would benefit from automated insights from images.

Model Diagram
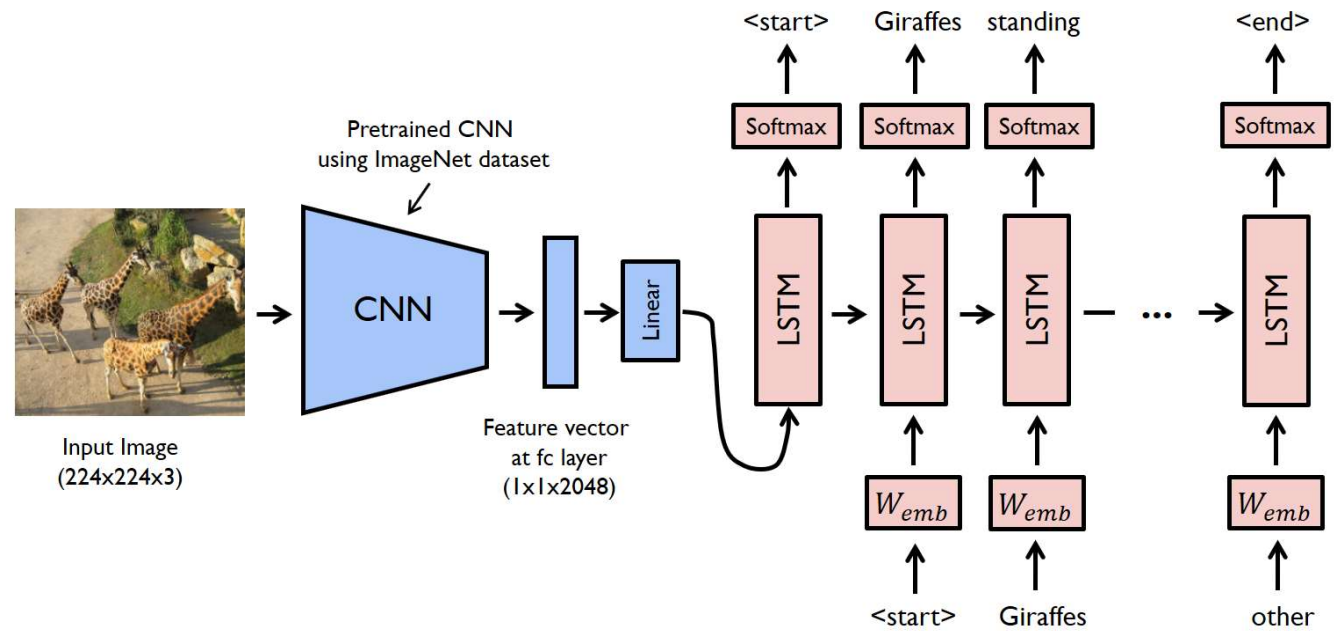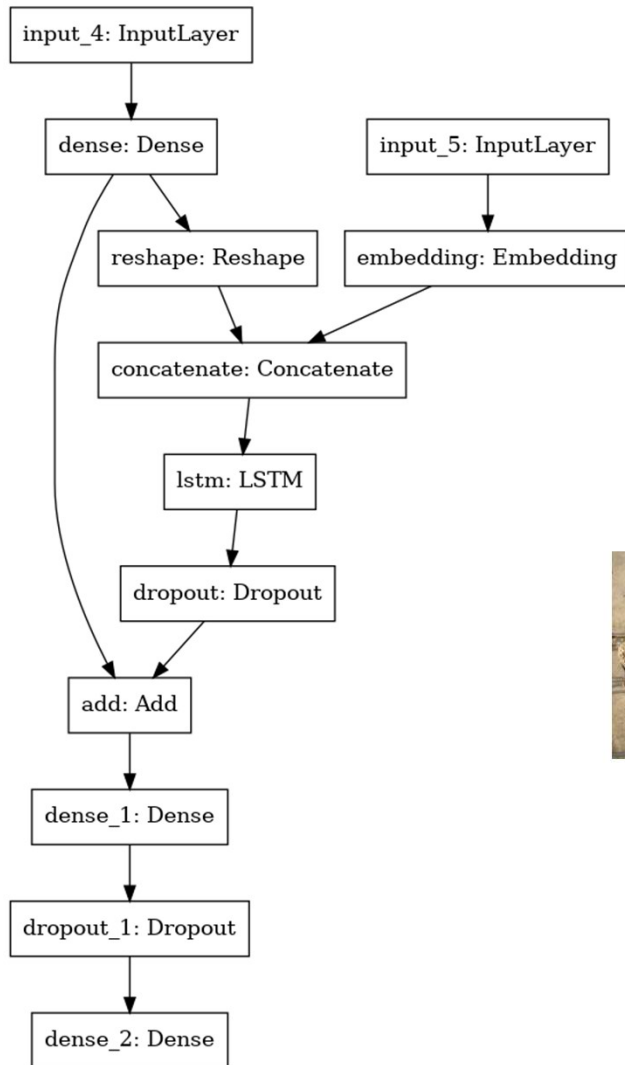
Future Work

- We will also implement Optical Character Recognition(OCR) which can be used for text translation for serving many different purposes.

- There will be scenarios where the user does not receive desired captions; therefore, to resolve this issue, we will enlist the user's assistance by adding the appropriate caption of the image that is correct according to the user, which can then be used to retrain the model, thereby improving the accuracy of our model.

- We can extend this software to serve wider audience by developing cross platform application.

- We can implement this feature for key frames of the videos.

Software and Tech Stack Used

- Python
- Machine Learning
- Jupyter Notebook
- Kaggle
- FastAPI
- React
- Javascript
- Visual Studio Code
- Git