Image caption generator application

A PROJECT REPORT

Submitted by

Vineet Sharma (19MIM10001)

Thevaprakash P (19MIM10003)

Shivanshu Rajput (19MIM10060)

Shrey Asthana (19MIM10066)

Kesavan R (19MIM10086)

BONAFIDE CERTIFICATE

Certified that this project report titled "**Image caption generator application "** is the Bonafide work of "**VINEET SHARMA (19MIM10001), THEVAPRAKASH P (19MIM10003), SHIVANSHU RAJPUT (19MIM10060), SHREY ASTHANA (19MIM10066)** and **KESAVAN R (19MIM10086)**" who carried out the project work under my supervision, certified further that to the best of my knowledge the work reported here does not form part of any other project / research work on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

ACKNOWLEDGEMENT

Primarily, I would like to thank the Lord Almighty for his presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to our instructor Pradeepthi Duggaraju, for continually guiding and actively participating in our project, giving valuable suggestions to complete the project work.

I would like to thank all the technical and teaching staff of the smart internz and smart bridge, who extended directly or indirectly all support.

Last but not the least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

Index

INTRODUCTION

The world around us is incredibly beautiful, full of wonders and to see them we have been provided with the eyes. And it is really disappointing to see many people with partially and complete blindness. It is our effort to improvise their life by providing the description of the image.

It is a tedious task to describe the content of the image captured using the properly English statement but it could have a significant impact by helping visually impaired people better understand their surroundings.

In the modern era of technology most of the cell phones have the built-in cameras to capture the surrounding, making it convenient to visualise the surrounding for the visual impaired person so that they can be benefitted and helped to overcome the problem that they have faced frequently.

PROBLEM STATEMENT

The problem is that the blind person must depend on someone or something to explore the

surrounding. The need of stick is mandatory for a blind person. We try to reduce the burden of

the blind person by providing him the description of the surrounding by just seeing the

surrounding in the form of an image.

SOLUTION TO THE PROBLEM

We are creating a web application where the user selects the image and the image is fed into the

model that is trained and generated caption will be displayed on the webpage. This generated

caption could be read aloud in the future for better aid to the visually impaired people.

LITERATURE SURVEY

Existing problem

For a blind person, it is really disturbing to not able to see the surrounding. In fact, they have to

depend on the other senses for perceiving the environment. The only way to overcome the

blindness is the transplant of the retina by which they can able to see the neighboring thing but

that is not affordable to each and everyone.

Another problem is that if the person is suffering from squint, then the transplant would also

does not affect the situation of the person.

Proposed solution

A method to somehow visualize the surrounding to the visual impaired person is our priority in

this project. We are creating a web application where the user selects the image and the image

is fed into the model that is trained and generated caption will be displayed on the webpage.

This generated caption could be read aloud in the future for better aid to perceive the

environment for the visually impaired people.

EXPERIMENTAL ANALYSIS

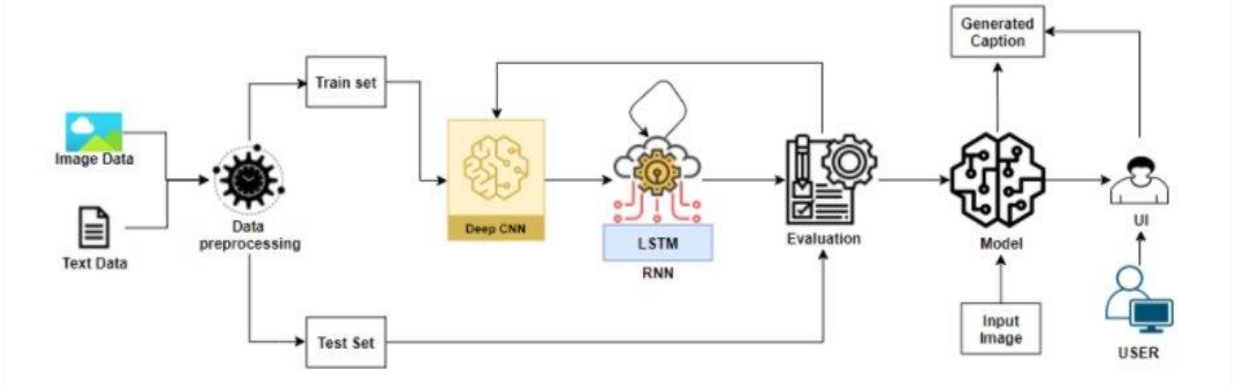
To accomplish our project whose main objective is to describe the image in the best

form to the viewer using the CNN-RNN model for the visual impaired person. We need

complete the task enlisted below: -

- Data Collection
  - Collect the dataset or create the dataset
- Data Preprocessing.
  - Import required Libraries
  - Extract features from each photo in the directory
  - Processing the text data or descriptions
- Model Building
  - Import the model building Libraries
  - Loading dataset for training the model
  - Tokenizing the Vocabulary
  - Define the Model
  - Define the CNN-RNN Model
  - Configure the Learning Process
  - Training the model
  - Save the Model
  - Testing the Model
- Application Building
  - Create an HTML file
  - Build Python Code

System architecture for the project



Hardware requirement

The laptop with a good broadband is required to accomplish the project.

Laptop must have a bare minimum of 8 gpu.

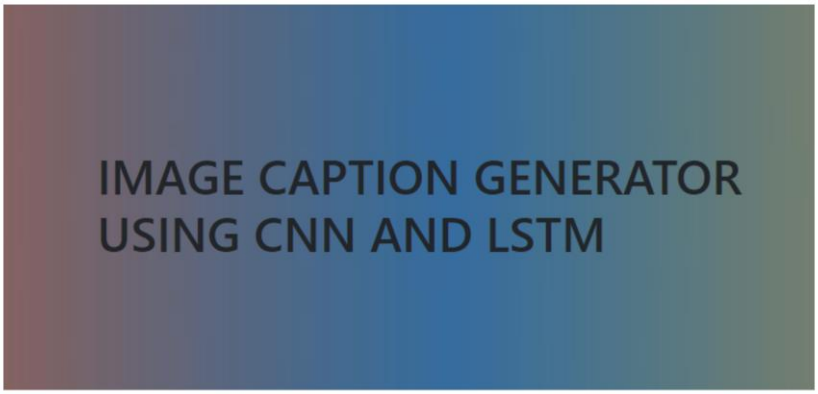
Software requirement

- Python 3.9
- Anaconda navigator
- TensorFlow version 1.14.0
- Keras 2.2.4
- Flask
- And other python libraries like NumPy, pandas, OpenCV, matplotlib and many more.

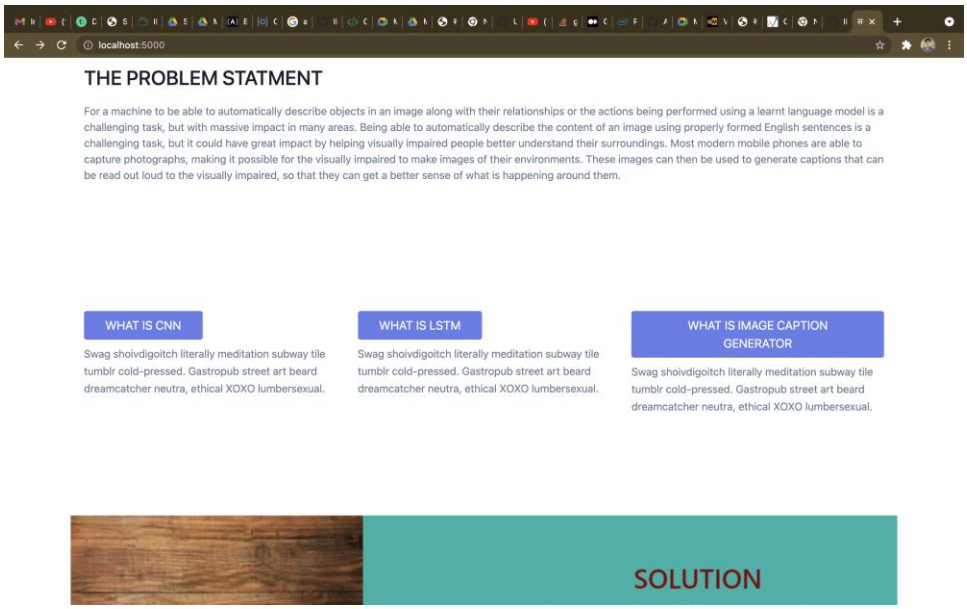
CONCLUSION

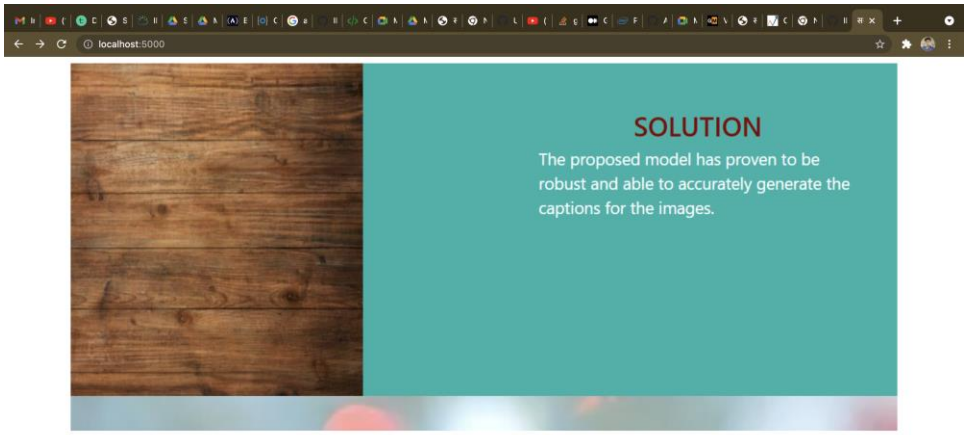
The following are the images of our project in which we are producing the captions for the

images uploaded.



THE PROBLEM STATMENT





NAME

For a machine to be able to automatically describe objects in an image along with their relationships or the actions being performed using a learnt language model is a challenging task, but with massive impact in many areas. Being able to automatically describe the content of an image using

NAME

For a machine to be able to automatically describe objects in an image along with their relationships or the actions being performed using a learnt language model is a challenging task, but with massive impact in many areas. Being able to automatically describe the content of an image using properly formed English sentences is a challenging task, but it could have great impact by helping visually impaired people better understand their surroundings. Most modern mobile phones are able to capture photographs, making it possible for the visually impaired to make images of their environments. These images can then be used to generate captions that can be read out loud to the visually impaired, so that they can get a better sense of what is happening around them.

Choose File : Upload

Choose file 99679241_adc853a5c0.jpg

Submit

startseq pelican

APPLICATION

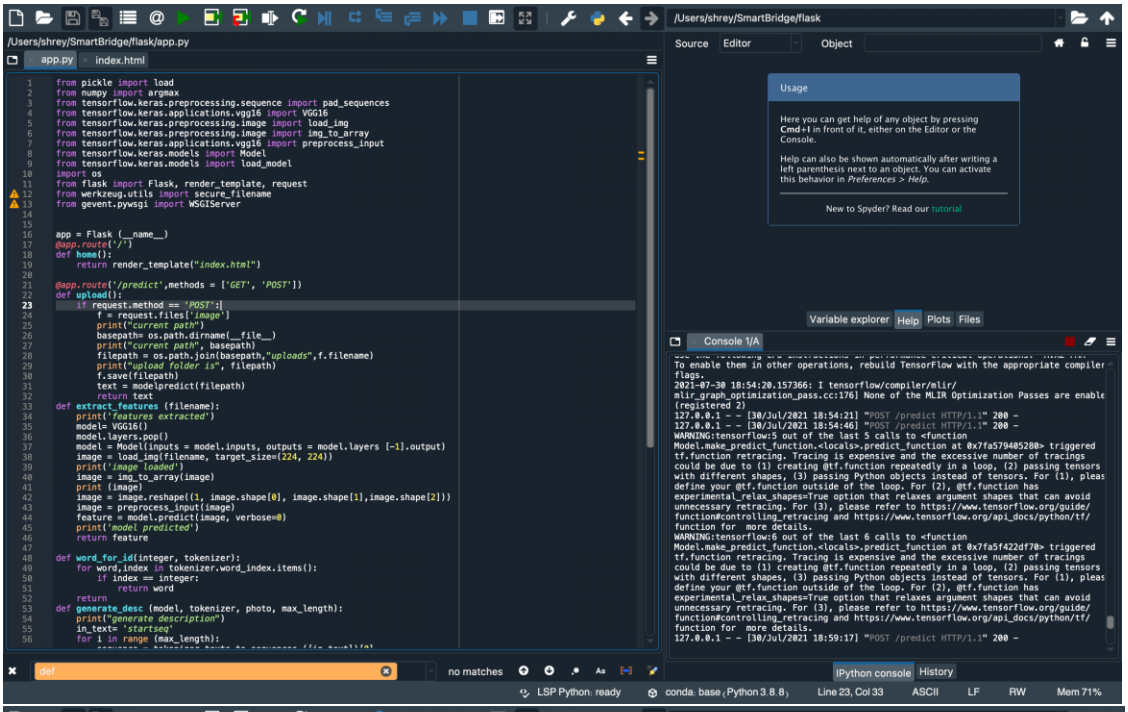
The following application could be used for better understanding of the uploaded images. With the voice assistance we can make a better application for the blind person to visualize the surrounding in a much better way.

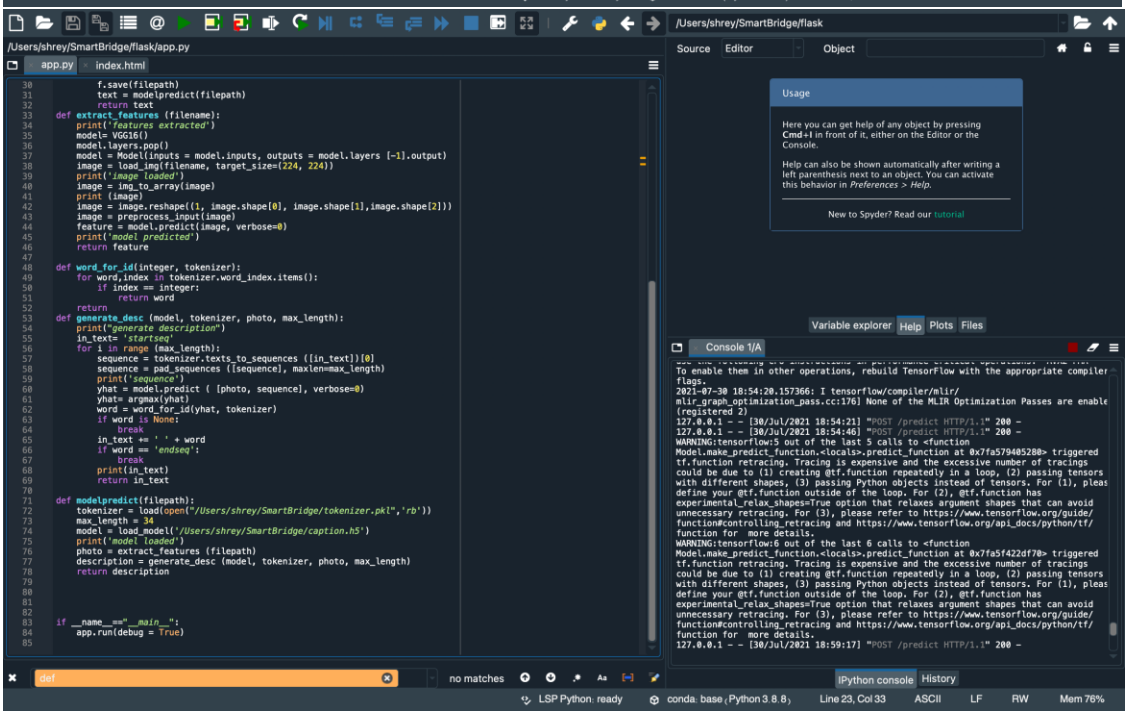
FUTURE SCOPE

There is always a scope of improvement in each and everything created. On the same principle there might be the future improvement in this project also. The usage of voice could be our next step to enhance the project. By the addition of the vocals would be boon to the blind person as they can here the description of the surroundings.

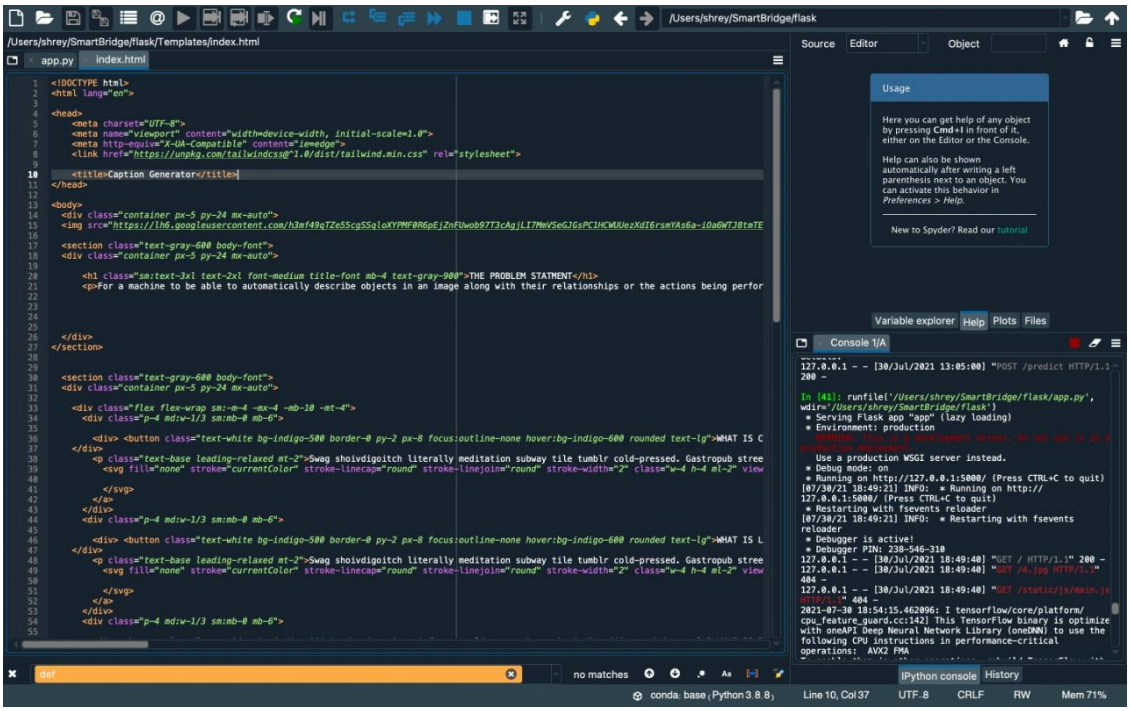
Next, we could launch the versions of this web application onto the different operating systems like android and iOS.

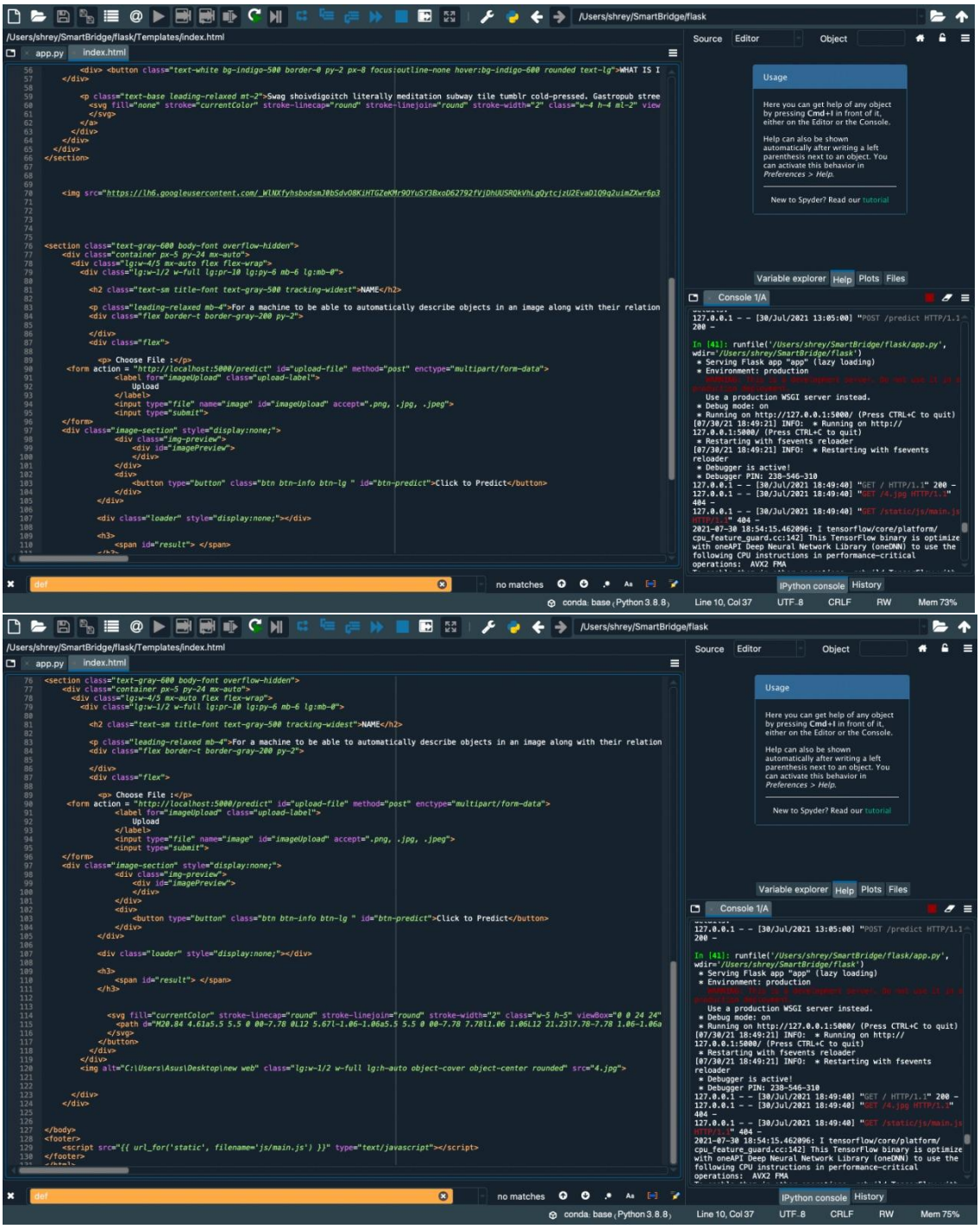
CODE SNIPPET





This is app.py file.





This is HTML code

BIBLIOGRAPHY

- https://smartbridge.teachable.com/courses/1450164/lectures
- https://www.math.ucla.edu/~minchen/doc/ImgCapGen.pdf
- https://realpython.com/python-web-applications/