# CSN-341 Computer Networks

# Image Captioning- Project Report

**Group Members:**

- Arjun Krishna (15115030)
- A.S.Akil Arif Ibrahim (15116001)
- Ashutosh Parija (15116010)
- Karan (15116028)
- Mohit Chaudhary (15116033)

## Introduction:

The application is basically a functional Google Chrome Extension that generates a caption for an image which describes what is contained in the image. The image captioning is done by a deep learning network which runs on a cloud server(Heroku). The caption generated is converted into speech by JavaScript's Web Speech API(SpeechSynthesisUtterance API).

### Audience Targeted:

- Visually impaired people
- All the people who want help in understanding the context of an image.

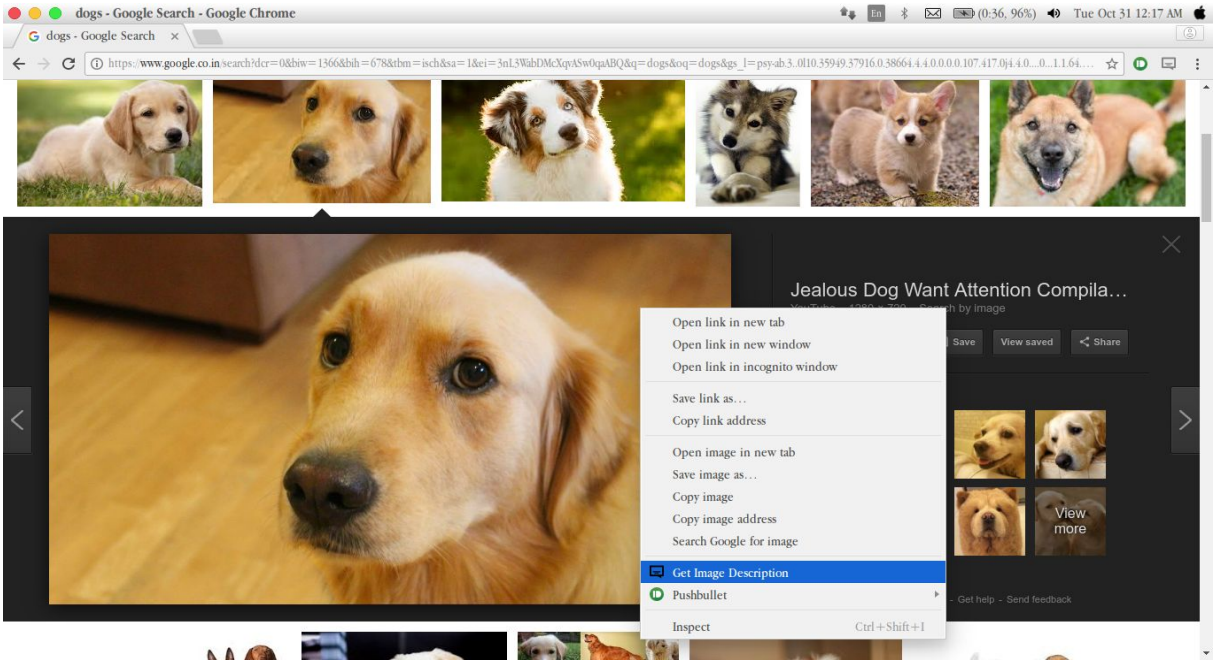
### Tools and Programming Languages Used:

- Python, HTML, JavaScript as Programming Languages.
- Heroku Platform is used for cloud deployment.

## Functioning:

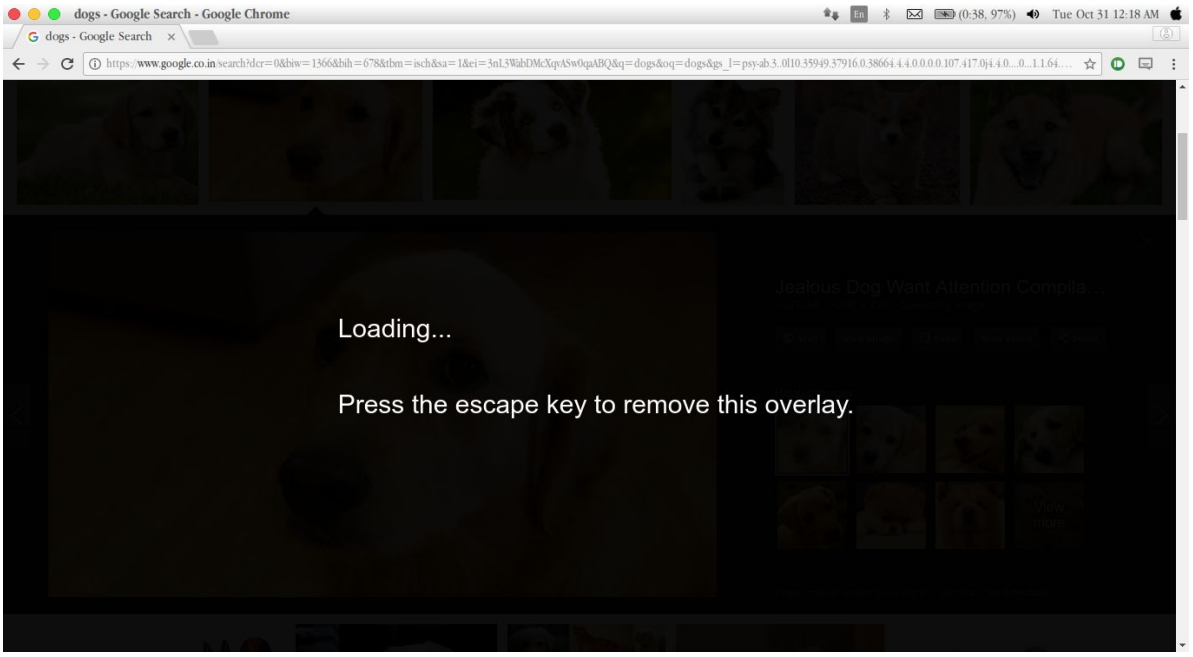
### Architecture of the application

When the user right clicks an image, there will be an option to view description of the image along with the other default options as it is shown in the image below. The Chrome extension provides this option.




Options shown when we right click on an image

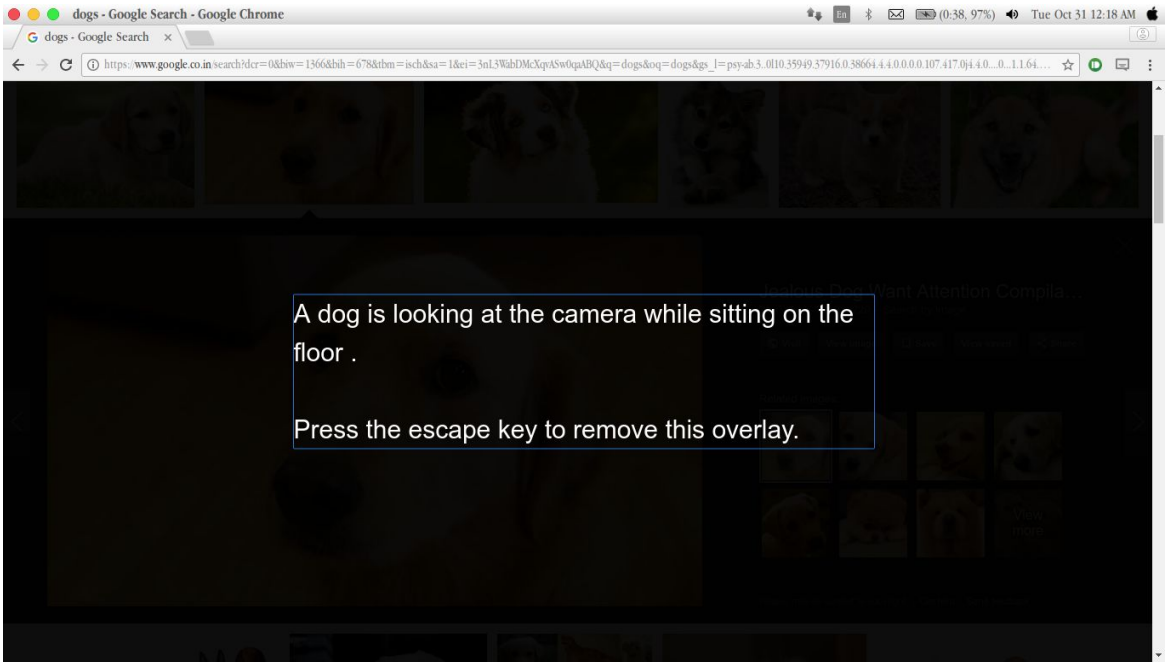
When this option is clicked, the url of the image is sent to the our program running on a cloud server,which is hosted on heroku, and an overlay is shown in the browser.



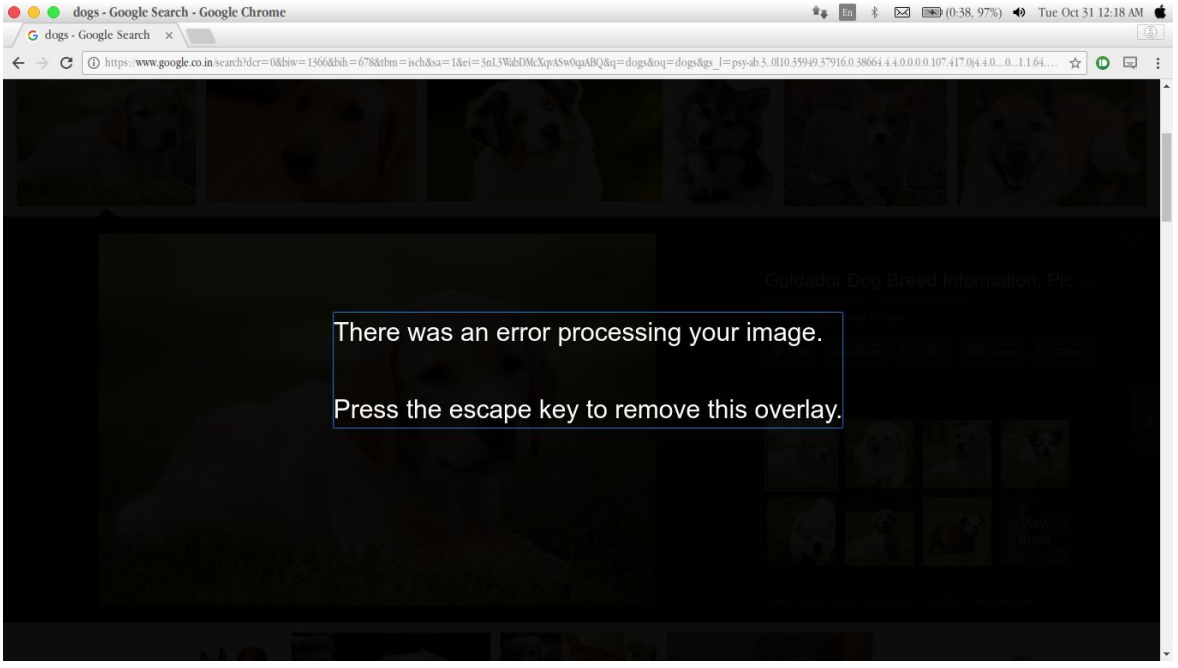

Overlay displayed after selecting the get caption option

This program captions the image using the Show and Tell model [1].This model uses NIC, an end-to-end neural network system that can automatically view an image and generate a reasonable description in English. NIC is based on a convolutional neural network that encodes an image into a compact

representation, followed by a recurrent neural network that generates a corresponding sentence. The model is trained to maximize the likelihood of the sentence given the image.
The caption generated is sent to the extension which is then displayed on the overlay, and is also generated as audio using SpeechSynthesisUtterance API. To exit the overlay, the user has to press Esc key. If there is any error in generation of caption, then a message will be displayed on the overlay, which states that there is some error as it is shown in the images below.

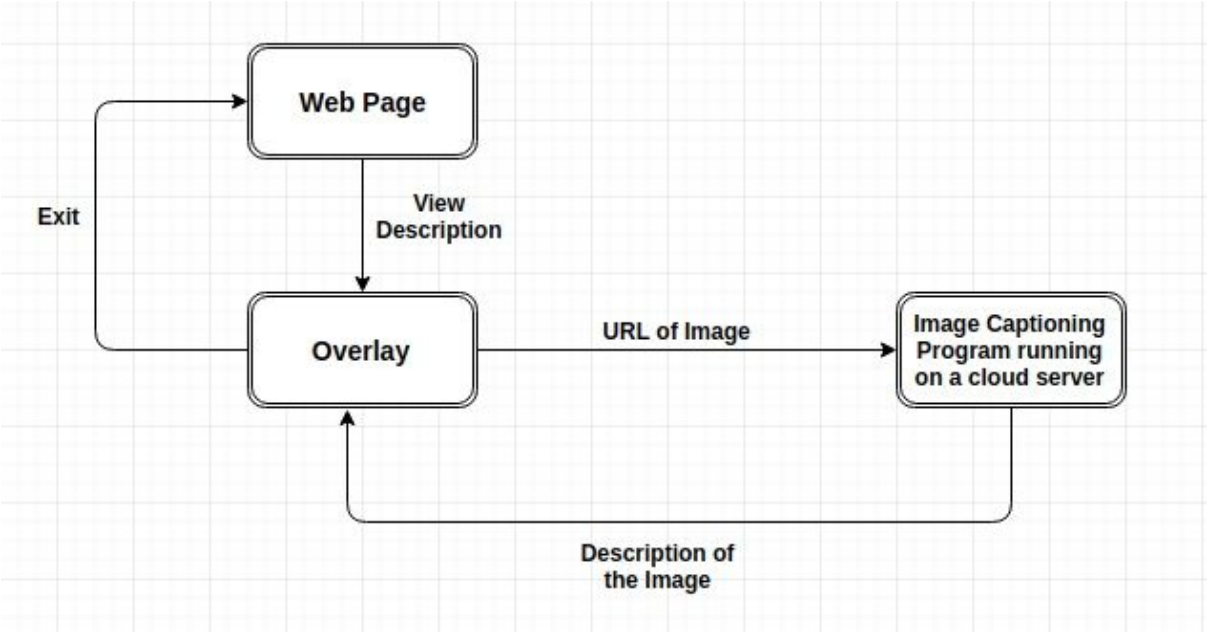



Response from the server when there is no error




Response from the server when there is an error

**Flowchart representation of the architecture**




## Modules and Functionality

**Modules:**

Chrome Extension-
JSON file for manifest
CSS file for styling
JavaScript file which provides the option to view image description only for images
JavaScript file which listens to the event of choosing view image description option, sends url of the image to image caption generator running on a cloud server, receives the description of the image and displays it on the overlay along with the audio of the description.

Image Caption Generator-
Python file which receives the url of the image from the JavaScript file of the extension and captions the image and returns the description of the image.
Python file which contains the Show and Tell model [1].
Dataset which contains image and its description.

## Limitations:

- Can't be used for protected images. Example- It can’t be used on the images in facebook as they are protected ones.
- The image caption generation also takes a considerable amount of time,the algorithm needs to be optimised in terms of both accuracy and time taken to generate the output.

## Improvements in the future:

- A better model that is trained on a more comprehensive image set can be used.

**References**

[1] Oriol Vinyals, Alexander Toshev, Samy Bengio, Dumitru Erhan, “Show and Tell: A Neural Image Caption Generator”, CVPR2015