Group 2

Members:

Harsh Choudhary- 17CH10016
Dharmender - 17CH30009
Abhijeet Deshmukh -17IE10002

Image Captioning Webapp for Visually Impaired

Introduction

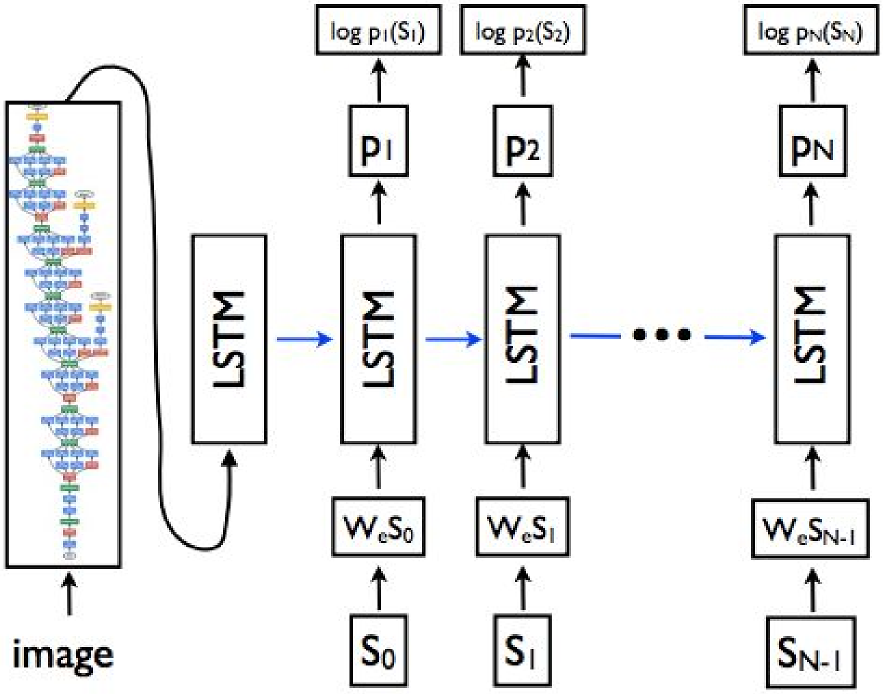
We designed this web app to help people with visual impairment by captioning images. The current user interface for feeding images to the app can be improved (difficult for a web app, couldn't proceed with the android app due to lockdown) so that people with visual impairments can easily operate it. We did the best that could be done within these circumstances.

Working

The **image** (**jpg** format only) is uploaded in the web app, which is sent to the backend deep learning model to convert images to text. The text describing the images is then read out so that a person can understand it. (The text description may not be accurate - training on larger datasets will improve performance).
Video running on the web app shows what we intended to make. (Watch demo_video)

Model Representation:

Text



Deep Learning Model:

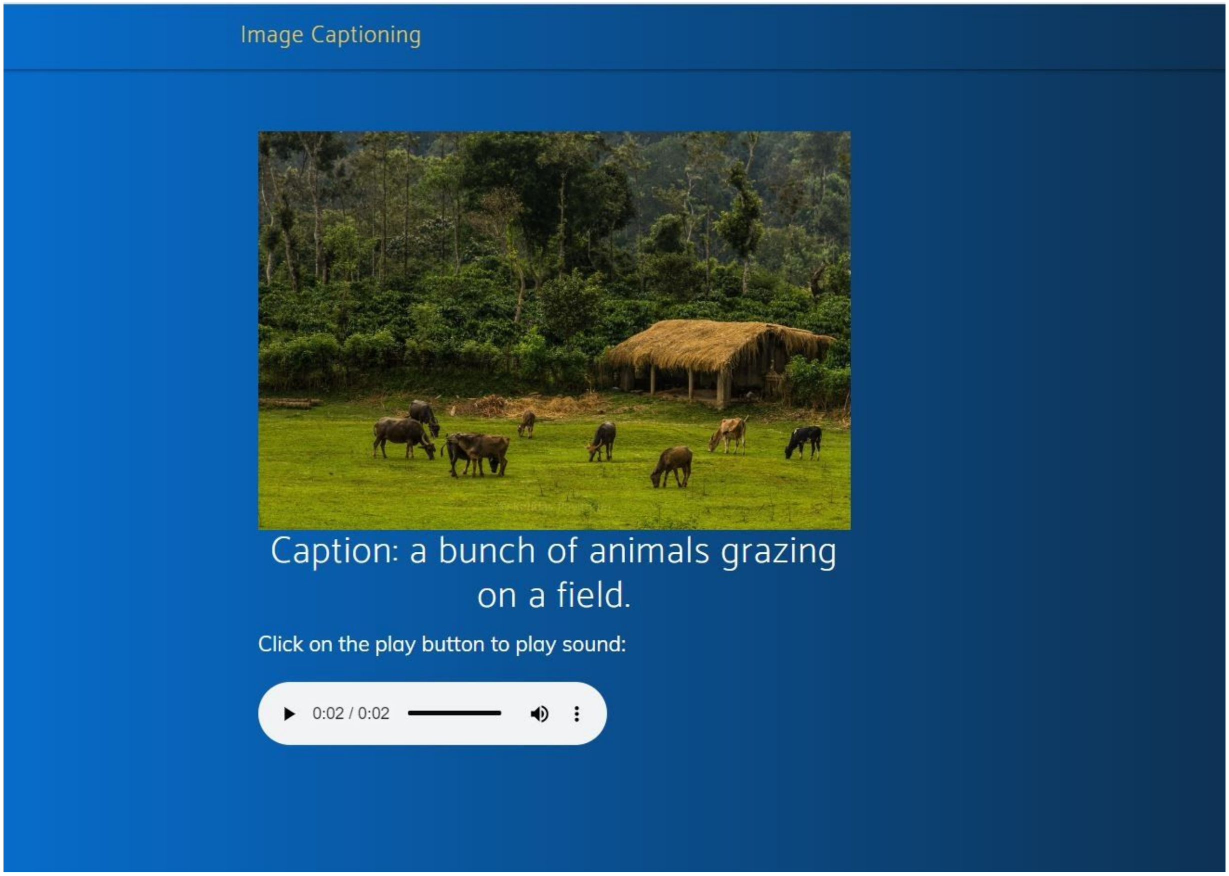
We used a subset of 30,000 captions from the MS-COCO dataset and their corresponding images to train our model.
Choosing more data would result in improved captioning quality.

We used InceptionV3 (which is pre-trained on Imagenet) to classify each image.

We limited the vocabulary size to the top 5,000 words (to save memory). Replaced all other words with the token "UNK" (unknown).

Model Details

- We extracted the image features from the lower convolutional layer of **InceptionV3** giving us a vector of shape (8, 8, 2048).
- Then squashed that to the shape of (64, 2048).
- This vector is then passed through the **CNN Encoder** (which consists of a single Fully connected layer).
- The **RNN** (here GRU) attends over the image to predict the next word.



Caption: a bunch of animals grazing on a field.

Click on the play button to play sound: