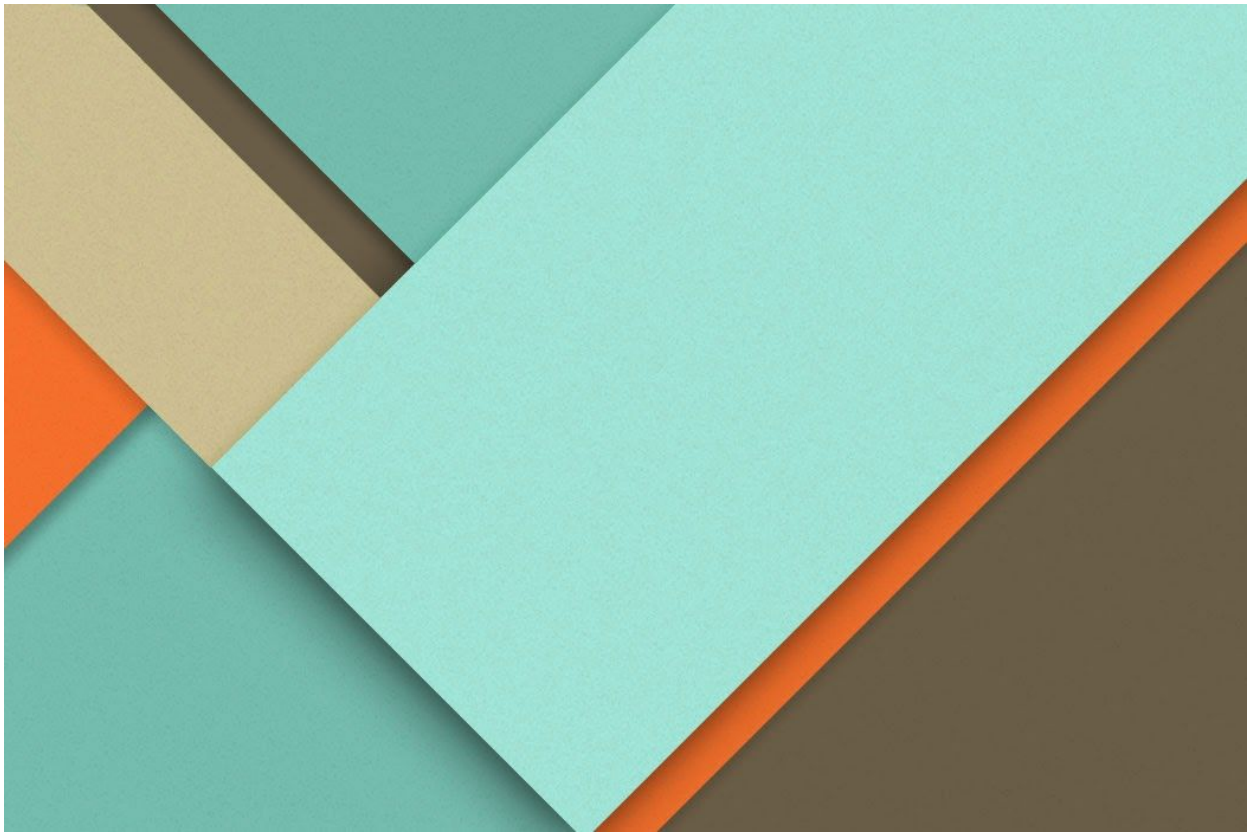

Scene Description In Videos For Blinds

Project Supervisor: Dr. Rupali Verma

TEAM MEMBERS:

Parikh Goyal (18103023)
Rajat Gupta (18103025)
Chaitanya Gupta (18103021)
Saiyam Goyal (18103030)

Motivation

Automatic image captioning is the task of automatically producing a natural-language description for an image. It has the potential to assist those with visual impairments by explaining images using text-to-speech systems. The project aims to extend the concept and apply it in videos, thereby providing the scene description of videos to visually impaired using text-to-speech systems, to help them better understand the video and its scenes.

Description

The project aims to develop a voice command activated web application that can accept videos from the local computer as well as Youtube video links, thereby working as a video player and provide users the feature of getting an on-demand description of scenes in videos. The web application will use an image captioning deep learning model to generate relevant captions for scenes that the users request and will deliver them through text-to-speech systems.

Problem Statement and Use Cases

Image Captioning is used to describe the context of an image, to tell useful information in the image as text, using deep learning models. It helps blind people to get a better understanding of not only videos but also web pages, blogs, and other images.

Image captioning in videos also finds its use case in making silent films more interesting by including the description of scenes as added audio to the movies.

Goals

1. To develop a model that could generate relevant scene-specific captions for different kinds of scenes in videos.
2. To develop a user-friendly web interface for playing videos and scene depiction that works with maximum support for visually impaired people.

Project Scope

The project mainly focuses on developing deep learning image captioning model to generate relevant scene description for various types of videos, and a web interface that can accept both local video files and Youtube video links to play and generate captions as audios. The project further aims to extend user comfort by incorporating voice commands and tries to better the image captioning model using the user feedback loop.

Specifications

The proposed project is further divided into four main tasks:

I. Image Captioning Model

The project requires an Image Captioning model to generate captions for images. The captions need to be of appropriate length and not too large. The model should be compact enough to give real-time like performance and should also be generic to generate relevant captions for different kinds of images belonging to a wide range of videos.

II. Web Interface

The project proposes to develop a web interface for video selection, playing, and scene depiction. Web interface ensures a good user experience. The web framework needs to interact with the user videos and extract the frames when the user requests for captions, and pass the frame to the deep learning model and output the caption generated by the model as audio.

III. Voice Commands

The project aims at making maximum possible user interaction with voice commands as possible, to ensure smooth working of application with visually impaired persons too.

IV. Feedback Loop For Continuous Learning

The project tries to include continuous learning features in the image captioning model, by using a feedback loop where users can feed in more relevant captions to the application to help in further learning of model and helping it generalize better.