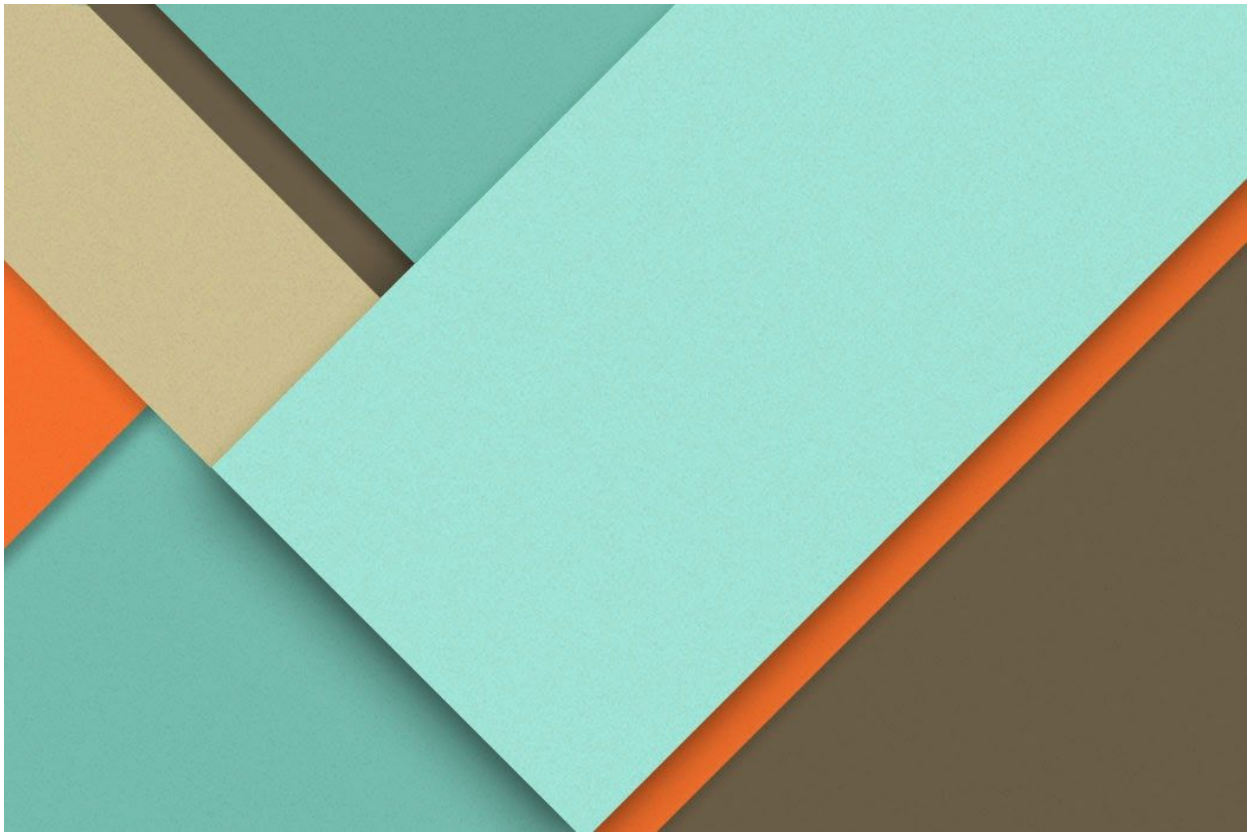

Scene Depiction In Videos For Blinds

15.08.2020

TEAM MEMBERS:

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

Overview

Scene Depiction in videos is the problem of describing the scene in the videos to visually impaired persons to make them understand better. The project proposes to develop a software that can help in generating appropriate captions for scenes in videos using deep learning techniques of Image captioning.

Description

The project aims to develop a web application that can accept videos from the local computer as well as Youtube video links and can play them on the application and provide users the feature of getting a description of scenes in videos on demand. The web application will use a pre-trained image captioning model to generate relevant captions for scenes that the user's request and will generate them as audios while playing videos. The project aims at developing the project with maximum assistance to the users by using audio instructions, enabling visually impaired user assistance while using the application.

Problem Statement and Use Cases

Image Captioning is a deep learning technique to describe the context of an image, to tell the useful information in the image as text. 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.