Neural Networks - Project Report

May 15, 2021

INTRODUCTION

The primary purpose of this project is to create a pipeline that downloads videos from Youtube and performs feature extraction based on requirements. This program allows us to create datasets with just 1 command line and can be used to train several types of neural networks models like video classification or to create a knowledge base.

INPUT ARGUMENTS

Since this program is written with CLI, there are several arguments that allow us to control what we need to accomplish.
The input arguments are explained below:

1. **link** - Takes in a link to a youtube video, the path of a text file containing several youtube links or keywords to perform a search on youtube.
2. **search_keywords** - A boolean flag that lets the program know that we have mentioned search keywords
3. **number_of_links** - An integer that tells how many videos to download from the search based on the keywords
4. **dataset_path** - Path where we would like our dataset to be downloaded
5. **video_resolution** - Resolution of the video we would like to download - 360p, 480p, 720p, 1080p
6. **load_pickle** - Load previously saved paths and information about downloaded videos from a pickle file
7. **create_directories** - Creates the directories for the files to be downloaded into, set to True if running the program for the first time.
8. **download_data** - Set to True if you would like to download the data or just proceed with feature extraction from the already downloaded data

LIBRARIES

- **Pytube** - Python library that gets information from Youtube API for a youtube video based on its video link or ID. Allows us to download video, audio, captions, title, etc.
- **Pytesseract** - Tesseract optical character recognition library for python that takes in an image and extracts recognised text from it. Used to convert Image to Text.
- **cv2** - Computer Vision library to read frames of a video file and to work with frames of the video.
- **googleapiclient.discovery** - Google API library for python to get subtitles of a specified video.

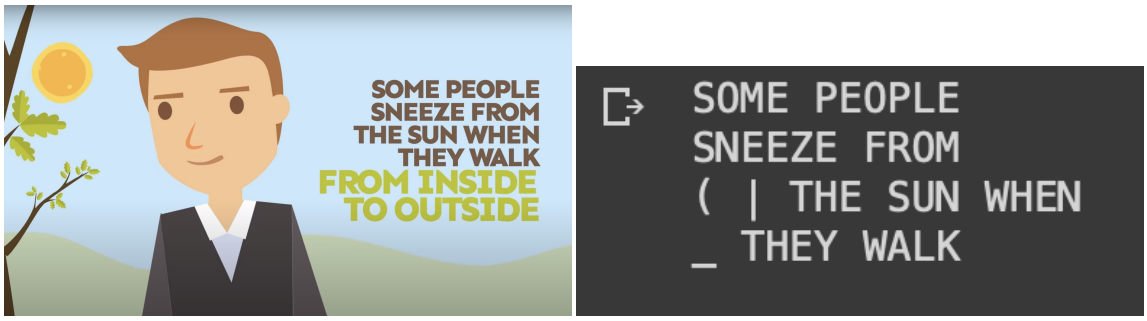
PIPELINE CLASS

The core of this program is in the Pipeline.py file where I have created a Pipeline Class which based on the arguments provided handles most of the heavy lifting and functionality. All the required imports are initialized in this class file and it can be imported into a main file and used as required.

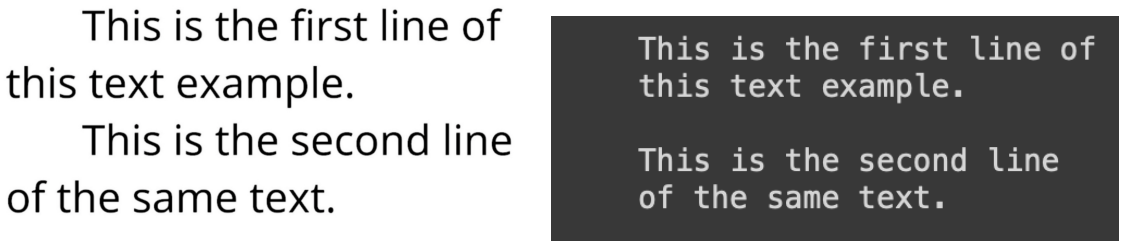
FEATURE EXTRACTION

We are extracting the text present in the video frames. Since we are working with K12 Educational videos, rather than the frames itself, the text in the frames is of more value to us. This allows us to convert large video files as our data to simple files of text which drastically reduces computational cost. Some examples of frames with text extracted from them is shown below. The input image is on the left and the output text is displayed on the right in an image.

Example 1



Example 2



Here we can see in an ideal black and white picture the performance of the library is very good but on a practical image it is a little more challenging. Even though it is not perfect, the library is good enough to give us the context of what is on the screen compared to having the entire video as a feature.

CONCLUSION

Since we know that a neural network is only as good as the data that it is being trained with, the most important part of creating and training models is the data collection process. This 1 line program makes it much easier to collect a dataset for training different types of models as mentioned before. This code can be extended to extracting text from the audio, I3D features from the videos to reduce feature size and so on.