Advertisement Detection in Audio-visual content

[007]Akash, [049]Himanshu, [124] Sundar.

Department of Computer Science and Engineering,

[1]National Institute of Technology,

Tiruchirappalli

Guided by Dr. Santhana Vijayan

Introduction

- In videos, for promoting sponsor companies' advertisement is often included
- The advertisement could be simply explaining their products or any unrelated video compared to main video can also be included
- The project attempts to identify the advertisement boundaries and create an application which skips advertisement part in the input video

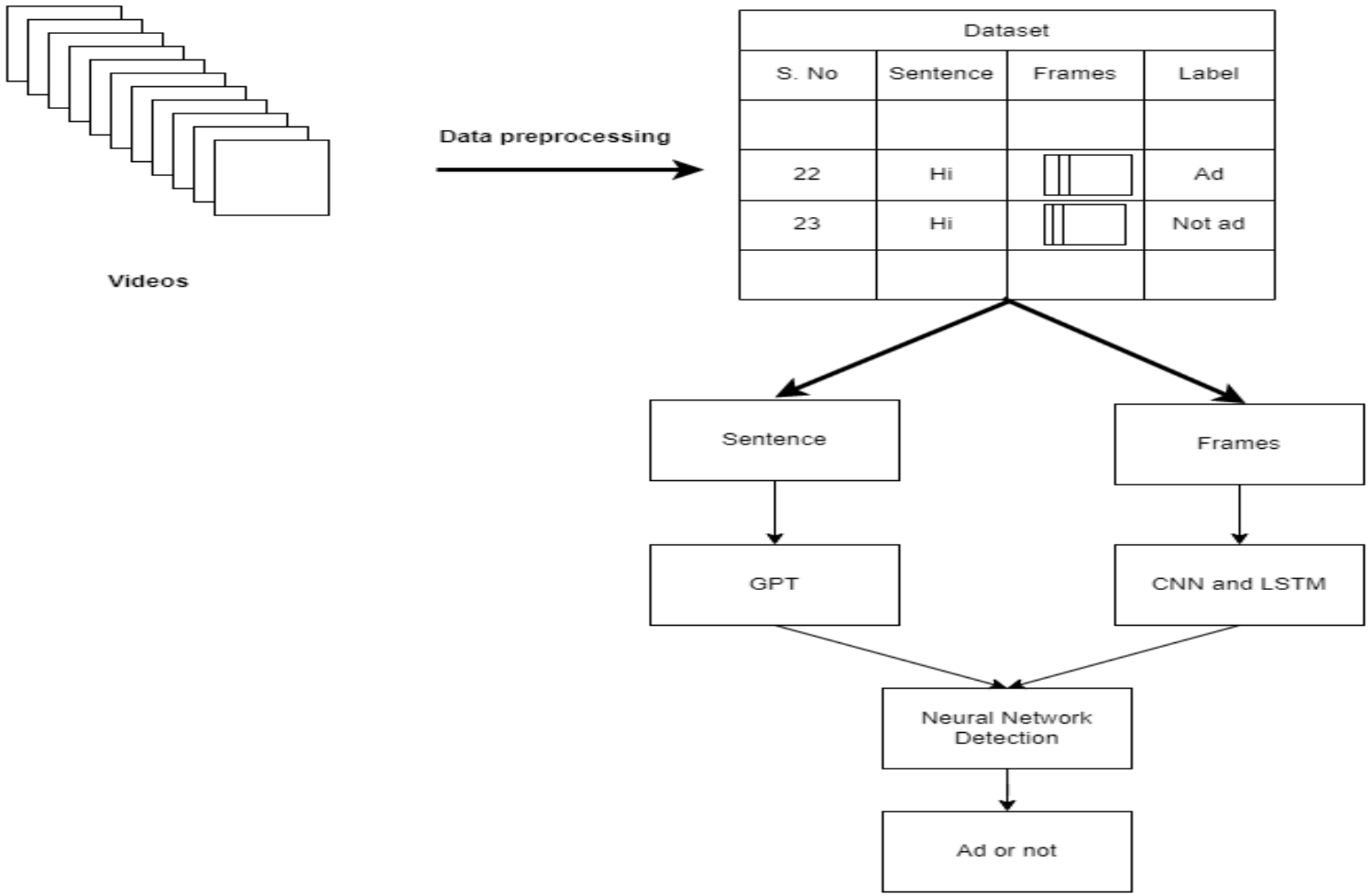
Existing Solutions

Authors	Methodology
Soumya Majumdar et al., [1](2022)	This paper detects advertisement in TV using only audio without frames. Advertisement is detected using Mel – Frequency Cepstral Co - effecient (MFCC) parameter.
Waseemullah, Najeed K et al., [2](2018)	This paper detects advertisement in a two step manner. As a first step, it analyses RGB mean of each frame in video and during the advertisement region it will have higher variance. Since some non - advertisement could also satisfy above condition, a secondary check based on whether frames are in same scene is used to classify.
Shervin Minaee et al., [3] (2022)	In this paper, advertisement is detected using video and audio. Both video and audio progress through two stream audio – visual convolutional neural network and results are used to identify advertisement.
Qian Xia et al., [4] (2012)	This paper identifies ad boundary using both audio and video. The frames are segmented and inter – frame difference is used for shot detection. It uses the fact that advertisement videos has higher short – time average energy. Combining both audio and video information, it detects advertisement on TV.

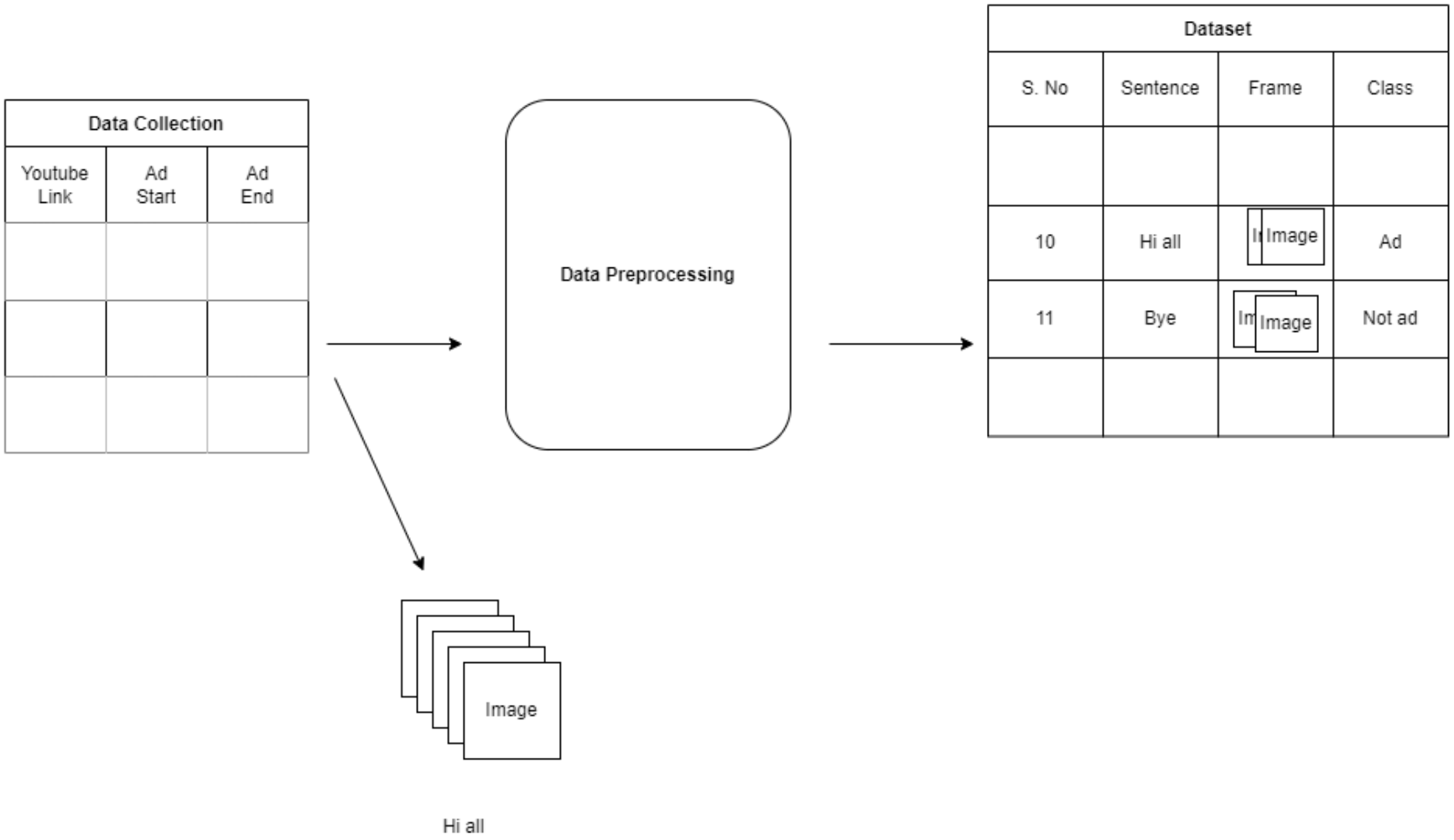
Limitations

Authors	Limitations
Soumya Majumdar et al., [1](2022)	The method of using only audio works poorly on implicit videos as advertisement will be shared on similar tone to video. It doesn't utilize the knowledge from frames.
Waseemullah, Najeed K et al., [2](2018)	The paper deals only with Television ads and it assumes that entire ad will be on same scene which may not be true. Processing some non – advertisements which passes first step leads to extra computations.
Shervin Minaee et al., [3] (2022)	This method doesn't utilize semantics of sentences and this model works on video which are completely either ad or not ad. It cannot identify advertisement boundary within the video.
Qian Xia et al., [4] (2012)	This paper detects advertisement on TV only and video boundary is detected based on change in frame. In implicit videos such change is minimal and might yield poor result.

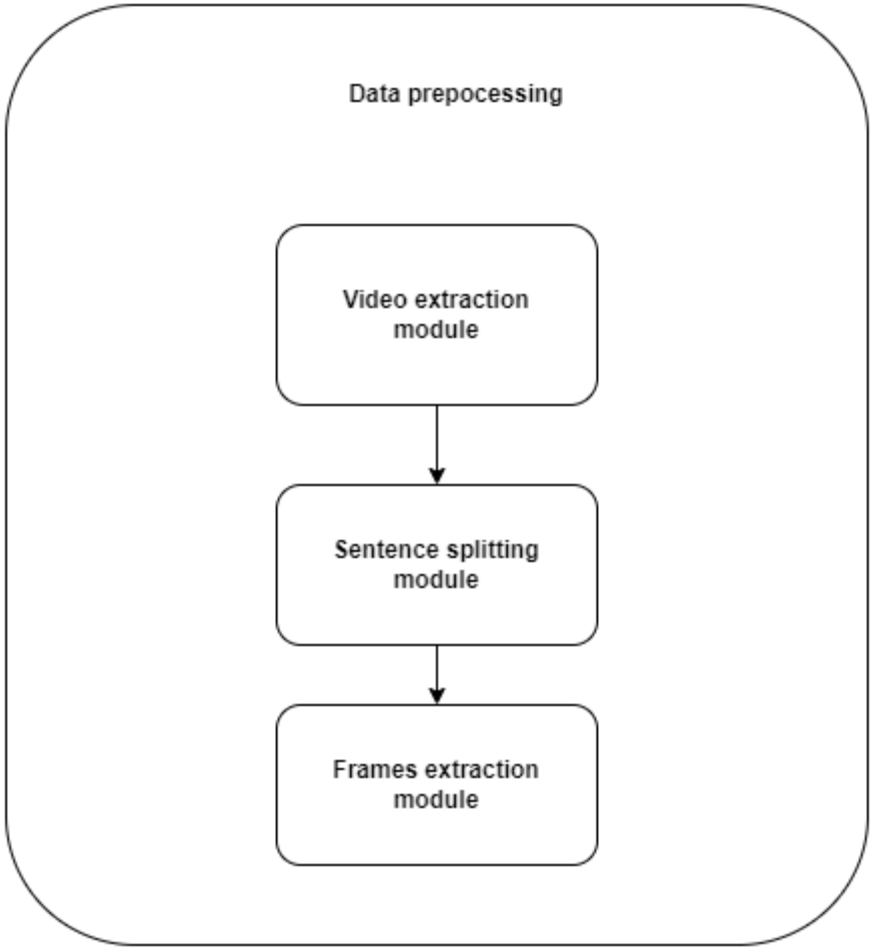
Block diagram of proposed methodology



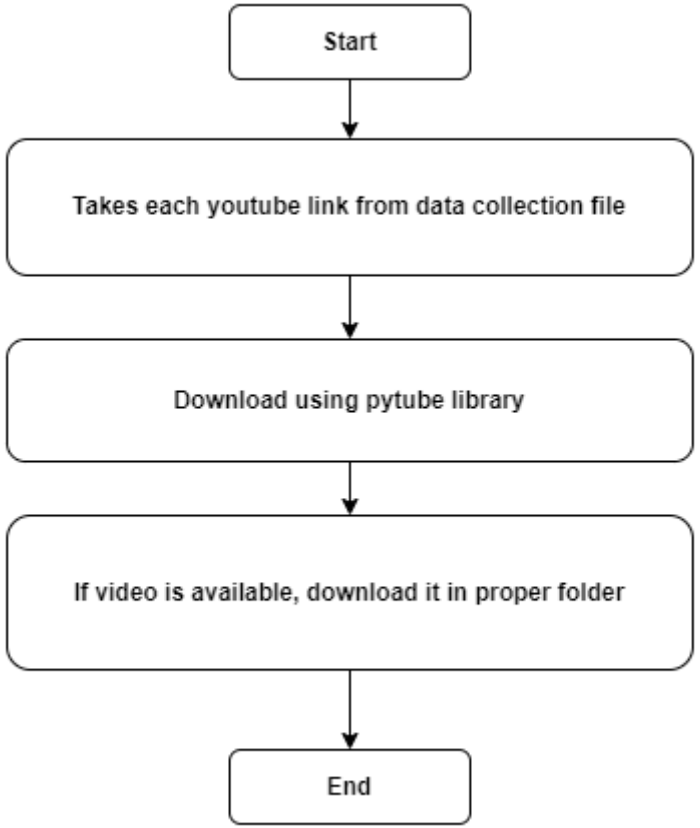
Workflow diagrams



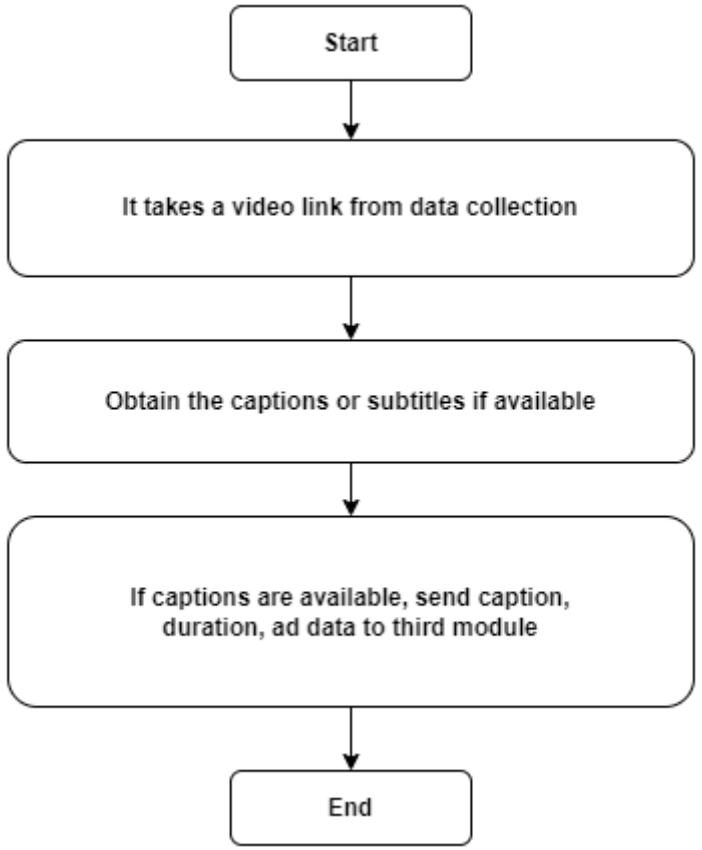
Workflow diagrams



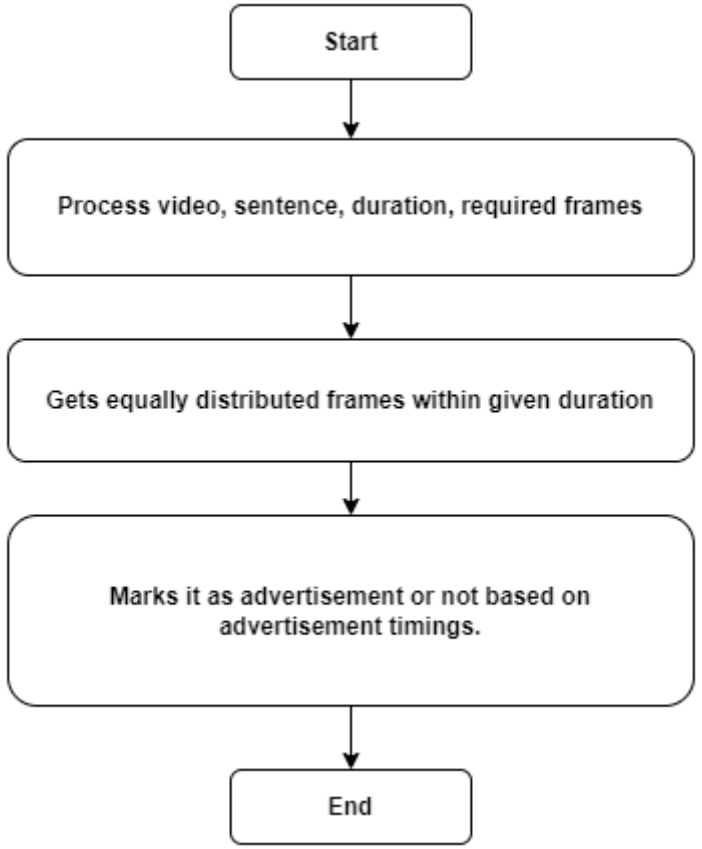
Workflow diagram – Video Extraction Module



Workflow Diagram – Sentence splitting module



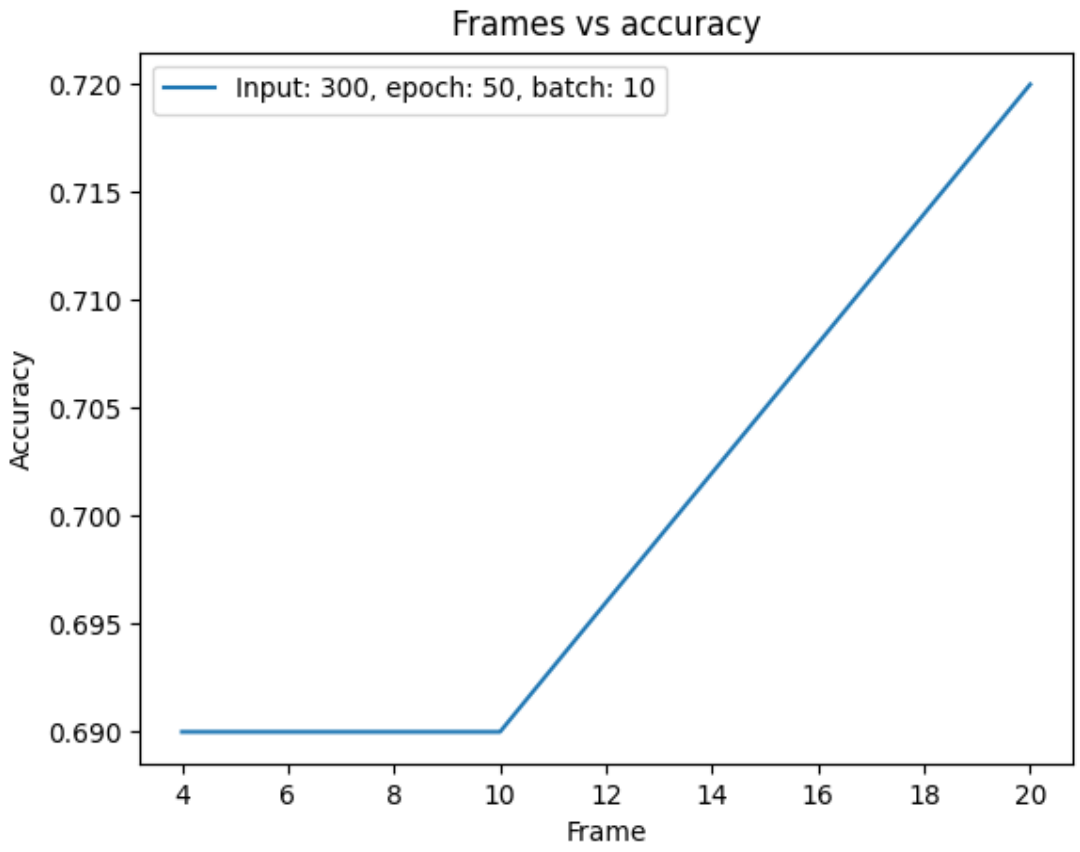
Workflow diagrams – Frames Extraction Module

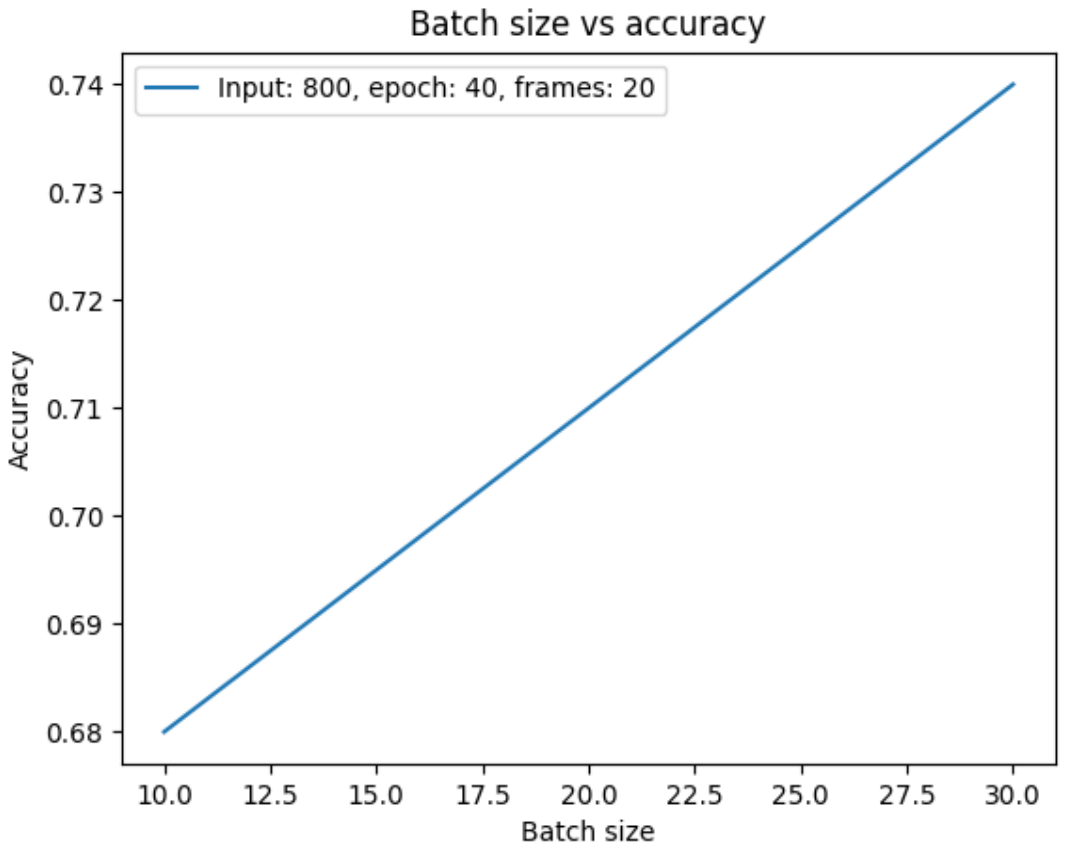


Implementation / Simulation environment

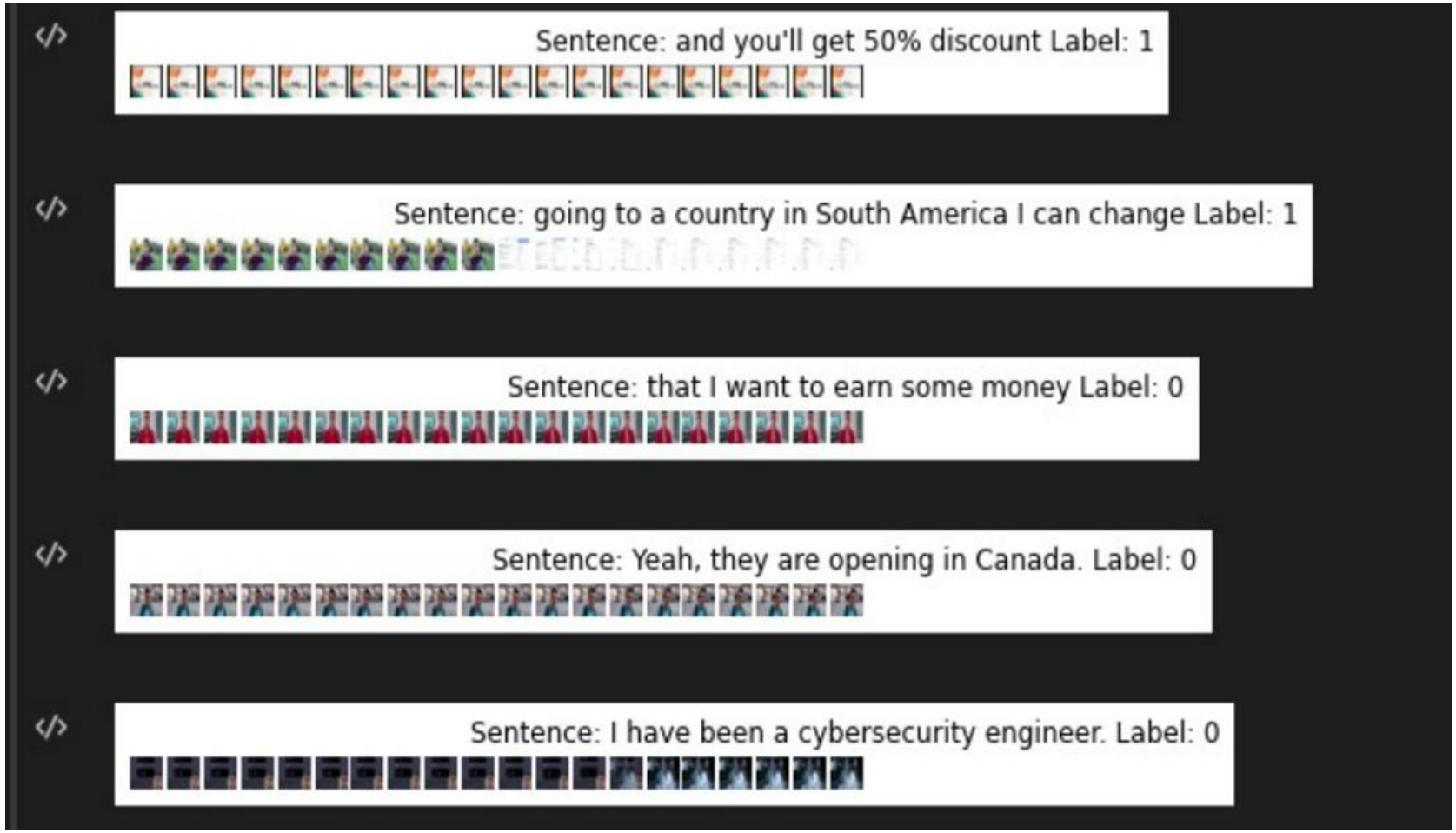
- Python
- Google Colab
- Jupyter Notebook
- Pytube – to download videos
- CV – to process frames

Performance Analysis





Results


Sentence: and you'll get 50% discount Label: 1
Sentence: going to a country in South America I can change Label: 1
Sentence: that I want to earn some money Label: 0
Sentence: Yeah, they are opening in Canada. Label: 0
Sentence: I have been a cybersecurity engineer. Label: 0

Results

```
full_model = tf.keras.Model(inputs=[input, mask, frameInput], outputs=[output_layer])

print(full_model.summary())
```

✓ 0.0s

Output exceeds the size limit. Open the full output data in a text editor

```
Model: "model"
_________________________________________________________________________________________
 Layer (type)                Output Shape              Param #     Connected to
=========================================================================================
 framesInput (InputLayer)    [(None, 20, 64, 64,  0           []
                              3)]

 time_distributed (TimeDistribu  (None, 20, 64, 64,   448        ['framesInput[0][0]']
 ted)                        16)

 time_distributed_1 (TimeDistri  (None, 20, 16, 16,   0          ['time_distributed[0][0]']
 buted)                      16)

 dropout (Dropout)           (None, 20, 16, 16,   0           ['time_distributed_1[0][0]']
                             16)

 time_distributed_2 (TimeDistri  (None, 20, 16, 16,   4640       ['dropout[0][0]']
 buted)                      32)

 time_distributed_3 (TimeDistri  (None, 20, 4, 4, 32  0          ['time_distributed_2[0][0]']
 buted)                      )

 dropout_1 (Dropout)         (None, 20, 4, 4, 32  0           ['time_distributed_3[0][0]']
                             )

...
Trainable params: 38,684,458
Non-trainable params: 124,440,576
```

Results

```
43/43 [==============================] - 98s 2s/step - loss: 0.4359 - accuracy: 0.8747 - val_loss: 0.4503 - val_accuracy: 0.8491
Epoch 13/40
43/43 [==============================] - 97s 2s/step - loss: 0.4279 - accuracy: 0.8747 - val_loss: 0.4461 - val_accuracy: 0.8491
...
Epoch 39/40
43/43 [==============================] - 99s 2s/step - loss: 0.1314 - accuracy: 0.9929 - val_loss: 0.2506 - val_accuracy: 0.8208
Epoch 40/40
43/43 [==============================] - 98s 2s/step - loss: 0.1312 - accuracy: 0.9835 - val_loss: 0.2481 - val_accuracy: 0.8019
```

```
full_model.evaluate([sentence_test_in, sentence_test_mask, framesTest], labelsTest)
```

[52] ✓ 6.7s

```
6/6 [==============================] - 7s 1s/step - loss: 0.2972 - accuracy: 0.8531

[0.2971917986869812, 0.8531073331832886]
```

```
print("original ad starting and ending time for video are " + str( tadStartOriginal)  + " and "+ str(tadEndOriginal))
print("predicted ad starting and ending time for video are "+ adPredStartTime+" and "+adPredEndTime)
```

[139] ✓ 0.0s

```
original ad starting and ending time for video are 5.07 and 5.41
predicted ad starting and ending time for video are 05:05 and 05:44
```

Reference

[1] Soumya Majumdar et al– “Detection of Advertisement Video Shots among Normal Shots Using MFCC Features of Audio” International Journal of Creative Research Thoughts (IJCRT), 2022

[2] Waseemullah, Najeed K, Umair Amin – “Unsupervised ads detection in TV transmission” Article Published in International Journal of Advanced Computer Science and Applications(IJACSA) , 2018

[3] Shervin Minaee et al – “Ad-Net: Audio-Visual Convolutional Neural Network for Advertisement Detection In Videos” Arxiv, 2022
https://doi.org/10.48550/arXiv.1806.08612

[4] Qian Xia et al – “Research on TV Advertisement Detection Base on Video Shot” International Conference on System Science , Engineering Design and Manufacturing Informatization, 2012