Visual to audio aid for visually impaired – Lecture 5

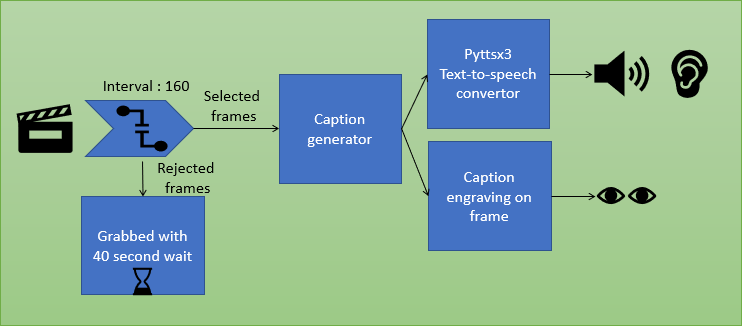
**Matriculation Number: 00503319, 00616219**

# Upcoming Conferences in nearby locations (europe)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Name | Date | Submission | Location | Price | Description |
| COLT Conference | July 9th, 2020 | July 8th | Graz University of Technology, Austria | $30 | A staple in AI events since 1988 |
| International Conference on Machine Learning (ICML) | July 12th, 2020 | July 13th to July 18th | Vienna, Austria | $25 | ICML is one of the leading international machine learning conferences in the world. |
| Intelligent Systems Conference (IntelliSys) | September 3rd, 2020 | 1-Feb-20 | Amsterdam, The Netherlands | 300 GBP | A truly global event featuring speakers and attendees from many countries around the world. |

# Project Progress

The progress till now involves capturing the visual input on specified interval and processing it for caption generation. The generated caption is then converted to speech and also displayed with visual input selection.



# Re-Defining the Image captioning module

The Image Captioning Collaboratory notebook contains raw text and blocks and are not appropriate to be ran as a module for the whole system. The code blocks were converted to functions and classes so that the whole Image Captioning module can be called as a single function call.

Checkpoints were also downloaded along with the pickles. They were extracted and referenced in the module to be loaded on first run to initialize the image captioning model

# Text to speech module

There are a number of text to speech modules publicly available but pyttsx3 is a unique module that gives access to its conversion muddle with offline capability. It is can be set up by installing pyttsx3 and importing the same in the program.

The engine needs to be initialized with various options for the sound. The engine object once initialized can be called to convert a string into speech.

# Initial variables and initialization Function

A suitable video is downloaded which has a frame rate of 25 frames per second. The frame selection interval is set to 160 meaning 6.4 seconds. The Old\_caption and New\_caption variables are declared for use in a global scope.

The Initialization function is written to initialize the image captioning module, it has the below components,

* checkpoint\_check()
* initialize\_inceptionv3()
* initialize\_tokenizer()
* initialize\_hyperparameters()
* define\_model\_components()
* load\_checkpoint()
* eng = pyttsx3.init()

The initialization function initializes the text to speech module pyttsx3 engine as well and references it as global object to be used from a different scope.

# Caption processing module

The caption once generated needs to be converted into speech using the text to speech engine previously initialized. The new caption is also logged into console for debugging purposes.

# Video Capturing mechanism

We have used the popular OpenCV library for working with video. OpenCV gives us the option for capturing video from a live device and also from a video file. For experimental purposes we here are letting openCV to use the video file referenced earlier as its source.

We are choosing every 160th frame to be processed and we are waiting for 20 milliseconds to process the next frame. For every frame that falls under the interval is just grabbed and dismissed with a wait of 40 milliseconds to keep up with human information processing limits.

For every frame we capture for processing the below tasks are carried out,

* we generate the caption based on the image
* we check if the old\_caption is empty in which case it is simply replaced
* the old and new captions are forwarded to the process\_caption function to process the speech output
* the new caption is also engraved on the frame captured and displayed for testing purposes

# Next Step – The Text similarity checking module

The captions generated have to be now compared for semantic similarity so that we do not repeat captions which essentially depict the same visual data.

The process\_caption function has been modified to incorporate a space for calling the trained comparator and accordingly the decision will be made if the audio is converted to speech and old\_caption is updated or not.