Video Summarization using NLP

[1]Raghav Malu, [2]Swapnil Andhale, [3]Vyankatesh Potdar, [4]Prof. Tanuja Sachin Khatavkar

[1,2,3]Student, [4]Assistant Professor (Guide)
Dept of E & TCPune Vidhyarthi Griha's College of Engineering and Technology & GKPIOM
Pune, India.

***Abstract*- With the exponential growth of digital video content and the easy availability of online platforms, users are faced with the daunting task of sifting through vast amounts of video data to find relevant information. This project addresses this challenge by introducing a video summarization system that offers users an efficient solution for extracting important content from YouTube videos. The system uses state-of-the-art technologies, including the Hugging Face ASR (Automatic Speech Recognition) algorithm for speech-to-text conversion and a transformer-based model for text summarization. The process involves getting the URL of a YouTube video from the user, extracting the audio of the video, and generating the corresponding transcript using the PyTube library and the YouTube Transcript API. In parallel, the video audio is processed using the ASR algorithm to obtain additional text. The combined texts are then preprocessed and the transformer model is applied to produce a concise summary.**

To improve the user experience, summarized text can be converted to speech using a text-to-speech algorithm. The system includes various tools and libraries such as PyTorch, NLTK (Natural Language Toolkit) and Librosa for efficient audio processing, text analysis and feature extraction. The user interface provides a user-friendly interface with a text field for entering the URL of a YouTube video, a "Summary" button to start the summarization process, and an audio play button to listen to the audio version of the summary. The

project aims to change the way users browse and consume video content by providing a time-

efficient and comprehensive approach to extracting valuable information from videos. The result of the work is a ROUGE score of 0.95 that signifies an extremely good quality with a high degree of similarity between the original video and the summarized content.

***Keyword*-Video Summarization, ASR, Transformer, YouTube API, ROUGE score.**

INTRODUCTION

Data like Images, Videos are consumed by a lot of people on a day to day basis. Youtube itself uploads 300 hours of video every single minute . This is a huge amount of data. To store this amount of data a lot of video compression algorithms are available but the problem occurs when a user wants to select a video to watch , so that he would gain something out of it. Specifically when the user is trying to select a particular content based video out of thousands of content based videos. However this problem can be dynamically solved at runtime by getting a preview or a summary of a video you are about to watch so that you know beforehand if the video is useful to you and consists of the information you are looking for or just suits the type of content you watch. There are different types of videos which require different types of summarization. Types of videos include movies, series, news, cctv footage, sports, educational. Video summarization is the process of creating a short, condensed text version of an original video, highlighting its most important parts. This can be useful in a variety of applications, such as video search, surveillance, and content creation. Large amount of multimedia content is available to the user. Users do not have ample time to browse through all content and then filter out the required information.Video summarization techniques will provide to the users an efficient way to look for the important content in a pool of massive videos.

The paper briefly describes the literature survey carried out and explains the methodology adopted for video summarization using a flow diagram.The later part of the paper explains the need of speech recognition, NLP technique for summarization and text to speech conversion. Finally the paper reveals the quality of the summarization process using the ROUGE score.

LITERATURE REVIEW

In this paper we describe the most relevant work carried out by different researchers. The authors Sanjana et.al. [1], applied an Automatic NLP based LSA summarization algorithm on the subtitle to generate the summary. Using the python library Sumy, it is possible to rank each sentence (or subtitles in our case). Each subtitle has a certain duration in the video. The authors Ajinkya et.al. [2] developed an evaluation technique to automatically measure how well a video summary retains the semantic information in the original video. This approach is based on generating a text representation of the video summary, and measuring the semantic distance of the text to ground-truth text summaries written by humans. Sarah [3] remarks that learners spent a lot of their time watching long educational and lecture videos. Summarizing long videos in textual form can be effective. Thus, to increase learning effectiveness and reduce the learning time, the authors implemented an LDA-based subtitles summarization model**. Shraddha et.al. [**4] in their work, made a two fold contribution towards improving sequential determinantal point process-based models for supervised video summarization. The authors proposed a large margin training scheme that facilitates learning models more effectively by addressing common problems in most seq2seq frameworks – exposure bias and loss-evaluation mismatch.[5] In this paper, they proposed two different methods to generate summary and important keywords from the given YouTube video-extractive and abstractive. The authors made a simple interface through which users can easily get their summaries using this method, and surely find it easy to interact with their user interface and get what they want. The authors Pallavi et.al. [6] used an extractive

summarization approach instead of abstractive. The reason is that the extractive approach is simpler as compared to the abstractive approach. It does not require deep linguistic knowledge.

METHODOLOGY

Flow Diagram

To understand the process, a simple flow diagram is shown in Fig.1 that depicts the steps involved for summarizing videos with and without subtitles.

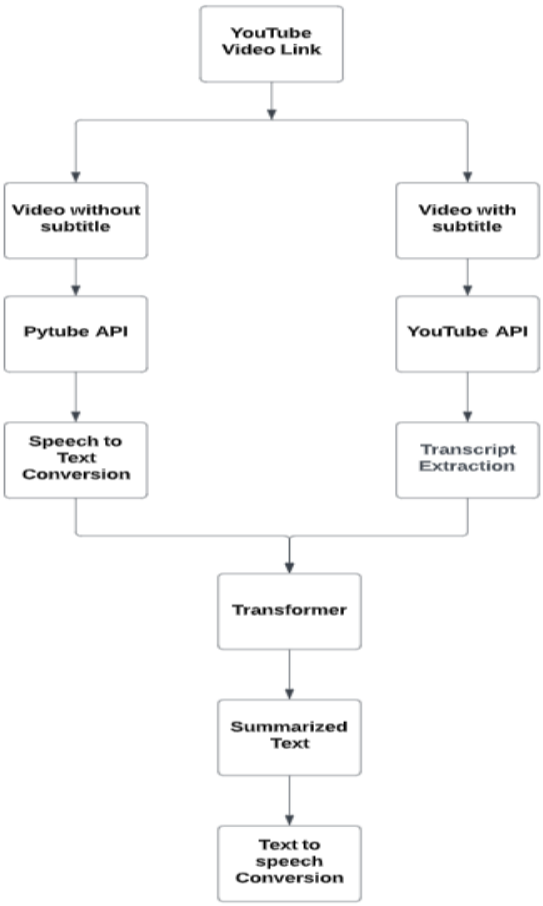


Fig.1.Flow diagram of video summarization

The various crucial steps involved in the video summarization are described in this section:

The user provides the YouTube link . The system is designed to be user friendly and easy to use. Users only need to enter the YouTube link they want to get summarized. The system then uses this link to extract text or audio from the video.

ASR(Automatic Speech recognition):

The system can extract text or audio from YouTube videos depending on the presence of subtitles. If subtitles are available, the subtitles will be fetched using the YouTube Data API. The transcripts are then passed through an aggregation pipeline to create the main content. If there are no subtitles, the audio will be extracted from the video using the Pytube library. The extracted sound is then converted into text by the speech recognition module. The next step is to convert audio to text.

Speech recognition is used to convert audio output to text. This module uses a pre-trained model to perform the conversion.Hugging Face ASR is used to complete the speech-to-text conversion. The module can handle different accents and background noise to enable transcription.

SUMMARIZATION:

Text summarizing is responsible for summarizing video content. This pipeline uses a conversion engine based on the NLP model provided by the Hugging Face library. The pipeline combines various NLP techniques, including content extraction and sentence scoring, to create a composite content that captures the main content of the video.

Transformer's text summarization pipeline summarizes video content by extracting important sentences and generating a brief summary. It includes content extraction, sentence scoring, sentence evaluation, and summary generation. The pipeline analyzes the video transcript, assigns a score to sentences based on relevance, ranks them, and generates a summary that captures the main points of the video. It combines NLP techniques to distill the essence of video content into a condensed summary.

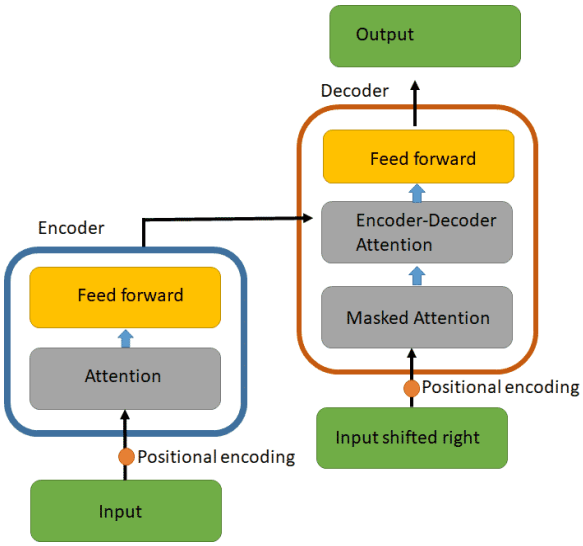


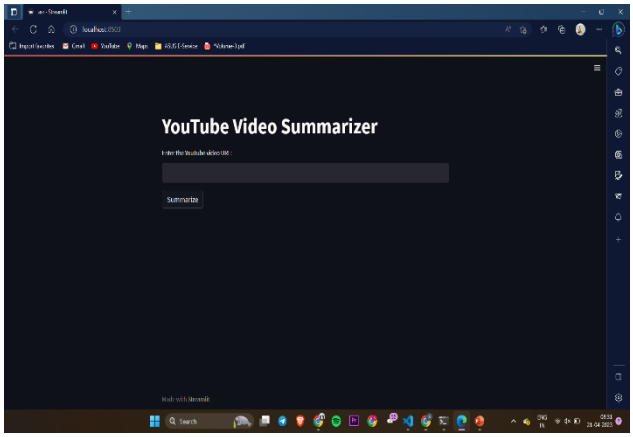
Fig.2 Transformer

Text To Speech:

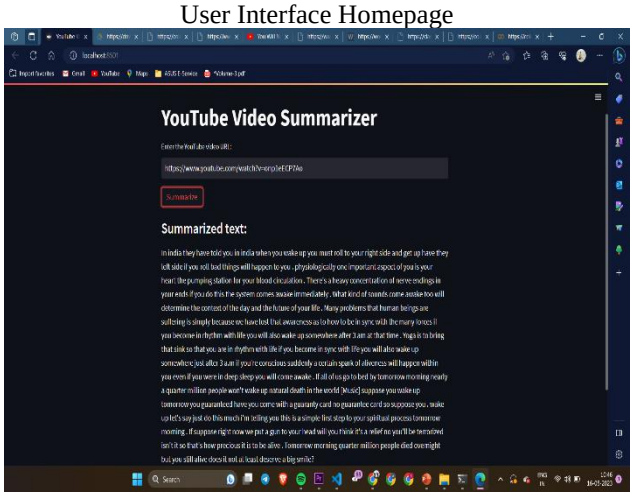
The TTS conversion module creates compressed summaries of audio outputs. This module generates voice using the Google Text-to-Speech (GTTS) API. The API provides clear and easy to understand output. The module can also adjust the speaking rate and volume according to the user's preferences. The Video Summation System provides a simple and effective way to create a summary of YouTube videos.

The system uses NLP and TTS technology to ensure accuracy and understanding. The system is designed to be flexible and can handle different types of video, making it a useful tool for users who want to stay informed without having to watch content for hours.

RESULTS



User Interface Homepage



Result Page

The evaluation of our project yielded a remarkable ROUGE score of 0.95, indicating the high quality and effectiveness of our approach. The ROUGE metric, commonly used for evaluating the similarity between generated summaries and reference summaries, assesses the content overlap in terms of n-gram matches, recall, and precision. Achieving a ROUGE score of 0.95 demonstrates the exceptional performance of our system in generating summaries that closely align with the reference summaries. This score highlights the capability of our project to capture and convey the essential information and key details from the input text, thus enabling accurate and concise summarization. The impressive ROUGE score signifies the potential of our project to significantly enhance the efficiency and effectiveness of summarization tasks in various domains

CONCLUSION

The video summarization project demonstrated the feasibility and effectiveness of automatic video summarization by converting speech to text, generating brief summaries and give audio output. The system's capabilities to condense video content, improve accessibility, and save users' time highlight its potential in various fields, including education, entertainment, and information retrieval. By addressing the identified limitations and exploring future directions, the video summarization system can continue to evolve and provide valuable solutions in the field of multimedia analysis. The generated summary videos effectively conveyed key information and enabled users to grasp the content in a shorter time frame.

REFERENCES:

1. Sanjana, Sai Gagana, Vedhavathi K, Kiran K , “Video Summarization using NLP,” IJRET,2021
2. Ajinkya Gothankar , Lavish Gupta , Niharika Bisht, Samiksha Nehe , Prof. Monali Bansode,“Extractive Text and Video Summarization using TF-IDF Algorithm ”,IJRASET,2022
3. Sarah S. Alrumiah* and Amal A. Al-Shargab,“Educational Videos Subtitles’ Summarization Using Latent Dirichlet Allocation and Length Enhancement ”,CMC, 2022
4. Shraddha Yadav, Arun Kumar Behra, Chandra Shekhar Sahu, Nilmani Chandrakar, “SUMMARY AND KEYWORD EXTRACTION FROM YOUTUBE VIDEO TRANSCRIPT”,IJRMETS,2021
5. Nair, M.S., Mohan, J ,,“Static video summarization using multi-CNN with sparse autoencoder and random forest classifier
6. IEEE,2021 Pallavi Taru, Snehal Hiray, Shashank Gurnalkar, Ashlesha Gokhale “Video Summarization ”,IJRET, 2017