YouTube Video Summarizer in Regional Language

1st Ranjana Jadhav
Department of Multidisciplinary Engineering
Vishwakarma Instituite of Technology,Pune
ranjana.jadhav@vit.edu

line 1: 2nd Prajwal Damre
Department of *Artificial Intelligence and data Science*
Vishwakarma Instituite of Technology,Pune
prajwal.damre21@vit.edu

line 1: 3rd Atharva Hire
Department of *Artificial Intelligence and data Science*
Vishwakarma Instituite of Technology,Pune
atharva.hire21@vit.edu

4thPriti Gosavi
Department of *Artificial Intelligence and data Science*
Vishwakarma Instituite of Technology,Pune
priti.gosavi21@vit.edu

5th Sanskruti Deshmukh
Department of *Artificial Intelligence and data Science*
Vishwakarma Instituite of Technology,Pune
sanskruti.deshmukh21@vit.edu

2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL) | 979-8-3503-9615-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICSADL61749.2024.00055

***Abstract*—The idea and features of the proposed YouTube video summarizer tool created especially for regional languages is presented in this research study. Generally, it is a challenging task for viewers, who prefer regional languages to access and understand videos that are primarily available in a different language in YouTube. The proposed YouTube video summarizer in regional language attempts to overcome this challenge by offering a service to summarize videos in regional languages, making them easier for viewers who might not understand the original language of the video to access and comprehend. The developed program analyses and extracts important information from the video's audio or subtitles using natural language processing (NLP) and machine learning algorithms. This YouTube video summarizer in local languages has a wide range of potential uses. It can improve non-English speakers' access to information, improve video browsing, and promote effective content consumption in localized language groups.**

Keywords: (YouTube video summarizer, Regional language, Video summarization, Natural language processing (NLP), Machine learning)

1.INTRODUCTION

The rapid growth of online video platforms, particularly YouTube, has revolutionized the way people consume and share information. Videos have become an integral part of our daily lives, covering a wide range of topics from education and entertainment to news and documentaries. However, with the overwhelming amount of video content available, it can be challenging for users to find relevant information efficiently. This challenge becomes even more significant for non-English speaking individuals who seek video summaries in their regional languages.

One of the primary challenges in summarizing videos in regional languages is the availability of accurate transcriptions. To address this challenge, the YouTube video summarizer utilizes speech recognition technology to convert the audio content into text. This transcription is then processed and analyzed to generate meaningful summaries. Moreover, automatic translation tools are integrated into the system to ensure that the transcribed text is translated into the target regional language, enhancing accessibility for non-English speaking users.While video summarization techniques have been extensively studied and developed for widely spoken languages like English, there is a noticeable gap when it comes to regional languages. Regional languages play a crucial role in connecting communities and preserving cultural heritage. It is essential to provide video summarization capabilities that cater to the diverse linguistic needs of these communities.

To bridge this gap, we propose a YouTube video summarizer specifically designed for regional languages. The objective of this system is to automatically generate concise and informative video summaries in regional languages, enabling non-English speaking users to access and comprehend video content more effectively.

The proposed YouTube video summarizer employs a combination of advanced technologies to transcribe, analyze, and summarize the audio content of videos. Natural Language Processing (NLP) techniques are utilized to extract and process textual information from the audio track. Machine learning algorithms are employed to identify key sentences and important segments that encapsulate the main ideas and content of the video.

The process of generating video summaries involves considering various factors such as relevance, saliency, and coherence. The system evaluates the importance and relevance of each segment within the video to select key sentences that effectively capture the essence of the content. The selected key sentences are then logically organized and condensed into a coherent summary that provides a comprehensive overview of the video.

II. RELATED WORKS

[1] introduces an unsupervised machine-learning approach for extractive Punjabi text summarization, addressing the gap in summarization techniques for regional languages. The proposed approach utilizes linguistic features and a novel sentence-scoring mechanism to automatically select important sentences from Punjabi text documents, generating concise and accurate summaries.
In paper [2] paper presents an extractive text summarization approach that utilizes sentence ranking techniques. The proposed method identifies and selects key sentences from a

979-8-3503-9615-7/24/$31.00 ©2024 IEEE
DOI 10.1109/ICSADL61749.2024.00055

Authorized licensed use limited to: ANURAG GROUP OF INSTITUTIONS. Downloaded on December 11,2024 at 12:18:25 UTC from IEEE Xplore. Restrictions apply.

text document based on their relevance and importance. By employing advanced ranking algorithms, the system effectively generates summaries that capture the essence of the original content. In paper [3] introduces a context-based text summarization system that leverages contextual information to generate comprehensive summaries. The proposed system incorporates advanced natural language processing techniques to analyze the context and relevance of the text, allowing for a more nuanced and accurate extraction of key information. In paper [4] presents an audio summarization approach specifically designed for podcasts. With the increasing popularity of podcasts, there is a need for efficient methods to summarize audio content and provide concise overviews. The proposed system utilizes advanced techniques such as speech recognition, speaker diarylation, and content analysis to extract key segments and generate summaries. In paper [5] The paper presents a news article summarization approach utilizing Transformers, a type of deep learning model known for its strong natural language processing capabilities. The proposed method leverages the power of Transformers to understand the contextual relationships and semantic structure of news articles, enabling the generation of concise and informative summaries. In paper [7] his research paper introduces a deep reinforcement learning framework for video summarization, incorporating semantic reward mechanisms. The proposed approach leverages the power of deep neural networks to learn policies for selecting key frames or segments from videos, while also considering the semantic relevance of the selected content. By using reinforcement learning techniques, the system is trained to optimize summary generation by maximizing the semantic rewards obtained from expert annotations or predefined semantic metrics. [8] presents an automatic video summarization approach that incorporates natural language processing and text fusion techniques to generate concise summaries with timestamps. The proposed method utilizes natural language processing to extract textual information from video transcripts and fuses it with video content analysis. By considering both visual and textual cues, the system identifies key moments in the video and associates them with relevant timestamps. [9] presents an automatic video summarization approach that incorporates natural language processing and text fusion techniques to generate concise summaries with timestamps. The proposed method utilizes natural language processing to extract textual information from video transcripts and fuses it with video content analysis. By considering both visual and textual cues, the system identifies key moments in the video and associates them with relevant timestamps. In paper [10] presents a comparative analysis of text-based video summarization techniques using deep learning approaches. The study focuses on evaluating and comparing the performance of different deep learning models for generating video summaries based on text information. The proposed techniques leverage the power of deep neural networks to process textual data, extract relevant information, and generate concise and informative video summaries.

III. METHODOLOGY

A. Collecting Transcript from the YouTube Video:

Enable automatic captions: YouTube offers an automatic captioning feature that uses speech recognition technology to create captions for videos. However, it is important to note that the accuracy of automated subjects can and still varies depending on factors such as audio quality, background noise, and speaker volume clarity can be a good starting point for essays.

Language support: YouTube's automatic captioning feature supports multiple languages. Make sure the video you want to summarize has subtitles in your preferred language. If automatic captions are not available in your preferred language, you may need to consider external captioning services or consider manually captioning the video.

Getting a transcript: If the video has automatic captions available, you can access it directly on the YouTube video page. Locate the "CC" (Closed Captions) button at the bottom of the video player and click on it to see available caption options. Select an automatic title, and the text will be displayed in the synchronous video playback.

Transcript copying: You can manually copy notes from a YouTube video page when the title automatically appears to hide the transcript. Select and copy the text in the header, paste it into a text editor or document, and save it for further processing.

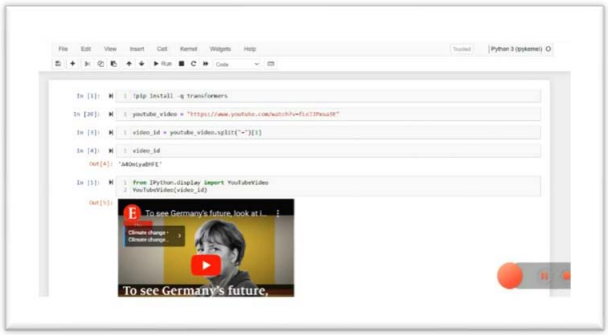

Fig1.Downloading YouTube Videos

TEXT SUMMARIZATION TECHNIQUE:

Install Hugging Face Transformers: The Hugging Face Transformer library provides pre-trained transformer examples for natural language processing tasks. You can use pip or to install the library in your Python environment.

Preprocess copy: Clean and pre-copy by removing unnecessary characters, symbols, and any unnecessary text, similar to pre-processing step.

Select pre-trained model: Face Turners Hugs Provides edited modifier models previously trained it appropriate for text summarizing Functions. Choose the model that suits your needs and resources. Popular examples of text abbreviations include BART, T5, and GPT-2.

Authorized licensed use limited to: ANURAG GROUP OF INSTITUTIONS. Downloaded on December 11,2024 at 12:18:25 UTC from IEEE Xplore. Restrictions apply.

Tokenization: Tokenize pre-processed copy into an appropriate format for insertion into a transformer model Hugging Face Transformers provides built-in tokenizers that handle this step successfully. These tokenizers divide the text into tokens, which are the basic units handled by the transformer model.

Data Collection & Cleaning: Gather YouTube video transcripts in the regional language and pre-process the text to remove noise.

Fine-tuning BART: Adapt a pre-trained BART model to understand the language specifics by training it with the pre-processed data.

Tokenization & Input Prep: Break down the cleaned text into tokens suitable for the fine-tuned BART model.

Summarization Execution: Use the model to create concise summaries for the regional language YouTube video content.

Evaluation & Refinement: Assess summary quality using metrics, refine the model based on feedback to improve accuracy.

Integration & Deployment: Implement the refined BART model into the YouTube video summarizer for generating accurate regional language video summaries.
d

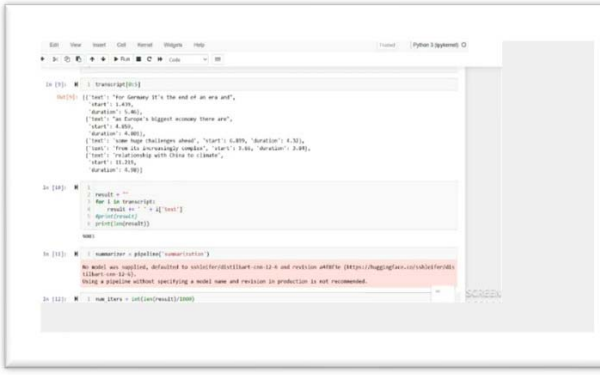

Fig2. Text Summarization

Generate summary: Assign the tokenized transcript to the selected transformer model and generate the summary. To summarize the selection, you can use pre-trained models such as sentence scoring algorithms or BART to select the most important sentences from the text. For abstract summarizing, you can use models like T5 or GPT-2 to create additional sentences that capture the essence of the video.
Understanding and summarising video content requires tokenization of pre-processed text in a local language. In order to do this, the text extracted from video transcripts or descriptions is broken down into smaller units, such as words, subwords, or characters, depending on the language's linguistic structure, using language-specific tokenizers or libraries. These tokens are the building blocks of analysis, allowing for the completion of tasks such as phrase recognition, information extraction, and the creation of succinct summaries that encapsulate the main points of the video content. To summarise videos more accurately and coherently, the summarisation algorithm can process and interpret regional language more effectively by tokenising the text.

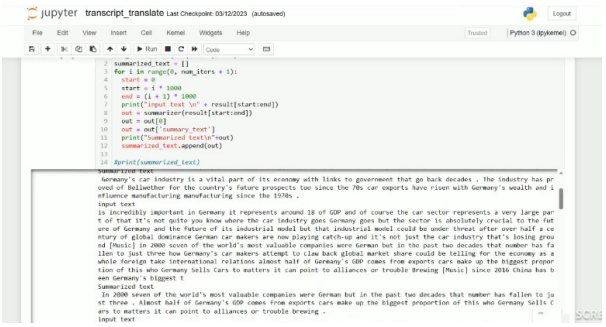

Fig.3 Generating Summary

Translate the summary in regional languages:

The code starts by running google trans library version 4.0.0rc1 using the pip package manager. This library acts as a Python wrapper for the Google Translate API, enabling language translation capabilities in code. Once the installation is complete, the required modules are imported, including the Google trans and Translator classes. To display the supported languages and their corresponding language codes, the code publishes the Google trans. LANGUAGES dictionary, which contains the language code-name mapping. This gives us a list of all the languages that can be translated. Next, an instance of the Translator class starts by passing Translator() to the translator variable. This step prepares the translation function.

Proceeding, the code recognizes the language of the summarized text variable. It converts the summarized text to a string format and then uses the detect() method of the translator object to determine the language. The known language code is stored in the Lang variable. This step is important because it helps identify the source language of the translation. When the language is found, the code uses the print() function to print the language code. This allows the source language of the translation to be identified. The code then prompts the user to enter the language code to which the text is to be translated. This input is stored in the Lang code variable, allowing users to specify the target language. Finally, the code translates using the translator object's translate() method.
The code extracts text in regional languages from YouTube video transcripts and converts it into a desired language using the translate() method from translator objects. Using language-specific models or libraries built into the translator object, this method takes the preprocessed regional language text as input and translates it. This method allows for multilingual processing of video summaries and smoothly translates regional language content into the desired language by calling it with the source text and the target language. This increases accessibility and reach for a variety of audiences.

Authorized licensed use limited to: ANURAG GROUP OF INSTITUTIONS. Downloaded on December 11,2024 at 12:18:25 UTC from IEEE Xplore. Restrictions apply.

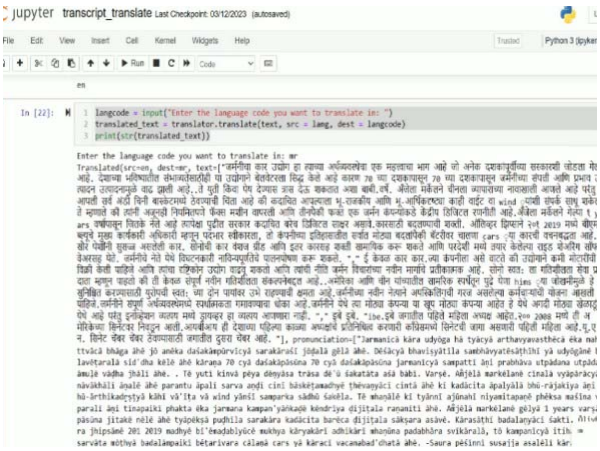

Fig.4 Video Synopsis in Vernacular(here it is in marathi)

The summarizer makes use of the refined BART model, which uses its sequence-to-sequence architecture to produce output after tokenizing and processing the regional language text taken from YouTube video transcripts. Because of this architecture, the model is able to extract important information from the video and understand its core ideas, creating succinct and logical summaries. The summarizer creates output summaries that capture the main ideas of the video by decoding the processed sequences using the language understanding and generation capabilities of the BART model. This makes it easier for viewers to understand and retrieve information in the local language.

IV. Results

In fig.1, The first step involves downloading YouTube videos that are in the desired regional language. This can be done using various tools or APIs available for extracting video content.
In fig.2, Once the videos are downloaded, the next step is to perform text summarization on the video transcripts. Text summarization techniques, such as extractive or abstractive summarization, can be employed to identify important sentences or generate concise summaries.
In fig.3, The extracted or generated summaries are based on the content of the video transcripts. The goal is to capture the key points and essential information in a condensed form, representing the essence of the original video.
As you can see in fig.4, Finally, the generated summaries are translated or converted into the desired regional language. This step ensures that the summarizer produces output in a language that is accessible and understandable to the target audience.

The results of a YouTube video summarizer in a regional language can vary depending on factors such as the quality of video transcripts, the accuracy of the text summarization techniques employed, and the effectiveness of the translation process. Evaluation metrics, such as ROUGE scores (Recall-Oriented Understudy for Gisting Evaluation), can be used to assess the quality of the generated summaries in terms of their similarity to human-authored summaries or reference summaries.

Metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation) are frequently employed when discussing accuracy scores for summarization tasks. For a summarization system, a good ROUGE-1 score (which measures the overlap of unigram tokens) might be in the range of 0.3 to 0.5, while a good ROUGE-2 score (which measures bigram overlap) might be in the range of 0.1 to 0.3.

However, the precise accuracy figures attained in a project can differ significantly depending on a number of variables, including the degree of regional language complexity, the variety of video content, the calibre of the training data, and the summarization model's tuning and optimisation. For regional language YouTube videos, for example, attaining ROUGE-1 scores above 0.4 or ROUGE-2 scores above 0.2 may indicate a highly accurate summarization system, but these

V. Conclusion

The YouTube Video Summarizer project stands as a significant tool, streamlining video content and conserving users' time and energy. By being available in a regional language, it garners more accessibility and reach across various linguistic demographics. This project demonstrates the ability to craft a concise, elegant summary that encapsulates the crucial facets of a video. It offers individuals the convenience of accessing their preferred videos in a condensed format, empowering them to grasp the essence of the content without investing extensive time in watching lengthy videos. Ultimately, it facilitates a user-friendly and efficient way to engage with diverse video content, enhancing accessibility and comprehension for a broader audience.

VI. Future Scope

"The future of YouTube Video Summarizer in regional languages holds immense promise. Advancements in AI algorithms are set to refine the summarization process, ensuring more precise and contextually relevant video summaries.Moreover, the potential inclusion of audio and video summaries alongside text summaries will offer a more comprehensive and multi-dimensional overview of content. Real-time summarization capabilities will enable quick insights into live events and ongoing discussions, catering to the dynamic nature of information dissemination.Collaborations with content creators will add value by integrating the summarization tool within their platforms, enhancing both content reach and user experience. Integration with multimedia platforms, coupled with language translation support, will further extend the tool's global utility.

Acknowledgment

We express our sincere appreciation to VIT Pune and Professor Ranjana Jadhav for their unwavering motivation and encouragement that inspired us to embark on this project. Their belief in our abilities and constant support have been

Authorized licensed use limited to: ANURAG GROUP OF INSTITUTIONS. Downloaded on December 11,2024 at 12:18:25 UTC from IEEE Xplore. Restrictions apply.

pivotal in our journey. We are truly grateful for their guidance, which has been a driving force in our project's success.

REFERENCES

[1] Kamal Deep Garg; Vikas Khullar; Ambuj Kumar Agarwal," ,Unsupervised Machine Learning Approach for Extractive Punjabi Text Summarization , September 2020

[2] J.N. Madhuri; R. Ganesh Kumar ,"Extractive Text Summarization Using Sentence Ranking", August 2019

[3] Rafael Ferreira; Frederico Freitas; Luciano de Souza Cabral; Rafael Dueire Lins; Rinaldo Lima; Gabriel França," , A Context Based Text Summarization System, June 2014

[4] Aneesh Vartakavi; Amanmeet Garg;Zafar Rafii, Audio Summarization for Podcasts, December 2021

[5] Rajesh Kumar Yadav; Rohit Bharti; Ritika Nagar; Sanchit Kumar, A Model For Recapitulating Audio Messages Using Machine Learning, August 2020

[6] Harivignesh S.; Avinash S.; Avinash V.; R. Kingsy Grace, Summarization of News Articles Using Transformers, February 2023

[7] Sakdipat Ontoum; Jonathan H. Chan, Automatic Text Summarization of COVID-19 Scientific Research Topics Using Pre-trained Models from Hugging Face, October 2022

[8] Haoran Sun; Xiaolong Zhu; Conghua Zhou, Deep Reinforcement Learning for Video Summarization with Semantic Reward, March 2023

[9] Ahmed Emad; Fady Bassel; Mark Refaat; Mohamed Abdelhamed; Nada Shorim; Ashraf AbdelRaouf," Automatic Video summarization with Timestamps using natural language processing text fusion, March 2021

[10] Rakhi Akhare; Subhash Shinde, Comparative Analysis of Text-based Video Summarization Techniques using Deep Learning, October 2022

Authorized licensed use limited to: ANURAG GROUP OF INSTITUTIONS. Downloaded on December 11,2024 at 12:18:25 UTC from IEEE Xplore. Restrictions apply.