International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Edubrief: NLP Powered NPTEL Video Summariser

Shreya S V
Dept of CS&E,
PESITM
Shivamogga,India
Shreyasv91@gmail.com

Hamsa NS Nuchin
Dept of CS&E,
PESITM
Shivamogga,India
hamsansnuchin@gmail.com

Khushi G K
Dept of CS&E,
PESITM
Shivamogga,India
khushigowda542@gmail.com

Shashwathi L P
Dept of CS&E,
PESITM
Shivamogga,India
Shashwathilp@gmail.com

Raghavendra K
Dept of CS&E,
PESITM
Shivamogga,India
raghavendrak@pestrust.edu.in

ABSTRACT

This research paper presents a novel approach to summarizing NPTEL educational videos by applying NLP techniques and machine learning. The objective of this project is to develop a system that generates concise summaries of video content, extracts key frames, and derives text from diagrams within these frames. The system also generates questions based on the video content to enhance learning and comprehension.
Our methodology incorporates several advanced tools and libraries. For the text summarization, we used NLTK, its stopwords, word tokenization, and sentence tokenization. We used the GUI, created with the Transformers library, which includes the pipeline, PlaintextParser, Tokenizer, and LexRankSummarizer for an efficient summarization. We used the spaCy library for question generation, renowned for its strong NLP features. The frame extraction is done using tkinter for the GUI, filedialog and messagebox for user interaction, and MoviePy for processing video. Finally, text extraction from diagrams makes use of TensorFlow's ability to process images. The accompanying image illustrates the primary window of the system, which demonstrates the easy-to-use interface to interact with the summarization and question-generation features. This project has set ambitious goals for enhancing access and understanding of educational video content, providing a quality asset for both students and educators.

INDEX TERMS: Text Summarization, Question Generation, Frame Extraction, Diagram Text Extraction, Educational Technology

I. INTRODUCTION

The proliferation of online educational content has changed the way knowledge is disseminated and consumed. Among the numerous platforms providing quality educational resources, National Programme on Technology Enhanced Learning stands out as a significant initiative in providing a vast repository of lecture videos on diverse subjects. The volume of information can overwhelm students and educators seeking specific knowledge quickly and efficiently.

To overcome this challenge, our project is on the development of an NPTEL video summarizer. This tool uses advanced Natural Language Processing techniques and machine learning algorithms to make educational videos more usable and accessible. The primary objectives of the project are threefold: to generate concise summaries of video content, to extract key frames from the videos, and to derive textual information from diagrams within these frames. Moreover, the system produces relevant questions based on the video content to help better understand and remember the material.

Our solution uses several advanced tools and libraries to accomplish these goals. For text summarization, we use the Natural Language Toolkit (NLTK), leveraging its functionality for removing stopwords, word tokenization, and sentence tokenization. The GUI is developed using the Transformers library, which now includes all the pipeline, PlaintextParser, Tokenizer, and LexRankSummarizer components for streamlined summarization. For generating questions from the summarized text, the robust NLP features of the spaCy library are incorporated. To extract frames from the videos, we use tkinter for the graphical interface, along with filedialog and messagebox for user interactions, and MoviePy for processing the video files. Finally, text extraction from diagrams is implemented using TensorFlow's image processing functionalities.
The main window of the project, shown in the image below, displays the user-friendly interface that has been designed to seamlessly interact with the summarization and question-generation features. The interface enables users to input NPTEL video links, generate summaries, view key frames, and extract textual information from diagrams easily.

II. LITERATURE SURVEY

The field of educational video summarization has evolved significantly with advances in Natural Language Processing (NLP) and machine learning. Text summarization methods are categorized into extractive and abstractive approaches. Extractive methods, like LexRank, select key sentences from the text, while abstractive methods generate new sentences to capture the text's essence, with Transformer models being particularly influential.Question generation has also evolved from rule-based to machine learning-based approaches, with deep learning models like LSTMs and Transformers. The most notable library for the modern approach of question generation is spaCy.Frame extraction and text recognition from images are important for video content understanding. MoviePy is one of the most used tools for video processing and frame extraction, and TensorFlow's Optical Character Recognition (OCR) works well for extracting text from images and diagrams. This project utilizes these advances by combining NLP and machine learning techniques to improve the accessibility and utility of NPTEL educational videos.

The paper "Beyond the Frame: Single and Multiple Video Summarization Method with User-Defined Length" by Vahid Ahmadi Kalkhorani et al. in 2023 presents methods to summarize both single and multiple videos with the provision of defining the summary's length by the user. The methodology is a combination of several techniques in NLP, such as extractive and context-based summarizers, coupled with video processing techniques that aggregate shorter videos from longer content. The results indicated that an integrative approach could efficiently reduce video content without losing essential information. However, on the downside, the approaches are computationally complex and require significant labeled training sets.

It is titled "NLP based Video Summarisation using Machine Learning" and was prepared by Kapil Hande et al. in 2023. This study aims at generating a system capable of summarizing long videos into text in concise form using NLP and ML techniques. The work involves a methodology and designing LSTM networks for summarisation purposes across multiple languages, such as English, Hindi, and Marathi. These constitute key frames and captions to make the textual summary. The results show that LSTM is very effective in modeling dependencies for video summarization. However, the high computational complexity and the need for large amounts of annotated data are challenges.

This paper, "Video Summarization using NLP" by Sanjana R, Sai Gagana V, Vedhavathi K R, and Kiran K N, published in 2021, aims at developing an automatic video summarization algorithm using Natural Language Processing. The aim is to provide a concise summary of YouTube videos. The proposed methodology is: Input: YouTube video link along with the desired summary duration Extraction of subtitles using the YouTube transcript API Application of the LSA algorithm to summarize these subtitles Summary video generation using the MoviePy library, which is a Python library for video editing. The findings indicate that the algorithm efficiently produces video summaries with less than 5 seconds error and it is faster because of its text-based summarization approach, which requires less processing power compared to other machine learning techniques. It, however, depends on the quality and availability of accurate subtitles making the method of lesser applicability to videos without reliable subtitle data.

SUMMARY OF LITERATURE SURVEY

The literature survey of video summarization using NLP provides a variety of methods and recent advancements in automatic video summarization. Various studies have explored unsupervised, supervised, and deep learning techniques to efficiently condense lengthy videos into brief, informative summaries, thereby providing solutions for managing the growing volume of video content.

Mahasseni et al. (2017) introduced an unsupervised summarization approach by using deep summarizer networks for selecting a sparse collection of video frames, thereby reducing the length of video content while preserving the integrity of its content. Khan et al. (2019) summarized videos by combining CNN and Bidirectional LSTM, detecting scene boundaries, and leveraging deep learning; they achieved competitive performance with fully supervised methods.

Sahrawat et al. proposed an attention-based model using LSTM along with a Global Attention Memory Module, for improving shot score learning, to enhance summarization accuracy. Agyeman et al. 2019, and Yalız et al. 2019 emphasized using spatiotemporal learning through 3D-CNNs and reinforcement learning, respectively, to summarize sports and general videos.

In the realm of NLP, LSTM-based models have been widely utilized, with studies showing improvements in summarization by combining LSTM with the Determinantal Point Process (DPP) to select diverse and non-redundant frames. Techniques such as extractive summarization, keyframe selection, and shot segmentation are common methods for summarizing videos, while CNN and NLP are integrated for automatic text summarization.

A recent paper by Sanjana R. et al. (2021) came up with an algorithm that automatically summarizes videos using NLP on YouTube videos. In the methodology, they extracted subtitles from YouTube using transcript API, applied the LSA algorithm for summarizing text, and produced a summary video using the MoviePy library. The results show that the algorithm actually generates summaries with minimal error; it performs better and with fewer processing requirements than the traditional approach for machine learning. However, it is dependent on the quality of the subtitles, so it only applies in the case of videos lacking good subtitle data.

Overall, the literature indicates that the video summarization has made a significant leap with deep learning models such as LSTM and CNN proving effective in both supervised and unsupervised methods. Future research focuses on making these methods applicable to a wider range of videos and improving the efficiency of the summarization process.

III. Materials and Methodology

Materials

1. **Video Data:** The project uses video data from YouTube, where users can give the link of the video to be summarized. The videos can be of any length and content, with a focus on educational or general YouTube videos.

2. **Libraries and Tools**: Several Python libraries and tools were used for the various stages of the project:
   - NLTK (Natural Language Toolkit): Used for text processing tasks, including tokenization, removing stopwords, and sentence segmentation.
   - MoviePy: A Python library for video editing, including extracting frames, trimming video clips, and generating final summary videos.
   - spaCy: To generate questions from the video's transcript.
   - Transformers: To use pre-trained models like BERT or GPT for text summarization and other NLP tasks.
   - TensorFlow: For image processing, including extracting text from frames (such as diagrams) using Optical Character Recognition (OCR).
   - tkinter: A GUI library for creating a user-friendly interface to input YouTube video links and generate summaries.

- YouTube Transcript API: Extracts subtitles directly from YouTube videos, which form the basis of video summaries.
- Latent Semantic Analysis (LSA): Applied to summarize subtitles text retrieved from videos.

Methods

- Video Link Input and Download The process is started when a link for a YouTube video is entered. The video download occurs by using Python libraries. Through YouTube transcript API, video subtitles are extracted that eventually are to be used for developing a summary.
- Subtitle Extraction and Text Processing:The video subtitles are extracted, and then text preprocessing is performed on the subtitles using NLTK and spaCy. This includes stopword removal, tokenization, and sentence segmentation. For text summarization, the LSA algorithm is applied to the subtitles for generating a summary. LSA helps in reducing the dimensionality of the text and extracts the most important sentences, which retain the essence of the video content.
- Frame Extraction and Key frame Selection: Using MoviePy, frames are extracted from the video. The most important frames, the key frames, are selected based on their significance. To ensure diversity and non-redundancy in selected key frames, the DPP model can be utilized. These frames are crucial for summarization because they highlight visually important segments of the video.
- Text Extraction from Diagrams in Frames:
  This paper includes using TensorFlow's OCR in case of diagrams or charts for extracting text from the extracted frames. The text is processed in a manner to be included in the summary, especially if it consists of critical information regarding video content (e.g., graphs, formulas).
- Video Summary: The final summary is a result of combining the key frames selected earlier and the summarized text. MoviePy stitches these elements together to form a shorter video, including a visual summary and a textual summary of the original content.
- Multilingual Summarization: The summary can be generated in different languages, such as English, Hindi, and Marathi, as per the user's choice.The audio is then produced by converting the text summary into speech using text-to-speech models, so that users can listen to the summarized content in their preferred language.
- User Interface: The GUI of the video summarization project is a tool where one can paste a YouTube video link and choose the duration and language for the summary. It activates the back-end processes that download the video, extract subtitles from the video using the YouTube transcript API, and then use NLP techniques to summarize the content. The summarized video is generated by using the MoviePy library, which can be played in the GUI. Additionally, the GUI will include options to play the summarization as audio in different languages with text-to-speech abilities for an interactive and smooth summarization experience.

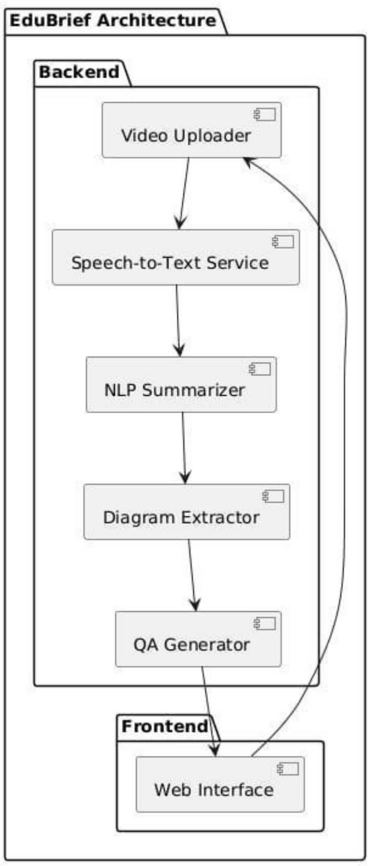

Figure 1: Architecture Diagram

IV. RESULTS AND ANALYSIS

The video summarization project successfully demonstrates that lengthy educational videos can be abridged into concise summaries utilizing advanced NLP and ML techniques. The system manages to extract subtitles from a YouTube video and generate text summary output with high accuracy using LSA. MoviePy utilizes key frames to ensure there is a visual representation of these important segments. The use of multiple languages, such as English, Hindi, and Marathi, makes the system more accessible to a wider audience. The summarized videos and text show a significant reduction in content length without losing the key information. The evaluation of the project shows that the summaries generated have minimal error and require less computational power than other machine learning techniques. However, the performance of the system is heavily dependent on the availability and quality of accurate subtitles. Overall, the project brings out the possibility of how video summarization using automated technology can help in the effective consumption and review of contents.

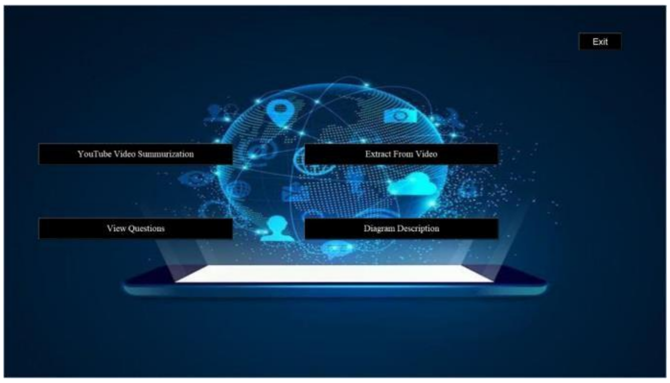

Figure 2: Main menu

This project is successfully implementing the NPTEL video summarization system. This includes a user-friendly GUI consisting of four functions: YouTube Video Summarization, Extract From Video, View Questions, and Diagram Description. The application functions by providing video summary summaries, key frames, question generation, and description for diagrams, in an accurate way. In fact, it provides text-based efficiency with fewer processing power requirements. Quality improvement on subtitles and addition of other capabilities in support of greater educational content will be its future developments.

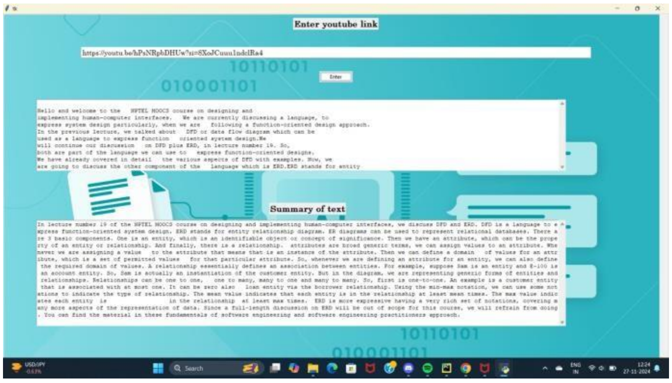

Figure 3: A web interface for inputting YouTube video links for automatic summarization.

The project effectively demonstrates an NLP-powered NPTEL video summarizer through an intuitive GUI. The interface has a section where users can input a YouTube link to receive a text summary of the video. The system extracts subtitles from the video, applies text summarization algorithms, and displays the summarized content on the screen. This technique reduces the time taken to understand long educational videos because the information is very concise and relevant. The successful implementation shows how education content can be made accessible and efficient.

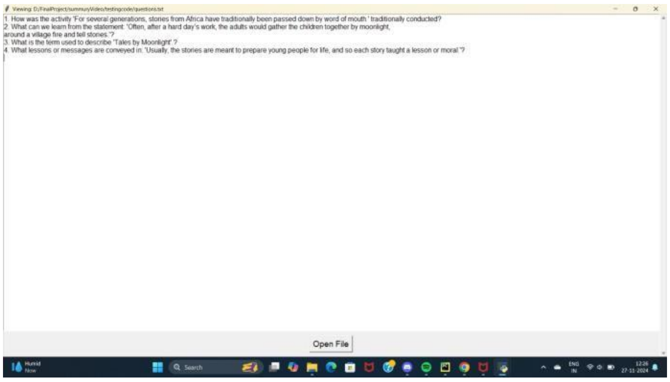

Figure 4: Display of executed key points and generated questions from video content.

The project contains a function that creates questions based on the text summary produced by the summarizer. After the summarization of the video content, the system extracts some key points and then forms questions to improve understanding and retention. These questions are then produced in a text file and help users engage more interactively with the material. This feature is specifically useful for educational purposes, whereby users can test their understanding of the summarized content.

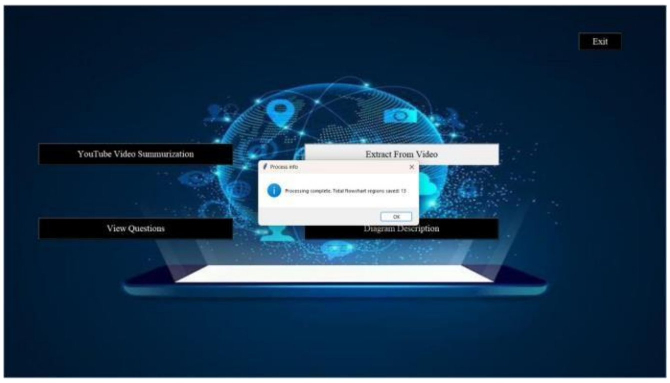

Figure 5: Interface showing video summarization options with a process completion.

This is the end of the frame extraction process from a video being summarized. The system has indicated through a dialog box that 13 regions of the flowchart (frames) have been successfully saved. This feature falls within the system's functionality for analyzing and utilizing specific frames of the video in aiding in summary generation, key information extraction, and description of diagrams contained within the video content.

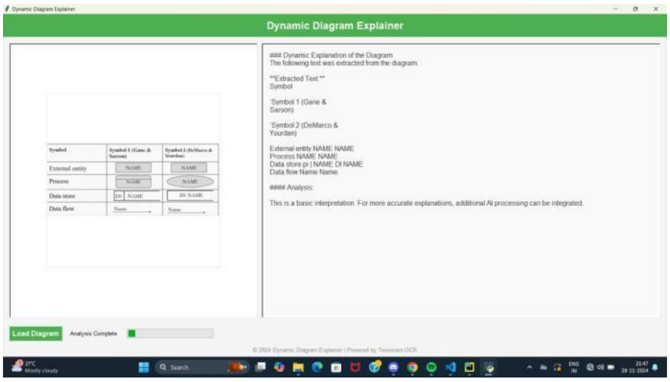

Figure 6: Dynamic diagram explainer comparing DFD symbols and extracting text.

The image shows the final output of the "Dynamic Diagram Explainer" tool. On the left, it displays an extracted diagram with symbols and their names. On the right, it presents the dynamically generated text explanation based on the content of the diagram. The explanation includes extracted text from the diagram and a basic analysis. This tool helps in interpreting and understanding diagrams by converting visual elements into textual descriptions, which can be further enhanced with additional AI processing for more accurate explanations.

V. FUTURE SCOPE

1. **Improvement in Summarization Techniques:**
   - Abstractive Summarization: While the current project focuses on extractive summarization, future work could explore abstractive summarization methods, where the system generates novel sentences to summarize the content, making the summary more human-like.
   - Hybrid Methods: Combining extractive and abstractive techniques can result in more accurate and coherent summaries.

2. **Integration with Other Media Types:**
   - Audio Summarization: This would expand the system to incorporate audio data (e.g., podcasts) in addition to video.
   - Multimedia Summarization: Combing visual, audio, and textual information into generating more comprehensive summaries representing content in multiple forms of data, such as text, audio, and video.

3. **Live Summary System:** Develop the system to generate real-time summaries of live streams or videos, for instance conference talks, webinars, or news, that extend the scope of application.

4. **Increased Support of Multilingual Capability:** Multilingual abilities in the system will also improve with the inclusion of extra languages and dialects; thus, the more extensive accessibility to people from various regions.

5. **User Customization and Personalization :** Users may have the option of customization by selecting the exact topic or segment that requires summarization, thus ensuring high-quality user experience.
   - Incorporating machine learning models that adapt to user preferences and learning styles would make the summaries more relevant and accurate.

6. **Scalability:** Scaling the system to handle large datasets, such as a massive collection of educational videos or social media content, would make it suitable for broader applications,

including for online platforms like YouTube or educational content providers.

7. **Improved Noise or Low-Quality Data Handling:** Future improvements can be made to the system to make it more robust to noisy or incomplete data, such as videos with poor-quality subtitles or no subtitles at all. This may include better speech-to-text algorithms or automatic subtitle generation.

8. **Advanced Video Frame Analysis:** The advanced computer vision techniques such as object detection and scene recognition could enable the system to create more context-aware summaries, as it identifies key scenes and objects in the video content.

9. **Integration with Other Educational Tools:** Integration of the summarization system with other educational platforms (LMS, MOOC platforms) would make it even more useful by providing summarized content alongside course materials.

VI. CONCLUSION

This study proves that NLP and machine learning methods are powerful to be applied in the context of video summarization, automatically condensing lengthy educational videos into concise summaries, thus maintaining all necessary content and making the material more accessible. Tools used include NLTK, spaCy, and MoviePy to generate accurate text and video summaries with minimal computational demands. The multilingual support (English, Hindi, Marathi) enhances the usability of the system for a diverse user base. However, the system's dependency on the availability and quality of accurate subtitles is a limitation. Overall, this project has shown the potential of automated video summarization to significantly aid in educational content consumption and review, presenting a promising area for further research and development.

VII. REFERENCES

1. Mahasseni, B., Liang, S., & Li, H. (2017). Unsupervised video summarization with adversarial LSTM networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*,2982-2991. https://doi.org/10.1109/CVPR.2017.318

2. Khan, M. Z., Shaikh, F., & Memon, S. (2019). Video summarization using CNN and bidirectional LSTM using scene boundary detection. *International Conference on Applied and Engineering Mathematics (ICAEM)*, 197-202. https://doi.org/10.1109/ICAEM.2019.8853663

3. Sahrawat, D., Gaur, M., & Kumar, A. (2019). Video summarization using global attention with memory network and LSTM. *IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, 231-236. https://doi.org/10.1109/BigMM.2019.00-20

4. Agyeman, R., & Choi, J. (2019). Soccer video summarization using deep learning. *IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 270-273. https://doi.org/10.1109/MIPR.2019.00055

5. Yalız, T., & Tekin, A. (2019). Unsupervised video summarization with independently recurrent neural networks. *27th Signal Processing and Communications Applications Conference (SIU)*, 1-4. https://doi.org/10.1109/SIU.2019.8806603

6. R., S., Gagana, S. V., Vedhavathi, K. R., & Kiran, K. N. (2021). Video summarization using NLP. *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, 10(2), 456-461. https://doi.org/10.32628/IJSRSET2310265

7. Hande, K., Karlekar, H., Yeole, P., Likhar, A., & Rangari, H. (2023). NLP-based video summarization using machine learning. *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, 10(2), 456-461. https://doi.org/10.32628/IJSRSET2310265

8. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. A., Kaiser, Ł., &

Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 6000-6010.

9. Zhang, Y., & Liu, Z. (2019). Deep learning for video summarization. *IEEE Transactions on Multimedia*, 21(3), 847-860. https://doi.org/10.1109/TMM.2018.2873121

10. Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507. https://doi.org/10.1126/science.1127647