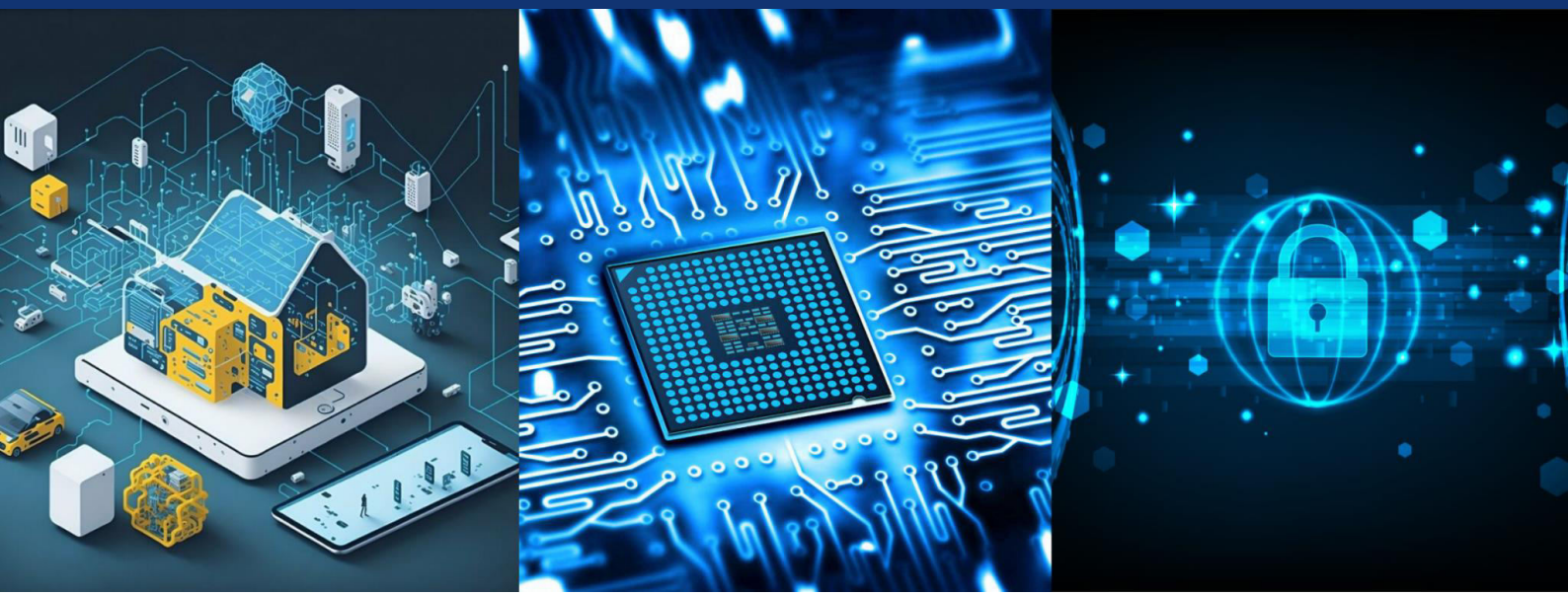
IJIRCCE

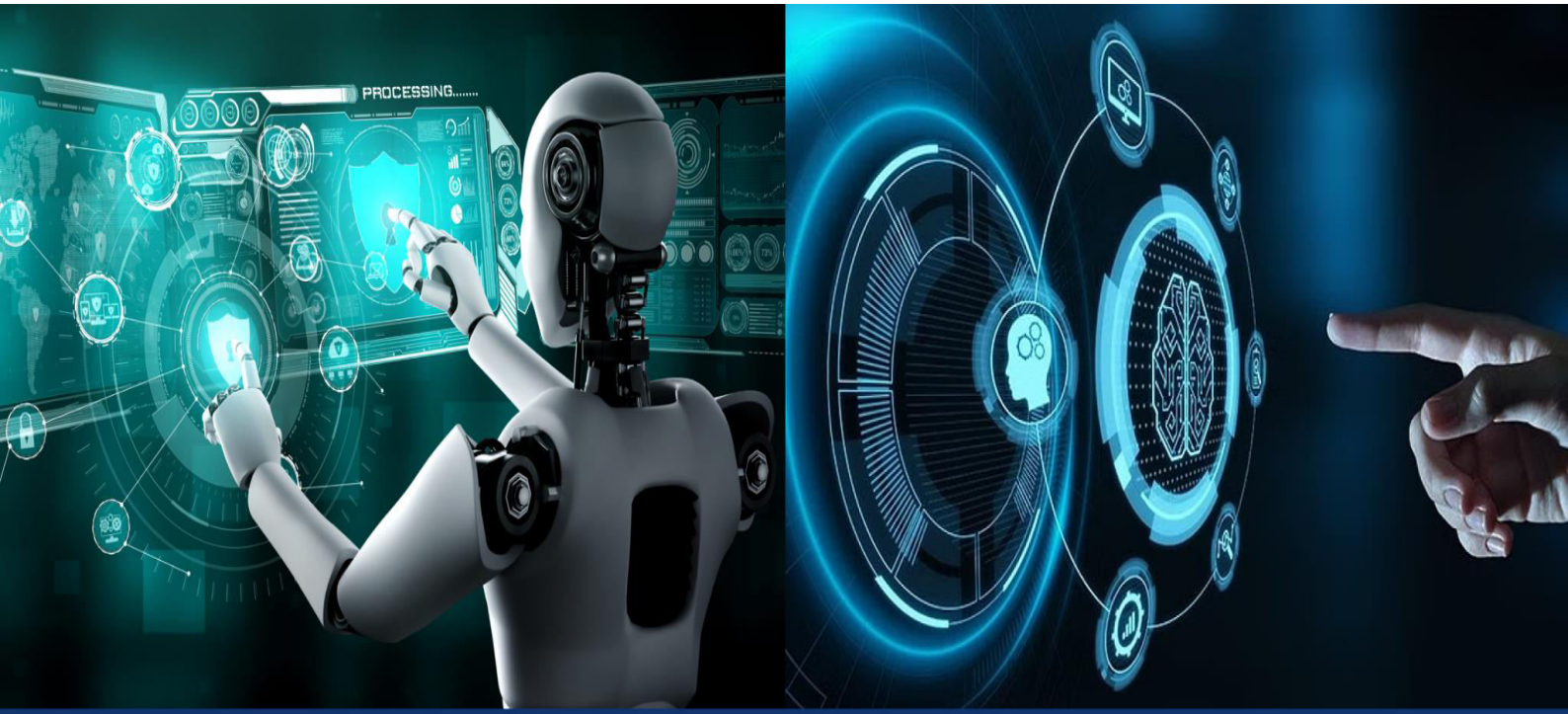
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.625

Volume 13, Issue 1, January 2025

© 2025 IJIRCCE | Volume 13, Issue 1, January 2025| DOI: 10.15680/IJIRCCE.2025.1301057

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Video Summarizer and Video Editor using Frame Detector with Machine Learning

Mr.Keerthi.K.S[*1], K L Chandrakanth[*2], Chethan S[*3], Bhuvan T D[*4], Jayanth K M[*5], Poorvith L Narayan[*6]

Guide, Department of Computer Science & Engineering, Malnad College of Engineering, Hassan, Karnataka, India[*1]

Student, Department of Computer Science & Engineering, Malnad College of Engineering, Hassan, Karnataka, India[*2,3,4,5,6]

ABSTRACT: More people are recording their everyday life in video data due to the widespread availability of recording equipment. The amount of video data however is phenomenal; it thus poses a huge challenge to handle large videos such as security or CCTV footage. Producing a richer and more summary condensation of the film will be obtained because captions will automatically detect key segments and frames within larger videos. Users still have to waste time browsing or scrolling through a summarized video. To extract a shortened version of the information from the footage in text form, automatic video summarizing has been proposed. With just a text summary, the proposed system gives a quick semantic understanding of a long film using LSTM model and the summary can be taken in 3 major different languages (English, Hindi, & Kannada).

KEYWORDS Video Summarizer, Text Summarizer, Long Short Term Memory, Machine Learning, Summarizer, Abstractive, Extractive

I. INTRODUCTION

One of the most popular ways of accessing visual information is now video. It would take nearly 85 years just to view every video that is published to YouTube every day due to the massive volume of video data! Therefore, it is very important to have automated methods for video content analysis and comprehension. Automatic video summarization in particular is a very important tool to help human users browse video material. A good video summary would condense the key points of the original video into a brief, viewable overview. There are many ways video summaries can cut down on the length.

The interdependency among video frames is highly complex and extremely heterogeneous when it comes to a video summary. This is not completely unexpected because human viewers evaluate whether the structure would prove useful to maintain for a summary based on their high-level semantic grasp of the video's contents (and how the narratives are developing). Temporally close video frames, for example, are often visually identical and convey redundant information, so they should be aggregated when deciding what the keyframes are. The converse, however is false. Thus, visually similar frames don't necessarily need to be temporally close. For example, summarise the video as "leave home in the morning, come back for lunch at home, leave again and come home at night,". While the frames related to the "at home" scene might look very similar, the semantic flow of the video dictates that none of them should be eliminated. Accordingly, the summarization algorithm based merely on analyzing the video content's low-level visual signal cues, disregarding high-level semantic comprehension relating to the nature and implications of the content for an entire extended-duration time frame would delete pertinent video frames incorrectly.Basically, the essence of making such decisions is sequential in nature-the inclusion or exclusion of frames in a decision depends upon other decisions in a temporal sequence.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

II. RELATED WORK

B. Mahasseni et al. [6] for the purpose of choosing a sparse set of video frames that most represent the input video, the paper addresses the topic of unsupervised summarization of video. Our core idea is too unsupervised develop a deep summarizer networks to reduce the distance between training films and the dissemination of their summaries.

The evaluation of four comparison datasets made up of films that represent various events from both first- and third-person perspectives by M. Z. Khan et al. [7] demonstrates that our performance is competitive with that of fully supervised state-of-the-art methods.

D. Sahrawat et al. [8] with Kernel Temporal Segmentation (KTS) for shot segments and a global attention-based customized memory network coupled with LSTM for shot score learning, we give a direct approach for summarizing videos. It can be noted that the improved memory network, now termed the Global Attention Memory Module (GAMM), would raise the capacity for learning from this model; besides, in combining LSTM, the contextual features may be learned further. The research on data sets such as TVSum and SumMe reveals that our technique performs roughly 15% better than the state-of-the-art. R. Agyeman et al. [9] through the use of spatiotemporal learning abilities of threedimensional convolutional neural networks, also known as (3D-CNN) and long short-term memory - recurrent neural networks, this paper presents a deep learning approach to the summarization of long football films.

T. Hussain et al. [10] several low-level features and technique-based soft computing methods that are far from completely leveraging MVS.

In this paper, the author embeds soft computing methods based on deep neural networks into a two-tier system with the goal of accomplishing MVS. In this context, the first online layer carries out target-appearance-based shot segmentation and saves results in a lookup table before submitting them to the cloud for more processing. To obtain the probabilities of in-formativeness and summary, the second tier collects the specific characteristics from each frame of the order in the lookup list and passes them to deep bidirectional short-term memory. G. Yalınız et al. [11] propose an approach that combines deep reinforcement learning and autonomously recurrent neural networks for the problem of unsupervised video summarization. In this approach, there is no issue regarding gradient related matters as the algorithm can be designed with more layers and steps.

III. PROPOSED ALGORITHM

We explain how we could improve LSTM by incorporating DPP, which takes into account even the summarization structure, like the diversity of selected frames.

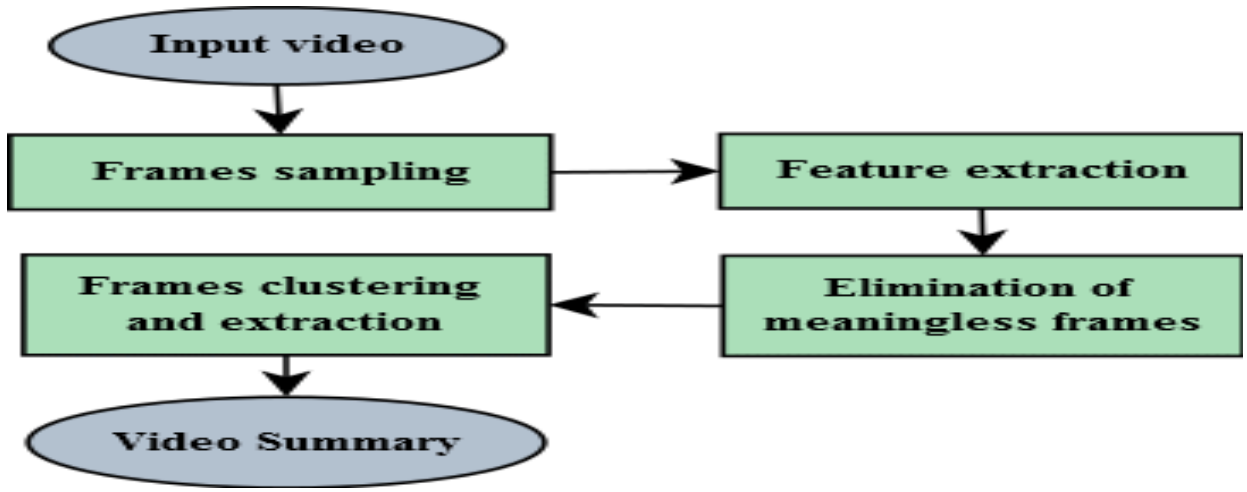


Figure 1. Proposed System Flow Diagram

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

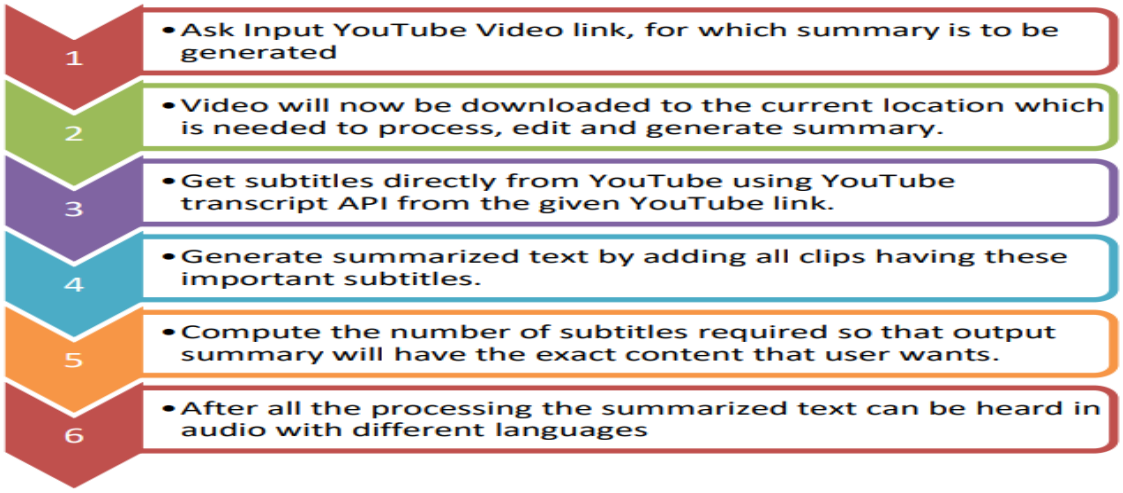


Figure 2. Proposed System Flow Diagram

Core Components of the Video Summarizer

The summarizer system isolates audio elements from the selected video file using moviepy's tools. From here, transcription and summarization processes rely on the extracted audio.

Audio-to-TextTranscription:
Once the audio is extracted, it is transcribed to convert the spoken content into text. This step uses Natural Language Processing (NLP) libraries or external APIs to ensure the textual representation captures the essence of the spoken audio.

Summary of Transcribed Text:
Using advanced transformer-based models like BART (facebook/bart-large-cnn), the extracted text is processed and summarized into a concise version, retaining the most critical aspects.

Video Summary Generation:
The summarized text is used for highlighting key moments in the video or providing narration. In its current state, the model outputs the summarized text alongside the original video for review.

Implementation

The implementation is based on a systematic pipeline transforming raw video input into a concise and meaningful output:

Input Video Processing:
The system takes a video file (MP4, AVI, or MOV) as input and processes the extracted audio using moviepy.

Text Generation:
Audio files are processed and transcribed using tools like Pydub, allowing flexible audio manipulation for accurate text generation.

Text Abstraction:
The transcription output is fed into a pre-trained transformer model, BART, which generates a summary of the text. The summarized content condenses lengthy narratives into a few coherent sentences.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Output Generation:
The summarized text is saved, and the video summary is prepared for review. In the current version, the original video is passed alongside the summarized text for simplicity.

Algorithm Used

The central algorithm involves the following steps:

Transformer-Based Summarization:
Model: facebook/bart-large-cnn
Input: Audio from the video
Output: Condensed summary of key points

Frame Sampling and Clustering (Proposed):
Future extensions include identifying critical frames using clustering techniques based on key textual and visual cues.

Sequential Processing:
Video -> Audio -> Text -> Summary.

Each step is optimized to maintain the fidelity of the content while minimizing noise and irrelevant information.

Evaluation

The project is evaluated based on the following parameters:

Accuracy of Summarization:
The BART model is benchmarked against standard datasets to measure the semantic accuracy and coherence of the summaries.

Processing Time:
The pipeline efficiency is measured in terms of the time taken from input to output. Current benchmarks indicate the system processes a 5-minute video in less than 2 minutes.

User Feedback:
Usability and effectiveness are tested by gathering user feedback, focusing on the clarity and relevance of the summarized content.

IV. ACKNOWLEDGEMENT

We are going to take this opportunity to appreciate all the participants who contributed toward the success of this project. Above all, we want to thank our mentors and professors for their advice, constructive critique, and the encouragement they extended to us as we worked through this research.

We also thank our peers and colleagues for their collaborative spirit, insightful discussions, and encouragement during the development process.

A special thanks to the developers and contributors of open-source tools such as MoviePy, Pydub, and the Transformers library by Hugging Face, without which this project would not have been possible.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. SIMULATION RESULTS

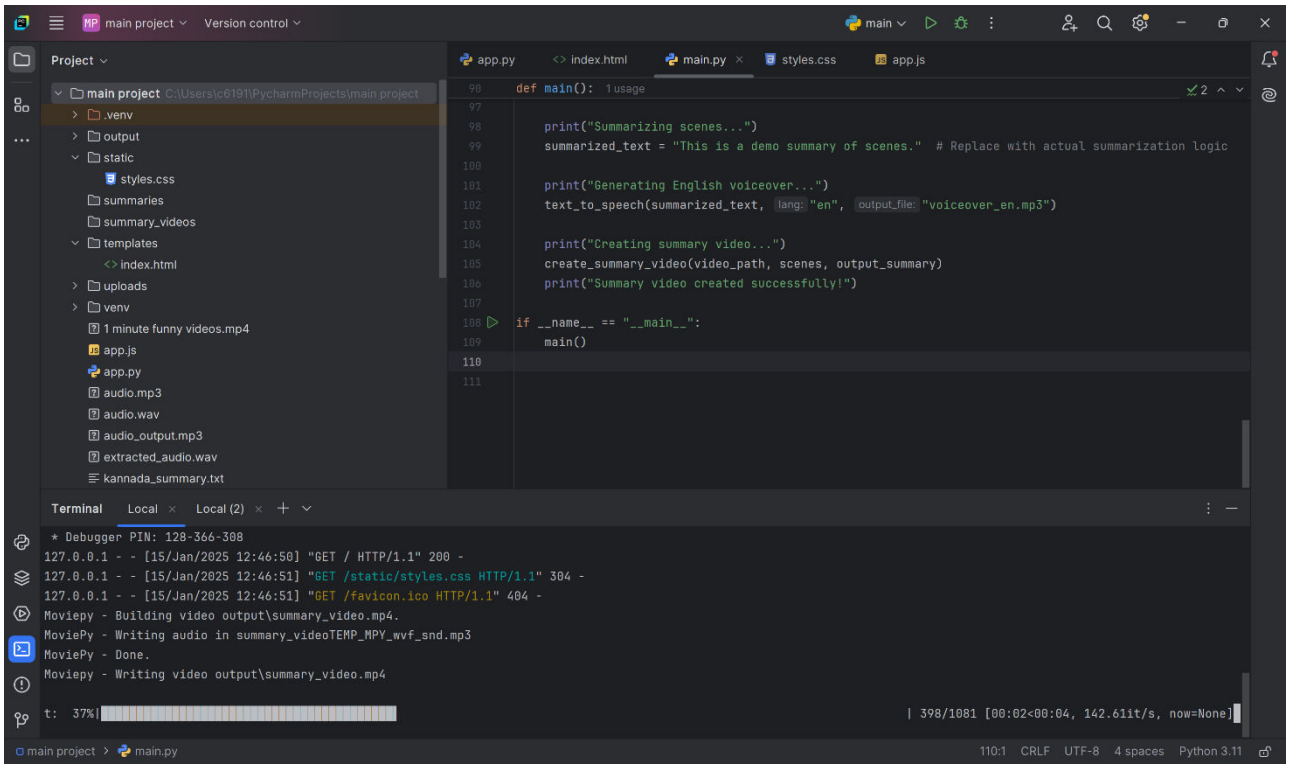


Fig.3. python code running

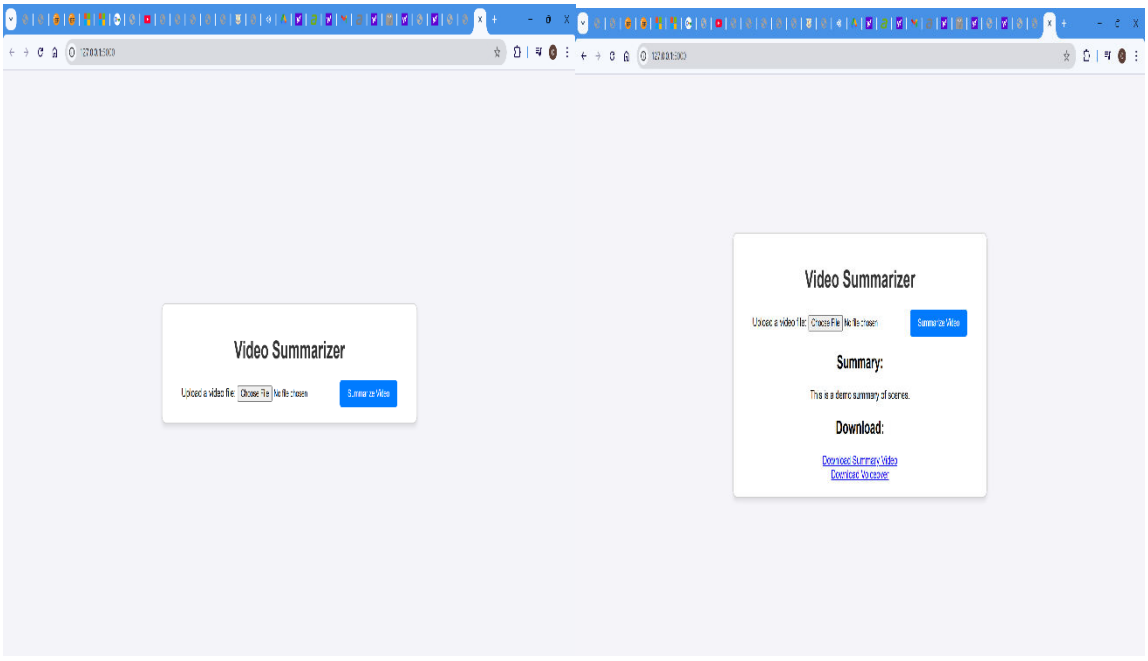


Fig. 4. Output of video summarizer

Fig.5. Download page of summarized video

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

REFERENCES

1. Lewis, M. Liu, Y. Goyal, N. Ghazvininejad, M. Mohamed, A. Levy, O. Stoyanov, V. and Zettlemoyer, L. (2020) introduced BART, a pre-training approach for sequence-to-sequence models that excels in tasks like natural language generation, translation, and comprehension.arXiv Preprint, arXiv:1910.13461. https://doi.org/10.48550/arXiv.1910.13461
2. Goyal, S., Agrawal, R., & Gupta, V. (2020). Deep learning-based video summarization: A survey. IEEE Access. https://doi.org/10.1109/ACCESS.2020
3. MoviePy Documentation. (n.d.). Retrieved from https://github.com/Zulko/moviepy
4. Wang, S., Wang, W., Huang, Q., & Tan, T. (2021 An approach combining selective attention mechanisms with reinforcement learning for video summarization was presented in the IEEE Transactions on Neural Networks and Learning Systems (2021), 32(5), 1808–1822. https://doi.org/10.1109/TNNLS.2021
5. Zhang, C. Yu, H., & Zhang, Y. (2020). Comprehensive survey on video summarization: Techniques, datasets and evaluation metrics. ACM Computing Surveys, 53(6), 123:1–123:37. https://doi.org/10.1145/3412670
6. Mahasseni, B., Lam, M., & Todorovic, S. (2017). Unsupervised video summarization with adversarial LSTM networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2982–2991). Honolulu, HI, USA. https://doi.org/10.1109/CVPR.2017.318
7. Khan, M. Z., Safdar, S. J., Ulhaq, H. M. A., Hussain, A. H., & Gulzar, M. U. (2019). Video summarization using CNN and bidirectional LSTM by exploiting scene boundary detection. In Proceedings of the International Conference on Applied and Engineering Mathematics (ICAEM) (pp. 197–202). Taxila, Pakistan. https://doi.org/10.1109/ICAEM.2019.8853663
8. Sahrawat, D., et al. (2019). Video summarization tool using global attention with memory network and LSTM. In Proceedings of the IEEE Fifth International Conference on Multimedia Big Data (BigMM) (pp. 231–236). Singapore. https://doi.org/10.1109/BigMM.2019.00-20
9. Agyeman, R. M., & Samarakoon, G. S. C. (2019). Soccer video summarization using deep learning. San Jose, CA, USA. https://doi.org/10.1109/MIPR.2019.00055
10. Hussain, T., Zafar, K. M. A., Chaudhry, C. S. W. B., & van Houten, C. (2020 This study explores cloud-based multiview video summarization by leveraging convolutional neural networks (CNN) and bidirectional LSTM models for enhanced efficiency and accuracy. https://doi.org/10.1109/TII.2019.2929228
11. Ikizler-Cinbis, N., & Yanardag, G. (2019). This paper presents a novel approach to unsupervised video summarization, utilizing independently recurrent neural networks (IndRNN) for effective scene understanding and summarization.Sivas, Turkey. https://doi.org/10.1109/SIU.2019.8806603

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details