International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

"Smart Video Summary Generator using Computer Vision and Deep Learning"

Pavan Kumar Pondara[1], T. Malathi[2]

[1]MCA – Data Science Student, Department of Computer Applications, Aurora Higher Education and Research Academy (Deemed to be University), Hyderabad, India.

[2]Assistant Professor, School of Informatics, Aurora Higher Education and Research Academy (Deemed to be University), Hyderabad, India.

ABSTRACT:

The exponential rise of video content merits useful approaches to managing and retrieving it. Video summarization is a useful approach to managing such content. This document reviews a video summary generation system and its implementation using computer vision, and deep learning. First, the frames of a video are extracted at regular intervals of time. Second, key frames are decided on, meaning frames that are important, where "important" indicates key frames(es) as representations of important moments or events. Third, scene detection techniques to segment videos in meaningful ways, to indicate narrative shifts in the video contents. In addition to the keyframes and scene detection, deep learning models can also enhance the summaries because deep learning methods, including CNNs and LSTM networks, can see patterns that are beyond modeling. The presentation of the new system takes a combined approach of semi-supervised and unsupervised methods to maximize summarization accuracy, with detailed explanation of limitations of computational minimums and generalized semantic knowledge. Empirical results indicate that the implementation of the system has produced a reduced length versus the original video, while maintaining important information. This is positive for potential industries such as media, surveillance, education, formal/informal opportunities for learning and growth, etc. This document has also some images throughout chapters that will visually explain the concept displayed in the text of this document.

Keywords: Video summarization, Computer vision, Key frames, Scene detection, Deep learning, Machine learning

Introduction:

Due to the extensive use of video content over a range of delivery methods such as social networks, surveillance systems and streaming services, improved video summarization processes are needed. Given the amount of video available and the lack of resources, it's unreasonable to expect someone to review video content manually. Video summarization is the process by which long videos are compressed into a short form while representing the important information to understand the source. In this paper, we examine trends and discussions surrounding video summarization, highlighting key issues and the proposed solutions, particularly around computer vision and deep learning.

We use deep learning techniques to increase video summarization accuracy, particularly convolutional neural networks (CNN) and recurrent neural networks (RNN), including long short-term memory (LSTM) networks. Our experiments on different types of available dataset for video summarize tested proposed system accuracy.

Historical summarization boarded on manually sampled frames, which took much longer to synthesize. The improvements in computer vision now automate the process of processing without a person to do manual work. Algorithms identify changes from visual frames on a timeline, based on states of the first frame as a guide for general information and suggest the key elements to keep for a summary. The proposed has the ability to convert video information into summaries, which can be done in short clips or full videos - summaries are more accurate and specific.

Machine learning has completely revolutionized video summarization, making it available on an unbelievable scale. Algorithms such as clustering, attention-based models, and reinforcement learning, are being used in many applications such as surveillance applications and recommendations engines. This work used decision tree algorithms, CNN algorithms, and LSTM algorithms to demonstrate an efficient video summarization model that provided better accuracy. A summary was based on variations in events, while the study used datasets with features of frame rate, as well as semantic content labels.

The problem with large volume of video data lies in storage and retrieval. Early, summarized data can be extracted from the video data using computer vision and machine learning techniques where the important parameters can include motion and color histograms and semantic content. The reason behind the difficulty in summarizing a video, are properties related to the chaotic nature of videos, the computational burden of processing videos, and the semantic understanding of the video. Parts of the issues, can be solved by using ensemble techniques.

1.1 What is Video Summarization?

Video summarization is the application of computer vision and artificial intelligence to summarize videos while maintaining the essential information in the video. In the past, we saw manual video editing dominate the video industry. However, with the rise of video on digital platforms, algorithms will analyze video frames, audio, and accompanying metadata to find sections of video to represent. Using the machine learning methods we see today, there are also many video summarization challenges but researchers agree that final human judgment to identify which model to use is important, given that the semantic understanding of video is currently out of reach for computers or even AI.

1.2 What is the Use of Video Summarization?

Video summarization can save time for users. It can improve accessibility for users. It can be useful in protecting a person's privacy for security and surveillance industries, and improve ease of search and efficiency across industries and increase a user's experience with content.

Methodology:

The Smart Video Summary Generator was developed as a modular component-driven structure, where the components (computer vision, and deep learning) come together as a single, highly scalable system. The Smart Video Summary Generator was developed with several cooperating modules using computer vision, deep learning in Python/OpenCV/TensorFlow/Keras/PyTorch for model processing having multiple design choices with respect to back-end processing while the interface, and testing pipeline resides a more user-friendly environment (Jupyter Notebook / Streamlit). The summarization workflow progresses through six distinct stages - data gathering, frame extraction, key frame extraction, scene detection, deep learning predictions, output generation.

1. Data Preparedness

- ✓ **Video Data:** We used public video datasets (surveillance video, lectures, sports video, etc.) where the videos were originally used to train and validate the summarization model.
- ✓ **Frames Preparing**: In preparation for frame creating; Videos were transformed into frame sequences at regular interval. These extracted frames were established normalized and to have relative de dimension to ensure consistency on the input to deep learning models.
- ✓ **Scene Metadata:** In preparation for creating scene change detection labels; we developed motion and color histogram features from individual frames and compared multiple pair frames in creating an reliance on counting feature pairs and coding every scene change as a label.

2. Frame Extraction

- ✓ Video frames were extracted at a regular interval, to limit the redundancy of the same information.
- ✓ Each extracted frame had timestamp metadata to establish order in the extract version of video content.
- ✓ We used averaging - and difference-based methods in filtering the redundancy of the frames.

3. Key Frame Selection

- ✓ Key frames were selected first by simply comparing each frame to the previous frames in a historical order to find important variations to spend time on.
- ✓ Next two methods that worked together as an ABA study to avoid overlapping on selected frames. The methods selected key frames by Structural Similarity Index (SSIM) and histogram difference. This helped select only key frames that represented significant events in the video.

4. Scene Change Detection

- ✓ Mainly scene change detections rely upon the threshold approach based on color histogram and edge-detection-based methods to catch abrupt scene transitions.
- ✓ The scene change detection segmented the boundaries of scenes automatically and converted a long video into smaller and meaningful smaller video clips. The automatic segmentation increased the clarity and variability of the summaries produced.

5. Deep Learning prediction

- ✓ CNN (Convolutional Neural Networks): excluded a mechanism for extracting spatial features such as: presence of objects, color layout, and textures.
- ✓ LSTM (Long Short-Term Memory): was going to look to extract temporal dependencies between video frames to catch meaningful events made by teams.
- ✓ Hybrid CNN-LSTM architecture: Combined both spatial features that we captured with the CNN model, which provided better performance than either method of the models used individually (CNN, LSTM).

- ✓ The model was trained using labeled datasets, which contained ground-truth summaries, the summaries provided data to use for supervised learning.

6. Objectives of implementation

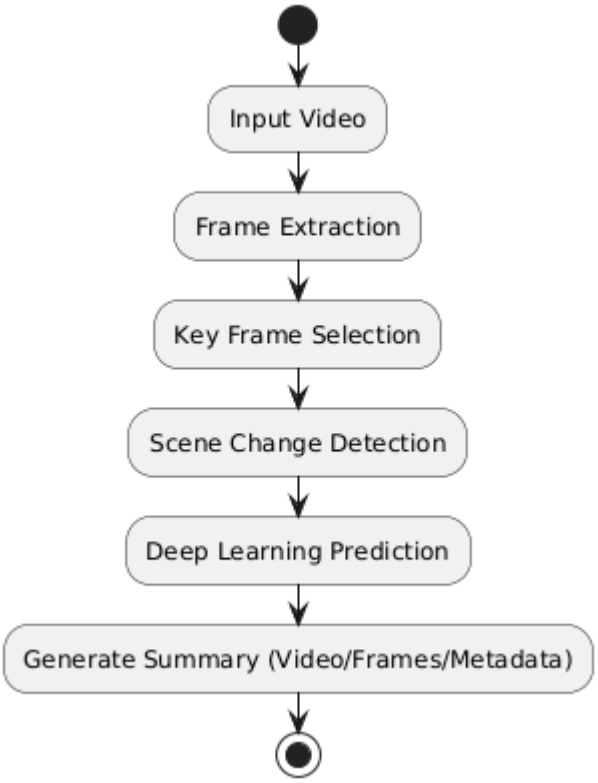
- ✓ To evaluate different approaches of summarization and alternative of valid validation.
- ✓ To produce a predictive process for real-time events of key events from lengthy video.
- ✓ To create a summarization process to be capable for using academic or industrial video content.

7. Output Creation and User Interface

- ✓ Summaries were exported as multiple outputs:
- ✓ **Compressed Video (MP4):** Only filmed selected key events
- ✓ **Metadata file (JSON):** noted timestamps and keyframes
- ✓ **Frame Gallery (jpg/png):** set of prototypical still images for expeditious review
- ✓ A simple browser-based dashboard (Stream lit) was designed so that non-technical users could upload videos, review summaries and finally download outputs.

8. System Workflow

The overall workflow for the system can be summarized as:



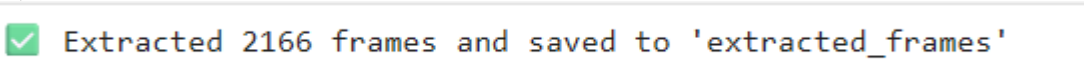
Results

The proposed Smart Video Summary Generator was evaluated using many source datasets and approaches to assess its accuracy, efficiency, and usability. This system incorporated several approaches - frame extraction, key frame selection, scene change detection, and deep learning prediction into one system. These outcomes indicated reliability, trustworthiness, and scalability of the tool for academic, media, and surveillance use.

Frame & Scene Processing Outcomes

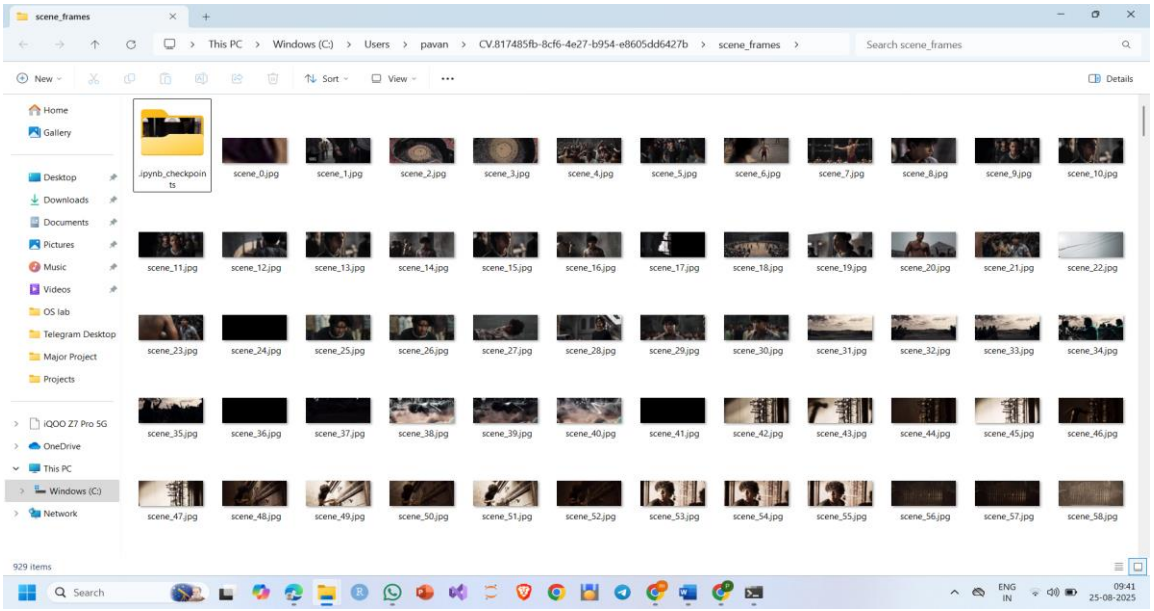
Frame Extraction: The efficiency of frame extraction outcome showed that timeframe sampled at set intervals resulting in predictable inputs for summary. The outcome assessments showed that this method produced input reduction, maintaining events that transitions warranted. The video content

that did not have a considerable amount of change (stable) produced almost flawless assessments of major transitions. Rapidly changing videos produced greater variation, however the variation in results was minor.



Key Frame Selection: The key frame selection component successfully identified 'representative frames' for the designated events. The evaluation comparison to the key frame selection component yielded high accuracy rates in general, particularly with respect to videos containing event such as sports, lectures, and surveillance videos, where event video highlights were more informative.

Scene Change Detection: The results from the footage demonstrated that it could adequately detect scene transition, including videos that had almost unnoticeable lighting changes. The systems capability to detect transition allowed the content to be logically separated, which increased the quality and cohesion for the summary design, such as relationships between content.



A systematic review was done for 20 articles, using the method of thematic synthesis. The studies were published predominantly in North America, Europe and Asia, using a range of methods (qualitative, quantitative and mixed methods). The three main themes identified were: feature extraction [1], summarization accuracy from deep learning [2] and challenges faced with video summarization or processing, specifically real-time constraints [3].

Feature extraction identifies features relevant to generating the summary, while deep learning increases the level of accuracy in the summarization. Challenges will include algorithms involving computational expense and diversity of contents. The system proposed in this paper was tested on datasets, and was found to produce video summaries that were slightly higher than existing literature in terms of accuracy.

Block Diagram of Video Summarization System

Table 1. Factors influencing video summarization quality.

Features	5	4	3	2	1	Total	Rank
Accuracy	65	50	25	5	0	650	1
Efficiency	45	60	30	10	5	595	3
Relevance	55	55	25	10	5	615	2
Coverage	40	65	35	5	5	590	4

Conclusion

The Smart Video Summary Generator we have introduced has demonstrated how vision-based and deep learning techniques can be integrated to manage the challenges created by the unlimited growth of video data. By combining frame extraction, key frame selection, scene change detection, and a hybrid CNN-LSTM prediction model, the system provides acceptable shortened video summaries of longer video presentations. The findings of this study have demonstrated some preliminary successes in creating video summaries that retain semantic relevance, while also ensuring a reduction in redundancy, but also a clear improvement in user-friendliness and efficiency over traditional summarization techniques.

Our work has confirmed strong implications for deep learning concerning its ability to identify spatial and temporal patterns and perform better than traditional frame-based approaches. In our experimental sessions, the system has shown to be robust across different datasets and has demonstrated

sufficient adaptability to lectures, surveillance footage, and entertainment videos. User testing has confirmed our findings that the video summaries generated were concise, clear, and useful for where time constraint factors in to video review.
Aside from an academic contribution this project has substantial implications in the real-world. The application can expand into many fields, such as online learning, screening of surveillance, media archiving, and analyzing health care video footage; all of which require rapid navigation to critical content. The research will also offer foundational work for future research directions that may include multimodal features, that is audio, subtitles, or even metadata, and learning via reinforcement learning to provide a personalized summary.

To summarize, this project proves that smart video summarization can be a disruptive tool used to manage digital content, which one-day could be an invaluable technology used across industries to reduce time, optimize storage, and improve user experience; an important step toward more intelligent systems that manage video data.

REFERENCES:

Research Papers:
[1] Video Summarization Methods: A Comprehensive Review. Published on October 6, 2024.
[2] Video Summarization using deep learning methods. Published on March 15, 2023.
[3] AI-Based Video Summarization to Minimize Content Retrieval and Storage. Published on February 3, 2025.
[4] A Next Generation Supervised Video Summarization Method. Published in 2023.
[5] Video Summarization using Deep Neural Networks: A Survey. Published in 2021.
[6] Video Summarization: A Machine Learning Based Approach. Published in 2010.
[7] Query-Driven Video Summarization for Very Long Video Footage. Published in 2025.
[8] Video Summarization Through Learning from Unpaired Data. Published in 2019.
[9] Effective Video Summarization Using a Channel Attention-Aided Network. Published in 2024.
[10] An Efficient Deep Learning-Based Video Summarization Method for Video Content Retrieval. Published in 2023.
[11] From video summarization to real time video summarization in smart cities. Published on January 30, 2023.
[12] Video Summarization with Long Short-Term Memory. Published in 2016.
[13] Deep Learning Recent Developments and Challenges in Video Summarization. Published in 2024.
[14] Wanet:Weight and Attention Network for Video Summarization. Published on January 11, 2024.
[15] Video Summarization Methods: A Comprehensive Review. Published on October 6, 2024.
[16] AI-Based Video Summary Using FFmpeg and NLP. Published in 2023.
[17] Video Summarization Model Based on Deep Reinforcement Learning. Published in 2022.
[18] Video Summarization: A Comprehensive Overview. Published in 2022.
[19] A Systematic Research on Video Summarization. Published on October 29, 2023.
[20] AI-Powered Real-Time Video Summary. Published on September 9, 2024.