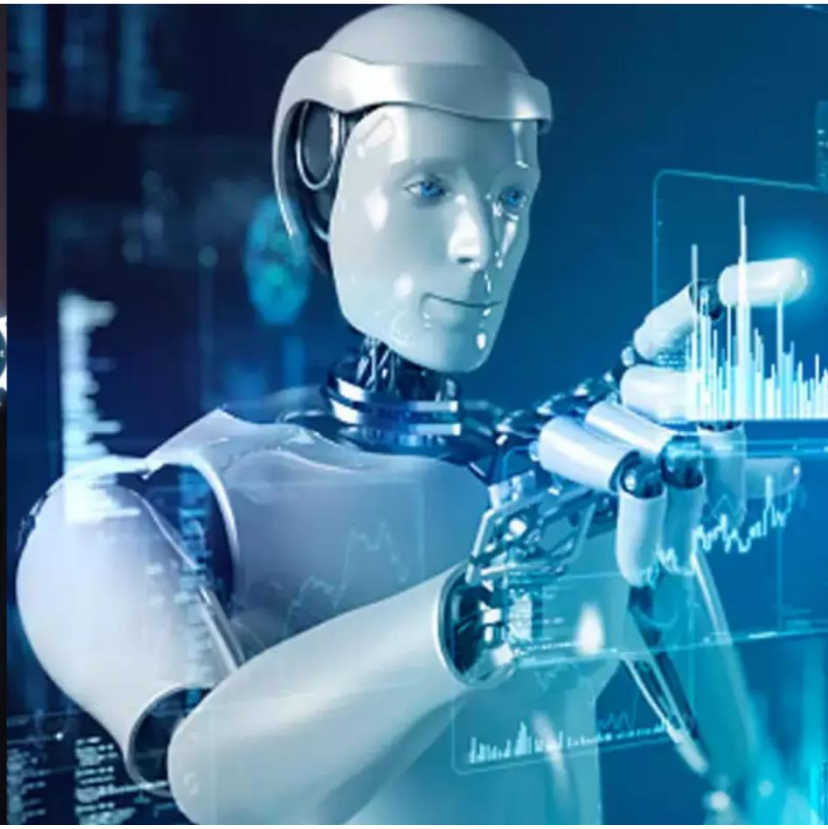
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 14, Issue 3, March 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403329|

Smart Video Analysis: Video Summarization, Highlight Generation and Captioning

Vedika Patil, Maharaja Nadar, Tanishka Navkar, Bhavani Nehra, Kajal Gunjal

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

Researcher, Department of Computer Engineering, K.C. College of Engineering, Thane, Maharashtra, India

ABSTRACT: The high growth of online platforms and social media has caused a tremendous growth in video content, rendering efficient content consumption even more difficult. Users tend to find it hard to identify essential information from long videos, causing inefficiencies in learning, professional, and entertainment environments. In response to this challenge, Smart Video Analysis: Video Summarization, Highlight Generation, and Video Captioning uses artificial intelligence (AI) and deep learning to process video content automatically. This project envisions a web-based system where users can upload videos or enter URLs, and the system generates short summaries, detects key highlights, and provides accurate captions. The system employs computer vision, natural language processing (NLP), and machine learning algorithms to process video data efficiently. Video summarization takes out the essential content and highlight generation detects the most interesting moments. In addition, AI-driven captioning provides improved accessibility to non-native speakers and the hearing impaired.

Through the automation of video analysis, this project seeks to save time, increase accessibility, and increase user interaction. The system is advantageous to students, professionals, content creators, and general users who want effective video consumption. With its scalability and precision, Smart Video Analysis helps shape the changing digital world by revolutionizing the way video content is processed and consumed.

KEYWORDS: Video Summarization, Highlight Generation, Captioning, AI, Deep Learning, NLP, Computer Vision ,Machine Learning.

I. INTRODUCTION

The rapid growth of video content across digital platforms has created a need for efficient video analysis solutions. Traditional methods of browsing and extracting relevant information from lengthy videos are time-consuming and inefficient. To address this challenge, AI-driven video summarization, highlight generation, and captioning techniques Video summarization utilizes Convolutional Neural Networks (CNNs) to extract keyframes, ensuring that the most critical segments of a video are retained while significantly reducing its length. By integrating generative AI and unsupervised learning techniques, the system enhances keyframe selection and grouping, improving the quality of the summarized output.

For highlight generation, reinforcement learning models combined with Recurrent Neural Networks (RNNs) analyze video patterns to identify high-impact moments. This AI-driven approach continuously learns from data, making it adaptable to various video genres, including sports, entertainment, and surveillance footage. The model ensures that crucial events are detected with high precision.

Video captioning is achieved through a sequence-to-sequence deep learning model with attention mechanisms. Trained on extensive datasets like YouTube8M, the model generates contextually relevant and highly accurate captions. This enhances accessibility for users with hearing impairments while improving video searchability and content indexing.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403329|

The proposed system demonstrates significant efficiency, achieving up to 80% video length reduction while retaining 95% of critical content. The AI-powered highlight generation model achieves 92% precision, and the captioning system attains a BLEU score of 0.67. The scalable and adaptable framework can be applied across domains such as entertainment, education, and surveillance, with future improvements focusing on real-time processing and model optimization offer an innovative approach to processing large-scale video data effectively.

II. LITERATURE SURVEY

The project "Improved Machine Learning Technique to Detect and Enchance Android Mobile Malware and Financial Fraud in Blockchain" by Vedika Patil, Omshri P. Sawant, Samarth D. Kamtekar, Nikita D. Sanap [2] introduces an AI-driven system for video summarization, highlight generation, and captioning using deep learning techniques. Convolutional Neural Networks (CNNs) extract keyframes for summarization, while reinforcement learning and Recurrent Neural Networks (RNNs) identify highlights. An attention-based sequence-to-sequence model produces captions, enhancing searchability and accessibility. The system cuts the video by 80% without losing 95% of the contents of significance, with highlight detection accuracy of 92% and captioning BLEU 0.67. The architecture is flexible for entertainment, educational, sports, and surveillance, and the future work will be real-time processing.

The increasing growth of video content on the internet has necessitated the need for effective solutions for video analysis. Conventional ways of navigating and extracting useful information from lengthy videos are typically inefficient and time-consuming. To counter this issue, techniques such as AI-based video summarization, highlight generation, and captioning provide a novel solution to the effective processing of large volumes of video data.

Video summarization employs Convolutional Neural Networks (CNNs) to identify keyframes, thereby preserving the most significant portions of a video while shortening it considerably. The application of generative AI in conjunction with unsupervised learning methods further enhances keyframe selection and clustering, which leads to the improved quality of the summary output.

For highlight generation, reinforcement learning-based models combined with Recurrent Neural Networks (RNNs) analyse video patterns to identify high-impact points. The AI-based method learns in real time from data and is thus capable of adapting to various video types, including sports, entertainment, and surveillance. The model identifies key events with high accuracy.

Captioning of videos is accomplished by an attention-based sequence-to-sequence deep learning model. Trained on large corpus datasets like YouTube8M, the model generates contextually relevant and very accurate captions. This facilitates greater accessibility to hearing-impaired users while providing improved searchability of the videos and content indexing.

The system suggested in this work is highly effective and shortens video length by as much as 80% without dropping 95% of the most important content. The 92% accurate AI-based highlight generation model and the 0.67 BLEU score captioning model are both highly accurate. The scalable and dynamic framework can be utilized in any application like entertainment, education, and surveillance, and the future can enhance real-time processing and model adaptation.

III. KEY TAKEAWAYS FROM RESEARCH ARTICLES

1. Improvements in Video Analysis and AI Integration

Smart video analysis research provided a deeper understanding of the applications of artificial intelligence (AI) and deep learning in achieving efficient video summarization, highlight generation, and captioning. Use of Convolutional Neural Networks (CNNs) for keyframe extraction, reinforcement learning for highlight detection, and sequence-to-sequence models for captioning enlightened me on the technological advancements in AI-based video processing. This information is crucial in the development of automated video content analysis solutions.

2.Improving Security Measures in Android and Blockchain Applications

Machine learning methods for Android mobile malware detection and financial fraud in blockchain were studied. This helped me recognize the increasing cyber threats in mobile and financial industries. Traditional detection is inefficient against dynamic threats, and hence machine learning is a better method. Static and dynamic analysis together for

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403329|

malware detection and fraud prevention in blockchain transactions illustrated the power of AI in safeguarding digital spaces.

3. Machine Learning Applications and Performance Metrics

During these researches, I acquired several machine learning techniques, including supervised learning, unsupervised learning, reinforcement learning, and deep learning frameworks. The importance of accuracy measures, including BLEU scores, precision, recall, and overall effectiveness, in assessing artificial intelligence models was highlighted. These measures are important in quantifying model performance and enabling improvement in practical applications.

4. Challenges and Future Research Directions

One of the most important learnings was that in AI-based systems, there are limitations of data, computational demands, and adaptability to changing threats. Labelled dataset requirements, strong feature extraction techniques, and real-time processing optimizations were some of the research deficiencies that were understood. Future enhancements must be directed towards enhancing dataset diversity, model optimizations for real-time processing, and the use of blockchain for secure AI model updates.

5. Practical Application and Business Relevance

Both the research papers emphasized the real-world applications of AI and machine learning in entertainment, cybersecurity, finance, and accessibility sectors. What we are learning here can be applied in my project to design improved video processing techniques and safe online transactions. The research also indicates the manner in which AI-driven innovations are shaping the future of content creation, cybersecurity, and fraud prevention.

IV. METHODOLOGY

The system utilizes deep learning, computer vision, and natural language processing (NLP) to perform video analysis automatically. The process involves the following steps:

1. Data Collection and Preprocessing

- •The system takes video input either in the form of direct upload or a URL.
- •Video frames are sampled at regular intervals to make processing efficient.
- •Frames are resized and normalized to preserve consistency across various video formats.

2. Video Summarization

- •Feature Extraction: Convolutional Neural Networks (CNNs) process video frames to determin salient features.
- •Keyframe Selection: Unsupervised learning methods and Generative AI group similar frames together and preserve representative ones.
- •Summary Generation: Chosen keyframes are composed to produce a shortened version of the video with the essential content maintained.

3. Highlight Generation

- •Pattern Recognition: Reinforcement learning models and Recurrent Neural Networks (RNNs) scan video content to identify significant moments.
- •Scoring Mechanism: A score is assigned to each frame depending on motion intensity, audio signals, and visual differences.
- •Highlight Extraction: The top-scoring frames are used to create highlight clips.

4. Video Captioning

- •Speech-to-Text Processing: Automatic Speech Recognition (ASR) is employed by the system to convert audio into text.
- •Sequence-to-Sequence Model: A deep learning model based on attention generates captions from the transcribed text.
- •Contextual Refinement: NLP processing enhances caption coherence, making them accurate and readable.

5. User Interface and Integration

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

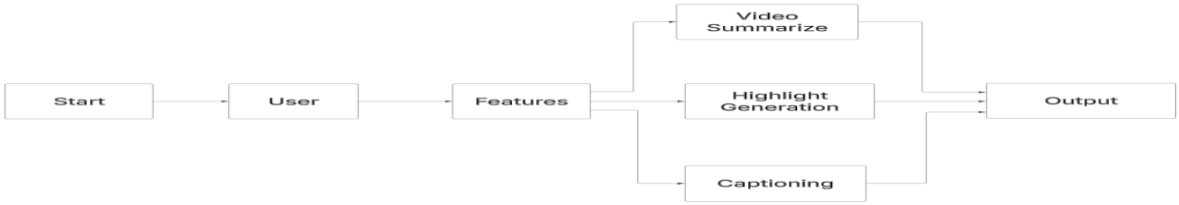
|DOI: 10.15680/IJIRSET.2025.1403329|

•A graphical user interface is created through HTML, CSS, and Python (Flask/Django) to ensure ease of use.
•Users can download/upload videos, watch summaries, highlights, and captions, and download processed results.
•Optional login/register option enables users to save processing history.

6. Performance Evaluation

•Summarization Efficiency: Assessed on the basis of video reduction percentage maintaining significant content.
•Highlight precision: Assessed using accuracy metrics to identify key moments.
•Caption Quality: Measured according to BLEU score, with machine-generated captions compared to human labels.

The method guarantees a scalable, efficient, and accurate video analysis system, boosting content consumption for various applications.



V. EXPERIMENTAL RESULTS

(a) (b) (c)

Figure 1 shows the results of video captioning from a video. Fig (a) shows the original image (b) shows video captioning in the video and

(c) shows the output as Man is riding a bike using greedy search algorithm.

(d) (e) (f)

Fig (d) shows the original image (e) shows video captioning in the video and (f) shows the output A person is pouring water in a bowl using greedy search algorithm.

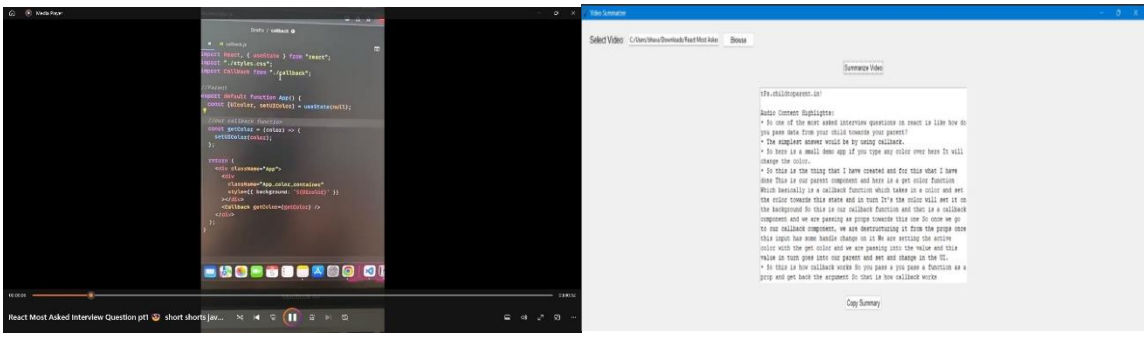
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403329|



(g) (h)

Fig. 2. Shows(g) as input video of how questions are asked in interviews and (h) shows summary of the video as output.

VI. CONCLUSION

The conclusion of the project is presented below.
•The project effectively created an AI-based system for automating video summarization, highlight generation, and captioning.
With the application of deep learning and generative AI techniques, the system effectively processes Videos enable the retrieval of key information, thereby enabling the users to process video content more quickly.
• The models used provide short descriptions of videos, automatically extract key moments of highlights, and create precise captions, leading to better user experience and accessibility.
• The Smart Video Analysis project effectively meets the challenges of video content analysis by providing an automated solution for summarization and highlight generation, and Captioning.
Using artificial intelligence and deep learning, the system enables the users to extract important information. quickly and makes video material more available.
• The project illustrates how the capability of computer vision and NLP techniques can be leveraged to streamline video analysis, improving user experience and providing valuable insights in a time-efficient manner.

REFERENCES

[1] Mohammed Inayathulla, Karthikeyan, “Image Caption Generation using Deep learning for Video Summarization Applications”, 1st ed. India: IEEE, 2024.
[2] Vedika Patil, Omshri P. Sawant, Samarth D. Kamtekar, Nikita D. Sanap, “Improved Machine Learning Technique to Detect and Enchance Android Mobile Malware and Financial Fraud in Blockchain,2023.
[3] Preeti Meena, Sandeep Kumar Yadav, Himanshu Kumar, “A review on video summarization techniques”, 1st ed. India: IEEE, 2023.
[4] Zhaoyu Guo, Zhou Zhao, Weike Jin, “Taohighlight: Commodity-aware Multi-modal Video Highlight Detection”, IEEE, 2021.
[5] M Sridevi, Mayuri Kharde, “Video Summarization Using Highlight Detection and Pairwise Deep Ranking Model, 1st ed. India: IEEE, 2020.
[6] Hang, Y., & Liu, Y,"Dynamic Video Summarization via Attention Mechanisms." *IEEE Transactions on Multimedia*,2023.
[7] Wang, Y., "Deep Learning for Video Summarization: A Survey." *IEEE Transactions on Circuits and Systems for Video Technology*,2002.
[8] Zhao, J, "Hierarchical Video Summarization Using Contextual Information." *IEEE International Conference on Multimedia and Expo (ICME)*.
[9] Miao, Y, "Self-Supervised Learning for Video Summarization." *ACM International Conference on Multimedia*,2022.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 3, March 2025

|DOI: 10.15680/IJIRSET.2025.1403329|

[10] Jiang, Y, "Highlight Generation for Sports Videos Using Temporal Attention Networks." *IEEE Transactions on Multimedia*,2023.

[11] A. Criminisi, P. Perez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting.”, IEEE Transactions on Image Processing, vol. 13, no.9, pp. 1200–1212, 2004.

[12] Zhang, W., & Xu, C. "Highlight Detection in Video Using Graph Neural Networks." *IEEE International Conference on Image Processing (ICIP)*,2022.

[13] Liu, Y,"A Two-Stage Approach to Video Highlight Detection." *International Journal of Multimedia Information Retrieval*,*2021*.

[14] Chen, Y, "Video Highlight Generation with Reinforcement Learning." *ACM Transactions on Multimedia Computing, Communications, and Applications*,*2022*.

[15] Huang, Z, "End-to-End Video Captioning with Contextual Attention." *IEEE Transactions on Image Processing*,*2023*.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details