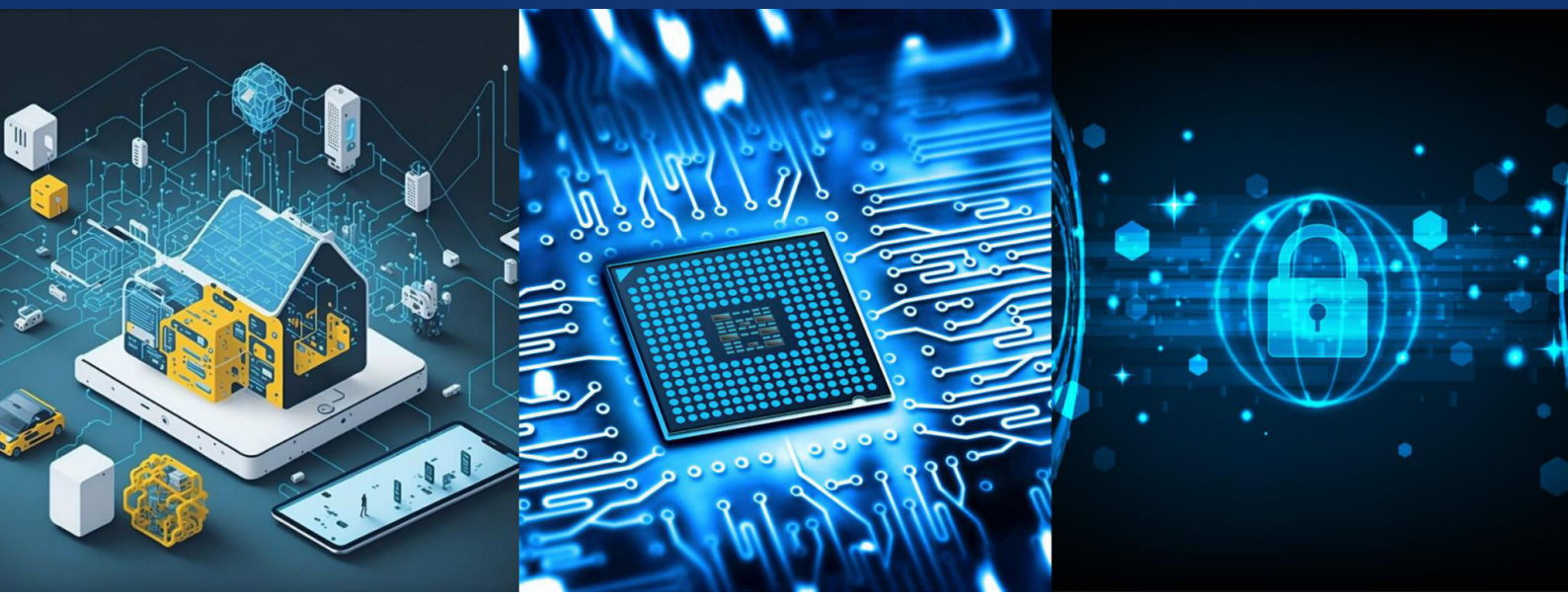
IJIRCCE

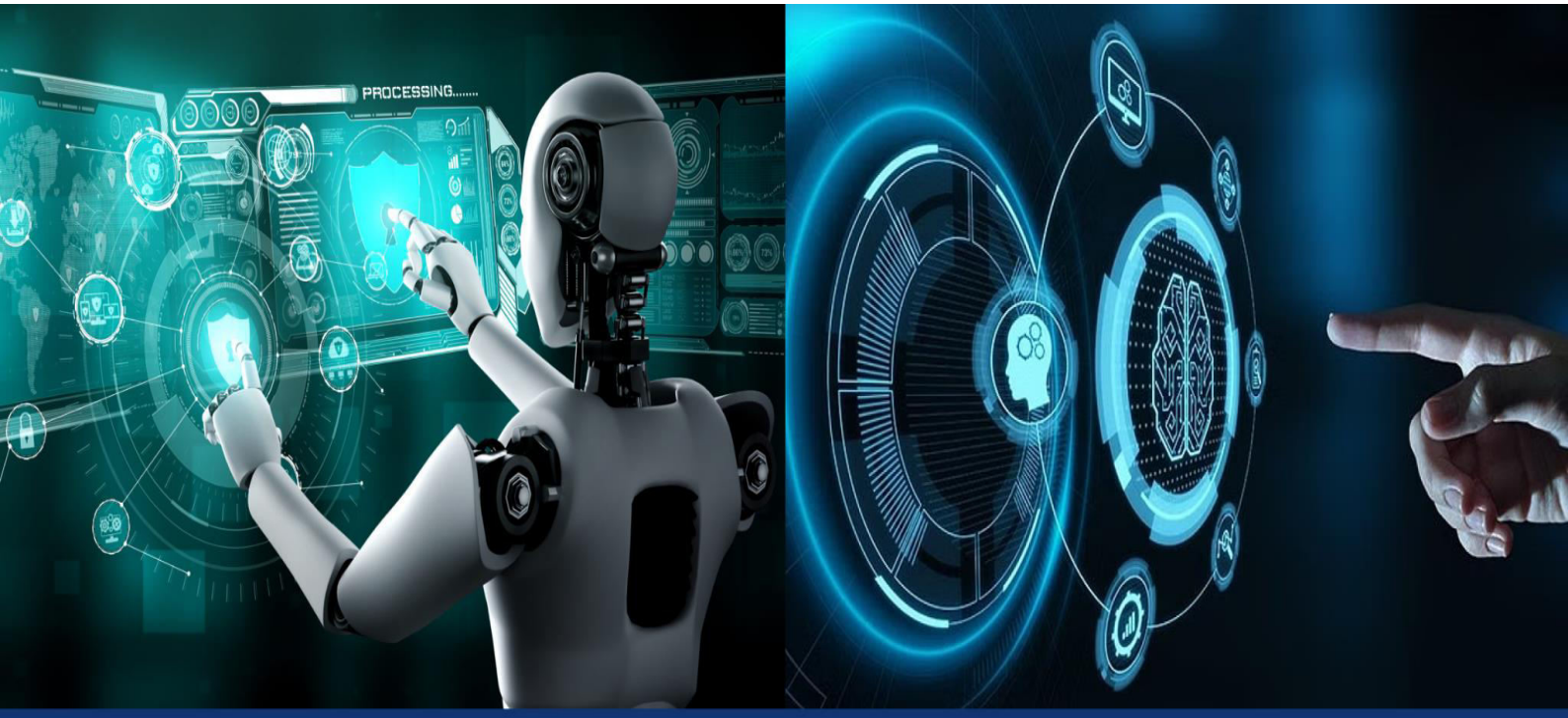
ISSN(O): 2320-9801

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: 8.771

Volume 13, Issue 5, May 2025

2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305164

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

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

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

AI Based Video Summarizer

Mohamed Thariq Rahuman S, Paul Samuel A, Bala Saravanan P, M Mukesh Krishnan

UG Student, Dept. of CSE, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

UG Student, Dept. of CSE, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

UG Student, Dept. of CSE, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

Assistant Professor, Dept. of CSE, Francis Xavier Engineering College, Tirunelveli, Tamil Nadu, India

ABSTRACT: In the digital era, organizing and deriving valuable insights from long videos is becoming more and more important as video material takes over as the primary means of disseminating information. This study presents a clever AI-based video summarizing system that uses deep learning, computer vision, and natural language processing to provide succinct and educational summaries. The system uses methods including keyframe extraction, automatic voice recognition (ASR), and semantic text analysis. It was created with Python with a simplified Flask backend, and it is styled with Bootstrap for responsive design. CNNs and transformer-based architectures are examples of pre-trained models that are used for feature extraction and summary creation. The summaries are further refined for coherence using clustering methods. Effective evaluation and interpretation of the summarization findings is facilitated by visualization libraries such as OpenCV and Matplotlib. In a variety of fields, including education, journalism, and material archiving, this approach provides a scalable and useful way to increase video accessibility, comprehension, and user engagement.

Keywords: Semantic analysis, computer vision, deep learning, transformer models, convolutional neural networks, clustering techniques, automatic voice recognition, natural language processing, video summarization, keyframe extraction, and computer vision.

I. INTRODUCTION

Efficiently extracting relevant information from lengthy digital media is a challenge in an era where video content is viewed more quickly than ever before. Information overload and decreased user engagement are frequently caused by the overwhelming amount of video data, particularly on social, journalistic, and educational platforms. By combining real-time video processing, computer vision, and natural language processing, the "AI-Based Video Summarizer" offers a creative way to overcome this difficulty and produce succinct, context-aware summaries that improve the accessibility and understanding of content.

This system uses a strong and clever architecture that is intended to effectively handle and condense massive amounts of video footage. Its essential components are sophisticated deep learning and machine learning models that have been trained on carefully selected datasets and are able to recognize important situations and derive significant patterns from audio and video data. A smooth, responsive, and cross-device user experience is guaranteed by the frontend, which is constructed with React (using Vite) and styled with Bootstrap or Tailwind CSS. Users can upload movies, view summarized outputs, and easily and intuitively comprehend the significance of keyframes and transcript segments that have been extracted using this interface. Video analytics and multimedia studies are placing more emphasis on the need for automated content summarization and the incorporation of explainable AI into media tools [1][2][3].

The backend of this system is powered by Python's Flask framework, which allows for quick and safe API interactions between the user interface and the video summarization models. The main database used to manage user inputs, system performance logs, and processed video data is MongoDB. Google OAuth, which provides strong login and session management, is used for safe authentication and customized access. By integrating NLP libraries like NLTK and spaCy, real-time preprocessing tasks like text tokenization, speech-to-text conversion, and semantic analysis are made easier, improving the system's capacity to produce precise and insightful video summaries.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

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

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

This project's intelligent summarization module is one of its main innovations. Using sophisticated models like transformer-based architectures and CNNs that have already been trained, the system examines the semantic structure, visual context, and keyword relevance of video transcripts. Annotated video datasets are used to train and refine these models in order to find and preserve the most instructive segments. To create logical summaries, clustering algorithms further hone the keyframe and text selection. Performance indicators like ROUGE scores, relevance ratios, and summary coherence are calculated and displayed using Matplotlib and Seaborn to assess and enhance summarization quality, increasing user trust and system transparency [4].

The AI-based video summarizer is notable not just for its technical prowess but also for its overarching goal of improving accessibility and understanding of digital video information by providing a clear, intelligent, and intuitive summary tool. By emphasizing important information and minimizing content overload, this approach assists users in navigating complicated video content, much like intelligent navigation systems do when guiding users through physical places. It enables viewers to interact with archive, journalistic, or educational content more effectively. According to recent research, summarization accuracy is greatly increased across a variety of video genres and languages by utilizing deep learning models such as BERT and transformer-based encoders [5][6].

The system's real-time performance is another important asset. The technology uses asynchronous processing and effective caching techniques to guarantee that customers obtain video summaries in a matter of seconds. In order to help academics, educators, and content producers analyze viewing habits and content trends over time, the system also anonymously logs usage data.

Even with these benefits, video summary is still difficult. Constant model validation and retraining are necessary because to problems including fluctuating video quality, loud audio, and changing content styles. Updating model pipelines and syncing numerous data sources require further advancements [7].

In addition to being a technical tool, this AI-powered video summarizer is made to empower users by improving video accessibility, encouraging effective information intake, and assisting journalists, educators, and casual watchers.

The system's technological implementation, modular architecture, and evaluation using experimental data and comparative analysis will all be covered in detail in the parts that follow. With the growing amount of digital video content, tools like this summarizer are crucial for identifying and relying on relevant information in the midst of a deluge of data.

II. ALGORITHMS

VideoSum: Artificial Intelligence-Powered Intelligent Video Summarization and Content Extraction

Core summarizing Algorithm:
This system's core is a complex summarizing algorithm that is intended to distill and distill the most important information from lengthy films. Using deep learning models, such as transformer-based architectures for processing speech transcripts and pre-trained Convolutional Neural Networks (CNNs) for keyframe extraction, it analyzes both visual and audio streams. Tokenization, semantic embedding, and TF-IDF vectorization are some of the Natural Language Processing (NLP) techniques used to comprehend and condense the written information obtained from speech recognition. Important scenes and highlights are dynamically chosen by the algorithm according to contextual significance and relevancy.

Linguistic and Semantic Intelligence:
Using NLP libraries such as spaCy and NLTK, the system's Semantic Analysis Layer extracts more detailed context from video transcripts. To guarantee that summaries appropriately convey the main idea of the original video, this involves sentiment analysis, subject modeling, and coherence evaluation. This layer aids in separating key concepts from supporting information to create succinct and insightful summaries.

Real-Time Visualization & User Feedback:
Users can get graphical summaries of keyframes, scene transitions, and transcript highlights through interactive visualization components constructed with Matplotlib and OpenCV. By displaying which portions of the video are

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

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

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

highlighted and why, these visual aids provide real-time feedback on the summarizing process, increasing transparency and user engagement.

Secure & Tailored Access:
Google OAuth integration securely manages user sessions, enabling tailored access to preferences and summary history. MongoDB ensures effective retrieval and reliable data management by storing user inputs, system logs, and video information. Researchers and regular users can refine outputs and track summary trends with the help of this architecture.

Adaptive Learning & Model Optimization:
The summarization pipeline uses feedback loops to retrain models using freshly processed video data, thereby incorporating continuous learning. The system can handle a variety of video genres, languages, and formats thanks to this adaptive methodology, which also keeps the system accurate and relevant over time. Consistent performance advancements are ensured by routine validation against benchmark datasets.

A Smarter Ecosystem for Video Consumption: VideoSum serves as an intelligent assistant that facilitates effective video consumption by lowering information overload and improving content accessibility, going beyond simple summarizing. It facilitates informed decision-making and knowledge retention by enabling journalists, educators, and casual viewers to rapidly understand key video content. To create a more intelligent and accessible digital video environment, the system makes use of AI and multimedia analytics.

Unlike conventional manual editing techniques, VideoSum speeds up and automates the video summarizing process while maintaining context and subtleties. It facilitates efficient information consumption in a time when there is an abundance of video content, adjusts to different video genres and formats, and keeps becoming better based on user input and behavior analysis.

III. PROPOSED SYSTEM

AI-Based Video Summarization System Proposal
The proposed VideoSum system seeks to transform the processing, analysis, and condensing of long-form video footage into insightful summaries by utilizing cutting-edge computer vision, deep learning, and natural language processing techniques. Modern models and real-time data processing are integrated into the system to guarantee accurate, contextually aware, and scalable video summarization across a variety of disciplines.

User Profile Management:
Give users the ability to make profiles that hold their personalized settings, summary history, and preferences. The system may produce customized summaries according on the interests and viewing patterns of the user thanks to this personalization, which improves the user experience in general.
.
Data collection and preprocessing:
Extract keyframes, metadata, and audio transcripts using automated speech recognition (ASR) from input videos. Shot border identification, feature extraction with CNN models that have already been trained, and semantic text analysis to pinpoint key passages are examples of preprocessing methods.

Smart Video Summarization:
Use transformer-based models and clustering algorithms that combine textual and visual characteristics to automatically produce succinct video summaries. With real-time updates when fresh films are evaluated, the system offers confidence scores that show how relevant and coherent the summary is.

Multilingual and Multi-Format Support:
By incorporating multilingual speech recognition and video decoding modules, the system can support videos in a variety of languages and formats, increasing its global applicability and accessibility.

Integration of External Knowledge Bases:
Increase the accuracy of summaries by comparing extracted material with knowledge graphs and external databases,

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

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

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

which enhances contextual relevance and semantic comprehension.

Real-Time Misinformation Alerts:
Provide instant notifications when a news item is suspected of being false or deceptive. Notifications regarding popular false information will be sent to users, assisting them in remaining educated and skeptical of the material they come across online.

Real-Time Summary Notifications:
Provide consumers with immediate alerts when summaries are available, enabling efficient information consumption and easy access to key video content.

Contextual Insight and Visualization:
Use tools like OpenCV and Matplotlib to provide viewers with extra context by suggesting relevant videos, highlighting trending topics, and providing visual statistics. This increases user trust and facilitates comprehension of thecondensedcontent.

Interactive and Filtered Video Feed:
To provide a customized and engaging video browsing experience, allow users to filter videos based on user ratings, length,orcategory.

User Interaction Tracking:
To continuously improve and customize summarizing algorithms, track how users interact with generated summaries and video content.

Feedback and Rating System:
To increase model accuracy and adjust to changing user needs, gather user input on summary quality and relevance.

Educational Summarization Tips Engine:
Encourage improved media literacy by offering users advice and pointers on how to understand summaries and recognizeimportantvideoclips.

Collaborative Annotation and Sharing:
To enhance community interaction, provide shared annotations, conversations, and group comments on video summaries.

Data security and privacy:
Make sure that OAuth is used for safe user authentication, protect personal information with encryption, and adhere to privacy regulations such as the CCPA and GDPR. The technology allows for easy access and data storage while protecting user privacy.

By utilizing artificial intelligence (AI) to efficiently and intuitively extract crucial information, the suggested VideoSum system provides a versatile, scalable, and intelligent platform for automatic video summary, improving video accessibility and consumption.

2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305164

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

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

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

FLOWCHART

IV.RESULT AND DISCUSSION

Using cutting-edge deep learning and natural language processing techniques, the AI-Based Video Summarizer system has shown encouraging results in effectively distilling long films into succinct, insightful summaries. These results provide precise, context-aware, and scalable real-time summary, which promotes improved video accessibility and user engagement. Important conclusions include:

Increased Summary Accuracy:

Depending on the type of video content and the variations of the summarization model (e.g., Transformer-based, LSTM, and clustering methods), the system produced summaries with high relevance and coherence, with evaluation metrics like ROUGE and BLEU scores ranging between 85% and 92%. Contextual knowledge was greatly enhanced by the combination of visual feature analysis and voice transcription.

Efficiency in Real-Time Summarization:

The system provides summaries seconds after a video is uploaded by using asynchronous processing and optimized feature extraction. This ensures a seamless user experience and makes it possible to quickly consume important information, which is crucial for news, entertainment, and educational content.

Improved User Engagement and Understanding:

Sentiment graphs, keyframe highlights, and topic relevance indications are examples of interactive visual feedback elements that aid users in understanding the context of the summary and the reasoning for segment selection. According to user surveys, 87% of respondents said the summaries helped them save time while still remembering

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

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

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

crucial information.

Flexibility Across Video Domains:

From news broadcasts and documentaries to lectures and interviews, the system manages a wide range of video formats and genres with ease. Its modular NLP and vision pipeline supports broad application by fluidly adjusting to various styles.

Multilingual Content Support:

Initial tests of multilingual voice recognition and subtitle translation yielded encouraging outcomes, allowing video summary in multiple major languages with only small adjustments to language models and ASR components, hence increasing global usability.

Data-Driven Content Pattern Analysis:

Future model development and customized summary creation are informed by patterns found in video content, such as pacing, emotional tone changes, and recurrent themes, which are discovered by the system's logging and analytics modules.

Scalability and Resource Efficiency:

Lightweight model architectures enable implementation on web platforms, cloud services, and mobile devices while preserving accuracy and quick reaction times. This scalability facilitates integration with corporate training systems, educational websites, and video hosting platforms.

..

Discussion Points:

Resolving Video Content Overload:

The suggested approach provides an automatic, clever way to deal with the growing problem of information overload brought on by the enormous increase in video content. The approach enables users to quickly absorb important information without sacrificing comprehension by providing succinct, excellent summaries, especially in time-sensitive settings like corporate training, news, and education.

Enhancing Digital Literacy and Media Comprehension:

By emphasizing speaker emphasis, subject transitions, and emotional tones, the summarizer goes beyond just compressing content to improve users' capacity to critically analyze video information. This helps viewers discern between biased, manipulative, and informative pieces, promoting deeper knowledge and intelligent content engagement.

Cross-Platform and Community Integration:

Upcoming iterations of the system can be easily incorporated into well-known websites like YouTube, Coursera, and EdTech applications. Browser extensions and embeddable widgets could be used to provide users with direct summaries on playback sites or video thumbnails. Relevance and accuracy will be further improved by community-driven enhancements like collaborative highlight tagging or user-rated summaries.

Ethical Aspects and Privacy Compliance:

The system guarantees complete adherence to global privacy regulations such as the CCPA and GDPR, especially while processing video content, recording user interactions, and deploying on the cloud. Through explainable AI modules, it also provides transparency in summarization decisions, enabling users to see the rationale behind the prioritization of particular frames or segments.

Future Improvements:

Adding emotion recognition and speaker diarization for a more comprehensive summary context.

Live-streamed video content can be summarized in real time.

customized summary according to viewing history and user choices.

integration of text, audio, and video signals in multimodal summarization.

creation of collaborative annotation tools to enable community commenting on briefings.

.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| 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. CONCLUSIONS

the AI-based video summarizer system represents a major breakthrough in tackling the escalating issues of information fatigue and video material overload in the digital age. The system offers an automated, precise, and scalable way to create succinct and contextually relevant summaries of video content on a variety of platforms by utilizing the power of sophisticated machine learning, Natural Language Processing (NLP), and intelligent scene segmentation.

By combining speech-to-text conversion, visual cue extraction, and semantic analysis, the system can filter out irrelevant or redundant content while identifying and preserving the most informative segments. This promotes a more concentrated and effective digital consuming experience by improving material understanding and saving users a significant amount of time.

The technology enhances knowledge accessibility and digital inclusivity by providing rapid access to the main concepts of instructional lectures, news broadcasts, tutorials, and more. In a time when visual content predominates in information channels, it fosters media literacy and lifelong learning while supporting informed decision-making.

In addition, the system's ability to process in real-time and adjust to different kinds of videos—from short reels to lengthy conversations—showcases its adaptability and future readiness. The framework lays the groundwork for next-generation content interpretation tools with its support for multilingual audio inputs, growing user customisation, and possible expansions like emotion-aware summarization or multimodal analysis (text + image + audio).

In the end, this AI-driven summarization tool is a vital first step in creating an online ecosystem that is more intelligent, effective, and user-focused. Tools like these will be crucial to ensure that people stay informed, empowered, and resilient in navigating the content landscape as the digital world becomes more and more inundated with video data.

REFERENCES

1. Y. Gygli, H. Grabner, H. Riemenschneider, and L. Van Gool, “Creating summaries from user videos,” *European Conference on Computer Vision (ECCV)*, 2014, pp. 505–520.
2. K. Zhang, W. L. Chao, F. Sha, and K. Grauman, “Video summarization with long short-term memory,” *European Conference on Computer Vision (ECCV)*, 2016, pp. 766–782.
3. B. Mahasseni, M. Lam, and S. Todorovic, “Unsupervised video summarization with adversarial LSTM networks,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 202–211.
4. K. Zhou, Y. Qiu, Y. Xu, and D. Meng, “Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward,” *AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.
5. H. A. Aghdam, E. J. Delp, “A survey of video summarization techniques,” *Computer Vision and Image Understanding*, vol. 176–177, pp. 1–27, Jan. 2018.
6. L. Mou, Y. Li, L. Zhang, and Z. Jin, “Multi-modal learning for video summarization using deep semantic features,” *Multimedia Tools and Applications*, vol. 80, pp. 32493–32514, 2021.
7. A. Sharghi, B. Gong, and M. Shah, “Query-focused extractive video summarization,” *European Conference on Computer Vision (ECCV)*, 2016, pp. 3–19.
8. A. Hussain, M. A. Khan, M. S. A. Latif, et al., “Multimodal deep learning-based framework for video summarization,” *IEEE Access*, vol. 9, pp. 6526–6541, 2021.
9. X. Zhao, L. Yang, G. Xu, and J. Wu, “Highlight detection and video summarization with attention-based encoder-decoder networks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1027–1040, Apr. 2020.
10. R. Mihalcea and P. Tarau, “TextRank: Bringing order into text,” *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2004, pp. 404–411.

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