IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Literature Review on Different LTSM-based Video Summarization Methods

[1]T.Gaikwad, [2]R.Pachlor
[1]iMtech CSE Student, [2]Professor
[1,2]Computer Science and Engineering Department,
[1,2]MIT Art Design and Technology University, Pune, India

Abstract – One of the most pressing issues in multimedia analysis in the present day digital age is video summarising (VS). Many other deep learning–based VS strategies have been proposed. To analyse, extract, and derive information from long videos in the quickest time feasible, they are ineffective. Extensive research and analysis of a wide range of deep learning methods were conducted to identify the root causes of problems that are connected with different deep learning strategies when it comes to identifying and summing the crucial events in such films. Several deep learning algorithms have been studied and investigated for their event detection and summarization capabilities. Selecting Intervals for Keyframes The detection of events, their categorization, and their aggregation are all tasks that fall under the purview of individual activities. Further, the limitations of each category are broken very thoroughly. The use of a deep network to spot low activity on a wide range of publicly available information raises some concerns as well. It is suggested that practicable approaches be employed to assess and improve video summaries generated from such data. Also included is a list of apps that have been recommended by the literature. Several deep learning methods that might be implemented in future studies are also discussed. The lecture continues with a discussion of possible future directions for further research in VS leveraging deep learning approaches.

Keywords: **Event summarization; Critical information in videos; Surveillance systems; Video analysis; Multimedia analysis.**

I. Introduction

Video summarization is the most advanced technological aspect that effectively utilizes the quality summarization of long video into a shorter format as regards effective analysis. Video summarisation proves to be beneficial and effective in grasping information intuitively and comprehensively. This research report in this context explores and identifies the challenges and issues with the contextual developments and approaches in video summarization techniques. Video summarization techniques have been considered a greater aspect of recent technological advancement with the effective and quality assurance in the summarisation results. Videos often involve a sequential hierarchical structure which is considered to be a complex issue in the existing summarization method relating to the degradation of the quality of the summarization. This report by exploring the different contextual research articles evaluates the effectiveness and significance of methodological and framework implementation in achieving greater significance in the video summarisation aspects.

II. Discussion

Considering the effective utilization of video summarization approaches in modern applications, the advancements have been attributed to the related development and inclusive growth in the rapidly growing field. With the changing dimension and increasing consumption of video content, it leverages the development of the most effective methods and techniques as regards video summarization as regards understanding, and developing quick navigation in the process.

III. Types of Video Summarization

While considering the types of most relative video summarization aspects, there are two contextual types including extractive and abstractive summarisation which reflect on the most related understanding and knowledge attribution of technological advancements. The extractive video summarisation defines the segmentation of frames for the greater representation of content in providing a structural summarization. It relates often to the various contextual aspects such as color and texture for greater and more efficient video summarization.

In the study by Agyeman *et al.* 2019, the authors effectively utilized the deep learning approach as regards the combined aspect of 3D - CNN and LSTM, and RNN so far as the summarisation of football videos is concerned. The authors demonstrated the summarization of the football movies and collected long videos and measured with the MOS as regards identifying the greater summarised version of videos. Though the model was initially developed for the football analysis, it can be demonstrated and applied to several other contextual sports representations and analysis of video summarisation aspects.

Video summary is one of the many industries where CNN (Convolutional Neural Network) technological advances have been found useful. Video summary is the process of extracting the most important details and headlines from long films so that viewers may rapidly understand the material without viewing the full thing (Agyeman *et al.* 2019). Approaches based on CNN have been successful in completing this goal. With the effective type of video summarization approaches of shot segmentation, predictive memorability, and entropy scores, Muhammad *et al.* 2020, effectively articulated the greater aspect of CNN-based video summarization with the effective demonstration of keyframe selection. The article by Sharma and Sungheetha (2021) well-defined the surveillance using video summarisation with effective CNN and SVM models. The article demonstrates the methodologies and techniques as regards the processing and analysis of video frames with the systematic approaches of dimension reduction, anomaly detection, and feature extraction contributing to the summarization process.

The authors Yoon *et al.* 2023 in this regard conducted experiments so far as evaluating the proposed framework of accuracy and computational efficiency are concerned with the effective utilization of publicly available data and information. The proposed framework regards the video summarization and activity recognition system in technological advancement by understanding the key activities, deep autoencoder, deep learning, and IoT sequential video summarisation and data analysis approaches. CNN models are well-suited for video summarization because they are excellent at understanding spatial characteristics and patterns from pictures. An already-trained CNN model is used to process video frames that have been extracted from the video (Hussain *et al.* 2019). Each frame of a video is captured by the CNN, and it subsequently identifies keyframes or significant portions. Considering the existing framework of the MVS aspects and functions, the authors proposed a more agile and inclusive framework articulating cutting-edge applications of artificial intelligence and IoT-enabled devices.

In addressing the issue of short segmentation in video summarisation, Zhao *et al.* 2018 proposed a framework of HSA-RNN as regards achieving effective summarization. Considering the approach of video summation, the datasets of SumMe, TVsum, CoSum, and VTW have been utilized which demonstrated the most contextual results and findings in generating high-quality summarisation.

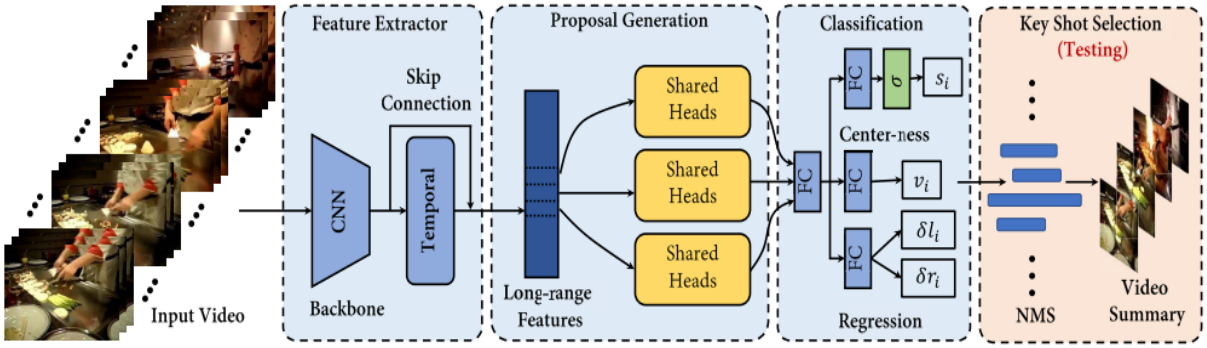

figure 1:anchor based video summarisation (source: https://ojs.aaai.org)

The study of Yoon *et al.* 2023 well articulated the unsupervised video summarisation aspects and deep reinforcement learning approaches provided a greater perspective of the long video shortening in ensuring the generated summarization more efficiently and consistently. In the article, Xiong *et al.* 2015**,** explored the challenges and contextual aspects of deep reinforcement learning for video summarization**.** The proposed method incorporated long-term dependency, auxiliary summarisation, and reward functioning for the greater summarisation of the deployment of mobile devices

The article by Archana and Malmurugan (2021) well defines the most contextual video summarization process in the real-time application of different aspects of the MOLE VS methodology in the process. The paper integrated the most contextual types of video summarization processes which include the frequency domain of edge detection for the comprehensive investigation of the precision and processing time of the video summarization aspects. Considering the findings and outcomes from the proposed MOLARS frameworks, the authors demonstrated a greater performance and development of overall efficiency and excellency of the video summarization process.

IV. Techniques of Vidoe Summarization

In the context of articulating the most effective techniques for video summarisation, recent developments have been identified as keyframe extraction, shot segmentation, feature-based segmentation, and deep learning approaches. Keyframe extraction involves the technique of various methodological implications with the clustering and graph-based algorithm representation for the inclusive informative frames in the video summarization. Shot segmentation involves the sequential segmentation of video frames in short and continuous shots with the analysis of the detection of motion and a histogram-based approach. Whereas feature-based summarisation identifies the key aspects

of the visual and audio representation in obtaining the effective extraction approach. The most advanced approach of deep learning has revolutionized the field of video summarization with the implied neural networks in optimizing visual features.

The most related research papers in the context of video summarization include the article by Agyeman *et al.* 2019, which effectively utilizes deep learning strategic implications in the soccer video summarization techniques. The authors effectively implemented 3D conventional neural networks in the soccer video summarization enabling spatiotemporal features and dynamics of the game frame distribution. In this study paper, the authors effectively summarised the modeling of football footage in sequential videos. The structural detection system on the soccer data with the effective utilization of ResNet-based 3D modeling. The combined approach of 3D-CNN and LSTM proved to be effective as regards the spatial and temporal information in the effective determination of the video summarization process.

The research study of Muhammad *et al* 2020 strategically utilized and implemented convolutional neural networks as regards the video summarisation aspects with effective resource utilization. The authors more effectively utilized video summarization approaches such as shot segmentation, picture memorability, and entropy score by understanding the characterization of CNN ensuring a greater impact on the technical implications of the video summarization techniques. The authors successfully employed both memorability and entropy in the keyframe of the video shot segmentation.

The study paper of Sharma and Sangeetha (2021) effectively approaches the discussion of video processing by classifying the combined aspect of CNN and SVM architecture. The proposed model and framework provide a greater aspect of video surveillance using the CNN and SVM architecture for video summarization with the condensing of long videos into segmented parts and effective content analysis.

In the research paper, Liu *et al.* 2023 strategically proposed unsupervised video summarization techniques with the effective utilization of deep reinforcement learning, interpolation, temporal consistency reward, and graph-level characteristics. The proposed framework reinforces the learning problem with the maximization of keyframes generated through a natural sequence of the network. Combining the CNN networks the author represents reward functioning with the inclusion of diverse keyframes. Considering the technical aspects, and articulated results, the proposed framework and methodological aspects prove to be effective for short-frame video summarization.

In this context of video summarisation aspects, the study of Haq *et al.* 2020, provides deep summarisation models based on the GRU and LSTM approaches to address the issues of lengthy dependencies in video segmentation. The process of video summarising, which entails reducing a lengthy film to a smaller summary video that contains the most crucial and instructive material, has been effectively implemented using recurrent neural networks (RNNs). RNNs, especially Long Short-Term Memory (LSTM) networks, have demonstrated success in modeling sequential data, making them appropriate for jobs involving video interpretation (Agyeman *et al.* 2019). The model represented in the proposed framework identified an unsupervised framework of deep reinforcement learning in the various parts of the summarisation aspects.

In the article, Yoon *et al.* 2023 effectively represented the most contextual usage of unsupervised video summarization with keyframes integration and reward function in achieving successful video summarisation for the long video summarization in a shorter version. The authors determined the proposed method by leveraging higher quality and performance of video shortening in the process.

The study of Archana and Malmurugan (2021) provided the most effective and novel video summarization approach with precision, recall, F-score, and processing time for comprehensive and inclusive metrics evaluation. The findings and results as regards the application of most related evaluation techniques and metrics provide the greater beneficial aspect for the optimized video summarisation aspects.

V. Evaluation metrics of video summarization

In accessing the greater effectiveness of video summarisation the most contextual metrics that are widely utilized and implemented in the video summarization approach include coverage, redundancy, representativeness, user satisfaction, and F-measure. All these metrics of evaluation in the video summarization aspect prove to be effective and relative in understanding and overall knowledge attribution. Coverage ensures the effective representation of content in the original video. F-measure is considered to be the wider usage in evaluating the quality and relevancy of comprehensive video summarization.

The study paper of Archana and Malmurugan (2021) effectively utilized the most contextual evaluation metrics in the video summarisation aspects which include processing time, F-score, precision, and recall as regards the proposed effectiveness of the LSTN, RNN, and MOL RVS for summarising the information in the video segmentation process. The sequence-to-sequence approach, which is composed of an encoder-decoder structure, has become a popular architecture for summarising videos using RNNs. For the purpose of preserving time-related information, the encoding RNN analyses the input frames as well as encodes those to create a fixed-dimensional representation (Lin *et al.* 2022). After decoding the encoded illustration and creating the chosen or created messages, the decoder RNN next creates the summary animation.

The evaluation metrics that have been effectively utilized and implied in the study of Xiong *et al.* 2015 well represented the experimental results as regards the proposed multiple inter-symbol obfuscations enhancing the wireless networks. Information-theoretic secrecy, computational secrecy, and BER evaluation metrics provided secure and confidential communication.

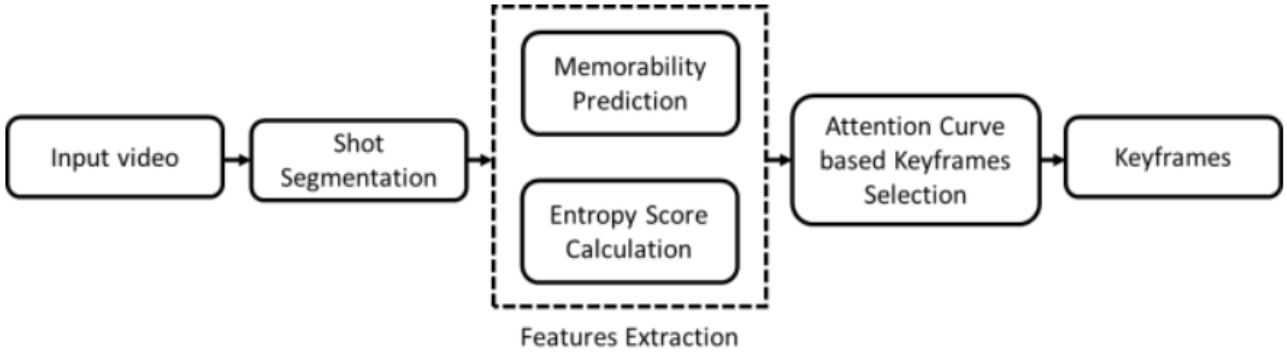

figure 2: framework of video segmentation (source: https://www.sciencedirect.com)

In the context of evaluating the video summarization techniques, the study of a Model Based on Deep Reinforcement articulated the most effective evaluation metrics such as F-score and explored dataset results in accessing the performance and effectiveness of the proposed method compared to other approaches.

Another contextual research paper by Yoon *et al.* 2023 defined the most articulated and effective evaluation metrics for the proposed unsupervised video summarisation methodological approach with the reinforcement of learning with interpolation. The authors strategically compared and evaluated the existing methods with the exploration of datasets and methods in accessing the video summarisation aspects in assessing the performance of the process. An appropriate loss function is used for conditioning the RNN model, frequently combining reconstruction loss, which gauges how similar the initially captured and produced frames are, and relevance score loss, which incentivizes the neural network to concentrate on crucial frames. In the field of image processing and video processing, the study paper of Sharma and Sangeetha (2021) utilized the common and existing evaluation metrics of video surveillance and machine learning approaches which include accuracy, efficiency, and loss as regards the proposed CNN-SVM architecture of video summarization.

With the specific evaluation of most related metrics, the study of Agyeman *et al.* 2019, defines the contextual MOS, precision, F-score, and user satisfaction in evaluating the video summarization process for the assessing quality, and effectiveness of the video processing aspects.

VI. Application of video summarization

Video summarization in a modern context has multifaceted implications and utilization as regards maintaining the important contents and information of the long videos in the shorter version. Across various domains and areas, video summarization seems to have different usage and implications. This approach of video summarisation defines the improved efficiency, time savings, and overall enhancement of video experience in the video summarization process. In the practical context, video summarization is identified to have a usage in surveillance, media, video production, sports analysis, e-learning and medical analysis approach, and traffic management. Sports video summary, security video analysis, and social networking material summarization are just a few of the areas where RNN-based video summarization techniques have produced promising results (Yu *et al.* 2020). They give customers a simple method to understand the essential points of long movies and shorten the amount of time needed for video viewing by automating and effectively condensing lengthy recordings into summaries. In this regard, the exploration of contextual analysis of several articles would provide a greater perspective and overview of the video summarisation usage and implementation in real-world examples.
The author Agyeman *et al.* 2019, provided the application of video summarization approaches in football matches in shorter versions with the summarised version. With The proposed model of 3D CNN and LSTM frameworks, the author argues that spatial learning allows viewers a quick grasp of the video presentation and summarised version.

The application of video summarisation as demonstrated by Muhammad *et al* 2020 provides efficient usage in surveillance with the effective utilization of shot segmentation addressing the challenges of administrative datasets. The authors discussed the potential application in wireless networks and IoT-enabled devices for intelligent monitoring with the resource-based summarization of videos. Whereas the article by Sharma and Sangeetha (2021) does not directly reflect any specific usage of video summarisation, it focuses on the accuracy and efficiency of the surveillance of abnormal incidents in the systems.

The proposed video summarisation network combining transformer and CNN by Yoon *et al.* 2023 averages a more inclusive and deep reinforcement of learning with accurate and efficient video summarisation in the easier representation of large numbers of videos. Another contextual study by Model-Based on Deep Reinforcement provides the greater implications and application of video summarisation in performance refinement of suitability for mobile devices. There are several benefits to utilizing CNN for video summaries. CNN models are trustworthy for locating pertinent material in videos since they have proven to be very accurate in tasks including object identification, scene interpretation, and feature extraction. Additionally, they are able to provide insightful and well-organized summaries thanks to their capacity to record geographical and temporal information. The study aims to explore the coherent and representative usage of video summarization with the deep reinforcement learning model. Though the article of Xiong *et al.* 2015 does not define any specific video summarization aspect but articulates enhancing wireless communication approaches in protecting the confidentiality and secrecy of the proposed model of MIO and its effectiveness.

In the study of Archana and Malmurugan (2021)**,** the author well articulated the effective utilization of video summarisation in real-time streaming situations. In this regard, the proposed method of the MOLARS provided the improved and effective utilization of video summarisation in the various contextual streaming scenarios. The detection of the F-score, recall, and precision, the proposed method promised to have a greater impact on the processing efficacy and overall summarization of video solutions.

VII. Conclusion

The detailed discussion in the context of exploring the video summarisation aspects in the various industrial applications and effective integration of types and methodologies as regards the implementation of video processing time and efficiency. In this context, the exploration of various contextual articles and authors related to the determination of an effective proposed framework and methodological aspect guides the development of comprehensive and inclusive understanding and knowledge attribution. The findings and outcomes from the detailed review of the various articles contribute immensely to the field of video summarization with the effective utilization of hierarchical structure and adaptiveness promising the most contextual solutions in a concise manager. By articulating the challenges and problems associated with short segmentation and video summarisation the research paper outlines the effective evaluation metrics and process of summarisation in improving the overall video processing aspects. The reviewing of the articles contributes to effective video summarization by addressing the challenges as regards the concise understanding of content analysis, computer patterns, and vision recognition. The thorough discussion of the technical aspects and methodological framework employed in the various literary articles provided the effectiveness of the proposed approach in the quantitative evidence of the video summarization process. Considering the review of the types of summarisation, evaluation of metrics, and application of video summarisation, this report evaluates the effectiveness and significance of methodological and framework implementation in achieving greater significance in the video summarization process.

VIII. References

[1] Agyeman, Rockson, Rafiq Muhammad, and Gyu Sang Choi. "Soccer video summarization using deep learning." 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, 2019.

[2] Alexoudi, Panagiota, Ioannis Mademlis, and Ioannis Pitas. "Escaping local minima in deep reinforcement learning for video summarization." (2023).

[3] Archana, N. and Malmurugan, N., 2021. RETRACTED ARTICLE: Multi-edge optimized LSTM RNN for video summarization. *Journal of Ambient Intelligence and Humanized Computing*, *12*(5), pp.5381-5395.

[4] Cai, Sijia, et al. "Weakly-supervised video summarization using variational encoder-decoder and web prior." Proceedings of the European conference on computer vision (ECCV). 2018.

[5] Haq, H.B.U., Asif, M. and Ahmad, M.B., 2020. Video summarization techniques: a review. *International Journal of Scientific & Technology Research*, *9*, pp.146-153.

[6] Hussain, Tanveer, et al. "Cloud-assisted multiview video summarization using CNN and bidirectional LSTM." IEEE Transactions on Industrial Informatics 16.1 (2019): 77-86.

[7] Hussain, Tanveer, et al. "Multiview summarization and activity recognition meet edge computing in IoT environments." IEEE Internet of Things Journal 8.12 (2020): 9634-9644.

[8] Liu, Hanchi, et al. "Application of Deep Learning-Based Object Detection Techniques in Fish Aquaculture: A Review." Journal of Marine Science and Engineering 11.4 (2023): 867.

[9] Muhammad, Khan, Tanveer Hussain, and Sung Wook Baik. "Efficient CNN based summarization of surveillance videos for resource-constrained devices." Pattern Recognition Letters 130 (2020): 370-375.

[10] Sharma, Rajesh, and Akey Sungheetha. "An efficient dimension reduction based fusion of CNN and SVM model for detection of abnormal incident in video surveillance." Journal of Soft Computing Paradigm (JSCP) 3.02 (2021): 55-69.

[11] Xiong, Tao, et al. "MIO: Enhancing wireless communications security through physical layer multiple inter-symbol obfuscation." IEEE transactions on information forensics and security 10.8 (2015): 1678-1691.

[12] Yoon, Ui Nyoung, Myung Duk Hong, and Geun-Sik Jo. "Unsupervised Video Summarization Based on Deep Reinforcement Learning with Interpolation." Sensors 23.7 (2023): 3384.

[13] Yu, Tianshu, Yikang Li, and Baoxin Li. "Rhyrnn: Rhythmic rnn for recognizing events in long and complex videos." Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16. Springer International Publishing, 2020.

[14] Zhao, Bin, Xuelong Li, and Xiaoqiang Lu. "Hsa-rnn: Hierarchical structure-adaptive rnn for video summarization." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[15] Zhou, Kaiyang, Yu Qiao, and Tao Xiang. "Deep reinforcement learning for unsupervised video summarization with diversity-representativeness reward." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 32. No. 1. 2018.

[16] Zhu, Wencheng, et al. "Dsnet: A flexible detect-to-summarize network for video summarization." IEEE Transactions on Image Processing 30 (2020): 948-962.

[17] Lin, Jingxu, Sheng-hua Zhong, and Ahmed Fares. "Deep hierarchical LSTM networks with attention for video summarization." Computers & Electrical Engineering 97 (2022): 107618.