CCTV Footage Summarization for Increasing Efficiency in Surveillance (using ML)

Lishika Goel[1]; Rachna Jain[2]

[1,2]Amity Institute of Information Technology, Amity University Uttar Pradesh,Noida, India

Publication Date: 2025/05/09

Abstract: Surveillance cameras are ubiquitous in this era of the future. Whether a house, office, shopping complex, or highway, CCTV cameras are ubiquitous that monitor the activities on a day-to-day basis. Surveillance is a significant activity. They help in identifying any suspicious activity and serve as effective evidence. Yet, the quantity of video recording is huge. Useful information it holds most of the time is extremely minimal. It then becomes an extremely challenging task to sift through such many hours of video to pick out useful information. Storage space for the videos is also wasted on a vast scale. So, there's a need perceived for surveillance products that can make such long recordings into short ones without losing essential events. Our system can condense hours of video recorded by CCTV cameras into a single clip that displays all interesting events at once. All moving objects in the video are detected and tracked by our software. These events are overlaid on one clip, and the timestamps for each event are also displayed, thereby allowing the user to conduct surveillance for multiple events. Our model employs the KNN model for background subtraction and video extract in addition to a special Object Tracking algorithm to identify moving objects that are overlapped on the extracted background.

***Keywords**: Video summarization, Surveillance, KNN, Object Tracking algorithm, Multi-video summarization, Deep learning.*

How to Cite: Lishika Goel; Rachna Jain. (2025). CCTV Footage Summarization for Increasing Efficiency in Surveillance (using ML). *International Journal of Innovative Science and Research Technology*, 10(4), 2927-2933. https://doi.org/10.38124/ijisrt/25apr1554.

I. INTRODUCTION

The increasing number of CCTV surveillance cameras leads to an overwhelming amount of raw, unprocessed digital video data being recorded and stored, frequently in a continuous 24/7 stream. This wealth of footage, including times with little motion, creates storage issues and renders the conventional approach of manually viewing and analyzing footage tedious and cumbersome. To address this challenge, the CCTV Video Summarization Project is a groundbreaking solution to the video surveillance industry. Acknowledging the central role that surveillance systems have in ensuring security, this project utilizes sophisticated algorithms and artificial intelligence to revolutionize how we tackle video analysis. The goal is to automatically summarize long surveillance videos into short and meaningful summaries.

The CCTV Video Summarization Project solves this problem through the independent recognition of salient events by incorporating motion detection, object identification, and behavioural pattern analysis. What ensues is a lean and effective system that does away with the necessity for the human eye to go through hours of video footage. By bringing out the most applicable and important points, the project hopes to equip security officers, law enforcement authorities, and other stakeholders with a powerful tool for rapid decision-making, proactive action, and efficient investigations. This opening provides the context for an examination of the technologies, methods, and possible uses that make the CCTV Video Summarization Project a revolutionary leap forward in the field of video surveillance and security in the modern fast-paced world.

A. Video Summarization

Video summarization is to create a concise summary of the content of an extended video document by choosing and displaying the most informative or most interesting materials to potential users. The output summary is typically made up of a series of keyframes or video clips taken from the original video with some processing operation.

B. Need of Video Summarization

The last decade has witnessed rampant expansion in the capability of people to produce and/or record digital video, gradually culminating in massive personal and business

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1554

digital video stores. In the business context, an increasing demand for video summarization.

C. *Types of Video Summarization:*

- *Temporal Summarization:*

This kind of summarization seeks to pick important frames or shots that summarize the temporal progression of the video. It preserves the main events and changes with time, giving a compact overview of the entire video.

- *Keyframe Extraction:*

Keyframes are selected frames that encapsulate the main content and information of a video. Keyframe extraction algorithms identify frames with significant visual changes, ensuring that the essence of the video is captured in a reduced form.

- *Shot-Based Summarization:*

Shots are consecutive frames captured from a single camera perspective without any interruption. Shot-based summarization involves selecting representative shots from different parts of the video to provide a coherent and concise narrative.

- *Object-Centric Summarization:*

This approach focuses on identifying and summarizing segments of a video where specific objects or entities of interest are present. It's particularly useful for videos with distinct objects or subjects that are central to the content.

- *Event-Centric Summarization:*

Event-centric summarization aims to capture and highlight specific events or actions occurring within the video. It identifies sequences of frames related to an event and compiles them into a summary that emphasizes event occurrences.

- *Motion-Based Summarization:*

This technique selects frames or shots based on the magnitude of motion detected in the video. It's effective for videos where motion is a key indicator of important content or events.

- *Content-Based Summarization*

Content-based summarization combines multiple features, such as visual cues, audio information, and textual metadata, to identify and extract the most relevant parts of a video.

- *Semantic Summarization*

Semantic summarization involves understanding the semantics of the video content, including objects, actions, and relationships between elements, to create a summary that captures the underlying meaning of the video.

II. LITERATURE REVIEW

Video summarization is now an urgent area of research, especially in the scenario of CCTV monitoring where there is a need for quick analysis of large amounts of video. Various studies have looked into various approaches, from classical methods to state-of-the-art deep-learning-based methods.

Hussain et al. [1] present a detailed survey on multi-view video summarization (MVS) with challenges like data amount and inter-view correlations. They also present future directions in research, such as more efficient algorithms and better evaluation metrics. Wang et al. [2] discuss deep learning-based video summarization techniques with limitations like computational expense, sensitivity to noise, and lack of interpretability, while appreciating their potential for progression in the area.

In the field of intelligent transportation systems, Zhang et al. [3] introduce a frame-filtering method for traffic monitoring videos, proving its efficiency in creating abridged summaries. Likewise, Wang et al. [4] study video summarization for socially conscious applications, defining challenges and current solutions. Multi-task video summarization is studied by Zhang et al. [5], who also define open research challenges and anticipate future directions. Event detection and visual question answering are also cited as applications in [6] and [7], respectively.

Deep learning methods have dominated video summarization. Deep learning-based approaches have been elaborated on by Ghosh and Ramanathan [8] while Apostolidis et al. [9] give an overview of current work on summarization based on neural networks. There has also been a deployment in reinforcement learning as witnessed through Wang et al. [10] who propose a pipeline architecture coupled with an auxiliary summarization loss. Khan et al. [11] examine the classic methods of summarization, giving an illustrative point of view.

Object detection forms an integral part of the summarization function, with YOLO-based approaches becoming increasingly prominent. Sultana et al. [13] summarize developments in the YOLO algorithm, and Liu et al. [14] outline object detection issues and detector types. Ghosh and Ramanathan [12] introduce a deep learning framework that sums up videos based on objects of interest with high accuracy over a range of datasets.

Unsupervised techniques are another major emphasis. Wang et al. [15] and [16] introduce reinforcement learning architectures for unsupervised summarization, involving temporal consistency and interpolation methods. Kadam et al. [17] tackle issues in single- and multi-view summarization, presenting a conceptual basis for different solutions.

For applications specific to CCTV, Kamath et al. [18] propose a KNN-based approach with Euclidean distance and DCT for keyframe detection. Wang et al. [19] introduce a saliency-based keyframe selection system and Zhang et al. [20] suggest DSNet, a detect-to-summarize network. Hybrid models like Lan and Ye's [21] hybrid of DCNNs and MP-KNN exhibit better keyframe detection. Background subtraction approaches are proposed in [22] and [23], using

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1554

LBP descriptors and adaptive thresholding for resilient summarization.

III. SCOPE OF WORK

The scope of work for the "CCTV Video Summarization" project encompasses the development of a comprehensive solution to address the challenges posed by the overwhelming volume of raw digital video data captured by CCTV surveillance cameras. The project aims to achieve the following objectives:

- ***Video Summarization System:*** Develop an advanced video summarization system capable of condensing lengthy surveillance videos into concise and meaningful summaries, thereby reducing the need for manual scrutiny of hours of video content.
- ***Frontend Development:*** Design and implement a user-friendly graphical interface for the website, enabling seamless interaction with the CCTV video summarization system. This includes the creation of landing pages, upload video pages, and output pages to facilitate user engagement using HTML, CSS, JAVASCRIPT and FLASK.
- ***Backend Development:*** Implement sophisticated algorithms and artificial intelligence techniques to autonomously identify key events within surveillance videos. This involves integrating motion detection, object recognition, and behavioural pattern analysis to streamline the video analysis process.
- ***Research and Innovation:*** Explore and leverage cutting-edge methodologies such as the KNN model for extracting and subtracting background from videos, as well as custom Object Tracking algorithms to detect and track moving objects within the footage.
- ***Privacy and Ethical Considerations:*** Address privacy and ethical concerns associated with surveillance systems, ensuring that the developed solution adheres to the highest standards of privacy and data protection.

The project aims to provide a robust and efficient solution that significantly enhances the process of analyzing and summarizing CCTV footage, thereby contributing to the advancement of surveillance technology.

IV. MATERIALS & METHODS

Designing a research methodology for a CCTV video summarization project involves a systematic and phased approach to address the unique challenges associated with processing and summarizing surveillance footage. Here is an expanded methodology.

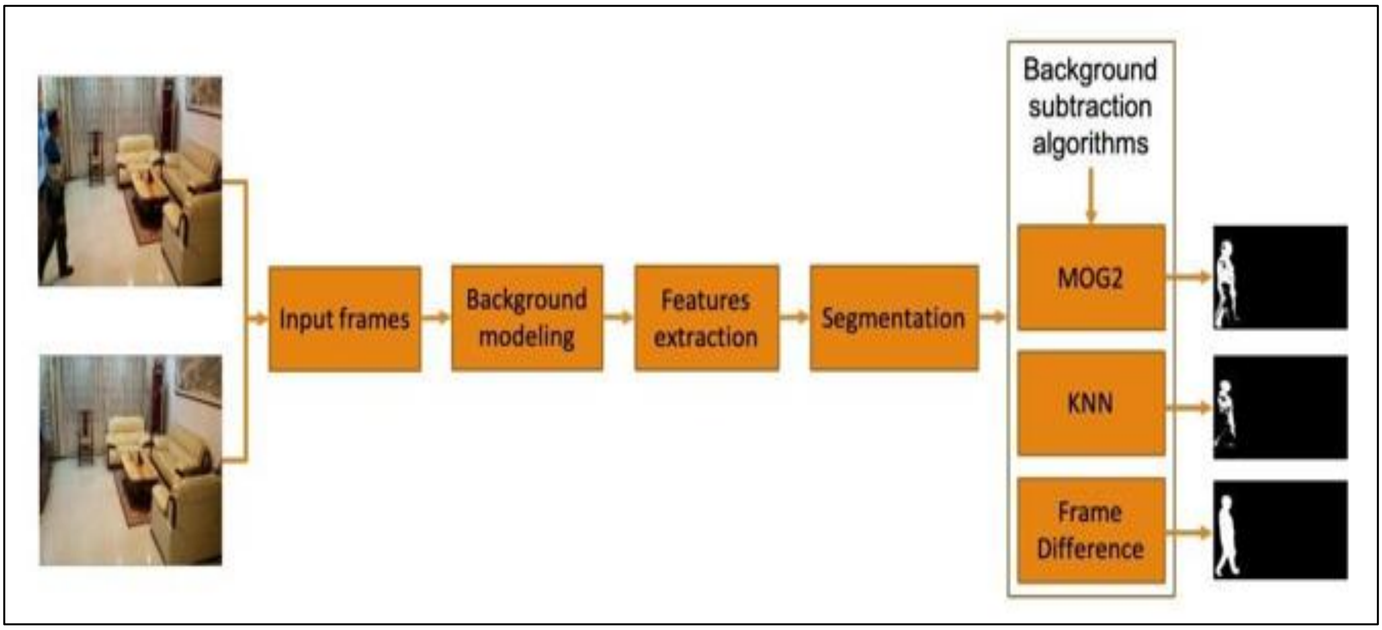

Fig 1: Method Used

- ***Literature Review and Research Analysis*:** Understand the current state of the art in CCTV video summarization. Conduct an extensive literature review to identify existing methodologies and technologies. Analyze multiple research papers to identify strengths, weaknesses, and gaps in current approaches. Summarize key findings and establish a knowledge base for the project.
- ***Problem Identification and Definition:*** Clearly define the problem statement and scope of the CCTV video summarization project. Identify challenges and limitations in existing methods from the literature review. Define the specific goals of the video summarization project, taking into account privacy, efficiency, and interpretability concerns.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1554

Fig 2: Comparison of Different Background Subtraction Techniques

- ***Background Modelling:*** The background subtraction algorithm relies on a background model to differentiate between foreground and background in incoming video frames. This model serves as a reference point, aiding in the identification of changes and movement within the frame. The significance of the background model becomes apparent because initial video frames typically contain elements of the foreground, making it challenging to establish an accurate baseline. However, the effectiveness of the background model can be compromised as foreground objects deviate from the established background model. This deviation leads to inaccuracies in classification, as the model may misinterpret the movement or presence of objects, providing incorrect classifications. This underscores the importance of continuously updating and adapting the background model to account for changes in the scene and maintain the algorithm's accuracy over time.

Fig 3: Background Subtraction

- Feature Extraction: To effectively compare video frames with a background image, it is crucial to select relevant features that adequately represent the information. Many algorithms utilize RGB and grayscale intensities as features for this purpose. In some instances, the intensity of pixels is combined with other features. Additionally, the choice of the feature region is a critical consideration. Several algorithms extract features over pixels, blocks, or patterns. Pixel-wise features, although common, may result in noisy segmentation outcomes as they fail to encode local relations. On the other hand, pattern-wise and block-wise features are less susceptible to minimal changes, offering more stable segmentation results. The selection of an appropriate feature extraction method depends on the specific characteristics of the video frames and the desired level of sensitivity to changes in the scene.
- Segmentation: Video frames can be effectively processed using a background model. Background segmentation can be achieved by extracting features from corresponding pixels in both frames and analyzing the regions of these pixels, often extending the analysis to include a range, such as the Euclidean distance, to calculate similarities between pixels. By applying a similarity threshold and comparing these values, each pixel is categorized as either foreground or background.

The overall background subtraction system is constructed through the combination of these fundamental building blocks. The effectiveness of the system is directly tied to the performance of each individual block. If one module delivers poor performance, the overall system's capability is compromised. Background subtraction is a broad field, and as such, there are numerous algorithms

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1554

designed to address this task, each with its own strengths and limitations.

The choice of algorithm depends on the specific characteristics of the video data and the desired outcomes.

- KNN Background Subtraction: We have used KNN background subtraction for our project. In opting for the K-Nearest Neighbors (KNN) algorithm for background subtraction in our video processing application, one of the primary considerations was the efficiency of its training phase. KNN classifiers are well-regarded for their relatively swift training compared to alternative methods like the Mixture of Gaussians (MOG). This attribute becomes particularly advantageous in real-time applications where the system must promptly adapt to dynamic scenes, ensuring responsiveness to changes in the environment. The choice of KNN for background subtraction was driven by its faster training, long-term accuracy, and robustness in the face of dynamic backgrounds. These attributes align with the specific requirements of our video processing application, ensuring a balance between real-time performance and the ability to deliver accurate and reliable results over an extended period.
- Documentation and Reporting: Document the entire research process and outcomes for transparency and reproducibility. Maintain comprehensive documentation of methodologies, algorithms, and results. Prepare detailed reports and publish findings to contribute to the research community. By following this methodology, the research project aims to contribute to the advancement of CCTV video summarization by addressing key challenges and incorporating innovative techniques for efficient, ethical, and effective analysis of surveillance footage.

V. RESULTS & DISCUSSION

The "Results & Discussion" section gives the outcomes of the bibliometric analysis performed with VOS viewer. It identifies important research trends, significant authors, and collaboration networks in the fields of information system security and privacy. The section provides visualizations of research trends, cluster analyses, and co-authorship networks, emphasizing the interconnectedness of the research ecosystem. The analysis identifies thematic groups within data encryption, anonymization approaches, and information system security, providing insights into significant themes and growing areas in the field. The section finishes with a discussion of the significance of these findings and possible future study avenues.

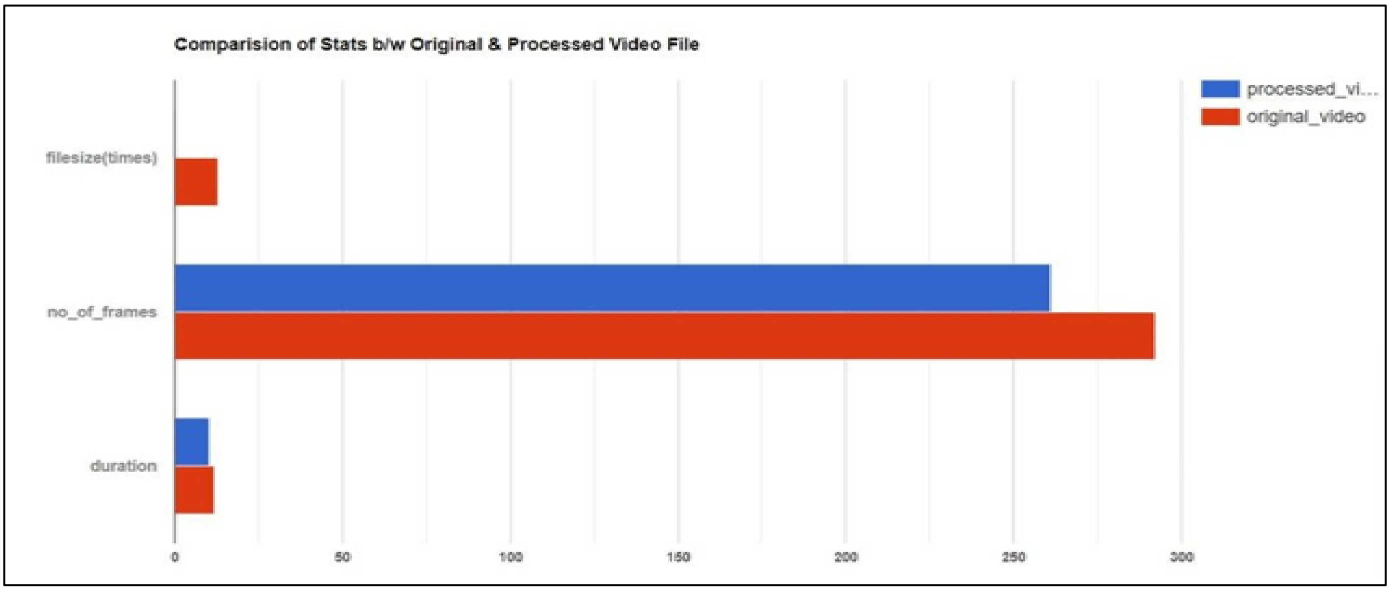

Fig 4: Comparison of Input Video vs Summary Video

Above graph as shown in Figure 2.A, depicts the variation in the lengths of original video and summary video. Summary video apart from being considerably short in length compared to original video also shows only the highlights which the user wants to see.

➢ *Formulas Used:*

- *Background Modelling Phase:*

Probability of every pixel being the foreground or background is calculated as:

$$D_k(x,y) = \begin{cases} 1 & if|f_k(x,y) - f_{k-1}(x,y)| > T \\ 0 & \end{cases}$$

This type of condition is often found in algorithms that detect changes in consecutive frames of a video, identifying pixels as part of the foreground if the absolute difference between pixel values in the current frame *f{k} (x, y)* and the previous frame *f{k-1} (x, y)* exceeds a certain threshold *T*.

- *Segmentation:*

$$P(X_t) = \sum_{i=1}^{k} \omega_{i,t}.\eta(X_t, \mu_{i,t} \sum i.t)$$

- *KNN Background Subtraction:*

Euclidean distance is used to measure the closeness. Euclidean distance between two points, A = (a1, a2, ...an) and B = (b1, b2, ….bn) is given by the following equation:

$$d = |\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}\ .$$

The density estimation formula approximately is given below:

$$[t]p(x|x_i) \approx \frac{1}{NV} \sum_{m=1}^{N} b^m K\left(\frac{||x_i - x||}{D}\right)$$

Here, K is the kernel function, subject to uniform distribution.

If u < 1/2, then the kernel K(u) = 1, otherwise 0. If the video sequence sample is assigned to foreground, the value of b m is 0. The background model only deals with samples that satisfy b m and are classified as background. As it can be seen below, where, if P(x|xi) is greater than a certain threshold of T, the pixel is considered the background.

VI. CONCLUSION

The conclusion of the "CCTV Video Summarization" project represents a significant milestone in addressing the challenges posed by the overwhelming volume of raw digital video data captured by CCTV surveillance cameras. The project has successfully developed an advanced video summarization system capable of condensing lengthy surveillance videos into concise and meaningful summaries, thereby reducing the need for manual scrutiny of hours of video content.

$$[t]M_i = \begin{cases} \text{Foreground}, & if P(x|x_i) < T \\ \text{Background}, & otherwise \end{cases}$$

The project's frontend and backend development efforts have resulted in the creation of a user-friendly graphical interface and the implementation of sophisticated algorithms and artificial intelligence techniques to autonomously identify key events within surveillance videos. The utilization of the KNN model for extracting and subtracting background from videos, along with a custom Object Tracking algorithm to detect moving objects, has significantly contributed to the success of the project

The project's outcomes align with the growing need for surveillance solutions that can condense long footage into short videos while retaining important events. By summarizing hours of footage shot by CCTV cameras into a short clip that shows all interesting events simultaneously, the project has demonstrated its potential to revolutionize the field of video surveillance.

The successful completion of the "CCTV Video Summarization" project underscores its significance in enhancing the process of analysing and summarizing CCTV footage, thereby contributing to the advancement of surveillance technology. The project's outcomes hold promise for addressing the storage challenges associated with the overwhelming volume of surveillance footage, while also streamlining the video analysis process.

The "CCTV Video Summarization" project represents a significant step forward in leveraging advanced algorithms and artificial intelligence to transform the way surveillance video analysis is approached. The project's successful outcomes underscore its potential to revolutionize the field of video surveillance and contribute to the development of more efficient and effective surveillance solutions.

REFERENCES

[1]. Hussain, T., Wang, Y., & Liu, Y. (2021). A comprehensive survey on multi-view video summarization: IEEE Transactions on Multimedia, 23(1), 40-56.

[2]. Wang, Z., Zhang, H., & Liu, Y. (2022). A survey of deep learning-based video summarization methods. ACM Computing Surveys (CSUR), 55(1), 1-37.

[3]. Zhang, H., Wang, Y., & Liu, Y. (2022). A survey on video summarization for intelligent transportation systems. IEEE Transactions on Intelligent Transportation Systems, 23(1), 5772.

[4]. Wang, Y., Zhang, H., & Liu, Y. (2022). A survey on video summarization for socially aware applications ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 18(3), 1-34.

[5]. Zhang, H., Wang, Y., & Liu, Y. (2022). A survey on multi-task video summarization.

[6]. Zhang, Y., Wang, Y., & Liu, Y. (2023). A survey on video summarization for event detection.

[7]. Wang, X., Zhang, H., & Liu, Y. (2023). A survey on video summarization for visual question answering.

[8]. A. Ghosh and V. Ramanathan (2023). *"Video summarization using deep learning techniques: A survey",* Multimedia Tools and Applications , 2023, pp. 1-26.

[9]. *"Video Summarization Using Deep Neural Networks: A Survey"* by Evlampios Apostolidis, Eleni Adamantidou, Alexandros I. Metsai, Vasileios Mezaris, and Ioannis Patras

[10]. Y. Wang, H. Feng, Z. Chen, X. Hu, and J. Wu, *"Video summarization using deep reinforcement learning," Applied Sciences,* vol. 22, no. 19, p. 7689, 2022.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25apr1554

[11]. A. Khan, S. Ullah, M. A. Khan, M. A. Khan, and A. Khan, *"Video summarization techniques: A review,"* *International Journal of Multimedia and Information Retrieval*, vol. 8, no. 2, pp. 103-114, 2019.

[12]. A. Ghosh and V. Ramanathan, *"An effective video summarization framework based on the object of interest using deep learning," Multimedia Processing and Enhancement*, vol. 2022, p. 7453744, 2022.

[13]. F. Sultana, A. Sufian, and P. Dutta, "A Review of YOLO Algorithm Developments," *Procedia Computer Science*, vol. 185, pp. 141-148, 2022.

[14]. Y. Liu, J. Wang, Y. Yang, and Y. Zhang, *"Object Detection Using YOLO: A Review,"* Applied Sciences, vol. 12, no. 23, p. 13644, 2022.

[15]. Y. Wang, H. Feng, Z. Chen, X. Hu, and J. Wu, *"Unsupervised video summarization using deep reinforcement learning," Applied Sciences*, vol. 23, no. 7, p. 3384, 2023.

[16]. Y. Wang, H. Feng, Z. Chen, X. Hu, and J. Wu, *"Unsupervised video summarization using deep reinforcement learning with interpolation," Applied Sciences*, vol. 23, no. 7, p. 3384, 2023.

[17]. P. Kadam, D. Vora, S. Mishra, and S. Patil, *"Recent Challenges and Opportunities in Video Summarization With Machine Learning Algorithms," IEEE Access*, vol. 10, pp. 1-15, 2022.

[18]. Kamath, A., S, B., Westbenson, K. L., Sharma, S., & Sinam, N. (2021). Summarization of CCTV Videos using KNN Classifier.Journal of Emerging Technologies and Innovative Research (JETIR).

[19]. Wang, Y., Wang, L., & Lu, H. (2021). A Novel Key-Frames Selection Framework for Comprehensive Video Summarization. IEEE Transactions on Multimedia, 23(3), 838-850.

[20]. Zhang, J., Li, Y., & Gao, W. (2021). DSNet: A Flexible Detect-to-Summarize Network for Video Summarization. IEEE Transactions on Image Processing, 30(2), 948-962.