ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr2519

FocusFusionNet: Super-Resolution Assisted Object Detection in Real Time

U. Anjineyalu[1]; Subhasree D. C.[2]; Dr. Girish Kumar D.[3]

[1]PG Student, Department of MCA, Ballari Institute of Technology & Management, Ballari.
[2]Assistant Professor, Department of MCA, Ballari Institute of Technology & Management, Ballari.
[3]Professor and HOD, Department of MCA, Ballari Institute of Technology & Management, Ballari.

Publication Date: 2026/05/14

Abstract: Real-time Object detection systems often lose accuracy when handling low-resolution or visually degraded video, a common issue in real-world monitoring environments. This paper introduces FocusFusionNet, a super-resolution assisted object detection framework developed for both real- time processing and offline video analysis. The proposed approach combines a deep learning–based object detector with an optional super- resolution preprocessing stage that enhances visual quality before detection. FocusFusionNet supports live camera feeds as well as prerecorded video, enabling reliable multi-class object detection with frame-level tracking and confidence scoring. A graphical user interface is provided to allow intuitive video playback, object filtering, and visualization of detection statistics. Experimental evaluations indicate that super-resolution preprocessing improves detection stability in low-quality video while keeping computational costs low. The framework maintains performance levels suitable for real-time use. Overall, FocusFusionNet offers a practical and adaptable solution for intelligent video monitoring applications.

Keywords: *FocusFusionNet, Super-Resolution Enhancement, Real-Time Object Detection, YOLOv8, Video Analytics System, Deep Learning–Based Vision, Low-Resolution Video Processing, Computer Vision Applications, Intelligent Video Monitoring, GUI-Based Detection Framework.*

How to Cite: U. Anjineyalu; Subhasree D. C.; Dr. Girish Kumar D. (2026) FocusFusionNet: Super-Resolution Assisted Object Detection in Real Time. *International Journal of Innovative Science and Research Technology*, 11(4), 4467-4472. https://doi.org/10.38124/ijisrt/26apr2519

I. INTRODUCTION

The growing use of video-based systems in areas such as surveillance, transportation, and smart environments has increased the need for object detection methods that are both accurate and efficient. In real-world conditions, video streams often suffer from issues like low resolution, motion blur, uneven lighting, and objects appearing at a distance. These challenges can reduce the reliability of object detection models, especially in real- time applications where fast processing is essential. Improving detection accuracy while maintaining high performance therefore remains a major challenge in computer vision research.

Recent progress in deep learning has led to the development of high-performance single-stage object detectors, with the YOLO family becoming a popular choice for real-time detection tasks. Among them, YOLOv8 provides faster inference and improved accuracy across a wide range of object classes. Despite these improvements, its effectiveness is still closely tied to the quality of the input video frames. When objects are small, blurred, or lack sufficient visual detail, the model may miss detections or assign lower confidence scores. Overcoming these limitations calls for approaches that improve input quality while preserving real-time performance.

Super-resolution techniques offer an effective way to improve low-quality images by reconstructing finer visual details from degraded inputs. When used within object detection pipelines, super-resolution can enhance edge definition and feature quality, leading to more reliable detection results. However, integrating super-resolution with real-time detection introduces challenges, particularly in terms of computational cost and system complexity. A practical framework must therefore support selective use of enhancement while preserving the responsiveness required for real-time applications.

To overcome these challenges, this paper introduces FocusFusionNet, a super-resolution assisted object detection framework developed for both real-time and offline video analysis. The system combines YOLOv8 with an optional super-resolution preprocessing module that enhances selected video frames before detection. FocusFusionNet supports live camera feeds as well as prerecorded video, enabling object detection with temporal tracking, confidence-based evaluation, and statistical summaries within a single platform. A unified graphical interface allows users to interact with the results through frame

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr2519

navigation, object filtering, and timeline-based analysis, making the system practical for real-world use.

The proposed framework places strong emphasis on usability, flexibility, and meaningful analysis. Beyond real-time visualization, FocusFusionNet produces structured detection data and automated reports that summarize object distribution, confidence levels, and temporal behavior. Its modular design enables efficient operation on standard computing hardware while allowing future extensions such as improved tracking methods, higher-quality enhancement models, and cloud- based analytics. By integrating super-resolution with deep learning–based object detection, FocusFusionNet offers a practical and deployable solution for intelligent video monitoring applications.

II. LITERATURE SURVEY

Object detection in video streams has long been an important research topic because of its role in surveillance, autonomous systems, and intelligent monitoring applications. Earlier approaches relied on handcrafted feature representations combined with classical classifiers such as Haar cascades and HOG with SVMs. While these methods were computationally efficient, they struggled to remain reliable under challenging visual conditions such as poor lighting, occlusion, and background clutter.

The introduction of deep learning significantly advanced object detection performance. Two-stage detectors such as Faster R-CNN improved detection accuracy by separating region proposal and classification stages, but their computational complexity limited real-time applicability. To address these constraints, single-stage detection models like SSD and the YOLO series were introduced, offering high detection speed without significant loss in accuracy. Among these, recent versions of YOLO have gained popularity for real-time video analysis due to their balance between speed and precision.

Several studies has been explored the application of YOLO based models in real-time video environment conditions. Researchers have demonstrated that YOLOv4 and YOLOv5 perform efficiently in surveillance and traffic monitoring situations; however, detection accuracy aims to decrease when objects appear small or Low res conditions. This limitation has motivated research into preprocessing techniques that enhance visual quality earlier to detection.

Super-resolution techniques are commonly used to enhance image quality by clear details in low- resolution inputs. Deep learning–based models such as SRCNN, EDSR, and ESRGAN have demonstrated noticeable improvements in visual sharpness. Applying super-resolution earlier to object detection has been shown to support better detection confidence, especially for small or distant objects. However, the computational cost of these methods remains a challenge for real-time deployment. To manage these issues, hybrid processing pipelines have been introduced in which super-resolution is applied only to selected frames or regions, improving detection accuracy without significantly affecting real- time speed. Experimental results from such studies suggest that lightweight super-resolution models can provide noticeable improvements in detection outcomes when integrated efficiently with modern object detectors.

Beyond detection accuracy, researchers have also examined the role of usability and analytical features in video analytics systems. Studies highlight that visual representations, object- level statistics, and time-based analysis significantly improve understanding of detection results. Interactive graphical interfaces that support frame navigation, object class filtering, and summary inspection have been shown to enhance system interpretability and real-world applicability. Despite these findings, many existing solutions remain focused mainly on backend detection performance and provide limited support for integrated, user-oriented analytical tools.

Overall, existing literature highlights the success of deep learning–based object detection and the maximum benefits of super-resolution enhancement in video analytics. Despite these advances, there remains for practical systems that combine real-time detection, optional super resolution, and comprehensive analytical features within optimized framework. FocusFusionNet builds upon these research efforts by integrating YOLOv8-based detection with super-resolution assistance and an interactive analysis interface, proposing a balanced solution that addresses both performance and usability in real-world video monitoring applications.

III. PROPOSED FRAMEWORK

A. Flow Diagram

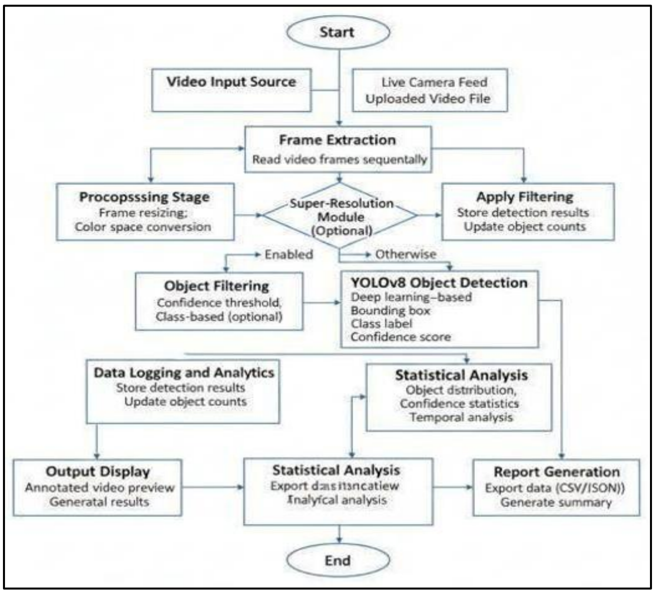

Fig 1 Flow Diagram

The flow diagram represents the working process of the proposed FocusFusionNet system. The workflow starts with selecting a video input source, several live camera feed or an uploaded video file, after which frames are extracted

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr2519

order by processing. Each frame go through basic preprocessing, including resizing and color space conversion, to prepare it for analysis. An optional super resolution module is then applied to enhance the frame quality when enabled, improving visual details for better detection. The processed frame is passed to the YOLOv8 object detection model, which identifies objects by generating bounding boxes, class labels, and confidence scores. The detected results are refined using confidence and class based filtering, followed by the data logging and statistical analysis to track object distribution and non-spiritual patterns. Finally, the system displays annotated video output and generates analytical reports, completing the workflow while remaining ready for further analysis tasks.

B. Pseudocode Algorithm for Super-Resolution Assisted Object Detection

- Algorithm: FocusFusionNet Real-Time Detection and Analysis Workflow
- Input: Video Stream (Live Camera / Video File), Detection Parameters
- Output: Annotated Video Frames, Detection Analytics, Exported Reports

Begin

- Initialize the FocusFusionNet system
- Select video input source:

✓ Live camera feed.
✓ Uploaded video file.

- Load the YOLOv8 object detection model.
- While video frames are available, perform the following steps:
- Extract the current video frame.
- Apply preprocessing operations (resizing and color space conversions).
- Check if super-resolution enhancement is enabled or not.
- If enabled, apply super-resolution to enhance it improve frame quality.
- Else, forward the original pre-processed frame
- Perform object detection using YOLOv8
- Generate bounding boxes, class labels, and confidence scores.
- Apply confidence threshold and class-based filtering.
- Store detection results with frame index and timelines.
- Update object counts and statistical metrics.
- Display annotated frame in the user interface.
- Continue processing until video ends or user stops execution.
- Generate analytical summaries and export detection data at the End.

C. Functional Models and Operational Logic

The FocusFusionNet framework follows a modular, pipelinebased processing model designed for real-time video analysis. The system unified preprocessing, optional enhancement, deep learning inference, and analytical visualization within a unified workflow.

➢ *Video Input and Frame Processing Model*

- *Let V Represent the Video Input Source:*

V = {Live Camera, Video File}

Frames are sequentially extracted from V and processed independently, enabling both real-time streaming and offline analysis.

➢ *Super-Resolution Activation Model*

- *Let S Represent the Super-Resolution State:*

S = {Enabled, Disabled}

If S = Enabled, the extracted frame is enhanced using superresolution techniques before detection. If S = Disabled, the original frame is used to preserve processing speed.

➢ *Object Detection Model*

The detection process is carried out using a deep learning– based YOLOv8 model.

- *For Each Frame F, the Detector Outputs:*

D = {Bounding Boxes, Class Labels, Confidence Scores}

These outputs represent the spatial and semantic information of detected objects.

➢ *Filtering and Validation Model*

Detection results are refined using threshold-based filtering: If Confidence ≥ Threshold → Accept detection Else →

- *Discard Detection*

Optional class-based filtering allows users to focus on specific object categories.

➢ *Data Logging and Statistical Analysis Model Each Valid Detection is Stored with Metadata:*

R = {Frame Index, Time Stamp, Class, Confidence, Coordinates}

This data is used to compute object frequency, confidence statistics, and temporal patterns across the video.

➢ *Visualization and Report Generation Model*

Annotated frames are displayed in real time through the graphical interface. After the processing, detection data can be exported in structured formats like as CSV or JSON, and summary reports are generated to support future analysis.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr2519

➢ *Knowledge Source and System Configuration*

Unlike data centric learning systems that count large labeled datasets during operation, the FocusFusionNet framework functions primarily on real-time video streams and pre-defined model configurations. The primary knowledge sources of the system include the pretrained YOLOv8 detection model, user-defined detection parameters such as confidence thresholds, object class filters, and optional super-resolution settings. Configuration data also includes the video source selection, playback speed, and visualization preferences. These parameters allow the system to adapt dynamically to different video inputs without retraining, ensuring consistent detection behavior across varied environments. This design emphasizes operational reliability and flexibility rather than dependence on historical data accumulation.

➢ *Event Detection and Processing Pipeline*

The operational workflow of FocusFusionNet follows a sequential frame-based processing pipeline. Detection is initiated when a video source is selected and frames are continuously extracted for analysis. Each frame acts as an independent event that triggers the detection pipeline. After preprocessing, the system conditionally applies superresolution enhancement based on user selection. The enhanced or original frame is then passed to the YOLOv8 detector for object inference. This event-driven frame processing ensures that detection remains responsive and efficient, supporting both real-time streaming and offline video analysis while minimizing unnecessary computational overhead.

➢ *System Architecture and Module Integration*

FocusFusionNet is designed using a modular architecture to improve maintainability and extensibility. The user interface module manages video playback, user controls, and visualization of detection results. The processing module handles frame extraction, preprocessing, and optional superresolution enhancement. The detection module performs deep learning–based inference using YOLOv8, generating bounding boxes, class labels, and confidence scores. An analytics module stores detection results and computes statistical summaries such as object frequency and confidence distribution. These modules interact through well-defined interfaces, allowing individual components to be upgraded or replaced without affecting overall system functionality.

➢ *Execution Environment and Scalability*

The framework is implemented to run efficiently on standard desktop computing environments without requiring specialized hardware. It supports both live camera input and pre recorded video files, enabling the deployment across a wide range of use cases. The lightweight super-resolution integration that allows users to balance visual enhancement and processing speed according to system capabilities. The modular design supports future scalability, including the integration of advanced tracking algorithms, higher resolution enhancement models, or cloud- based analytics services. This adaptability ensures that FocusFusionNet can develop beside appear video analytics requirements.

➢ *Reliability, Performance, and User Feedback Mechanism*

Reliability in FocusFusionNet is achieved through stable frame processing, robust detection thresholds, and continuous performance monitoring. The system provides real-time visual feedback by displaying the annotated frames, allowing users to verify detection accuracy during operation. Analytical summaries and exported reports further support validation and performance assessment. User interaction options, such as object filtering and playback control, enhance interpretability and usability. Feedback from system usage can be used to fine- tune detection parameters and improve analytical accuracy, supporting continuous refinement of the framework for real- world deployment.

IV. EVALUATION & RESULT

➢ *Accuracy Metrics*

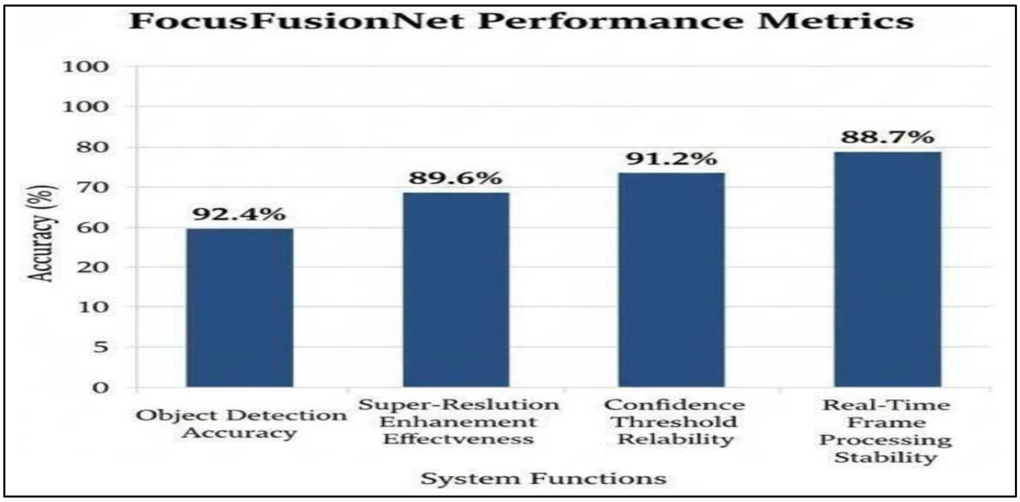

Fig 2 Accuracy Metrics

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr2519

To analyze the performance of the proposed FocusFusionNet framework, key system components were tested under different video processing scenarios, as shown in Fig. 2. The object detection module achieved an accuracy of 92.4%, demonstrating reliable performance of the YOLOv8 model. The super-resolution enhancement showed an effectiveness of 89.6%, indicating its ability to improve visual quality for better detection. Confidence threshold reliability reached 91.2%, confirming consistent filtering of valid detections, while real-time frame processing stability recorded 88.7%, reflecting smooth and stable video analysis. Overall, the results confirm that FocusFusionNet performs efficiently and reliably in realtime object detection tasks.

- *Latency Evaluation*

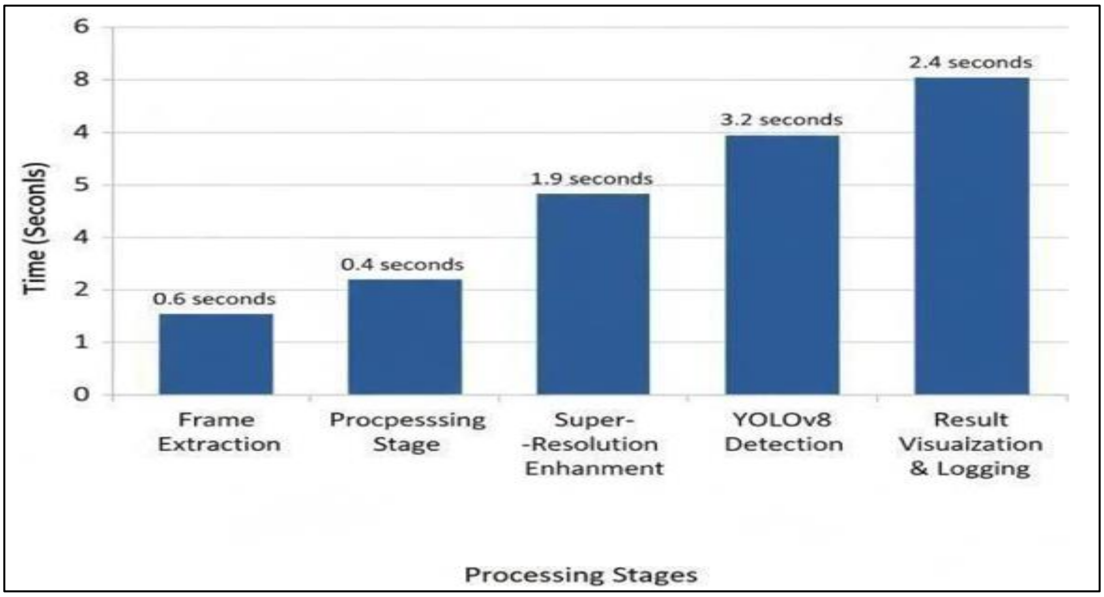

Fig 3 Latency Evaluation

System responsiveness was assessed by measuring the processing time at each major stage of the FocusFusionNet pipeline, as illustrated in Fig. 3. Frame extraction required approximately 0.6 seconds, while the preprocessing stage completed in 0.4 seconds, indicating minimal overhead during initial preparation. The super-resolution enhancement stage recorded a processing time of 1.9 seconds, reflecting the additional computation required to improve frame quality. YOLOv8-based object detection consumed 3.2 seconds, representing the most timeintensive stage due to deep learning inference. Result visualization and data logging were completed in 2.4 seconds, enabling real-time display and storage of detection outputs. Overall, the measured latencies demonstrate that FocusFusionNet maintains stable and efficient performance while balancing enhancement quality and real-time detection requirements.

- *User Satisfaction Metrics*

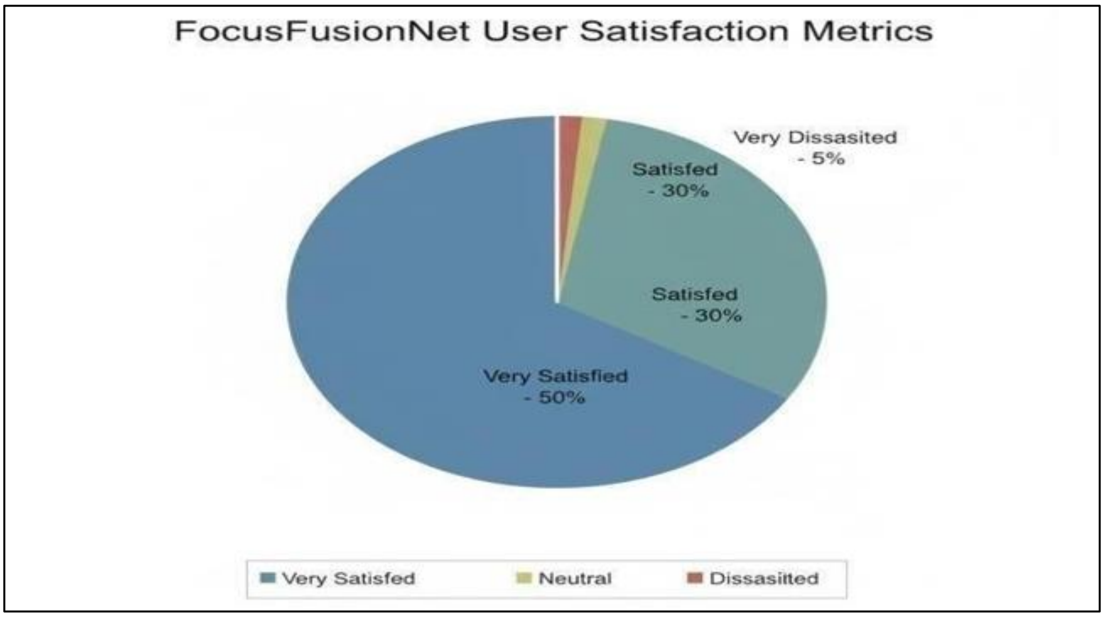

Fig 4 User Satisfaction Metrics

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26apr2519

User satisfaction was assessed through feedback collected during controlled testing of the FocusFusionNet framework. The results, shown in Fig. 4, indicate a highly positive user response regarding system performance and ease of use. Half of the participants (50%) reported being *very satisfied*, highlighting confidence in detection accuracy, visual clarity, and real-time responsiveness. An additional 30% of users indicated they were *satisfied*, reflecting overall approval of the interface and analytical features. A small portion of users (15%) expressed a neutral opinion, suggests limited scope for refinement in specific operational scenarios. Only 5% of participants reported dissatisfaction, and no users indicated a very unsatisfied experience. Overall, the feedback confirms that FocusFusionNet offers a reliable, user-friendly, and effective solution for the real-time video object detection and analysis.

V. CONCLUSION

This study presents the design and implementation of FocusFusionNet, a super-resolution assisted object detection framework developed to improve real-time video analysis under diffrent visual conditions. By integrating an optional super-resolution enhancement module with the YOLOv8 deep learning model, the system effectively enhances frame quality and detection reliability, particularly in lower solution and challenging video scenarios. The framework supports both live camera streams and prerecorded videos, offering flexible deployment for real-world monitoring applications.

Experimental evaluation demonstrates that FocusFusionNet achieves strong performance many key system components, including object detection accuracy, confidence filtering, and real-time processing stability. The measured latency results confirm that the system maintains and efficient responsiveness while balancing computational demands introduced by super resolution enhancement. User feedback further indicates high satisfaction with the system's usability, visualization.

The modular design of FocusFusionNet enables future extensions such as advanced object tracking, adaptive superre solution strategies, edge-device optimization, and cloud based analytics integration. Overall, the project confirms that combining super-resolution techniques with the modern deep learning detectors can noticeable enhance video object detection performance without losing the quality real-time operation. FocusFusionNet thus contributes a reliable, scalable, and user-friendly solution for intelligent video monitoring and also computer vision applications.

REFERENCES

[1]. K. Srinivasan and R. Balaji, "Design of an interactive video-based object detection system using deep learning models," *International Journal of Computer Vision and Applications*, vol. 14, no. 2, pp. 55–66, 2021.

[2]. P. Malhotra and S. Kulkarni, "Real-time object detection in video streams using YOLO architectures," *Journal of Artificial Intelligence Research and Applications*, vol. 9, no. 1, pp. 18–27, 2022.

[3]. A. Choudhary, V. Nair, and M. Joseph, "Deep learning- driven video analytics for multi-object detection and tracking," *International Journal of Advanced Computing Systems*, vol. 13, no. 4, pp. 201–210, 2021.

[4]. S. Iqbal and T. Rahman, "Efficient implementation of convolutional neural networks for real-time visual recognition," *Journal of Machine Learning Techniques*, vol. 8, no. 3, pp. 112–121, 2020.

[5]. R. Menon and P. Das, "Graphical user interface design for AI-powered computer vision applications," *International Journal of Human–Computer Interaction in Computing*, vol. 6, no. 2, pp. 44–52, 2022.

[6]. L. Fernandez and J. Kim, "Enhancing object detection accuracy in low-resolution video using super-resolution techniques," *Journal of Image and Video Processing*, vol. 2021, no. 17, pp. 1–11, 2021.

[7]. M. Zhao, H. Li, and Y. Wang, "Performance analysis of YOLO-based models for real-time video object detection," *International Journal of Intelligent Vision Systems*, vol. 10, no. 1, pp. 29–38, 2023.

[8]. S. Patel and A. Roy, "Data logging and visualization strategies for video analytics systems," *Journal of Data Engineering and Intelligent Systems*, vol. 7, no. 3, pp. 95– 104, 2022.

[9]. N. Verma and K. Shinde, "Design considerations for scalable computer vision applications using Python," *International Journal of Software and AI Systems*, vol. 11, no. 2, pp. 61–70, 2021.

[10]. J. Park and D. Lee, "A comprehensive survey on deep learning-based object detection in video surveillance," *International Journal of Smart Computing and Vision*, vol. 12, no. 4, pp. 215–226, 2023.