IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154137

An Intelligent Video Surveillance System for Multi-Threat Detection and Real-Time Alerting

Aniket Ligam[1], Avadhoot Katte[2], Atul Shelar[3], Niharika Dasari[4], Rakesh Suryawanshi[5]

Department of Computer Engineering, A.C. Patil College of Engineering, Kharghar, India[1-5]

Abstract: Traditional CCTV surveillance systems rely heavily on continuous human monitoring, which is inefficient, error-prone, and unsuitable for large-scale deployment, as operators may miss critical events due to fatigue, overlapping camera feeds, poor lighting condi-tions, and the difficulty of tracking multiple screens simultaneously. To overcome these limitations, this paper presents an Intelligent Video Surveillance System (IVSS) capable of detecting multiple safety and security threats in real time. The proposed system integrates object detection, weapon detection, fire and smoke detection, fall detection, crash detection, and face recognition with watchlist matching into a unified pipeline. A YOLO-based model is employed for fast and accurate detection of persons and suspicious objects, while dedicated modules analyze fire patterns, abnormal human posture, sudden motion changes, and identity verification against a watchlist database. The system supports both live camera streams and recorded video input, making it flexible for various surveillance scenarios. To improve reliability and reduce false alarms, confidence-based filtering and temporal consistency checks across consecutive frames are applied before generating alerts. Detected incidents are stored as evidence in the form of annotated frames or video clips for further analysis. The modular architecture enables flexible deployment by allowing individual detection components to operate independently or as part of an integrated system. The proposed IVSS is well-suited for security-sensitive and safety-critical environments, providing en-hanced monitoring efficiency, reduced human workload, and faster response to potential threats.

Index Terms: Intelligent Surveillance, YOLO, Face Recognition, Watchlist Matching, Fall Detection, Fire Detection, Real-Time Monitoring, Object Detection

I. INTRODUCTION

Video surveillance has become a critical component of modern safety, security, and monitoring systems. It is widely used in public infrastructure such as transportation systems, hospi-tals, educational institutions, commercial buildings, industrial environments, and residential areas. The primary goal of surveillance systems is not only to record visual data but also to detect suspicious activities, prevent incidents, and support rapid response mechanisms.

Traditional CCTV systems rely heavily on human operators to continuously monitor multiple video feeds. However, this approach is inefficient and unreliable in real-world scenarios. Human operators often experience fatigue, reduced attention span, and cognitive overload when observing multiple screens simultaneously. As a result, critical events such as fire out-breaks, accidents, suspicious behavior, or unauthorized access may be missed.

In addition, the rapid growth of surveillance networks has led to a massive increase in video data. Continuous manual monitoring of such large-scale data is impractical. This has created a strong demand for intelligent surveillance systems that can automatically analyze video streams, detect abnormal events, and generate alerts in real time. The proposed system focuses on a multi-threat detection approach.

Unlike traditional systems that detect only a single type of event, this system integrates multiple detection capabilities including fire detection, weapon detection, fall detection, crash detection, object detection, and face recognition with watchlist matching.

Organization: The rest of the paper is structured as fol-lows. Section II presents the literature review, discussing existing approaches and recent advancements in intelligent video surveillance systems. Section III describes the proposed system architecture, including the overall design and modular framework for multi-threat detection. Section IV explains the methodology adopted for video acquisition, preprocessing, detection, tracking, and alert generation. Section V discusses the implementation details and technologies used in the sys-tem. Section VI presents the experimental results along with performance evaluation and analysis. Section VII concludes the paper by summarizing the key findings and contributions. Finally, Section VIII highlights the future scope and potential enhancements of the proposed system.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154137

II. LITERATURE REVIEW

The field of video surveillance has evolved significantly over time. Early systems relied on traditional computer vision techniques such as background subtraction, frame differencing, contour detection, and optical flow analysis. These techniques were effective in simple environments but struggled in real-world conditions due to sensitivity to lighting changes, shad-ows, noise, and dynamic backgrounds.

To overcome these limitations, machine learning techniques were introduced, followed by deep learning approaches. Con-volutional Neural Networks (CNNs) revolutionized video anal-ysis by enabling automatic feature extraction and improving detection accuracy.

Among object detection models, YOLO has become one of the most widely used approaches due to its ability to perform detection in a single forward pass. This makes it extremely fast and suitable for real-time applications [1]. Later versions such as YOLOv3 and YOLOv4 improved accuracy and robustness [2], [3].

Face recognition has also seen major improvements. FaceNet introduced embedding-based recognition, where faces are rep-resented as feature vectors in a high-dimensional space [4]. ArcFace further improved recognition accuracy by enhancing feature separation [5]. Face detection models such as MTCNN and RetinaFace provide reliable detection under challenging conditions [6], [7].

For activity recognition, fall detection is typically based on posture analysis, motion patterns, and temporal changes. Pose estimation techniques improved the accuracy of such systems [10]. Similarly, crash detection relies on sudden motion anal-ysis and trajectory changes.

Fire detection has evolved from simple color-based detection to deep learning-based approaches that can identify flame and smoke patterns more accurately. These methods are more robust and suitable for real-world deployment.

Tracking algorithms such as SORT and DeepSORT play an important role in maintaining object identity across frames [8], [9]. This helps in reducing false alarms and improving reliability.

Despite these advancements, most systems focus on a single task. Real-world surveillance requires multi-threat detection. The proposed IVSS addresses this gap by integrating multiple detection modules into one system.

III. PROPOSED SYSTEM / SYSTEM ARCHITECTURE

The proposed Intelligent Video Surveillance System (IVSS) is designed using a modular and scalable architecture that enables efficient processing of multiple surveillance tasks in real time. The system follows a structured pipeline in which video data flows through a sequence of interconnected layers, each responsible for a specific function such as acquisition, preprocessing, detection, tracking, and alert generation. This layered approach improves flexibility, maintainability, and adaptability, allowing the system to operate effectively in diverse real-world environments.

The architecture supports both live video streams, such as RTSP or HTTP feeds, and recorded video files as input sources. Each frame from the input stream is processed sequentially through multiple modules, enabling simultaneous detection of various safety and security threats. The modular nature of the system allows individual components to function independently or as part of a combined pipeline, making the framework highly customizable based on application require-ments.

At the initial stage, the input layer is responsible for acquiring video data from cameras or stored files. It ensures continuous frame capture and provides a stable input stream for further processing. This layer is capable of handling multiple camera feeds simultaneously, which is essential for large-scale surveil-lance systems.

Following this, the preprocessing layer prepares raw frames for analysis. It performs operations such as resizing frames to a fixed resolution, normalizing pixel values, reducing noise, and adjusting brightness or contrast. These steps are crucial for maintaining consistent input quality and improving the per-formance of deep learning models, especially in challenging conditions such as low lighting, motion blur, or environmental noise.

The detection layer serves as the core of the system and uses a YOLO-based deep learning model to identify objects within each frame. It detects entities such as persons, vehicles, bags, and weapon-like items, and outputs bounding boxes, class labels, and confidence scores. Non-maximum suppression is applied to eliminate duplicate detections, ensuring accurate localization of objects.

In parallel, the fire detection layer focuses on identifying flame and smoke patterns. Unlike general object detection, this module analyzes specific visual characteristics such as color distribution, texture, and motion patterns associated with fire. Advanced models are used to distinguish actual fire events from similar visual elements like lights or reflections, and temporal smoothing is applied to ensure stable detection across consecutive frames.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154137

The face recognition layer performs identity analysis by de-tecting faces, aligning them, and extracting feature embeddings. These embeddings are compared with stored watchlist data using similarity measures such as cosine distance. When a match exceeds a predefined threshold and remains consistent over time, the system generates an alert for watchlisted indi-viduals. This functionality is particularly important for security monitoring and access control applications.

The fall detection layer analyzes human posture and motion over time. It tracks changes in body orientation, position, and movement speed to identify abnormal patterns. A fall is typically detected when there is a rapid transition from an upright posture to a horizontal position. By relying on temporal analysis rather than single-frame observations, the system reduces false alarms caused by normal actions such as sitting or bending.

The crash detection layer is designed to identify sudden impact-like events, such as vehicle collisions or abnormal disturbances in the scene. It evaluates motion intensity, trajectory changes, and scene disruptions to detect potential crash events. This module is particularly useful in traffic monitoring and industrial safety applications.

To improve reliability, the tracking and fusion layer maintains object identity across frames using algorithms such as SORT or DeepSORT. Tracking allows the system to analyze object behavior over time, while fusion combines outputs from differ-ent modules to make more informed decisions. For example, detecting a weapon near a tracked individual increases the confidence of a potential threat. Temporal verification is also applied to confirm detections across multiple frames, reducing false positives.

Once an event is confirmed, the alerting layer generates notifications and stores evidence. This includes saving annotated frames or short video clips, recording timestamps, and identifying the source camera. Alerts can be displayed on a monitoring dashboard or sent through external communication channels such as email or messaging systems, ensuring timely response to detected incidents.

Finally, the dashboard layer provides an interactive interface for users to monitor live feeds, review detected events, and analyze stored evidence. It enables security personnel to manage incidents efficiently and respond appropriately. The dashboard enhances usability and serves as the central control point of the system.

Overall, the proposed architecture enables efficient and reliable multi-threat detection within a unified framework. Its modular design ensures scalability, allowing additional detection mod-ules to be integrated without affecting existing components. By combining detection, tracking, and temporal analysis, the system achieves high accuracy and reduced false alarm rates, making it suitable for deployment in real-world safety and security environments.

IV. METHODOLOGY

The methodology adopted for this system focuses on designing a scalable, efficient, and privacy-aware video surveillance solution. The development process is divided into multiple stages, including video acquisition, preprocessing, detection, tracking, fusion, and alert generation.

A. Video Acquisition and Preprocessing

The system reads input either from a live camera feed or from a recorded file. Each frame is resized to a fixed resolution to ensure consistent model input. Additional preprocessing such

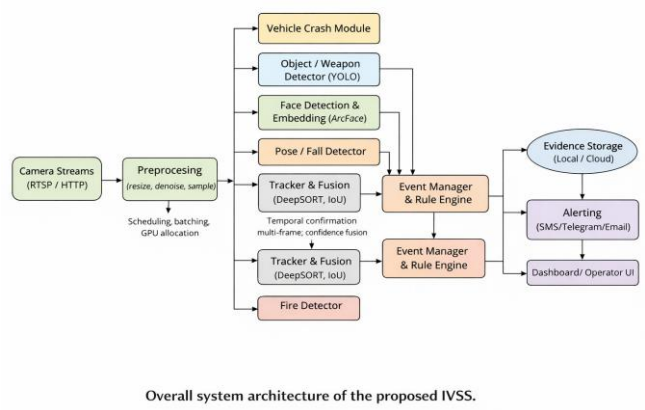

Fig. 1: Overall system architecture of the proposed IVSS

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.154137

as normalization, noise reduction, and brightness correction is applied when needed to improve performance in low-light or unstable conditions. These preprocessing steps help the detection models work more reliably and maintain consistent output quality.

B. Detection Modules

The core detection stage uses multiple specialized models. A YOLO-based detector is used for persons, objects, and weapon-like items. A fire-detection model analyzes flame and smoke patterns. A face-recognition model extracts embed-dings and compares them with watchlist templates. A fall-detection model evaluates body posture and movement. A crash-detection model looks for sudden abnormal motion and impact-like scene changes.

Each module produces a confidence score, and the final decision is based on verification across consecutive frames. This improves robustness and reduces single-frame mistakes.

C. Tracking and Temporal Verification

The system tracks objects and persons across frames so that detections are not treated as final unless they remain consistent over time. This is important for events such as smoke, falls, and crashes, where a single frame is often insufficient. Tempo-ral verification helps reduce false alarms caused by momentary motion, occlusion, or lighting variation.

D. Alert Generation

Once an event is confirmed, the system stores the detected frame or clip, records the time and camera ID, and updates the dashboard. Alerts can also be extended to email or message notifications in future versions. This ensures that security per-sonnel can review incidents quickly and respond appropriately.

V. IMPLEMENTATION / TECHNOLOGIES USED

The project is implemented using Python, OpenCV, and deep learning frameworks. Flask is used to build the monitoring dashboard. The detection pipeline can be extended with a database for event logging and cloud storage for evidence management.

A. Object and Weapon Detection

A YOLO-based detector is used as the main object detection engine because it provides a practical balance between speed and accuracy. In this project, the detector can be trained or fine-tuned on relevant classes such as person, bag, vehicle, and weapon-related categories. During inference, the model outputs bounding boxes, confidence values, and class labels. Non-maximum suppression is applied to remove duplicate detections.

For weapon detection, a stricter confidence threshold is usually preferred because false alarms in this category can have serious operational consequences. The model can be fine-tuned using custom images collected from the project dataset and augmented with flipping, scaling, and color changes to improve generalization.

B. Fire Detection

Fire detection is handled separately because fire and smoke of-ten have different visual characteristics from ordinary objects. A dedicated model analyzes each frame or frame window to identify flame-like or smoke-like regions. To reduce flickering detections, the system can use temporal smoothing or frame voting before declaring a fire event. This makes the alert more stable and reliable in real-time use.

C. Fall Detection

Fall detection is implemented by analyzing the posture and motion of a person across multiple frames. The system tracks changes in body orientation, centroid position, and movement speed. If a person changes from an upright posture to a near-horizontal posture with a rapid downward motion in a short time, the event may be classified as a fall. To avoid false alarms, the detection is confirmed only after temporal consistency is observed across consecutive frames.

D. Crash Detection

Crash detection is designed for vehicle or impact-like events. The module detects sudden abnormal motion, impact-style movement, or scene disruption patterns that may indicate a crash. This is useful for road surveillance, parking areas, and transport-monitoring scenarios. Like the fall module, crash detection benefits from temporal analysis rather than single-frame detection alone.

E. Face Recognition and Watchlist Matching

Detected faces are cropped from the video frame and aligned before feature extraction. The face-recognition model

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.154137

produces an embedding vector for each face. These embeddings are compared with stored watchlist embeddings using cosine sim-ilarity or another distance metric. If the similarity crosses a defined threshold and remains consistent across frames, the system raises a watchlist alert.

This step is useful in security-sensitive environments where known individuals, restricted persons, or suspicious targets need to be identified quickly. To reduce incorrect alerts, the system should store watchlist images in a consistent format and use multiple reference images per person when possible.

F. Tracking, Fusion, and Alert Generation

Because a single frame may contain uncertain detections, the system combines results over time. Tracking helps maintain the identity of objects or people across frames, while fusion combines outputs from different modules. For example, a weapon detected near a tracked person is more important than a single isolated weapon detection.

When an event is confirmed, the system performs the following actions:

- saves the annotated frame or video clip,
- stores event time and camera ID,
- displays the alert on the dashboard,
- optionally sends notification through email or other chan-nels.

VI. RESULTS AND DISCUSSION

The proposed Intelligent Video Surveillance System (IVSS) was evaluated using live and recorded video inputs cap-tured under different environmental conditions such as indoor and outdoor settings, varying illumination, partially occluded scenes, and different camera angles. The main purpose of the evaluation was to verify whether the system could de-tect multiple threat categories in real time, maintain stable performance across frames, and generate timely alerts with reasonable accuracy.

The results show that the modular pipeline is capable of processing video continuously while identifying key incidents such as fire, weapon presence, fall events, crash-like motion, and watchlist face matches. The use of temporal confirmation across consecutive frames helped reduce false alarms and im-proved the reliability of the final alert generation. In addition, the fusion of detection, tracking, and recognition modules made the system more suitable for practical surveillance scenarios than a single-purpose detection model.

A. Dataset and Testing Setup

The evaluation was carried out using a combination of public datasets and custom video samples collected for the project. The testing setup included:

- public datasets for general object detection and face-related evaluation,
- custom video clips collected for fire, weapon, fall, crash, and watchlist scenarios,
- day and night samples to verify robustness under different visibility conditions,
- different camera positions and viewing angles to simulate real surveillance deployment,
- both clear and partially occluded scenes to test the stability of the pipeline.

B. Performance Measures

The system was analyzed using standard performance mea-sures commonly used in computer vision and surveillance applications:

- **Precision**: the proportion of detected events that are actually correct,
- **Recall**: the proportion of actual events that were success-fully detected,
- **F1-score**: the harmonic mean of precision and recall, used to measure balanced performance,
- **mAP**: mean Average Precision, used mainly for object and weapon detection evaluation,
- **FPS**: frames processed per second, used to measure real-time performance,
- **Latency**: delay between event occurrence and alert gen-eration.

These metrics are important because a surveillance system must not only be accurate, but also fast enough to respond in real time. In safety-critical applications, even a short delay may reduce the usefulness of the alert. Therefore, both detec-tion quality and processing speed are necessary for practical deployment.

C. Observed Outcome

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.154137

The proposed system successfully identifies multiple event types within a single unified monitoring pipeline, demonstrat-ing its capability to handle complex real-world surveillance scenarios. It effectively detects persons and general objects, highlights weapon-like items, recognizes known individuals through watchlist-based face matching, identifies fire and smoke patterns, and detects abnormal human posture as well as crash-like motion events. The integration of these diverse de-tection modules within a single framework allows the system to provide comprehensive situational awareness rather than focusing on a single threat category. Experimental observations indicate that the system maintains stable performance across different environmental conditions, including variations in lighting, partial occlusions, and dynamic backgrounds.

The system also demonstrates efficient real-time performance, making it suitable for continuous monitoring applications. Its modular design allows it to be adapted to different environ-ments such as hospitals, public spaces, industrial facilities, and transportation systems. Overall, the proposed IVSS not only improves detection accuracy and operational efficiency but also enhances the overall effectiveness of modern surveillance systems by combining automation, intelligence, and reliability into a single integrated solution.

D. Result Table

The sample frames shown in Fig. 2 demonstrate how the system annotates different event types in video input. Each detection module produces bounding boxes, labels, or recog-nition output depending on the task.TABLE I: Representative Result Format for the Project

Module	Metric	Observation
Object Detection	mAP / FPS	Real-time detection of persons and objects with stable frame processing
Weapon Detection	Precision / Recall	High-priority alerts generated only after confidence-based verification
Fire Detection	Recall	Reliable flame and smoke identifica-tion with temporal smoothing
Fall Detection	Recall	Abnormal posture and rapid down-ward motion detected across consec-utive frames
Crash Detection	Recall	Sudden impact-like motion identified in surveillance footage
Face Recognition	Precision / Recall	Watchlist matching performed us-ing face embeddings and similarity checking

(a) Weapon detection (b) Fall detection

(c) Fire detection (d) Watchlist match

(e) Crash detection (f) Object detection

Fig. 2: Sample annotated frames from the proposed IVSS

DOI: 10.17148/IJARCCE.2026.154137

VII. CONCLUSION

The successful engineering of this Intelligent Video Surveil-lance System demonstrates that multiple surveillance tasks can be integrated into a single modular framework. The system combines object detection, weapon detection, fire detection, fall detection, crash detection, and face recognition with watchlist matching into a unified pipeline that supports real-time monitoring and alert generation.

The results show that the proposed design is suitable for environments where both safety and security are important. Temporal verification, confidence filtering, and modular fusion improve reliability while reducing false alarms. This makes the system more practical than conventional manual monitoring and more adaptable to real-world deployment requirements.

VIII. FUTURE WORK

While the proposed Intelligent Video Surveillance System (IVSS) provides a strong and effective foundation for real-time multi-threat detection, several improvements can be made in future versions to enhance its performance, scalability, intelligence, and security for real-world deployment.

One important area of improvement is end-to-end crypto-graphic security. Since surveillance systems handle sensitive video data and security-related information, protecting this data from unauthorized access is essential. Future versions of the system can incorporate secure communication protocols such as Transport Layer Security (TLS) to encrypt data trans-mission between cameras, servers, and client applications. In addition, techniques such as certificate pinning can be used to prevent interception attacks, while encryption of stored video clips and alert data can ensure that sensitive information remains protected even if storage systems are compromised.

Another key enhancement is the development of a cloud-based dashboard with multi-tenancy support. Currently, the system can be extended to a cloud architecture that allows central-ized monitoring and remote access through web browsers or mobile devices. This would enable multiple users to access the system simultaneously while maintaining role-based access control to ensure that each user has appropriate permissions. Multi-tenancy support would allow the system to manage multiple organizations, locations, or departments within a single platform, making it suitable for large-scale deployments such as smart cities, enterprise environments, and distributed surveillance networks.

Future work can also focus on automated event response mechanisms. At present, the system primarily detects events and generates alerts, but further improvements can enable automatic actions based on detected threats. For example, the system could send real-time notifications through email, SMS, or mobile applications, automatically generate incident reports or tickets, and trigger emergency responses such as alarms or system shutdowns in critical situations. This level of automation would reduce dependence on manual intervention and significantly improve response time.

Finally, improving privacy controls is an important consid-eration for future development. Since features such as face recognition and watchlist matching involve sensitive personal data, it is necessary to ensure that privacy is maintained. Future versions of the system can incorporate privacy-preserving tech-niques such as face anonymization or masking when identifica-tion is not required. Secure storage mechanisms, strict access control policies, and detailed audit logs can help track data usage and prevent misuse. Additionally, ensuring compliance with data protection regulations and ethical guidelines will be essential for responsible deployment.Overall, these enhancements will make the proposed IVSS more secure, scalable, intelligent, and suitable for real-world applications, while also addressing important concerns related to privacy and data protection.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[2] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018.

[3] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv:2004.10934, 2020.

[4] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *CVPR*, 2015.

[5] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive Angular Margin Loss for Deep Face Recognition," in *CVPR*, 2019.

[6] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[7] J. Deng, J. Guo, J. Zhu, and Y. Zafeiriou, "RetinaFace: Single-stage Dense Face Localisation in the Wild," arXiv:1905.00641, 2019.

[8] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple Online and Realtime Tracking," in *IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3464–3468.

[9] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," in *ICIP*, 2017, pp. 3645–

DOI: 10.17148/IJARCCE.2026.154137

3649.

[10] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields," in *CVPR*, 2017, pp. 7291–7299.

[11] T.-Y. Lin et al., "Microsoft COCO: Common Objects in Context," in *European Conference on Computer Vision (ECCV)*, 2014, pp. 740–755.

[12] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A Face Detection Benchmark," in *CVPR*, 2016, pp. 5525–5533.

[13] X. Han et al., "Fire and Smoke Detection with Burning Intensity Estimation," arXiv preprint arXiv:2410.16642, 2024.

[14] E. R. Daniel, "Wildfire Smoke Detection with Computer Vision," arXiv preprint arXiv:2301.05070, 2023.

[15] H. Ghahremannezhad et al., "Real-Time Accident Detection in Traffic Surveillance Using Deep Learning," arXiv preprint arXiv:2208.06461, 2022.

[16] M. Hussain et al., "YOLOv5, YOLOv8 and YOLOv10: The Go-To Object Detectors for Real-Time Applications," arXiv preprint arXiv:2407.02988, 2024.

[17] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2015.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *CVPR*, 2016, pp. 770–778.

[19] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convo-lutional Neural Networks," in *ICML*, 2019, pp. 6105–6114.

[20] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv:1704.04861, 2017.

[21] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *ICLR*, 2021. .