

Accident and Fight Detection Using Artificial Intelligence

Abhijith Chandran*, Ajay Janardhanan#, Akhil Siji Chandy &, Amaljith KK@, Jyothis Joseph +

Department of Computer Science and Engineering, College of Engineering Kidangoor, Kottayam, Kerala, India

*abhijithchandran3000@gmail.com, #ajayjanardhanan30@gmail.com, &akhilsiji7@gmail.com,

@amaljitkk09@gmail.com, +jyothis@ce-kgr.org

Abstract—The rapid expansion of CCTV surveillance systems has led to an overwhelming volume of video data, making manual analysis of critical incidents such as road accidents and violent activities inefficient and time-consuming. This paper presents an intelligent video analysis system that automatically detects accident and fight-related anomalies from CCTV footage to support post-incident investigation. The system utilizes a custom-trained model based on YOLOv8 to identify anomalous events within video streams. Upon detecting an accident, vehicle number plates are extracted using optical character recognition (OCR) and linked with temporal information. A multi-camera analysis approach is then employed to trace the movement of the involved vehicles across different camera feeds, enabling reconstruction of their route using timestamp correlations and camera metadata. Experimental evaluation demonstrates that the proposed system effectively detects anomalous events and accurately extracts relevant vehicle information, significantly reducing manual effort in surveillance analysis. The system provides a practical and scalable solution for traffic monitoring, law enforcement, and urban safety applications.

Index Terms—broadcast engineering, manuscript template, article preparation, scientific publishing, IEEE style

I. INTRODUCTION

the rapid growth of urbanization and smart city initiatives, CCTV surveillance systems have become widely deployed for monitoring public spaces, traffic flow, and ensuring safety. These systems continuously generate massive volumes of video data, making manual monitoring and analysis both inefficient and impractical. Critical incidents such as road accidents and violent altercations often go unnoticed or require significant human effort and time to investigate, especially when events span across multiple camera feeds. This creates a need for intelligent systems that can automatically analyze surveillance footage and assist in identifying and tracking such events.

Recent advancements in computer vision and machine learning have enabled automated analysis of video data for tasks such as object detection, event recognition, and activity analysis. Models based on convolutional neural networks, particularly the YOLO (You Only Look Once) family, have shown strong performance in real-time object detection tasks due to their speed and accuracy. These models can be adapted and fine-tuned for specific applications, making them suitable for detecting complex events in surveillance footage.

In this work, we propose an intelligent CCTV video analysis system designed to detect accident and fight-related anomalies

and assist in post-incident investigation. The system utilizes a custom-trained model based on YOLOv8 to identify anomalous events in video streams. Upon detecting an accident, the system extracts vehicle number plates using optical character recognition techniques and associates them with temporal metadata. Furthermore, a multi-camera analysis module is employed to trace the movement of vehicles involved in the incident across different camera feeds. By leveraging timestamp correlations and camera location information, the system reconstructs the route of the vehicles, enabling efficient tracking and analysis.

The proposed system aims to reduce the dependency on manual surveillance, improve the efficiency of incident analysis, and provide a practical tool for traffic monitoring and law enforcement. Unlike conventional approaches that focus solely on detection, this work emphasizes both event identification and post-event tracking, making it a comprehensive solution for intelligent surveillance systems.

II. RELATED WORKS

A variety of research efforts have been dedicated to enhancing the landscape of automated surveillance and object recognition

Siddique et al. [2] explored real-time video surveillance automation by evaluating deep learning classification techniques that detect hostile activities using spatiotemporal features. To overcome the prediction flickering of single-frame image classifiers and the high computational cost of traditional depth-sensor methods, the researchers tested lightweight models that can be easily integrated into IoT environments, specifically ConvLSTM, LRCN (with custom CNN and VGG-16 feature extractors), CNN-Transformer, and C3D. The models were trained and tested on a dataset of 350 clips containing both violent behaviors and similar non-violent actions (like hugs, handshakes, and high-fives) to effectively minimize false positive alarms. The experimental results yielded test accuracies of 83.33% on the custom LRCN, 80% on both ConvLSTM and C3D, 76.67% on the CNN-Transformer, and 70% on VGG16-BiLSTM, demonstrating the viability of these models for automated, real-time anomaly detection.

Salman et al. [3] proposes an advanced anomaly detection framework for surveillance videos by combining efficient keyframe extraction and explainable deep learning. The authors introduce SilentFrm, a method that selects important

frames using frame difference analysis to reduce redundancy and computational cost, and XAI-Inv3, an enhanced InceptionV3 model integrated with explainable AI techniques such as Grad-CAM and guided backpropagation. This combination enables both high-accuracy detection and interpretability of model decisions, which is critical for real-world surveillance systems. The approach achieves strong performance across multiple datasets, reaching up to 99% accuracy in violence detection and demonstrating scalability for large-scale CCTV environments.

Kim et al. [4] introduced GTA-Crime, a synthetic dataset for fatal violence detection using GTA5-generated CCTV-like footage. The key contribution is a data generation pipeline + adversarial domain adaptation to bridge the gap between synthetic and real-world data. This improves performance on rare events like shootings and stabbings, which are hard to collect in real datasets.

Zhang et al. [5] proposed a three-stage video analytics framework for extracting high-resolution traffic data from CCTV footage. The system integrates object detection, 2D homography-based perspective transformation, and trajectory reconstruction to estimate vehicle count, speed, and acceleration. A YOLOv5-based detector and tracking algorithms such as BYTETrack were used to generate vehicle trajectories. The model achieved low error rates (4.5% in counts and $\pm 10\%$ MSE in speed), demonstrating accurate real-world traffic analysis using surveillance cameras.

Fahim et al. [6] developed a fine-tuned YOLOv9-based vehicle detection system for intelligent transportation applications. The model was trained on a region-specific dataset containing diverse vehicle classes and achieved a high detection performance of 0.934 mAP at IoU 0.5. The study highlights the importance of adapting object detection models to local environments and proposes deploying the system on CCTV networks for real-time traffic monitoring and data-driven decision making.

Xi et al. [7] proposed a hybrid deep learning framework combining Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) for traffic accident detection from CCTV footage. GANs were used to generate synthetic data to address dataset imbalance, while CNN, fine-tuned CNN, and Vision Transformer models were used for classification. The approach achieved up to 95% accuracy, demonstrating improved detection performance and robustness in complex traffic scenarios.

Manu et al. [8] proposed a vision-based fire detection system using CCTV and image processing techniques. The system employs color-based segmentation, edge detection, and Haar-cascade classification to identify fire regions and trigger alerts in real-time. Implemented on a Raspberry Pi, the model achieved high detection accuracy (98.89%) and demonstrated a cost-effective solution for integrating anomaly detection into existing surveillance systems.

Singh et al. [9] presented a real-time anomaly detection system using CNN and RNN architectures for automated surveillance. The model extracts spatial features using CNN and captures temporal patterns using RNN (LSTM) to classify activities as normal or anomalous. The system detects multiple

abnormal events such as violence, theft, and accidents, reducing the need for manual monitoring and improving real-time threat detection in CCTV systems.

III. METHODOLOGY

This section describes the overall architecture and workflow of the proposed intelligent CCTV surveillance system. The system is designed as a multi-stage pipeline that processes video input, detects anomalies and objects of interest, performs identity recognition, and reconstructs movement paths across multiple cameras.

A. System Overview

The proposed system follows a modular pipeline consisting of four major stages: data acquisition, preprocessing & detection, multi-camera search & recognition, and visualization. The workflow is as follows: CCTV Input \rightarrow Frame Extraction \rightarrow Anomaly Detection \rightarrow Plate Recognition \rightarrow Multi-Camera Matching \rightarrow Route Mapping \rightarrow Dashboard

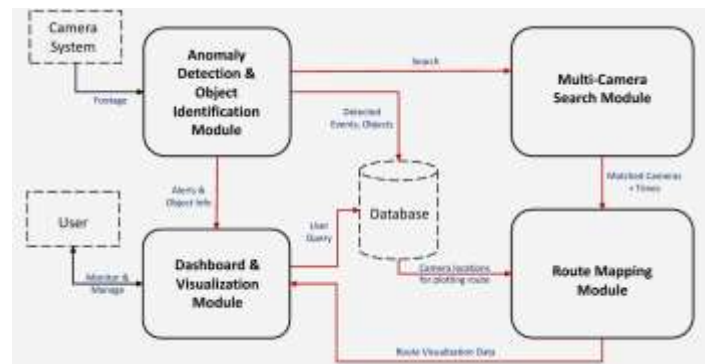


Fig. 1. System Architecture

1) *Anomaly Detection & Object Identification Module*: Analyzes video footage to detect unusual incidents like accidents and fights. Once an accident is identified, it extracts and identifies vehicles associated with the event for further analysis or tracking.

- Processes pre-recorded CCTV footage.
- Detects anomalies using identification models.
- Identifies and isolates key objects linked to the anomaly.
- Stores detected anomaly data and object embeddings in the database.
- Sends alerts and object information to the Dashboard & Visualization module

2) *Multi-Camera Search Module*: Enables the system to locate a detected vehicle across multiple CCTV cameras, either triggered automatically by anomaly detection or manually by user search queries.

- Receives an object image/embedding.
- Searches for visually similar objects across stored CCTV frames and embeddings.
- Retrieves matching cameras and corresponding timestamps.
- Stores search results and logs in the database.
- Sends matched results to the Route Mapping module for movement reconstruction.

3) *Route mapping module*: Reconstructs the movement path of an object based on multi-camera search results. It uses known camera locations to create a simplified visual map of the object’s route before and after an anomaly.

4) *Dashboard and visualization module*: Allows the control room operator or authorized user to monitor anomalies, initiate searches, and view generated routes in an intuitive, interactive environment.

- Displays alerts, anomalies, and detected objects in real time.
- Retrieves and displays search results and route visualizations.
- Handles user queries and forwards commands to appropriate modules.
- Logs user actions and system activity into the database.

B. Model Training

The model was trained using the YOLOv8 framework, adopting a transfer learning approach to adapt a pre-trained object detection model for surveillance-based anomaly detection. The YOLOv8 Nano variant (yolov8n.pt), pre-trained on the COCO dataset, was selected due to its lightweight architecture and efficient training performance. A custom dataset was constructed by integrating multiple publicly available surveillance datasets, followed by a preprocessing pipeline that standardized the data into YOLO format. To address the lack of annotations in certain data sources, an auto-annotation strategy was employed using a pre-trained YOLO model, with additional filtering applied to maintain label quality and control dataset size.

The training process involved fine-tuning the pre-trained YOLOv8 model on the prepared dataset using the Ultralytics training pipeline. The framework handled data loading, batching, and loss computation internally, enabling efficient optimization of the model parameters. The entire training workflow was implemented in a Python-based environment, ensuring modularity and flexibility for future improvements in dataset construction and training configuration.

IV. RESULTS AND DISCUSSION

The proposed surveillance model was evaluated on a validation set of 1,219 images spanning three incident classes — accident, normal, and fight — using standard object detection metrics: precision, recall, F1-score, mean Average Precision at IoU threshold 0.50 (mAP@0.50), and mAP@0.50–0.95. The evaluation was performed using the trained YOLOv8n weights on a Tesla T4 GPU, with an input resolution of 640×640 and a batch size of 16. The per-class and overall results are summarised in Table [I].

TABLE I
 PERFORMANCE METRICS OF THE PROPOSED SURVEILLANCE MODEL

Class	Precision	Recall	F1-Score	mAP@0.50
Accident	99.67%	99.40%	99.54%	99.50%
Normal	91.08%	91.80%	91.44%	96.50%
Fight	84.94%	91.77%	88.22%	95.20%
Overall	91.90%	94.32%	93.07%	97.07%

The model achieved an overall mAP@0.50 of 97.07% and an mAP@0.50–0.95 of 87.70%, demonstrating strong and consistent detection performance across all evaluated classes. The overall precision and recall of 91.90% and 94.32% respectively indicate that the model maintains a well-balanced trade-off between false positives and false negatives, which is critical in a real-time surveillance context where both missed incidents and false alarms carry operational consequences.

Among the three classes, the accident class achieved the highest performance across all metrics, with a precision of 99.67%, recall of 99.40%, and F1-score of 99.54%. This can be attributed to the use of exclusively human-verified annotations for the accident class during the data preparation stage, which ensured high label quality. Auto-annotated accident images were deliberately excluded from training based on observations that noisy auto-generated labels for this class degraded model accuracy in preliminary experiments.

The normal class achieved a precision of 91.08% and recall of 91.80%, yielding an F1-score of 91.44%. The fight class recorded the lowest precision at 84.94%, though its recall remained high at 91.77%, giving an F1-score of 88.22%. The relatively lower precision for the fight class reflects the inherent visual ambiguity between fight and normal scenes — both involve human figures in motion — making precise localization more challenging for the model.

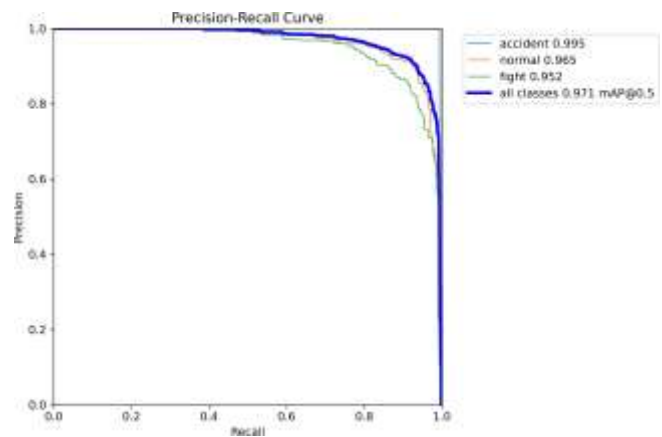


Fig. 2. Precision-Recall curve for each incident class at IoU threshold 0.50. All three classes achieve area under the curve above 0.95, with the accident class reaching 0.995.

The Precision-Recall curve shown in Figure 2 confirms these observations. All three class curves remain tightly clustered near the top-right corner of the plot, indicating consistently high precision even at high recall values. The accident class achieves an area under the curve of 0.995, while normal and fight achieve 0.965 and 0.952 respectively, all well above the 0.95 threshold generally considered indicative of strong detection performance.

The F1-Confidence curve shown in Figure 3 reveals that the model reaches its peak mean F1-score of 0.93 at a confidence threshold of 0.384. The deployed system uses a threshold of 0.60, which sits on the plateau of the accident class curve where F1 remains near its maximum, ensuring that accepted detections are reliable while filtering out low-confidence noise.

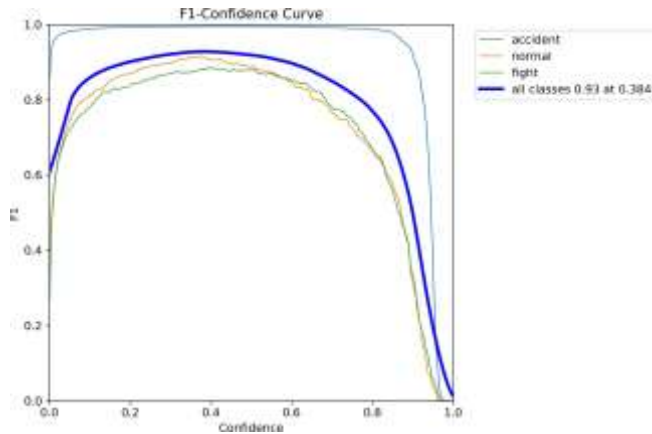


Fig. 3. F1-Confidence curve showing the trade-off between F1-score and confidence threshold for each class. The model achieves a peak mean F1-score of 0.93 at a confidence threshold of 0.384.

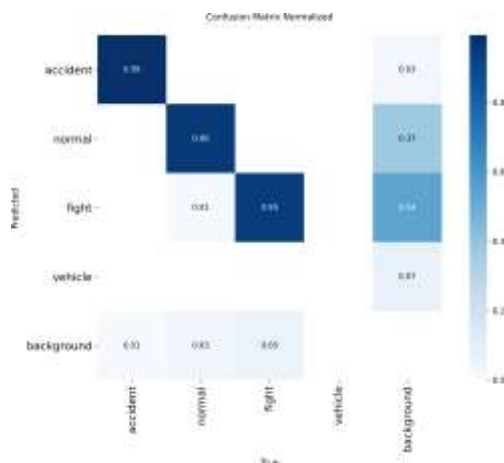


Fig. 4. Normalised confusion matrix of the proposed model. Diagonal values represent true positive rates per class: accident (0.99), normal (0.96), and fight (0.95).

The fight and normal classes show a gradual decline beyond confidence 0.50, which is consistent with the greater visual complexity of those scenes.

The normalized confusion matrix shown in Figure 4 further validates the model’s classification behavior. Correct predictions dominate the diagonal, with the accident class achieving a true positive rate of 0.99, the normal class 0.96, and the fight class 0.95. The only notable off-diagonal pattern is observed in the background column for the fight class, where a false negative rate of 0.54 against background is recorded. This indicates that the model occasionally fails to produce a bounding box for fight events when subjects occupy a small or ambiguous region of the frame. This is a known limitation of single-stage object detectors like YOLO on scenes that lack a clearly localised subject, and can be addressed in future work by incorporating temporal context across frames, which the sliding-window detection gate in the deployed pipeline partially compensates for.

It should also be noted that the vehicle class, while present

in the trained model, was excluded from quantitative evaluation in this study due to the unavailability of independently labelled vehicle validation data. The vehicle class serves primarily as a supporting detection layer for the licence plate extraction phase of the system and does not directly affect incident classification accuracy. Overall, the results demonstrate that the fine-tuned YOLOv8n model is highly effective for automated surveillance incident detection, achieving near-perfect performance on the accident class and strong performance on the more challenging normal and fight classes, with an inference speed of 3.5ms per image making it well-suited for real-time deployment.

V. CONCLUSION

This paper presented an intelligent CCTV video analysis system for automated detection of road accidents and violent incidents, with post-incident vehicle tracking across multiple camera feeds. The system combines a custom-trained YOLOv8n model for anomaly detection, OCR-based licence plate extraction, and a multi-camera route reconstruction module. Experimental evaluation on 1,219 images demonstrated strong overall performance, achieving a mAP@0.50 of 97.07%, precision of 91.90%, recall of 94.32%, and an inference speed of 3.5ms per image, confirming its suitability for real-time deployment. The sliding-window detection gate further enhances reliability by filtering single-frame false positives. Future work will focus on improving fight detection through temporal modelling, expanding validation coverage to all four trained classes, and extending the system to support live video streams and edge device deployment.

REFERENCES

- [1] Abdul Aziz A.B and Aindri Bajpai, “Attire-based anomaly detection in restricted areas using yolov8 for enhanced cctv security,” 2024.
- [2] T. S. Labib Ahmed Siddique, Rabita Junhai and Tanvir Rahman, “Analysis of real-time hostile activity detection from spatiotemporal features using time distributed deep cnns, rnns and attention-based mechanisms,” 2023.
- [3] S. A. F. A. Muhammad Salman, Naveed Abbas and Tanzila Saba, “Enhancing surveillance anomaly detection with keyframes and explainable inception model,” Egyptian Informatics Journal 31, 2025.
- [4] H. Seongho Kim, Sejong Ryu and Je Hyeong Hong, “Gta-crime: A synthetic dataset and generation framework for fatal violence detection with adversarial snippet-level domain adaptation,” 2025.
- [5] A. Linlin Zhang, Xiang Yu and Y. Abdul Rashid Mussah, “Application of 2d homography for high resolution traffic data collection using cctv cameras,” 2023.
- [6] Shahriar Ahmad Fahim, “Finetuning yolov9 for vehicle detection: Deep learning for intelligent transportation systems in dhaka, bangladesh,” 2024.
- [7] Y. Y. Zhenghao Xi, Xiang Liu and Yangyu Zheng, “Integrating generative adversarial networks and convolutional neural networks for enhanced traffic accidents detection and analysis,” 2025.
- [8] R. Manu YM and Shashikala S V, “Anomaly alert system using cctv surveillance,” 2022.
- [9] S. Virender Singh and Dr. Pooja Gupta, “Real-time anomaly recognition through cctv using neural networks,” 2020