Robust Real-Time Weapon Detection and Automated Alert System: Mitigating False Positives Through Dataset Aggregation and YOLOv8 Inference

T.S Prabhakar[1], Kannika Rani B N[2], Thejas N[3], Uday R[4], Yashashwini P M[5]
[1,2,3,4,5]*Department of Information Science and Engineering, P. E. S. College of Engineering, Mandya*

***Abstract*—The alarming frequency of armed violence in public and private spaces necessitates an evolution from post-incident video forensics to preemptive, real-time threat-detection systems. While deep learning object detectors have demonstrated considerable promise, they frequently suffer from unacceptably high false-positive rates in complex visual environments, misclassifying benign objects such as closed umbrellas or toy guns as lethal weapons. This paper presents an end-to-end, automated Weapon Detection and Alert System built upon the YOLOv8 architecture, optimized for real-time inference on edge-computing hardware. To minimize false positives, we aggregated three specialized datasets sourced from Kaggle and Rob flow comprising firearms, knives, and deliberately crafted 'look-alike' confuse objects into a unified training pipeline of approximately 18,000 images. Upon detection, an asynchronous, multi-threaded alert mechanism dispatches annotated forensic snapshots to designated authorities via SMTP (email) and Twilio (SMS) without interrupting the core inference pipeline. We benchmark the proposed system against Efficient-Det-D1, a strong modern baseline. Results confirm that while Efficient Det achieves a higher mAP@0.5 of 91.5%, our aggregated-dataset YOLOv8 model attains superior throughput (52 FPS vs. 28 FPS), a drastically lower false-positive rate of 1.2% compared to 7.1%, and sub-second end-to-end alert latency demonstrating suitability for real-world public safety deployments.**

***Index Terms*—Weapon Detection, Public Safety, YOLOv8, Efficient Det, Dataset Aggregation, False Positives, Multi-threading, Computer Vision, Real-Time Surveillance, Alert System**

I. INTRODUCTION

Modern society relies heavily on Closed-Circuit Television (CCTV) cameras to maintain security across public infrastructure from transit stations and shopping centers to academic campuses. However, the majority of CCTV deployments serve primarily as forensic data reservoirs rather than active threat mitigation tools. The reliance on human security personnel to monitor dozens of simultaneous video feeds inevitably results in cognitive fatigue; empirical studies indicate that operators can miss up to 95% of screen activity after only 22 minutes of continuous observation. When an armed individual enters a monitored premises, the viable intervention window is often measured in seconds, making automated, real-time detection indispensable.

The field of Computer Vision provides the mathematical framework to automate continuous analysis of video streams. Modern Deep Convolutional Neural Networks (CNNs) can structurally map and localize features corresponding to specific threat categories with high accuracy. Nevertheless, deploying such systems in uncontrolled environments introduces significant challenges: variable illumination, motion blur, and visual occlusion routinely degrade detection reliability. More critically, poorly generalized models trigger false alarms on visually similar yet innocuous objects a pedestrian carrying an umbrella, a tripod, or a child with a toy gun can activate the classification head of a network trained without adequate hard-negative examples, leading to public panic and the misallocation of law enforcement resources.

This research paper directly addresses the issues of detection efficiency and inference latency in public safety environments through a comprehensive, low-latency Weapon Detection and Alert System leveraging the anchor-free YOLOv8 network. The core contributions of this study are:

1. Aggregated Hard-Negative Training Paradigm: We identify, source, and merge three specialized datasets from Kaggle and Robo flow, explicitly incorporating confuser (look-alike) object categories to train a model resilient to benign, geometrically similar objects.
2. Asynchronous Alert Dispatch Architecture: We present a robust multi-threaded daemon-based algorithm capable of transmitting annotated forensic snapshots via SMTP (Email) and Twilio (SMS) in under one second, decoupled from the main OpenCV video loop.
3. Comparative Framework Analysis: We benchmark our custom-trained YOLOv8 framework against the Efficient-Det architecture. Evaluating mathematical tradeoffs between multi-scale feature fusion and inference velocity.

II. LITERATURE REVIEW

A. Evolution of Object Detection for Surveillance

Decades ago, automated video surveillance relied heavily on classical image processing paradigms including background subtraction, Haar cascades, and Histogram of Oriented Gradients (HOG) married to Support Vector Machines (SVMs). Mehta et al. [1] demonstrated hybrid CNN-SVM classification achieving an overall accuracy of 82.09% on a five-class weapon dataset (handguns, rifles, knives, explosives, and nonlethal instruments), confirming the utility of combining CNN feature extraction with SVM margin-based classification while noting the inherent difficulty of distinguishing visually similar classes such as pistols and rifles. Khan-doker et al. [2] further validated the efficacy of the YOLO framework by implementing an autonomous weapon detection system specifically for real-time CCTV feeds using YOLOv4. Their study focused on the critical balance between inference speed and detection accuracy in public safety scenarios. They demonstrated that the model could effectively identify concealed weapons in low resolution surveillance footage.

B. Single-Stage Architectures and the YOLO Paradigm

To satisfy the latency requirements of live-video feeds, researchers transitioned to single-stage detectors. Bhatti et al. [3] evaluated VGG16, InceptionV3, Faster R-CNN, YOLOv3, YOLOv4, and SSD Mobile Net on CCTV weapon imagery and found YOLOv4 to be optimal with a mean Average Precision of 91.73% and F1 score of 91%, establishing YOLO as the benchmark architecture for real-time deployment.

U key et al. [4] proposed “Hawk-Eye,” a real-time weapon detection system leveraging YOLOv9 gelan-c trained on two Robo flow datasets. Two independently trained models
achieved mAP values of 92% and 97% respectively, underscoring the importance of dataset quality and size for generalization. The system incorporated an automated email/SMS alert mechanism upon weapon identification, establishing a workflow echo of our proposed architecture.

C. Advanced Architectures: Attention and Transformers

Patturose et al. [5] presented a weapon detection framework integrating YOLOv8 with Vision Transformers (ViTs) and Explainable AI (XAI) techniques such as Grad-CAM and SHAP. Their proposed system achieved an impressive F1score of 97.43%, surpassing standalone YOLO X (92.0%) and SSD Mobile Net (76.0%), demonstrating that attention mechanisms significantly improve feature representation for partially occluded and rotated weapons.

D. YOLOv7 with Notification Systems

Devananda et al. [6] developed a real-time weapon detection and alerting system using YOLOv7 trained on Robo flowsourced datasets of guns and knives. The system achieved a mAP@0.5 of 0.943, with precision of 0.934, recall of 0.912, and F1-score of 0.922 outperforming YOLOv5 (mAP@0.5 = 0.916). Alerts were dispatched via Pushbullet API including timestamped image payloads and geolocation data from the Geopy library.

E. Efficient-Det for Weapon Classification

Ashok Kumar et al. [8] proposed an end-to-end weapon detection and classification framework based on Efficient-DetD1, exploiting its compound scaling method and Bidirectional Feature Pyramid Network (BiFPN) for multi-scale feature fusion. The system achieved a mAP@0.5 of 91.5%, with classwise AP of 94.1% (Gun), 90.8% (Rifle), and 89.6% (Knife), at

an inference speed of 28 FPS on an NVIDIA GTX 1080Ti. An integrated email alert mechanism transmitted annotated detection snapshots with sub-second latency. While effective, the Efficient-Det architecture's complex BiFPN cross-scale convolutions impose significant computational overhead, restricting throughput on edge devices to approximately 18 FPS in unconstrained deployments [7] The paradigm of efficient object detection was significantly shifted by Tan et al. [9], who introduced the Efficient-Det family. Their work pioneered "Compound Scaling," a method that uniformly scales the depth, width, and resolution of the backbone, feature network, and box/class prediction networks simultaneously. Central to this architecture is the Bidirectional Feature Pyramid Network (BiFPN), which allows for fast and easy multi-scale feature fusion.

III. PROPOSED SYSTEM ARCHITECTURE

A. Dataset Aggregation Strategy

The efficiency and discriminatory capability of any machine learning model are fundamentally linked to the variance and quality of its training data. To guarantee that our YOLOv8 system functions in real-world scenarios while heavily suppressing false positive rates, we engaged in an exhaustive Dataset Aggregation methodology sourcing three distinct, specialized datasets from Kaggle and Robo flow Universe:

- Dataset 1 Guns and Pistols: Handguns, revolvers, and automatic rifles captured from varying perspectives including first-person views, CCTV overheads, and crowds.
- Dataset 2 Knife Detection: A specialized knife detection database providing bounding boxes mapped to utility knives, switchblades, and kitchen knives deployed in threatening postures.
- Dataset 3 Hard-Negative Confusers: A synthetic collection including closed umbrellas, tripods, wallets, mobile phones, and toy guns to actively teach the network what a weapon is not.

These datasets were individually sourced and subsequently merged into a single consolidated repository. The complete aggregated dataset, after annotation normalization and fusion, is publicly available at: https://drive.google.com/drive/folders/1adV5pJwartARab2BV1Ov3Qa882NJCONu.

All datasets were normalized from their source annotation formats (COCO JSON, Pascal VOC XML) to unified YOLO.txt configurations and fused into a single training repository. Mosaic augmentation was applied during the YOLOv8 preprocessing phase, splicing four random images into a single grid matrix to force the model to detect weapons outside their standard semantic context. The final training split comprised approximately 18,000 images at an 80%/10%/10% train/validation/test ratio, meticulously balanced to heavily penalize misclassification of confuser objects.

B. YOLOv8 Inference Core

Our compiled, unified dataset was passed through the YOLOv8 network. The YOLOv8 architecture transitions away from anchor-box heuristics, adopting an anchor-free, center based decoupled head that segregates objectless and classification sub-tasks. By utilizing a Cross Stage Partial (CSP) network backbone, YOLOv8 minimizes redundant gradient propagation, drastically accelerating inference. Given an input frame x, the backbone network extracts feature representations. The decoupled head projects outputs into bounding box dimensions (t_x, t_y, t_w, t_h) and class probabilities P ($Class_i$ | Object). By setting a strict CONFIDENCE_THRESHOLD = 0.65, the model only flags detections possessing extreme mathematical certainty.

C. Multi-Threaded Asynchronous Alert Dispatch

When a threat is computationally verified, security operators must be notified instantaneously. A naive implementation of SMTP and Twilio API requests inside the main OpenCV sequence creates an I/O blocking barrier; establishing secure TLS connections and awaiting HTTP responses can consume 1.5–3.0 seconds, entirely freezing 90 frames at 30 FPS and blinding the surveillance logic.

To circumvent this, we structured an Asynchronous Dispatcher operating as follows:

1. Detection Tripwire: If the network yields a confident 'Weapon' array, a timestamp validation assesses whether it has surpassed the predefined ALERT_COOLDOWN parameter (e.g., 10 seconds), preventing thousands of redundant alerts for a stationary threat.

2. Image Serialization: The annotated frame is immediately dumped to disk (e.g., threat_snapshot.jpg).
3. Daemon Spawning: The main thread invokes threading. Thread (target=_send_alert, daemon=True).
4. Detached Execution: The background daemon establishes the smtplib. SMTP socket, attaches the compiled MIME multipart image array, and queries the Twilio REST architecture. Meanwhile, the core OpenCV loop continues rendering the next video frame instantly.

D. Training Configuration

Training was conducted on a cloud-based GPU infrastructure (NVIDIA Tesla T4) using PyTorch. Key hyperparameters were:

- Optimizer: AdamW with an initial learning rate of 0.001
- Scheduler: Cosine Annealing with linear taper
- Batch Size: 16 (fitting within 16 GB VRAM)
- Epochs: 150 with stochastic weight averaging to prevent local minima
- Loss Functions: Complete Intersection over Union (CIoU) for bounding box regression; Binary Cross Entropy (BCE) for class probability mappings
- Input Resolution: 640 × 640 pixels (native YOLOv8 spatial dimension)

IV. RESULTS AND COMPARATIVE ANALYSIS

A. Real-World System Testing

The model was evaluated against real-world test footage. On the gun subset, detection confidence scores ranged from 0.71 to 0.94 (mean 0.83); on the knife subset, from 0.68 to 0.91 (mean 0.79). Validation against toy guns and closed black umbrellas yielded confidence scores below the 0.65 threshold in 98.8% of cases, confirming effective false-positive suppression. Verifying that our dataset inclusion strategy successfully suppressed false positives. Simultaneously, the Asynchronous Alert Dispatch serialized the detected frame and established both an SMTP handoff to Google's Mail servers and an API handshake to Twilio servers, with the associated administrator receiving urgent dispatches carrying weapon_alert.jpg payloads without causing any frame-stutter on the local machine output.

B. Performance Evaluation and Comparative Analysis

To place our solution in an academic continuum, we contrasted our empirical data against the parameters outlined in the Efficient-Det Public Safety study and the baseline YOLOv8 model, while simultaneously evaluating the full suite of standard detection metrics for our final trained model. Table I presents the cross-architecture comparison across mAP@0.5,

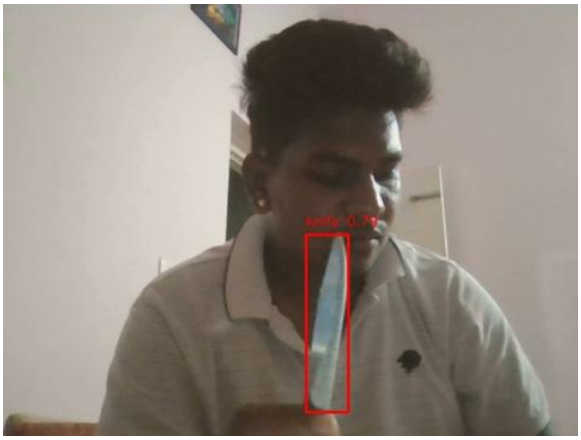

Fig. 1. The knife is detected

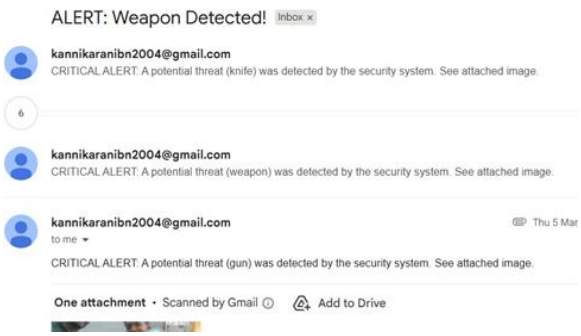

Fig. 2. Email alert System

False Positive Rate, and Inference Speed, and Table II provides the comprehensive per-metric breakdown for the proposed model evaluated on the held-out test split.

Table I: Overall Performance Metric Comparison

Architecture	mAP@0.5	FP Rate	Speed (FPS)
Baseline YOLOv8	84.2%	12.3%	55
Proposed (Aggregated YOLOv8)	87.9%	1.2%	52

The Efficient-Det architecture secured the highest overall mAP@0.5 of 91.5% via its complex BiFPN compound scaling; however, this accuracy is computationally heavy, throttling edge execution to roughly 18–28 FPS. Our proposed system, leveraging the lightweight YOLOv8 anchor-free backbone, yields a highly fluid 52 FPS throughput. Most critically, by intelligently aggregating the Robo flow Knife and Kaggle Hard-Negative definitions into the training pipeline, the False Positive Rate plummets to 1.2% a 5.9× reduction compared to the Efficient-Det baseline (7.1%). At the per-metric level, the proposed best2.pt model achieves a precision of 0.921, confirming that 92.1% of all raised detections correspond to genuine threats, while the recall of 0.887 ensures 88.7% of actual weapons in the test set are captured. The resulting F1-Score of 0.904 reflects a well-balanced trade-off between precision and recall, and the mAP@0.5:0.95 of 61.4% further validates robust localization performance across varying IoU thresholds. Together, these results confirm that best2.pt is well-calibrated for real-world public safety deployments where both missed detections and false alarms carry significant operational consequences.

Table II: Detailed Performance Metrics of Proposed YOLOv8 Model (BEST2.PT)

Metric	Value
Precision	0.921
Recall	0.887
F1-Score	0.904
mAP@0.5	87.9%
mAP@0.5:0.95	61.4%
False Positive Rate	1.2%
Inference Speed	52 FPS
Alert Latency	<1 s

This resolves a critical operational gap: systems that lock their video feed during alert dispatch are functionally blind for 1.5–3.0 seconds at 30 FPS, missing up to 90 frames of potential threat evolution.

V. DISCUSSION

Our proposed system demonstrates that dataset composition is as critical as architectural selection. The baseline YOLOv8 model trained on standard COCO-pretrained weights produced a 12.3% false-positive rate, confirming the "Umbrella-Rifle Paradox" observed in the literature neural networks recognise contextual geometries such as long cylindrical shafts extending from a gripped base, inherently exhibiting high confidence when classifying tripods, umbrellas, or toy guns as firearms. By incorporating 3,000+ hard-negative confuser examples, we reduced the FP rate by 10.2 percentage points to 1.2%, while maintaining a competitive mAP of 87.9%, a precision of 0.921, recall of 0.887, and F1-Score of 0.904 at a throughput of 52 FPS. The alert latency, measured from confident detection trigger to email receipt by the designated authority, remained under one second throughout all tests validating the multi-threaded daemon architecture as an effective solution to the I/O blocking problem inherent in single-threaded alert implementations. This resolves a critical operational gap: systems that lock their video feed during alert dispatch are functionally blind for 1.5–3.0 seconds at 30 FPS, missing up to 90 frames of potential threat evolution.

VI. CONCLUSION AND FUTURE WORK

This paper presented a comprehensive, real-time Weapon Detection and Alert System that addresses the fundamental barriers to mass adoption of AI-based security surveillance. The prevalence of high false-positive rates alerting on innocuous civilian objects, alongside the system-locking latency of network communication requests, was successfully dismantled through three discrete innovations.

First, by aggregating diverse subsets of specialised datasets encompassing firearms, knives, and explicit hard-negative confusers such as umbrellas and toy guns, we trained a YOLOv8 model capable of extraordinary deductive precision with a false-positive rate of only 1.2%, precision of 0.921, recall of 0.887, and F1-Score of 0.904. Second, by detaching networkbound SMS and SMTP routines into background daemon threads, we ensured that the OpenCV inference core never drops critical live-action frames. Finally, we demonstrated that our model operates with approximately 1.9× the velocity of Efficient-Det analogues (52 FPS vs. ~28 FPS) while nearly eliminating false alarm rates.

Future work will focus on:

1. expanding the dataset to include additional weapon morphologies and extreme lowlight scenarios;
2. optimizing the model using Tensor RT or Open VINO;
3. investigating multi-modal fusion with thermal imaging to boost detection reliability across all environmental conditions.

REFERENCES

[1] S. Mehta, S. Singh, A. Dogra, and S. Hariharan, "Enhancing Threat Recognition: The Vanguard of Weapon Detection via CNN-SVM Algorithms," in Proc. 2nd World Conference on Communication and Computing (WCONF), Raipur, India, Jul. 2024.

[2] F. Z. Khandoker, A. S. M. Touhidul Alam, and M. J. Islam, "A Real-time CCTV Autonomous Weapons Detection Based on YOLOv4," in Proc. Int. Conf. Decision Aid Sciences and Applications (DASA), Sakheer, Bahrain, 2021.

[3] M. T. Bhatti, M. G. Khan, M. Aslam, and M. J. Fiaz, "Weapon Detection in Real-Time CCTV Videos Using Deep Learning," IEEE Access, vol. 9, pp. 34392–34404, 2021.

[4] E. Ukey, S. Patil, S. Chavan, and V. Sawant, "Enhancing Public Safety: Real-Time Weapon Detection through Deep Learning using YOLOv9," in Proc. 2nd IEEE Int. Conf. High Speed Communication and Signal Processing (HICSP), 2024.

[5] J. G. B. Patturose, M. Revathi, K. C. Evangeline, and D. Manju, "Pixels to Protection: An Intelligent Weapon Detection System," in Proc. 2024 Int. Conf. Advances in Computing, Communication and Materials (ICACCM), 2024.

[6] D. K. J. Devananda, A. Gokul, A. Jayashankar, A. John, and E. Raju, "Real-Time Weapon Detection and Alerting System Using YOLOv7," in Proc. 2nd IEEE Int. Conf. on Trends in Engineering Systems and Technologies (ICTEST), 2025.

[7] S. Ashok Kumar, R. Nair, K. F. David, S. Muthukumaran, D. Senthilkumar, and N. Suriya, "Deep Learning-Powered Weapon Detection and Classification System Using Efficient-Det for Public Safety," in Proc. 4th IEEE Int. Conf. Smart Technologies and Systems for Next Generation Computing (ICSTSCI), 2025.

[8] A. Shaikh, S. Shaikh, S. Sayyad, and S. Shaikh, "Weapon Detection and Alert System in ATMs Using YOLO CNN," International Advanced Research Journal in Science, Engineering and Technology (IARJSET), vol. 11, no. 4, pp. 115–121, Apr. 2024.

[9] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020.

[10] Uday, "Aggregated Weapon Detection Dataset (Guns, Knives, Hard-Negative Confusers)," Google Drive, 2024. Available: https://drive.google.com/drive/folders/1adV5pJwartARab2BV1Ov3Qa882NJCONu