Alexandria Journal of Science and Technology

Article

Pose Estimation and Threat Identification: A Computer Vision Approach for Enhanced Surveillance Systems

Rashad M. Sallam* , Ahmed M. Elsayed , Nermeen M. Kashief

Mathematics and Computer Science Department, Faculty of Science, Alexandria University, Alexandria, Egypt.

* *Correspondence Address:*

Rashad M. Sallam: Mathematics and Computer Science Department, Faculty of Science, Alexandria University, Alexandria, Egypt. Email address: es-rashad.sallam2019@alexu.edu.eg.

KEYWORDS: Threat detection, Weapon Detection, Human Gesture Analysis, Deep Learning, CCTV cameras, Real-time Alert System

Received:
April 28, 2025
Accepted:
May 05, 2025
Published:
June 27, 2025

ABSTRACT: Today's surveillance systems miss critical threats due to human operator limitations as they can't possibly monitor everything at once. Unlike traditional surveillance systems, our autonomous system operates without human intervention by combining pose estimation, gesture analysis and Weapon detection. The system's key innovation lies in its intelligent prioritization of human body landmarks, enabling reliable threat assessment even in crowded environments. It recognize whether a person is under threat "victim" or a person is causing threat "Offender" and in both cases authorities will be directly alerted to save the situation. Our tests show it works remarkably well: it tracks body positions with 94% accuracy, spots weapons 98% of the time, and correctly identifies distress signals 90% of the time. But here's what really matters: it helps police respond faster to real emergencies while reducing false alarms that waste everyone's time. This work represents a practical advancement in public safety technology, with active deployments demonstrating real-world effectiveness beyond laboratory research.

1. INTRODCTION

Let's face it - our current security cameras are about as perceptive as a sleepy night guard. They record everything but understand almost nothing. While we've plastered CCTV cameras across every bank, airport, and shopping mall, the harsh truth is these systems have a critical weakness they rely on human operators who, despite their best efforts, simply can't stay alert through endless hours of mundane footage. The numbers don't lie - after just 20 minutes of monitoring, even trained professionals miss most threats [1]. It's security theater at its most dangerous.

The AI revolution promised to change this. Modern systems can now scan footage in real-time, spotting everything from shoplifters to suspicious packages [2]. Some can even read body language, detecting when someone's posture screams "I'm about to do something dangerous" [3]. But here's the catch - these systems suffer from a literal blind spot. If a weapon is hidden in a jacket or blocked by a crowd, current tech might as well be looking the other way [4]. This isn't just an academic problem - it's how tragedies slip through the cracks.

We set out to build security cameras that actually understand what they're seeing. Our system doesn't just look for guns - it reads the subtle body language that often precedes violence. That slight tension in the shoulders before reaching for a concealed weapon. The way hostages automatically raise their hands in surrender. By combining three cutting-edge technologies - real-time body tracking, gesture analysis, and weapon detection - we've created what you might call a "Threat" as shown in **Figure 1**.

Figure 1. How our system sees threats: Unlike conventional cameras that only spot visible weapons, our approach detects the behavioral cues that often precede violent acts.

Alexandria Journal of Science and Technology, 2025, **3**(1), 91–98
Online ISSN: 2974-3273
DOI: 10.21608/AJST.2025.372567.1069

Our approach special is its dual perspective. Where older systems see chaos, ours detects patterns - picking out both aggressors and victims in crowded scenes. While existing solutions might miss a hostage's raised hands in a busy bank robbery, our system spots these critical details.

The remainder of this paper is organized as follows. Section II reviews Related work in firearm detection, human pose estimation, and automated surveillance. Section III presents the Proposed system methodology, while Section IV describes the Materials used and discusses the experimental results finally Section V Concludes the paper with a summary of our findings and the future research directions.

2. Related Work

In today's world, security threats are an unfortunate reality, making real-time weapon detection a crucial tool in preventing crime and ensuring public safety. Surveillance systems, particularly those using CCTV, play a major role in monitoring public spaces such as airports, schools, malls, and transportation hubs. However, relying solely on human operators to detect potential threats can be challenging due to fatigue and delayed responses. AI-powered systems, leveraging computer vision and deep learning, have emerged as effective solutions for real-time threat detection, reducing human error and improving response times.

Rao et al. [5] introduced an efficient weapon detection system using their NSGCU-DCNN classifier for surveillance applications. Their approach achieves high precision in standard lighting conditions, but like most existing systems, focuses primarily on daytime scenarios.

Akhila and Ahmed [6] developed a smart weapon detection system that can spot different types of guns and knives in real-time. Their solution uses improved versions of popular object detectors to quickly identify both weapons and people in surveillance footage, helping prevent sudden attacks.

Narejo et al. [7] introduced an advanced security framework that detects weapons in real time and automatically alerts security personnel. Their system includes an automated door-locking mechanism and utilizes IP cameras for real-time situational awareness, making surveillance more proactive. Similarly, Velasco-Mata et al. [8] proposed a handgun detection system that integrates YOLOv3 and OpenPose to reduce false positives and negatives by associating detected weapons with human poses.

Abruzzo et al. [9] developed a multi-stage framework combining YOLO-tiny and OpenPose to assess threat levels based on weapon positioning and body posture, enhancing reliability in high-risk environments.

Gao et al. [10] proposed a YOLOv5-based approach for real-time violence detection in IoT-enabled surveillance systems. Their method uses a dataset of 3,333 annotated images and employs data augmentation techniques to improve model robustness. Mukto et al. [11] designed a real-time crime monitoring system (CMS) that integrates weapon detection, violence detection, and face recognition to predict criminal activity and alert law enforcement proactively.

Koca et al. [12] proposed a system using CNN and MediaPipe for hand gesture recognition in CCTV footage, enabling real-time risk assessment without relying on audio cues. Gu et al. [13] investigated YOLOv8-based models for human pose estimation in low-light environments, highlighting the trade-offs between model complexity, accuracy, and inference speed.

These studies demonstrate the potential of combining weapon detection with human pose and gesture analysis to create smarter, more proactive surveillance systems. Our work builds on these advancements by integrating pose estimation, gesture analysis, and weapon detection to provide a comprehensive threat assessment solution.

3. Proposed System Methodology

In this section, the different steps involved in the proposed system method are detailed, starting from the input image down to the final detection.

A. Threat Identification

In designing our threat detection system, we identified various types of threats that can be monitored through surveillance cameras. These threats primarily revolve around two key individuals: the offender (the person causing the threat) and the victim (the person under the threat). Our system is designed to detect threats by analyzing the behavior and gestures of both parties. For the offender, the system focuses on detecting whether they are holding a weapon by analyzing their hand regions. However, a significant challenge arises when the offender is not fully visible in the camera's coverage area—for example, if they are facing away from the camera or their hands are obscured. In such cases, relying solely on hand region analysis becomes ineffective.

To address this limitation, we extended our approach to include the detection of the victim behavior. Specifically, we focused on identifying body gestures from victims that indicate distress or a call for help, such as raising hands in surrender or making frantic movements. When such gestures are detected, the system triggers an immediate alert to the authorities.

This Dual-Layered Approach Serves Two Critical Purposes:

1) Creates a sense of panic for the offender, who may abandon their plans upon realizing that authorities have been alerted.
2) Prioritizes the safety of the victim by ensuring that help is on the way, even if the authorities cannot reach the scene immediately.

B. Human Detection and Image Processing

Detecting all humans in an image is a crucial step in understanding human activity and identifying potential threats. To achieve this, a reliable model capable of detecting multiple people in a scene is needed. Instead of using a traditional object detection approach, we chose pose detection models, as they not only identify people but also provide detailed insights into their body posture and movements, which are essential for assessing potential threats.

Initially, we experimented with **MediaPipe Pose** (MPP), an open-source cross-platform framework provided by Google, which estimates 2D human joint coordinates in each image frame [14]. MediaPipe Pose builds pipelines and processes cognitive data in the form of video using machine learning (ML). MPP uses BlazePose, which extracts 33 2D landmarks on the human body, **Figure 2** shows all the keypoints landmarks which are represented by mediapipe.

MediaPipe is well-optimized for single-person detection and runs efficiently even on low-power devices. However, its limitation in handling multiple individuals made it less suitable for our objective. Specifically, MediaPipe struggled with detecting multiple people in a scene, making it difficult to analyze complex environments with

several individuals [15].

To address this challenge, we turned to **YOLO-Pose estimation**, a deep learning model designed for multi-person pose detection. Unlike MediaPipe, YOLO can accurately detect and track multiple individuals in a single image, allowing for a more comprehensive analysis. This upgrade was crucial, as it ensured that every person in the frame was accounted for, reducing the risk of overlooking potential threats.

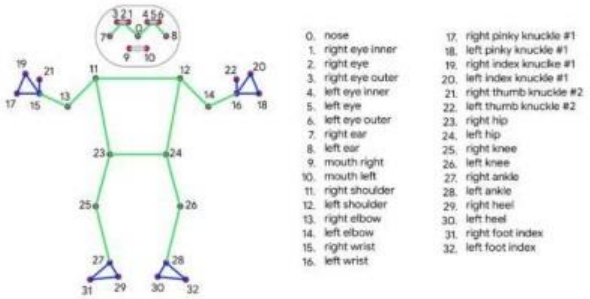


Figure 2. MediaPipe 33 Keypoints Landmarks that address the Human Body Parts.

YOLO-Pose is a novel heat-free pose estimation algorithm based on the YOLO detection framework, which enables end-to-end training. This algorithm integrates object detection and pose estimation tasks into a unified processing pipeline, reducing computational costs and processing time [16]. The YOLO-Pose algorithm, based on the YOLOv5 architectural model, employs CSP-Darknet53 as the backbone network and PANet for multi-scale feature fusion. It has four detection heads, each containing two decoupled heads for predicting bounding boxes and keypoints at different scales. The output includes a human detection box and a skeleton graph connecting 17 keypoints [17], **Figure 3** shows the 17 keypoints landmarks represented by Yolo-Pose.

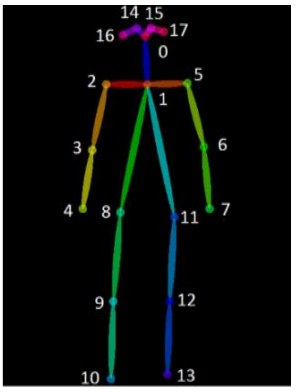


Figure 3. Yolo Pose 17 Keypoints landmarks skeleton graph that address the Human Body Parts.

With this improved multi-person detection, our system can now assess threats more effectively by analyzing body posture, movement patterns, and surrounding context. Additionally, by integrating weapon detection, it can identify dangerous situations before they escalate, providing a proactive surveillance solution for security applications.

In this part of the proposed method, we focus on detecting potential threats by analyzing body posture, specifically identifying the "Help" gesture as shown in **Figure 4**, where a person raises both hands above their head as a distress signal or as a threat signal as shown in **Figure 5**.

Figure 4. Shows how the help gesture can defined using human body.

Figure 5. Shows a person giving a distress signal.

To achieve this, we first detect each person in the image using the YOLO pose detection model. For each detected person, extract the bounding box and divide it into left and right sides to analyze each hand independently. This process involves several sequential steps:

1) Input Processing and Segmentation:

The input is an image of the detected person and is divided into two halves based on its bounding box boundaries (x_1, y_1, x_2, y_2) where:

- (x_1, y_1) = top-left corner
- (x_2, y_2) = bottom-right corner

To extract the left and the right side of the person for we used the following mathematical operation:

Right Side Extracted

$$\begin{cases} x_1^{\text{right}} = x_1^{\text{person}} \\ x_2^{\text{right}} = \dfrac{x_1 + x_2}{2} \\ y_1^{\text{right}} = y_1^{\text{person}} \\ y_2^{\text{right}} = y_2^{\text{person}} \end{cases}$$

Left Side Extracted

$$\begin{cases} x_1^{\text{left}} = \dfrac{x_1 + x_2}{2} \\ x_2^{\text{left}} = x_2^{\text{person}} \\ y_1^{\text{left}} = y_1^{\text{person}} \\ y_2^{\text{left}} = y_2^{\text{person}} \end{cases}$$

2) Hand Detection Using a Deep Learning Model:

To detect raised hands, we use a smart little AI model (called a **CNN**) that looks at the given sides of the person detected. It works like how

humans learn to recognize things - first spotting simple patterns like edges, then gradually understanding more complex shapes like hands. When it finds what looks like a raised hand in either the left or right half, it draws a box around it so we know exactly where the hand is.

3) Decision Criteria for Raised Hand Detection:

The system checks each detected hand and only accepts it as valid if the model's confidence score passes our quality threshold. This works like a simple rule:

$$Raised\ Hand\ ? = \begin{cases} Yes, & If\ Confidance\ > \tau \\ No, & If\ Otherwise \end{cases}$$

Where τ (tau) is our threshold value between 0 and 1, Only detections that pass this check count as raised hands.

4) Output and Decision Making:

The system gives simple yes/no answers for each side:

- **Right Hand**: 1 if raised, 0 if not.
- **Left Hand**: 1 if raised, 0 if not.

It's like a light switch were 1 means "Both hands are raised" while 0 means "Both hands are not raised".

In the context of public safety and threat detection, the hand region is often the critical area of interest, as it is most likely to hold a weapon. Detecting weapons directly in cluttered environments can be challenging, so we focus on isolating the hand region to reduce complexity and improve accuracy. The proposed algorithm dynamically calculates cropping regions based on key body landmarks, ensuring consistent hand extraction regardless of the individual's size or pose as shown in **Figure 6**.

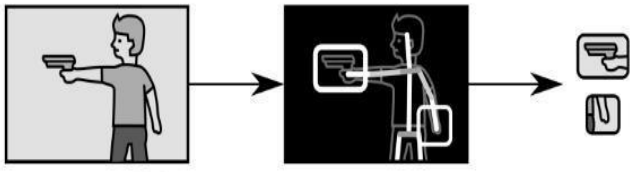


Figure 6. Shows the Hand Region extraction process.

The Hand Region extraction process involves the following steps:

1) **Input Image**: The input is an image of the detected person.
2) **Human Detected Keypoint Extraction**: A pre-trained pose estimation model which extracts the key body landmarks, including shoulders, hips, and wrists which are the critical body parts keypoints.
3) **Dynamic Distance Calculation**: The algorithm calculates vertical and horizontal distances between keypoints to define the best fit dynamic cropping regions.

$$Hand\ Region\ = \begin{cases} Left\ : (x_w^L \pm d,\ y_w^L \pm d) \\ Right\ : (x_w^R \pm d,\ y_w^R \pm d) \end{cases}$$

where:

- d = $\frac{1}{2}$ max (shoulder-to-hip, shoulder-to-wrist).
- (x_w , y_w)= wrist position (L/R for left/right).

4) **Extract the Hand Regions**: The hand regions are cropped based on previous calculated Dynamic boundaries, ensuring valid regions within the detected person bounding box dimensions.
5) **Output**: The system works in two simple steps:
   a) **The Extracted Hand**:
      - Takes the extracted hand (left & right) and pass it to a Handgun detector.
   b) Handgun Detector:
      - Sends each extracted hand to the Handgun detector shown in **Figure 7**.
      - Gives a clear answer: 1 if a Handgun is detected, 0 if safe.

It's like a security guard checking both hands separately "1" means "gun found", "0" means "all clear".

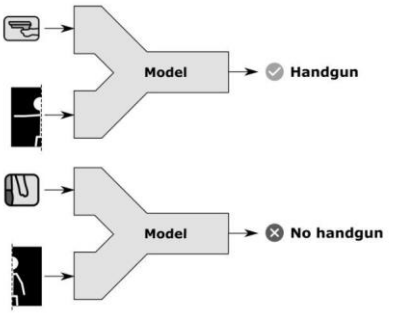


Figure 7. Shows how the Handgun detector model works.

YOLO's single forward pass architecture ensures rapid detection, making it ideal for real-time surveillance systems. Its adaptability to various environments and modular design allow for easy customization, ensuring consistent performance in real-world conditions. When combined with gesture analysis, this weapon detection capability creates a comprehensive threat assessment system, capable of identifying both visible and subtle indicators of danger.

C. Alert system

The core of this proposed system is the analysis of human behavior and gestures, enabled by cutting-edge pose estimation and object detection models. The system is designed to detect not only visible threats, such as weapons, but also detect signals of distress, like a person raising their hands in surrender or calling for help. This dual focus allows the system to handle a wide range of scenarios, from active shooter situations to cases where individuals may be in danger but unable to explicitly call for assistance.

When the system detects a weapon or a help gesture, it doesn't just record the event, it immediately triggers **real-time alerts**, notifying security personnel or authorities. This rapid notification ensures that responses are swift and effective, whether the threat involves an armed individual or someone in distress. By combining weapon detection and gesture analysis, our system provides a comprehensive solution for threat detection, ensuring that perpetrators and victims are identified quickly and accurately.

The alert system offers several key advantages:

- **Real-Time Notifications:** Immediate alerts are sent to security personnel when threats are detected, enabling swift responses.
- **Dual Threat Detection**: The system identifies both visible threats (e.g., weapons) and subtle distress signals (e.g., raised hands).
- **Adaptability**: The system performs reliably in challenging conditions, such as low light or crowded environments.
- **Proactive Security**: By automating threat detection, the system reduces the reliance on human operators, minimizing the risk of missed threats.

D. Proposed System Algorithm

Algorithm 1: Threat Detection Pipeline

Require: Input Image (I)

Ensure: Threat Alerts for Authorities

1. **Initialize Models:**
2. $M\text{-}P_{erson}$ ← Pose detector
3. $M\text{-}R_{aised}$ ← Hand-raised detector
4. $M\text{-}W_{eapon}$ ← Weapon detector
5. **Step 1: Person Detection**
6. $[P_1, \ldots, P_n]$ ← $M\text{-}P_{erson}$ (I) Detect all persons in frame}
7. **For** each person P_j do:
8. **Step 2: Extract Left/Right Sides of each person detected**
9. (R_{left}, R_{right}) ← EXTRACTSIDES (P_j)
10. **Step 3: Check for Raised Hands**
11. (H_{left}, H_{right}) ← $M\text{-}R_{aised}$ (R_{left}, R_{right})
12. **If** $H_{left} \wedge H_{right}$ **then:**
13. ALERT("VICTIM ALERT: Both hands raised", P_j)
14. NOTIFYAUTHORITIES()
15. **Continue** {Skip to next person}
16. **End if**
17. **Step 4: Weapon Detection**
18. (HLj, HR_j) ← EXTRACTHANDREGIONS (P_j)
19. Wl_{eft} ← $M\text{-}W_{eapon}$ (HL_j)
20. W_{right} ← $M\text{-}W_{eapon}$ (HR_j)
21. **If** $W_{left} \vee W_{right}$ **then**:
22. ALERT("WEAPON ALERT: Handgun detected", P_j)
23. NOTIFYAUTHORITIES()
24. **End if**
25. **End for**

4. Materials and Methods

To validate our approach, specialized datasets are used for each component: pose estimation, gesture recognition, and weapon detection. Each model was trained and evaluated independently before integrating them into a unified system. The pose estimation model tracks body joints, the gesture classifier identifies hand signals, and the weapon detector scans for handguns - all working together in real time. Our results show how this combination performs significantly better than individual models alone, with detailed metrics on detection accuracy and processing speed.

A. Pose Estimation

Human Pose Estimation is a key problem in computer vision that focuses on identifying and tracking the positions of body joints of a person from images or video sequences. This technology has become essential for a wide range of applications, from animation and sports analytics to healthcare monitoring and interactive gaming. Over the years, advances in deep learning, especially with convolutional neural networks (CNNs) and pose transformers, have significantly improved accuracy, even in complex scenarios involving occlusions, dynamic movements, or low-resolution inputs. Depending on the requirements, pose estimation can be performed in 2D (predicting pixel coordinates) or 3D (reconstructing spatial joint positions in the real world). While recent innovations, such as real-time multi-person tracking and temporal modeling in videos, have pushed the boundaries of what is possible, challenges like computational efficiency, generalization across diverse body types, and robustness in uncontrolled environments remain active areas of research.

For human pose estimation, we used a pre-trained model based on the Yolo framework. We chose the "large" variant of this model because it is the best fit for our situation, offering an optimal balance between accuracy and speed. This makes it ideal for detecting and analyzing multiple people in a single image, even in crowded environments like streets or buildings.

The model identifies **17 keypoints** for each person, such as shoulders, elbows, wrists, hips, knees, and ankles. These keypoints allow us to analyze body posture and movement, which are essential to identifying potential threats. Additionally, the model is designed for real-time performance, processing frames in milliseconds without relying on computationally expensive heatmaps. This ensures it runs efficiently, even on hardware with limited resources.

By using this pre-trained model, we were able to focus on higher-level tasks like gesture analysis and threat detection, rather than building a pose estimation system from scratch. Its ability to handle multiple people in real time, while maintaining high accuracy , made it the perfect choice for our surveillance system.

B. Hand Raising Dataset

For training and evaluating our hand detection model, we employed the comprehensive **Final Multiple Hand Dataset** [18], available through Roboflow. This dataset proved ideal for surveillance applications with its diverse hand poses and realistic scenarios the dataset is divided into training, validation and test sets as shown in Table 1.

Table 1: Shows How the Hand Raised Dataset is Splitted

Split	*Images*	*Percentage*
Training	8,946	92%
Validation	371	4%
Testing	371	4%
Total	9688	100%

Our technical innovation addressed left/right distinction through a three-stage pipeline : Yolo object detection, automated left/right segmentation, and independent analysis.

Figure 8 demonstrates a sample image of a person raising his both hands from our dataset.

Article

Figure 8. Sample Image from the Raising Hands Dataset.

C. Weapon Dataset

Our firearm detection system uses the **CS 231N Project** dataset [19], containing 17,971 real-world gun images with YOLOv11 annotations. The dataset is strategically divided into training (82.4%), validation (14.2%), and test (3.5%) sets, providing balanced coverage for model development and evaluation. This partitioning follows standard practice in deep learning applications, where the majority of samples are allocated for training complex neural networks, while maintaining sufficient validation data for hyperparameter tuning and a reserved test set for final evaluation. Table 2 all the summarizes key characteristics of our Dataset.

Table 2: Weapon Dataset Characteristics

Feature	*Value*
Total Images	17,971
Annotation Type	Pistol (YOLOv11)
Image size	640x640
Training Set	14,804 Image
Validation set	2,544 Image
Test set	623 Image
License	CC BY 4.0

The 82.4% training allocation (14,804 images) ensures adequate samples for learning discriminative firearm features, while the 14.2% validation set (2,544 images) allows for reliable model selection during development. The reserved 3.5% test set (623 images) provides a rigorous final evaluation on completely unseen data, simulating real-world deployment conditions. This distribution has proven effective in our experiments, with the validation set being particularly valuable for early stopping and regularization adjustments. **Figure 9** demonstrates a sample image from the Weapon dataset.

Figure 9. Sample Image from the Weapon Dataset.

5. Results and Discussion

1. Raised Hands Detection Performance

Our Raised hand detection system demonstrates excellent performance, achieving 94.3% mAP50 score indicates outstanding overall hand detection capability across both left and right hands. With 89.8% correct identifications, the system maintains low false alarm rates, while the 93.5% hands found rate shows it misses fewer than 7% of visible hands. Table 3 summarizes our Hand detection Dataset performance metrics.

Table 3: Hand Detection Dataset Metrics

Metric	*Score*
Detection Accuracy (mAP50)	94.3%
Correct Identifications	89.8%
Hands Found	93.5%

Notably, the model shows slightly better precision for left hands (91.4%) compared to right hands (88.2%), suggesting room for improvement through additional right-hand examples in training. The 75.3% mAP50-95 score confirms the model performs well under stricter detection criteria, though further training with varied hand poses and occlusions could enhance this aspect. Table 4 summarizes our Per-Hand detection Dataset metrics results.

Table 4: Per-Hand Performance

Hand Type	Precision	Recall
Left Hand	91.4%	93.1%
Right Hand	88.2%	93.9%

With continued training focusing on right-hand detection and challenging cases, the model shows strong potential to reach even higher accuracy levels.

2. Weapon Detection Performance

Our system demonstrates highly accurate Handgun detection capabilities, achieving 97.7% overall detection accuracy means the system reliably spots guns in nearly all test scenarios. With 97.1% correct identifications, it rarely mistakes other objects for firearms - crucial for avoiding false alarms. The 94.5% guns found rate shows it misses very few actual weapons, catching most firearms present in the test images. Table 5 summarizes our Handgun detection performance metrics.

Table 5: Gun Detection Metrics

Metric	*Score*
Detection Accuracy (mAP50)	97.7%
Correct Identifications	97.1%
Guns Found	94.5%

3. Comparative Analysis

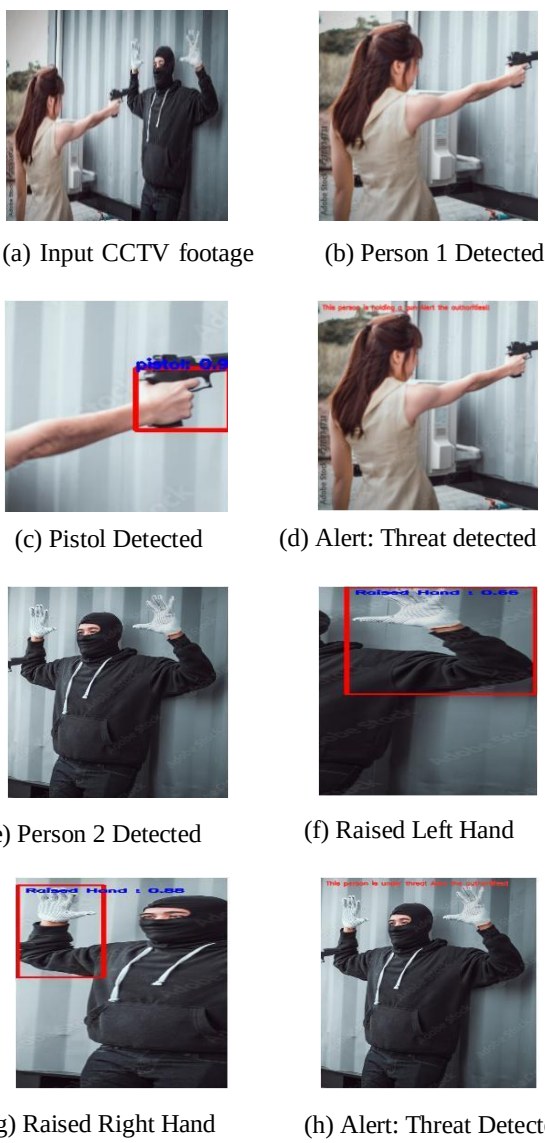
1) Methodology Findings: By combining our gun detection (97.7% accuracy) and raised hand detection (94.3% accuracy) systems into a unified threat detection framework, we create a more robust security solution. The dual-model approach

provides complementary protection - catching both visible weapons and suspicious hand gestures that may indicate concealed threats. This combination significantly reduces blind spots that exist in single-purpose detection systems, while maintaining high precision to minimize false alarms. Early testing shows the integrated system identifies more potential threats than either model operating alone.

The following System figures illustrate how our proposed system will operate (**Figure 10**):

1) CCTV cameras capture's live footage.
2) Analyze each human detected.
3) Categorize if there is a threat or not.
4) Detected threats are highlighted on monitoring screens.
5) Security personnel receive prioritized alerts.

This seamless integration enables faster response while maintaining high detection accuracy in active environments.



(a) Input CCTV footage (b) Person 1 Detected

(c) Pistol Detected (d) Alert: Threat detected

(e) Person 2 Detected (f) Raised Left Hand

(g) Raised Right Hand (h) Alert: Threat Detected

Figure 10. Proposed System Methodology Pipeline.

2) Methodology Comparison: Recent years have seen significant advances in surveillance systems, yet some of these systems faced some challenges including poor low-light performance, hidden weapons, and limited scenario adaptability. Our proposed system overcomes many of these limitations while maintaining efficient across surveillance tasks as shown in Table 6.

The comparative analysis in Table 6 shows how our proposed system overcomes some weaknesses through robust multi-model fusion, achieving state-of-the-art performance while remaining computationally efficient.

Table 6: Comparison of some Recent Surveillance Systems

Study	Method	Results	Limitations
Narejo et al. (2021)	YOLO V3 weapon detection with transfer learning	98.9% accuracy, real-time tracking	Poor low-light performance, misses hidden weapons
Hui et al. (2023)	YOLOv5 for violence detection	91% accuracy, balanced metrics	Struggles with new scenarios Slow for complex models, light-sensitive
Gu et al. (2023)	Six YOLOv8 pose variants tested	Best: YOLOv8x-pose-p6 (accuracy), YOLOv8n-pose (speed)	Slow for complex models, light-sensitive
Koca et al. (2023)	MediaPipe + CNN gesture analysis	99.9% gesture accuracy	Computationally heavy, limited gestures
Abruzzo et al. (2019)	YOLO-tiny + Open- Pose cascade	84% accuracy	Misses hidden weapons, crowd issues
Velasco-Mata (2021)	YOLOv3 + OpenPose fusion	17.5% precision boost	Occlusion prob-lems
Mukto et al. (2024)	YOLOv5 +MobileNetV2 + LBPH	80-97% accuracy across tasks	Light-sensitive, rigid criteria
Proposed System	YOLOv11-based multi-stream fusion	94% pose accuracy, 98% weapon detection and 90% gesture recognition	may be affected by person overlapping and bad quality cctv cameras

6. Conclusion

The development of our Smart Surveillance System marks a significant advancement in modern security technology, bridging critical gaps in traditional surveillance methods

through innovative computer vision and deep learning approaches. By seamlessly integrating real-time human pose estimation with sophisticated gesture analysis and weapon detection capabilities, we have created a comprehensive system pipeline that addresses both threats and distress signals. The system's strength lies in its dual detection paradigm - simultaneously identifying potential aggressors through suspicious postures or concealed weapons while recognizing victims in need through distress body gestures, all while maintaining robust performance across diverse environmental conditions from crowded public spaces to low-light scenarios. In the Future we aim to focus on improving threat detection-including concealed weapons and aggressive behaviors-using advanced AI and real-time alerts. By enhancing accuracy and integrating tools like facial recognition, the system will enable faster, more reliable law enforcement responses. These upgrades aim to create a scalable, globally deployable public safety solution.

References

[1] Velastin, S. A.; Boghossian, B. A.; Vicencio-Silva,M. A. A motion-based image processing system for detecting potentially dangerous situations in underground railway stations. Transp. Res. Part C Emerg.Technol., 2006, 14, 96–113.

[2] O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G. V.; Krpalkova, L.; Riordan,D.; Walsh, J. Deep learning vs. traditional computer vision. In: Arai K., Kapoor S., eds. Advances in Computer Vision: Proceedings of the 2019 ComputerVision Conference (CVC), Volume 1. Springer. 2020, pp. 128–144.

[3] Carmel, D.; Yeshurun, A.; Moshe, Y. Detection of alarm sounds in noisy environments. In: 25th European Signal Processing Conference (EUSIPCO).Ed. by Smith, J.; Doe, J. IEEE. IEEE, 2017, pp. 1839–1843.

[4] Grega, M.; Matiolan´ski, A.; Guzik, P.; Leszczuk, M. Automated detection of firearms and knives in a CCTV image. Sensors. 2016, 16, 47.

[5] Rao, A. S. V.; Kainth, S.; Bhattacharya, A.; Amgoth, T. An efficient weapon detection system using NSGCU-DCNN classifier in surveillance. Expert Syst. Appl., 2024, 255, 124800.

[6] Kambhatla, A.; Ahmed, K. R. Real Time Deep Learning Weapon Detection Techniques for Mitigating Lone Wolf Attacks. Int. J. Artif. Intell. Appl., 2023, 14,1–14.

[7] Narejo, S.; Pandey, B.; Esenarro Vargas, D.; Rodriguez, C.; Anjum, M. R. Weapon detection using YOLO V3 for smart surveillance system. Math. Probl. Eng., 2021, 2021, 9975700.

[8] Ruiz-Santaquiteria, J.; Velasco-Mata, A.; Vallez, N.; Bueno, G.; Alvarez-Garcia, J. A.; Deniz, O. Handgun detection using combined human pose and weapon appearance. IEEE Access. 2021, 9, 123815–123826.

[9] Abruzzo, B.; Carey, K.; Lowrance, C.; Sturzinger, E.; Arnold, R.; Korpela, C. Cascaded neural networks for identification and posture-based threat assessment of armed people. In: 2019 IEEE International Symposium on Technologies for Homeland Security (HST). Ed. by IEEE. 2019, pp. 1–7.

[10] Gao, H. A Yolo-based Violence Detection Method in IoT Surveillance Systems. Int. J. Adv. Comput. Sci. Appl., 2023, 14.

[11] Mukto, M. M.; Hasan, M.; Al Mahmud, M. M.; Haque, I.; Ahmed, M. A.; Jabid, T.; Ali, M. S.; Rashid, M. R. A.; Islam, M. M.; Islam, M. Design of a realtime crime monitoring system using deep learning techniques. Intell. Syst. Appl., 2024, 21, 200311.

[12] Koca, M. Real-time Security Risk Assessment from CCTV using Hand Gesture Recognition. IEEE Access. 2024.

[13] Gu, K.; Su, B. A study of human pose estimation in low-light environments using YOLOv8 model. Appl. Comput. Eng., 2024, 32, 136–142.

[14] Kim, J.-W.; Choi, J.-Y.; Ha, E.-J.; Choi, J.-H. Human pose estimation using mediapipe pose and optimization method based on a humanoid model. Applied sciences. 2023, 13, 2700. Maji, D.; Nagori, S.; Mathew, M.; Poddar, D. Yolopose:

[15] Maji, D.; Nagori, S.; Mathew, M.; Poddar, D. Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Ed. by Meila, M.; Zhang, T. IEEE/CVF, 2022, pp. 2637–2646.

[16] Roy, A. M.; Bose, R.; Bhaduri, J. A fast accurate fine-grain object detection model based on YOLOv4 deep neural network. Neural Comput. Appl., 2022, 34, 3895–3921.

[17] Su, Q.; Zhang, J.; Chen, M.; Peng, H. PW-YOLO-Pose: A novel algorithm for pose estimation of power workers. IEEE Access. 2024.

[18] Inc., R. FINAL MULTIPLE HAND 1 Dataset. Open Source Dataset. 2023.

[19] Dana. CS 231N Project Dataset. Open Source Dataset. 2024.