Hidden Weapon Detection and Tracking in Thermal Video Using Auto-Labeling and Deep Learning Approaches

Hassanein Yarob Saeed Albakaa * and Ali Abdulkarem Habib Alrammahi
Department of Computer Science, College of Computer Science and Mathematics, University of Kufa, Najaf, Iraq
Email: hassanein.albakaa@gmail.com (H.Y.S.A.), alia.alramahi@uokufa.edu.iq (A.A.H.A.)

Manuscript received June 24, 2025; revised August 1, 2025; accepted August 7, 2025
*Corresponding author

***Abstract*—This study proposes an automated framework for detecting and tracking concealed weapons in thermal video, aimed at real-time surveillance in high-security public areas like airports and stadiums. Due to the lack of relevant public datasets, thermal videos were recorded using an infrared camera in scenarios simulating concealed weapons. Frames extracted from the videos were automatically annotated using the GroundedSAM model, which aligns textual prompts ("gun", "knife") with image content, eliminating manual labeling. A YOLOv11n model was trained on over 6,200 labeled thermal images, achieving 81% precision, 82% recall, and 87% mAP@50. For tracking across frames, a Graph Neural Network (GNN) connected detected objects over time, 0.88 consistency for guns, and 0.66 consistency for knives. The integration of smart annotation, thermal-aware detection, and GNN tracking demonstrates strong potential for real-time, robust weapon detection in crowded, security-sensitive environments.**

***Index Terms*—GroundedSAM, autodistill, YOLOv11n, thermal video dataset, Graph Neural Network (GNN) tracking algorithm, special evaluation metrics for tracking algorithm**

I. Introduction

Continuous monitoring for concealed weaponry and dangerous objects has remained a significant problem in security constantly, especially in crowded areas (e.g., airports) and public venues (e.g., stadiums) where manual surveillance of each person is impossible [1, 2]. This has fueled an increasing requirement for automated, stand-off security systems to detect threats with sufficient precision to allow for their isolation or neutralization on time with minimum damage to civilian persons [1, 3].

Electromagnetic (EM) waves have long been utilized in non-contact security screening systems. Technologies such as X-rays, Millimeter Waves (MMW), and Terahertz (THz) waves are among the most employed modalities for detecting concealed objects in human subjects [1, 3, 4]. The internal X-ray images are more detailed, as X-rays have a high penetration depth. Still, their use is becoming more and more limited due to the health risks related to ionizing radiation exposure [5–7].

Although the wavelength of the IR thermography is between a few and ten micrometers, which is not the same as that of MMW and THz, security practices have certain special advantages. While it has some limitations regarding penetration under thick clothing, it offers better spatial resolution (and thus better visualization at a distance) [2, 8, 9]. Also, because iris recognition as a biometric system requires facial identification (due to the facial characteristics embedded in the iris), such IR-based imagery can provide enhanced privacy and offers fewer facial features implied in the form of the iris capture at passive presentation]10[. In passive thermography, the human body acts as the thermal source, and variations in heat distribution caused by concealed objects can be detected through thermal gradients on the surface of clothing [11, 12].

Even though IR has limited ability to perceive the surrounding environment, IR technology has exhibited great potential when integrated with Artificial Intelligence (AI) methods, especially Convolutional Neural Networks (CNNs), which have achieved good performance in several thermal imaging missions such as damage detection and object segmentation [13–16]. Nevertheless, IR-based systems may have inconsistency when they are manually made (especially in the case of layered clothing). Thus, adopting machine learning algorithms is important to increase the robustness and accuracy of detection results [17].

This study proposes a deep learning-based approach for concealed weapon detection and tracking in thermal video using CNNs. The researcher collected a custom dataset of thermal videos, and the training labels were automatically generated using the GroundedSAM model, which annotates thermal images based on textual descriptions such as "weapon" or "knife." These annotations were then used to train a YOLOv11n object detection model. The proposed framework enables efficient and automated threat identification, significantly reducing the reliance on human intervention in surveillance scenarios.

II. Related Work

The recent progress in deep learning, along with the advances in thermal imaging, has markedly enhanced the performance of detecting concealed weapons from challenging surveillance scenes. Several studies have

doi: 10.18178/ijeetc.14.6.403-411

investigated combining CNNs architecture with infrared data to detect hidden threats with high precision. The authors have also used thermal-RGB fusion, active segmentation, and logistic regression-based classification in a low-visibility environment. Despite encouraging performance, most current models do not support real-time tracking and depend on costly manual annotation. In this section, we introduce related work and emphasize the novelty of our combined auto-labeling GNN-based tracking framework.

Santos *et al*. [18] conducted a systematic review on deep learning-based weapon detection in surveillance footage, focusing on the methods employed, dataset characteristics, and challenges in automatic weapon detection. The review highlights several models, including Faster R-CNN and YOLO (You Only Look Once) architecture. The study discussed datasets that incorporate Realistic images and synthetic data, which have been shown to improve detection performance. The review notes that while various models demonstrated improvements in accuracy, specific numerical metrics were not consistently provided. However, it emphasizes that the performance of these models significantly decreases under challenging conditions, such as poor lighting and small weapon detection. The main challenges identified include Poor lighting conditions affecting detection accuracy.

Gosain *et al*. [19] presented an automatic weapon detection system based on image processing and machine learning that aims to be an alternative to classic X-ray techniques. The presented approach combines thermal/IR images and traditional RGB or HSV images by applying Discrete Wavelet Transform (DWT) to suppress noise while retaining essential features for weapon detection and extracting the facial features. The processed video feeds are then analyzed through a two-step pipeline: we make use of YOLOv6 to detect objects by having a focus on weapon-related segments to prevent false positives and then we employ a Convolutional Neural Network (CNN) based on VGGNet to classify firearms. This final weapon classification model is logistic regression, which has been trained on a custom 1,759 weapon image dataset. Tested on 32 random images, the system produced promising performance metrics, suggesting possible real-time surveillance and threat localization by correctly identifying concealed weapons, yet preserving facial details for quick threat elimination.

Ahmar *et al*. [20] used them as the benchmark to compare the performance of thermal imaging for object detection and tracking in degraded lighting context with state-of-the-art detection-based algorithms, namely, the Task-Aligned One-Stage Object Detection (TOOD) and the Varifocal Feature Network (VFNet), applied to thermal vs. RGB images. Their experiments on the City Scene RGB-Thermal MOT Dataset, which consists of paired thermal and visible images labeled manually from FLIR cameras, showed that the thermal-based trackers are much stronger than the RGB trackers, especially in low-light conditions. While precise accuracy measurements were not provided, the thermal models achieved high recall and strong detection across a range of conditions. In addition, a dynamic cut-off was also introduced within the tracking-by-detection pipelines, which uses the bounding box dimension as input to improve the association in multi-object tracking. These results highlight the usefulness of thermal imaging in reliable object detection and tracking in challenging lighting conditions.

Veranyurt *et al.* [21] proposed a deep learning-based framework for real-time detection and localization of concealed pistols using thermal imagery. The system combines two deep learning models: a fine-tuned VGG19 convolutional neural network for classification, which achieved an F1 score of 0.84 on the test set, and a fine-tuned YOLOv3 model for multi-task classification and localization, attaining a mean average precision (mAP) of 0.95 with bounding box detection in approximately 10 milliseconds. The authors created a custom dataset comprising thermal video recordings of multiple human models alongside publicly available thermal images to simulate various concealment scenarios. Their results demonstrate the effectiveness and robustness of integrating thermal imaging with advanced neural networks for enhancing security surveillance through accurate and rapid concealed weapon detection.

Khor *et al*. [2] investigated the use of infrared thermography combined with machine learning techniques for non-invasive detection and classification of concealed objects beneath clothing. The study employs Convolutional Neural Networks (CNN), specifically a transfer-learned ResNet-50 model pre-trained on ImageNet, fine-tuned with infrared images from controlled experiments simulating security checkpoint scenarios. Several image preprocessing techniques were applied to enhance object visualization, including principal component analysis, Chan-Vese active contour segmentation, and Fuzzy-c clustering. The optimized ResNet-50 classifier achieved Area-Under-Curve (AUC) values of 0.869 and 0.922 on datasets of 900 and 3082 images, respectively, with prediction errors reduced to 19.9% and 14.9% after threshold optimization. This work demonstrates the potential of combining thermal imaging and deep learning for effective, real-time concealed object detection in security screening applications.

Muñoz *et al*. [17] proposed a novel two-stage method for concealed handgun detection that integrates thermal imaging with deep learning techniques. The method first detects potential firearms at the frame level and subsequently verifies their association with detected persons, effectively reducing false positives and negatives.

A significant contribution is the development of a lightweight algorithm optimized for low-end embedded devices, facilitating deployment on wearable and mobile platforms such as chest-worn Android smartphones equipped with miniature thermal cameras. The study includes a tailored thermal dataset simulating controlled concealment scenarios for system validation. Experimental results demonstrate an mAP@50-95 of 64.52%, outperforming previous state-of-the-art approaches and confirming the method's effectiveness and scalability for real-world law enforcement and surveillance applications.

Muñoz *et al*. [22] introduced a novel two-stage concealed handgun detection method leveraging thermal imaging combined with deep learning. The approach first detects handguns at the frame level and then verifies their spatial association with detected persons to reduce false alarms. A lightweight algorithm optimized for low-end embedded devices enables deployment on wearable platforms such as chest-worn Android smartphones with miniature thermal cameras. The authors also developed a dedicated thermal dataset simulating various concealment scenarios, including clothing types and distances. Experimental results show promising detection performance with a balance between accuracy and computational efficiency, achieving a mean Average Precision (mAP) of approximately 64.5% on the test sets. This work demonstrates the feasibility of real-time, hands-free concealed weapon detection suitable for security applications in airports, public events, and law enforcement operations.

Torregrosa-Domínguez *et al*. [23] aimed at enhancing on-the-fly weapon detection in industrial environments, with special emphasis on small weapons, for which hiding is still tricky. The proposed work uses the current state-of-the-art object detection models, i.e., YOLOv5, YOLOv7, YOLOv8, and proposes Scale Match, which aims to improve the detection quality for weapons with one of the smallest aspect ratios. Authors created Disarm-Dataset as a collection of datasets, including new and previously obtained images, to cover complex scenes where weapons are not easily identified from simpler scenes with identifiable weapons.

Despite the numerous contributions to concealed weapon detection using thermal imaging and deep learning, most prior works, such as Alavi *et al*. [18], Veranyurt *et al*. [21], and Khor *et al*. [2], focused solely on detection without integrating object tracking mechanisms. While Muñoz *et al*. [17, 22] introduced a two-stage detection and association method, they focused on mobile deployment rather than scalable deep learning tracking architectures. Moreover, most existing literature relies on manually annotated datasets [18, 22], which can be resource-intensive. In contrast, our approach incorporates automatic labeling through GroundedSAM, enabling scalable dataset preparation. Furthermore, our system uniquely integrates YOLOv11n for lightweight yet accurate detection and leverages Graph Neural Networks (GNN) for robust multi-frame tracking, filling the gap in real-time, end-to-end concealed weapon detection and tracking in thermal videos.

III. Methodology

The proposed method consists of a multi-stage pipeline for detecting and tracking concealed weapons in thermal video streams. As illustrated in Fig. 1, the system begins with an infrared camera's thermal data acquisition, followed by frame extraction and automatic annotation using GroundedSAM. Preprocessed images are then used to train the YOLOv11n detection model. Detected objects are subsequently tracked across frames using a Graph Neural Network (GNN), which assigns consistent identities based on spatial and temporal features.

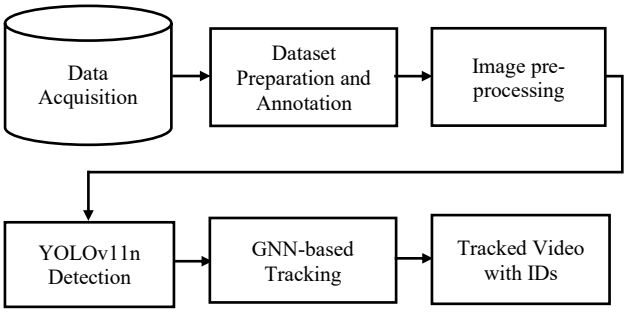


Fig. 1. Stages of the proposed method.

A. Data Acquisition (Thermal Imaging Equipment)

This study's thermal video data was exclusively captured using the Hikvision DS-2TD2617B-6/PA bi-spectrum network camera, as shown in Fig. 2. Although the device supports both thermal and optical imaging, only the thermal imaging capability was utilized to ensure consistency with the research objectives.

This camera provides thermal data for reliable object detection and analysis in low-visibility or concealed environments. The thermal sensor operates in the Long-Wave Infrared (LWIR) range with a resolution of 160 × 120 pixels and a Noise Equivalent Temperature Difference (NETD) of less than 40 mK, enabling the detection of subtle temperature variations across objects and surfaces.

The thermal lens has a focal length of 6 mm, optimized for mid-range detection. The camera offers connectivity with PC and NVR systems through USB and LAN interfaces and is manageable using Hikvision software tools such as iVMS-4200 and HikCentral. The camera was tripod-mounted for stability during experiments and simulations, as depicted in Fig. 2.

Fig. 2. Thermal camera mounted on tripod for stable video capture.

This setup enabled the collection of a reliable thermal dataset under controlled indoor and outdoor conditions, featuring human subjects carrying concealed objects and subjects without concealed items intended for subsequent detection analysis.

Fig. 3 illustrates the tools used to collect the recorded data, representing a mock-up of weapons concealed during the recording process. These tools include a field knife with a leather sheath and an automatic pistol. These models were displayed against a wooden background to highlight their physical characteristics and design details,

emphasizing that they were concealed under clothing during the data collection phase.

Fig. 3. Knife with a leather sheath and an automatic pistol.

B. Dataset Preparation and Annotation

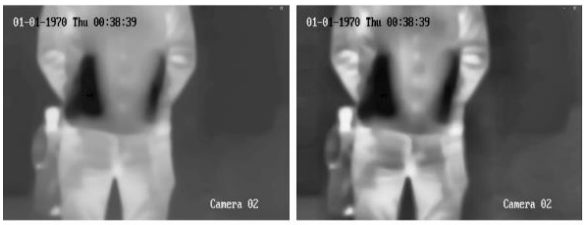
Thermal video streams were converted into a dataset of over 6,200 frames, extracted at a rate of one every five frames using the Supervision library. Automatic annotation was performed using GroundedSAM, guided by a caption ontology that defined "Gun" and "Knife" as target categories.

This process significantly reduced the need for manual labeling while generating accurate bounding boxes. The annotations were structured in YOLOv11n-compatible format and used for training.

C. Image Preprocessing

Each extracted frame underwent a preprocessing stage to enhance the visibility of thermal features and improve detection accuracy. The following techniques were applied:

- Contrast enhancement using CLAHE (Contrast Limited Adaptive Histogram Equalization) [24, 25] adaptively boosts local contrast and enhances thermal edges while minimizing noise amplification. This was especially useful for highlighting subtle temperature variations and object contours in thermal frames, as shown in Fig. 4 [26, 27].
- Background suppression using intensity thresholding techniques to remove low-temperature regions irrelevant to the detection task [28, 29].
- Normalization of thermal intensities, ensuring consistent pixel value distributions across varying lighting conditions and thermal gradients.



(a) (b)

Fig. 4. (a) Raw thermal image. (b) Enhanced using CLAHE to improve contrast and visibility.

D. Automatic Annotation Using GroundedSAM

GroundedSAM stands as an advanced framework in semantic segmentation, specifically designed to tackle the challenges posed by open-set object detection and adaptable segmentation across a wide range of visual domains. This innovative system combines two leading-edge components: Grounding DINO, an open-vocabulary object detector capable of zero-shot detection, and the segment anything model (SAM), a flexible segmentation architecture that generalizes effectively across various image types and tasks [30].

For the automatic annotation of guns and knives in thermal video, GroundedSAM works as follows:

1) Textual prompt input

- The user is asked to input text prompts: "gun" and "knife".

2) Frame-by-frame detection with grounding DINO

- Each thermal frame is feed into Grounding DINO.
- It leverages the prompts to do zero-shot detection, producing bounding boxes around guns and knives, even though those kinds of items weren't in its training data.

3) Pixel-level segmentation with SAM

- SAM takes the bounding boxes or textual guidance and performs fine-grained segmentation on each detected object.
- The result of this feature is a mask that segments the exact shape of each gun or knife in the image.

4) Automated annotation output

The output of each frame also includes the object label ("gun", "knife"), the bounding box, and the segmentation mask.

E. Model Training

The YOLOv11n model, a lightweight, real-time object detection architecture, was trained on the prepared thermal dataset. The training dataset comprised approximately 6,200 images extracted from thermal video streams under varied conditions to ensure robust generalization. The model was trained under the following settings: Image resolution: 640×640 pixels and number of epochs: 50. Training progress was closely monitored using Ultralytics' built-in tools, which provided comprehensive visualizations of performance metrics, including classification loss, localization loss, and objectivity loss.

The model demonstrated reliable performance under diverse thermal conditions in detection, as shown in Fig. 5. It consistently detected concealed weapons, although detection accuracy declined slightly when objects were hidden beneath thick fabrics. This was attributed to emissive differences between materials (e.g., cotton ≈ 0.67 vs. polyester ≈ 0.80). Despite this, the model maintained robust performance, particularly for firearms.

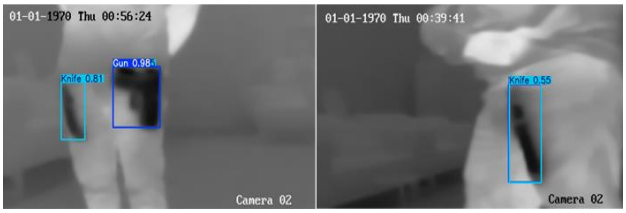


Fig. 5. The process of detecting the gun and knife.

F. Tracking Hidden Weapons

The next step in the weapons detection stage, which involves tracking hidden weapons across video frames, was implemented by integrating a Graph Neural Network (GNN) framework using PyTorch Geometric. Each

detection obtained by the YOLOv11n model was treated as a node within a constructed graph, where edges were established between temporally adjacent detections of the same object class (either "Gun" or "Knife").

The GNN was trained to predict Track IDs by evaluating spatial proximity and appearance similarities across detections, effectively associating objects over time and forming coherent trajectories. A tracked video (tracked_output.mp4) was generated after completing the tracking process. In this video, bounding boxes and corresponding assigned track IDs were visualized, illustrating the successful association of objects across frames under challenging thermal imaging conditions, where conventional appearance-based trackers often fail due to limited visual cues, as shown in Fig. 6.

Fig. 6. Tracking gun and knife across frames using GNN, with unique IDs and bounding boxes.

The tracking pipeline can be summarized as follows:

- *Detection Phase:* Object detection was performed on each video frame using the YOLOv11n model at an input resolution of 640 × 640 pixels. Some frames did not produce detections, primarily due to concealment or environmental factors.
- *Tracking Phase:* Detections were converted into a graph structure modeling object relations across frames. The GNN assigned consistent Track IDs based on learned spatial and appearance features. The tracked video was outputted and saved as tracked_output.mp4.
- *Postprocessing and Analysis:* Tracking results were analyzed to extract performance metrics evaluating tracking consistency and reliability. This approach allowed the system to maintain stable object identities even under partial occlusions or variable thermal conditions.
- *Object tracking performance:* The performance of object tracking was quantified using the following metrics. Results showed good gun up-to-down consistency, no false ID switches, and an average consistency score of 0.88. These results also demonstrate the system's capability to preserve consistent object identities over multiple frames, even while dealing with partially occluded and/or thermally different provided resized objects.

IV. Results Analysis

A. Analysis of Auto-Detection and Annotation Results

The automatic detection and annotation pipeline quality was quantitatively evaluated using several criteria. Results validate the ability of the system to detect concealed weapons in thermal imagery. We summarize the results for our detection accuracy in Table I, showing that the proposed approach achieves high precision and recall for the defined classes of weapons.

TABLE I: Weapon Detection Accuracy Metrics

Metric	Value
Precision	≈ 81%
Recall	≈ 82%
mAP@50	≈ 87%
mAP@50–95	≈ 85%

The precision-recall curves and loss plots over training epochs are visualized in Fig. 7, yielding intuitive knowledge about the convergence pattern of the model and the class-specific prediction accuracy.

The curves showed a continuous increase in precision and recall with the advancement of training, which means that the model learns well. With it, reduced training and validation loss could mean that the model was well learned and the overfitting was minimal.

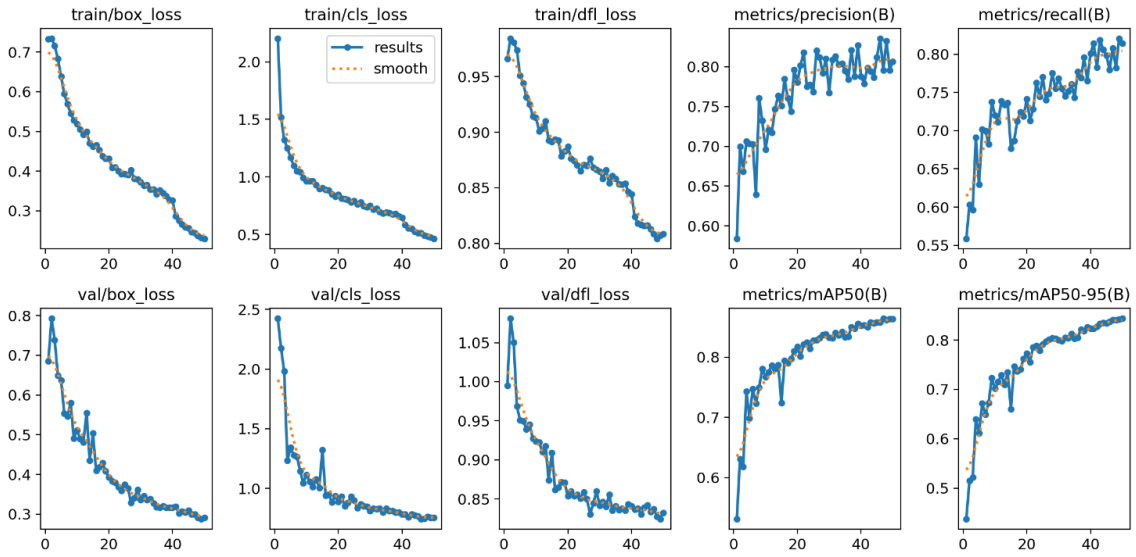


Fig .7. Precision-recall curves and loss graphs.

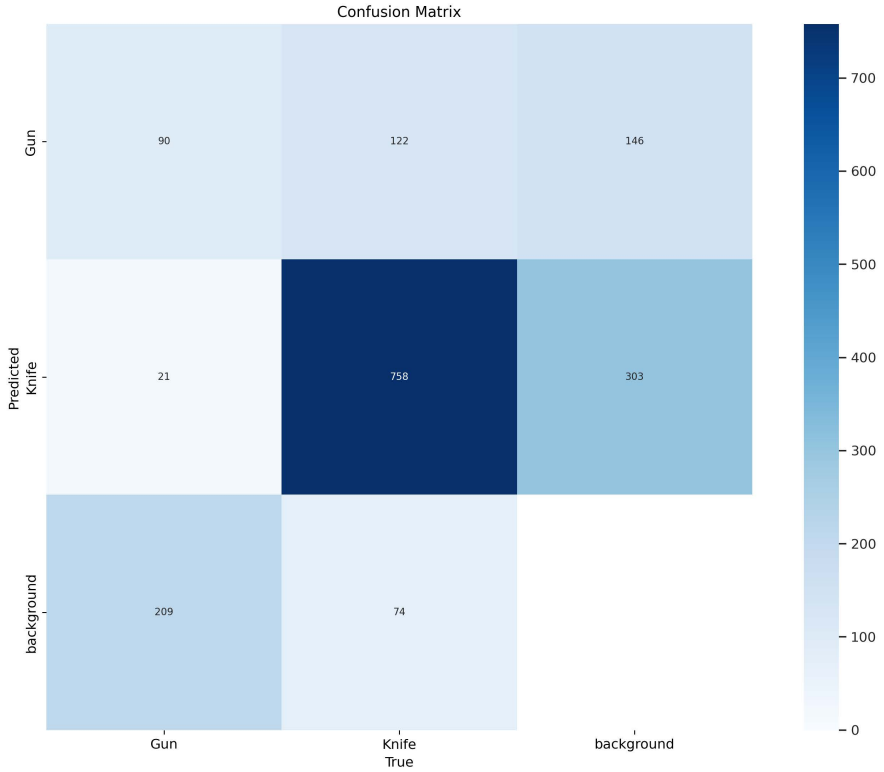


Fig. 8. Confusion matrix.

One way to evaluate the accuracy of detection and labeling is by using a confusion matrix, as shown in Fig. 8. This matrix indicates that the model can correctly classify objects labeled as "gun" and "knife," regardless of whether they are in the background. The confusion matrix highlights that knives were detected more reliably than guns, suggesting areas for targeted improvements in firearm detection accuracy by augmenting training samples or enhancing feature extraction techniques for guns under thermal imaging conditions.

Additionally, the overall trends observed in the precision-recall curves and loss graphs (Fig. 5) support the findings from the confusion matrix, showing steady improvements in class-specific accuracy throughout training.

B. Analysis Results: Hidden Weapons Tracking Performance

The tracking quality was quantitatively evaluated using several metrics, as summarized in Table II (Tracking Quality Summary). The results demonstrate high consistency in tracking guns, with zero ID switches and a high consistency score of 0.88, indicating robust temporal association.

TABLE II: TRACKING QUALITY SUMMARY

Metric	Gun	Knife
Total Detections	2179.0	3783.0
Unique Track IDs	1.0	2.0
Average Frames per Track	1927.0	1246.5
ID Switches	0.0	1.0
Track Fragmentation	81	60
Maximum Track Length (frames)	1927.0	2491.0
Consistency Score	0.88	0.66

In contrast, knife tracking exhibited slightly greater fragmentation and ID switching, with a lower consistency score of 0.66. This behavior is expected, given Knife due to its small size and weak thermal signature, leading to ID switches. Future work may improve this using multi-modal inputs or specialized tracking models. The system's processing efficiency was also evaluated to verify its suitability for real-time applications. The average processing times per frame were: Preprocessing: 2.3 milliseconds, Inference: 9.0 milliseconds, and Postprocessing: 0.7 milliseconds.

These latency results validate the real-time capability of the system and its potential suitability for use in real-time monitoring applications where fast detection and tracking are required.

C. Tracking Quality Evaluation

To evaluate the performance of the object tracking model, we employed several standard metrics commonly used in multi-object tracking (MOT) tasks. These metrics include Average Frames per Track, ID Switches, Fragmentation, and Consistency Score and are defined as follows:

Total detections (TD): Total Detections represent the cumulative number of times all objects are detected across all frames. It indicates how frequently the tracker identifies any object, regardless of identity. Higher values suggest increased object visibility and detection activity. This is a metric that can be calculated as Eq. (1):

$$TD = \sum_{i=1}^{N} d_i \quad (1)$$

where N is the total number of frames, and d_i is the number of detected objects in the frame i.

Unique track IDs (UT): This metric counts the number of distinct object IDs assigned during tracking. It reflects how well the tracker distinguishes between different objects. A higher count typically indicates greater object diversity or potential over-fragmentation. This metric can get it value as Eq. (2):

$$UT = \left|\{ID_j\}\right| \tag{2}$$

where $\{ID_j\}$ is a set of unique track IDs assigned during the tracking period and $|\{ID_j\}|$ is the a size of the set.

Average Frames per Track (AFT): This metric measures the average number of frames in which each object is successfully tracked, and this metric can be calculated as Eq. (3):

$$AFT = \frac{1}{UT}\sum_{i=1}^{UT} F_i \tag{3}$$

where UT is the total number of unique track IDs and F_i is the Number of frames in which the object I was tracked.

ID Switches (IDS): An ID switch occurs when the identity assigned to an object changes between frames, and this metric can be calculated as Eq. (4):

$$IDS = \sum_{i=1}^{UT} IDS_i \tag{4}$$

where IDS_i is the number of ID switches for track i.

Fragmentation (Frg): This indicates how often the tracking of an object is interrupted, and this metric can be calculated as Eq. (5).

$$Frg = \sum_{i=1}^{UT} (S_i - 1) \tag{5}$$

where S_i is the number of continuous fragments for the track i

Maximum Track Length (MTL): This metric represents the maximum frame duration for which any single object is continuously tracked. It highlights the tracker's capacity for long-term persistence. Higher values indicate robust tracking performance for at least one object. The metric can calculate as Eq. (6).

$$MTL = \max_{i=1}^{UT} F_i \tag{6}$$

where MTL is the maximum track length, UT is the total number of track IDs and F_i is the number of continuous frames for track i.

Consistency score: Unlike classical metrics that are computed individually, the Consistency Score enables a holistic view of tracking reliability. For example, good temporal coherence is characterized by a high average frame rate per track and zero ID switches, with low fragmentation. Compared to separate F scores, this score eases object-type comparison (guns vs. knife).

We have added a more detailed rationale and mathematical definition of the metric in the new "Tracking Quality Evaluation" section as (7), clearly interpreting its value in the comparison expressed in Table II.

$$Score = \frac{1}{3}\left(\frac{AFT}{MTL} + \frac{1}{1+IDs/TD} + \frac{1}{1+Frg/UT}\right) \tag{7}$$

where AFT is average frames per track, MTL is maximum track length, IDS is total ID switches, T is total detections, Frg is Total fragmentation count, and UT is the Number of unique track IDs.

These metrics provide a comprehensive view of the tracking quality, enabling the evaluation of the robustness and stability of the proposed system.

This analysis enables a highly systematic examination of the robustness and operational stability of the tracking pipeline for the "Gun" and "Knife" categories, as presented in Table II. The system consistently yielded high rates when tracking guns, zero ID switches, and a consistency even scoring above 0.88. For the knife, the tracking stability was slightly lower, with one ID switch and a consistency score of 0.66. However, this is still a reliable performance, given the complex motion and appearance variations of thermal video sequences.

Notably, the longest track we recorded for the knife (2,256 frames) highlights the model's ability to maintain long-term tracking across consecutive frames, even in adverse circumstances. Together, these metrics describe the effectiveness of the detection and tracking integration, confirming that the proposed pipeline is suitable for real-world thermal surveillance needs.

V. Conclusion

In this work, we proposed an integrated pipeline to detect and track concealed weapons in thermal video. Leveraging the YOLOv11n model for object detection and a Graph Neural Network (GNN) for robust object tracking, the proposed pipeline demonstrated promising results across various performance metrics. The detection phase achieved high levels of accuracy, with a precision 81% and recall of approximately 82% each, a mAP@50 of 87%, and a mAP@50-95 of 85%, as illustrated through the precision-recall curves and loss graphs Fig. 5. The confusion matrix analysis Fig. 6. revealed a higher classification accuracy for knives compared to guns, highlighting potential areas for targeted enhancement, particularly in firearm detection under thermal imaging conditions. Tracking performance was assessed through a detailed tracking quality summary (Fig. 8), demonstrating the system's ability to maintain object identities across video frames. The tracking module achieved zero ID switches and a consistency score of 0.88 for guns.

In contrast, knife tracking exhibited slightly higher fragmentation and identity switching. This behavior is expected, given That Knife Detection Is Challenging due to its small size and weak thermal signature. Furthermore, system performance metrics (Fig. 7) confirmed that the processing pipeline operates in real-time, with an average of 2.3 ms for preprocessing, 9.0 ms for inference, and 0.7 ms for postprocessing per frame. This real-time capability makes the system highly suitable for live surveillance applications requiring immediate threat detection and continuous object monitoring. Despite the system's robustness, certain limitations were observed, particularly in scenarios involving thick fabric concealment. Differences in material emissivity, such as those between cotton (~0.67) and polyester (~0.80), affected the thermal signature of concealed objects, resulting in minor reductions in detection accuracy.

VI. Future Work

To further enhance the proposed system's performance and expand its applicability, the following directions are suggested:

- *Dataset augmentation:* Enrich the training dataset with a broader range of weapon types, concealment materials, and environmental conditions.
- *Advanced feature extraction:* Integrate multi-modal data sources, such as fusing visible and thermal imagery, to improve detection under heavy occlusion.
- *Enhanced tracking algorithms:* Explore using more sophisticated tracking models like DeepSORT or ByteTrack to reduce ID switching and track fragmentation.
- *Model optimization:* Implement model compression techniques (e.g., pruning, quantization) to reduce computational requirements further while maintaining accuracy.
- *Deployment at scale:* Validate the system's performance on larger datasets and in real-world surveillance settings to ensure scalability and operational reliability.

Integrating deep learning-based detection with GNN-based tracking offers a robust framework for concealed weapon identification and monitoring, contributing significantly to advancements in thermal video surveillance technologies.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Hassanein Yarob S. Albakaa, the principal author who wrote all the articles, collected all the necessary information and presented it for easy understanding. Ali Abdulkarem Habib Alrammahi audited and reported on relevant research articles and guided the scope and focus of the review. All authors had approved the final version.

Acknowledgment

The primary researcher would like to thank the supervisor who followed up on writing and revising the research. In addition, we would like to thank everyone who contributed to providing the information needed to complete this research.

References

[1] N. R. Council, *Existing and Potential Stand-Off Explosives Detection Techniques*, Washington, DC, USA: National Academies Press, 2004.

[2] W. Khor, Y. K. Chen, M. Roberts, and F. Ciampa, "Infrared thermography as a non-invasive scanner for concealed weapon detection," in *Proc. Defence & Security Doctoral Symposium*, Cranfield Univ., UK, 2024. doi:10.17862/cranfield.rd.25028030.v2

[3] Y. Cheng, Y. Wang, Y. Niu, and Z. Zhao, "Concealed object enhancement using multi-polarization information for passive millimeter and terahertz wave security screening," *Opt. Express*, vol. 28, no. 5, pp. 6350–6366, 2020.

[4] M. Kastek, M. Kowalski, H. Polakowski, P. Lagueux, and M.-A. Gagnon, "Passive signatures of concealed objects recorded by multispectral and hyperspectral systems in visible, infrared and terahertz range," in *Proc. SPIE Active and Passive Signatures V*, 2014, vol. 9082, pp. 70–77. doi: https://doi.org/10.1117/12.2049803

[5] H. D. Barth, M. E. Launey, A. A. MacDowell *et al.*, "On the effect of X-ray irradiation on the deformation and fracture behavior of human cortical bone," *Bone*, vol. 46, no. 6, pp. 1475–1485, 2010.

[6] K. Faraj and S. J. Mohammed, "Effects of chronic exposure of X-ray on hematological parameters in human blood," *Comp. Clin. Path.*, vol. 27, no. 1, pp. 31–36, 2018.

[7] X. Liang, J. Y. Zhang, I. K. Cheng, and J. Y. Li, "Effect of high energy X-ray irradiation on the nano-mechanical properties of human enamel and dentine," *Brazilian Oral Research*, vol. 30, no. 1, 2016. doi: 10.1590/1807-3107BOR-2016.vol30.0009

[8] N. Kukutsu and Y. Kado, "Overview of millimeter and terahertz wave application research," *NTT Tech. Rev.*, vol. 7, no. 3, pp. 5–10, 2009.

[9] M. Minukas, "Developing an operational and tactical methodology for incorporating existing technologies to produce the highest probability of detecting an individual wearing an IED," M.S. thesis, Naval Postgraduate School, Monterey, CA, USA, 2010.

[10] M. Kowalski, A. Grudzień, N. Palka, and M. Szustakowski, "Face recognition in the thermal infrared domain," in *Proc. SPIE Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies*, vol. 10441, pp. 80–87, 2017. doi: https://doi.org/10.1117/12.2277534

[11] M. R. Dickson, "Handheld infrared camera use for suicide bomb detection: Feasibility of use for thermal model comparison," M.S. thesis, Kansas State Univ., Manhattan, KS, USA, 2008.

[12] Z. Xue, R. S. Blum, and Y. Li, "Fusion of visual and IR images for concealed weapon detection," in *Proc. 5th Int. Conf. Inf. Fusion*, vol. 2, 2002, pp. 1198–1205.

[13] Z. Wang, L. Wan, N. Xiong *et al*., "Variational level set and fuzzy clustering for enhanced thermal image segmentation and damage assessment," *NDT & E International*, vol. 118, 102396, Mar. 2021. doi: 10.1016/j.ndteint.2020.102396

[14] D. Wang, Z. Wang, J. Zhu, and F. Ciampa, "Enhanced pre-processing of thermal data in long pulse thermography using the Levenberg–Marquardt algorithm," *Infrared Physics & Technology*, vol. 99, pp. 158–166, Jun. 2019.

[15] B. Chen, W. Wang, and Q. Qin, "Robust multi-stage approach for the detection of moving target from infrared imagery," *Opt. Eng.*, vol. 51, no. 6, 067006, 2012.

[16] Z. Wang, G. Y. Tian, M. Meo, and F. Ciampa, "Image processing based quantitative damage evaluation in composites with long pulse thermography," *NDT & E International*, vol. 99, pp. 93–104, 2018. doi: 10.1016/j.ndteint.2018.07.004

[17] J. D. Muñoz, J. Ruiz-Santaquiteria, O. Deniz, and G. J. Bueno, "Concealed weapon detection using thermal cameras," *J. Imaging*, vol. 11, no. 3, p. 72, 2025.

[18] T. Santos, H. Oliveira, and A. J. C. S. R. Cunha, "Systematic review on weapon detection in surveillance footage through deep learning," *Computer Science Review*, vol. 51, 100612, Feb. 2024.

[19] S. Gosain, A. Sonare, and S. Wakodkar, "Concealed weapon detection using image processing and machine learning," *Int. J. Remote Imaging Appl. Sci.*, vol. 9, no. 12, pp. 1374–1384, 2021.

[20] W. A. Ahmar, I. Bekkouch, H. Khoukhi, and M. Daoui, "Multiple object detection and tracking in the thermal spectrum," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, USA, 2022, pp. 277–285.

[21] O. Veranyurt and C. O. Sakar, "Concealed pistol detection from thermal images with deep neural networks," *Multimed. Tools Appl.*, vol. 82, no. 28, pp. 44259–44275, 2023.

[22] S. Nath and C. Mala, "Thermal image processing-based intelligent technique for object detection," *Image Vis. Process.*, vol. 16, no. 6, pp. 1631–1639, 2022.

[23] Á. Torregrosa-Domínguez, J. A. Álvarez-García, J. L. Salazar-González, and L. M. Soria-Morillo, "Effective strategies for enhancing real-time weapons detection in industry," *Appl. Sci.*, vol. 14, no. 18, p. 8198, 2024.

[24] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed., San Diego: Academic Press, 1994, pp. 474–485.

[25] A. M. Jaber and F. A. O. Sari, "Enhancing of features for road crack image using EEcGANs," *Int. J. Electr. Electron. Eng. Telecommun.*, vol. 14, no. 2, pp. 108–114, Mar. 2025.

[26] C. K. Teo, "Digital enhancement of night vision and thermal images," M.S. thesis, Naval Postgraduate School, Monterey, CA, USA, 2003.
[27] L. K. Abood, "Contrast enhancement of infrared images using Adaptive Histogram Equalization (AHE) with Contrast Limited Adaptive Histogram Equalization (CLAHE)," *Int. J. Phys. Appl.*, vol. 16, no. 37, pp. 127–135, 2018.
[28] H. Li, X. Zhang, Y. Wu, and Q. Liu, "Thermal infrared-image-enhancement algorithm based on multi-scale guided filtering," *Front. Comput. Sci.*, vol. 7, no. 6, p. 192, 2024.
[29] A. A. H. Alrammahi, F. A. O. Sari, H. A. H. Shamsuldeen, and C. Science, "Analysis of the development of fruit trees diseases using modified analytical model of fuzzy c-means method," *Iraqi J. Environ. Sci.*, vol. 29, no. 1, pp. 358–364, 2023.
[30] J. E. Gallagher, A. Gogia, and E. J. Oughton, "A Multispectral Automated Transfer Technique (MATT) for machine-driven image labeling utilizing the Segment Anything Model (SAM)," arXiv preprint arXiv:2402.11413, Feb. 2024. https://arxiv.org/abs/2402.11413.

Hassanein Yarob Saeed Albakaa earned his bachelor's degree in software engineering from Imam Al-Sadiq University in Iraq in 2015. He is currently pursuing a master's degree in computer science at the University of Kufa, located in Najaf, Iraq. In addition to his academic endeavors, he serves as the cybersecurity manager at Najaf International airport. His research interests encompass the fields of computer vision and cybersecurity.

Ali A. H. Alrammahi received master's degree in information technology from Dr Babasaheb Ambedkar Marathwada University in Aurangabad, India, and a Ph.D. degree in data mining from Tambov State Technical University, Russia, in 2022. Currently, he works at the University of Kufa in Najaf, Iraq. His research interests include data analysis and data science.