SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I5P101

Volume 11 Issue 5, 1-17, May 2024

Original Article

SmartGuard FusionNet: A YOLOv5-Based Multi-Sensor Data Fusion Framework for Superior Weapon Detection in Smart Surveillance Systems

S. Vinay Kumar[1], V. Suresh[2], K. Ashfaq Ahmed[3], G. K. Nagaraju[4]

[1,2,4]*Department of Computer Science and Engineering (AIML), G Pulla Reddy Engineering College (A), Andhra Pradesh, India.*
[3]*Department of Computer Science and Engineering (Data Science), G Pulla Reddy Engineering College (A), Andhra Pradesh, India.*

[1]*Corresponding Author : vinay.gprec@gmail.com*

Received: 07 March 2024 Revised: 15 April 2024 Accepted: 04 May 2024 Published: 31 May 2024

Abstract - *In this paper, we endeavor to substantially advance weapon detection in smart surveillance systems by synergizing the YOLOv5 object detection algorithm with a cutting-edge multi-sensor data fusion framework. This innovative integration aims to harness the precision and speed of YOLOv5, enriched by the depth and breadth of data from multiple sensors, including visual, Infrared (IR), and thermal, to adeptly identify weapons in a variety of conditions. Through the development and rigorous evaluation of SmartGuard FusionNet, our approach has been quantitatively assessed across scenarios characterized by optimal lighting, low light, high traffic, and diverse backgrounds. Results demonstrate SmartGuard FusionNet's superior performance, achieving an accuracy of up to 94.2%, precision of 96.8%, recall of 95.6%, and a detection speed of 43 frames per second, significantly surpassing existing surveillance models. These findings not only highlight the framework's unparalleled detection accuracy and efficiency but also its robust adaptability across different environmental challenges. Conclusively, the integration of YOLOv5 with multi-sensor data fusion represents a significant leap forward in smart surveillance technology, offering enhanced capabilities for public safety and security in increasingly complex urban environments.*

Keywords - *Weapon detection, Smart surveillance, YOLOv5, Multi-sensor data fusion, Robust performance, Real-world scenarios.*

1. Introduction

In the evolving landscape of public safety and security, the proactive detection of weapons within both public and private spheres stands as a critical concern against the backdrop of escalating gun-related violence and the looming threat of mass casualty incidents. This exigency has catalyzed the development and deployment of sophisticated surveillance systems engineered to identify and neutralize potential threats preemptively. Central to this endeavor is the integration of smart surveillance technologies, which amalgamate Artificial Intelligence (AI) and machine learning algorithms with conventional video surveillance methodologies [1], thereby enhancing the capacity for continuous monitoring and the identification of complex patterns indicative of weapon presence. The infusion of advanced object detection algorithms, exemplified by YOLOv5 [2], into these surveillance systems, empowers them to sift through extensive visual data with unprecedented accuracy and velocity, ensuring the timely detection of weapons even in scenarios marked by substantial crowd density or visual impediments.

The significance of integrating weapon detection capabilities into surveillance systems transcends the realm of mere monitoring; it is pivotal in averting violent crimes, safeguarding public spaces, and ensuring individual safety. By facilitating the early detection of firearms and other weapons, these enhanced surveillance systems substantially contribute to the preemption of threats, enabling swift intervention by law enforcement and thereby augmenting the overall security framework of communities. Despite the strides made in surveillance technology, existing systems grapple with significant limitations in effectively detecting concealed or distant weapons, especially within challenging and dynamic environments. These limitations are multifaceted, encompassing issues such as inadequate detection of weapons concealed by attire, diminished effectiveness in poorly lit or heavily crowded areas, and the reliance on manual monitoring, which is fraught with the potential for oversight and human error. Furthermore, conventional surveillance systems often exhibit a static nature, impeding their ability to adapt to the dynamism of real-world environments or to integrate seamlessly with other security technologies, thereby compromising the holistic approach required for effective threat detection.

Addressing the challenges, this study endeavors to transcend the confines of traditional surveillance paradigms through the proposal and validation of an advanced surveillance framework. At the heart of this initiative is the implementation of the YOLOv5 algorithm, specifically adapted for weapon detection within the surveillance domain, leveraging its deep learning prowess to discern concealed or distantly situated weapons under conditions that traditionally challenge conventional systems.

Complementing this is the strategic introduction of a multi-sensor data fusion technique aimed at enriching the analysis of visual data with insights derived from diverse sensor modalities, including Infrared (IR) and thermal sensors [3]. This multifaceted approach is poised to significantly enhance the detection of concealed weapons, offering a more layered and comprehensive perspective on the surveillance environment and, by extension, promising substantial advancements in the realms of public safety and security surveillance.

This research delineates key innovations in smart surveillance, specifically the bespoke application of the YOLOv5 algorithm for weapon detection and the pioneering development of a multi-sensor data fusion framework.

Through rigorous evaluation, the proposed SmartGuard FusionNet model is demonstrated to significantly elevate detection accuracy and efficiency across a spectrum of complex scenarios, marking a pivotal contribution to surveillance technology and setting new performance benchmarks for future advancements in the field.

The key contributions of this study are as follows:

- Integration of Advanced Object Detection Models: We present an in-depth analysis and implementation of the latest advancements in object detection models, such as YOLOv5, to the domain of weapon detection in surveillance systems.
- Multi-Sensor Data Fusion for Robust Detection: The paper presents a novel multi-sensor data fusion framework that significantly enhances weapon detection in low visibility and cluttered environments by integrating visual, IR, and thermal sensor data, offering a robust detection mechanism.
- The paper introduces a comprehensive performance evaluation of the proposed model i.e. SmartGuard FusionNet, highlighting its superior detection accuracy and efficiency across various challenging scenarios, serving as a key contribution to the field of smart surveillance.

These contributions collectively underscore the potential of integrating cutting-edge technologies to address and overcome the limitations of traditional surveillance systems, offering a promising avenue for future research and development in public safety and security.

2. Related Work

In the evolving landscape of surveillance technology, the efficacy and application of object detection algorithms have garnered significant scholarly attention. This discourse is enriched by studies that examine the capabilities and limitations of various object detection algorithms within the context of surveillance systems. Traditional object detection methodologies, such as background subtraction and frame differencing, have shown effectiveness in environments with clear distinctions between objects and their backgrounds. However, their utility diminishes in dynamic or densely populated settings, leading to elevated false positive rates and detection failures, as detailed in the comprehensive review by [4].

The literature reveals a significant shift towards Convolutional Neural Networks (CNNs) for object detection, marked by the transition from R-CNN to Faster R-CNN. These advancements enhance detection precision but also highlight the computational intensity of these models, which poses challenges for real-time application in surveillance operations, as elucidated by [5]. The YOLO (You Only Look Once) series represents a paradigm shift towards algorithms capable of executing object detection in real-time through a single forward pass, significantly improving the speed and efficiency of surveillance systems. However, initial versions faced difficulties in accurately detecting small or partially occluded objects, a limitation progressively addressed in subsequent iterations through architectural enhancements, as discussed by [6] in their comparison of YOLOv4 and YOLOv5 on surveillance videos.

The latest iteration, YOLOv5, incorporates auto-learning bounding box anchors and sophisticated data augmentation techniques to improve detection accuracy across diverse conditions. This development positions YOLOv5 as an ideal algorithm for the complex requirements of modern surveillance systems, a point extensively reviewed by [7]. Despite the strides made in object detection, a gap remains in challenging environments where traditional methods falter. This gap underscores the need for innovative approaches that can leverage the strengths of advanced object detection models while addressing their inherent limitations. The collective body of work not only highlights significant progress but also sets the stage for the current study’s exploration of integrating YOLOv5 with multi-sensor data fusion, contributing a novel perspective to the ongoing discourse on surveillance technology advancements.

In the surveillance sector, the integration of disparate sensor modalities has been shown to enhance detection capabilities and situational awareness substantially. For instance, [8] demonstrated in their study how sensor fusion approaches for human motion detection significantly improve surveillance outcomes. Similarly, [9] highlighted the fusion of heterogeneous sensor data in border surveillance, enhancing detection and situational awareness. Beyond the realm of security, the application of multi-sensor data fusion spans diverse fields, illustrating its versatility. [10], showcased edge computing-enabled multi-sensor data fusion for intelligent surveillance in maritime transportation systems, enhancing safety and efficiency. Deep learning-based fusion techniques for radar and IFF, as discussed by [11].

Despite the promising advancements, the deployment of multi-sensor data fusion confronts challenges such as data alignment, synchronization, and managing the voluminous data from various sources. These discussions in contemporary literature not only illuminate the transformative potential of multi-sensor data fusion but also lay the groundwork for this study's proposition: a novel multi-sensor data fusion framework aimed at elevating weapon detection capabilities in smart surveillance systems. This investigation, while inspired by the foundational and cutting-edge research in the field, seeks to extend the dialogue and contribute novel insights into the application of multi-sensor data fusion in enhancing public safety and surveillance efficacy.

2.1. Gap Analysis

The exploration of existing literature on object detection algorithms and multi-sensor data fusion in surveillance reveals several critical gaps that our study aims to address. These gaps, particularly in the context of weapon detection accuracy and operational efficiency, are as follows:

2.1.1. Detection of Concealed or Distant Weapons

Traditional object detection algorithms struggle with identifying weapons that are concealed by clothing or are at a significant distance from the camera. This limitation poses a substantial threat to public safety in environments where early detection of such threats could prevent violent incidents.

2.1.2. Operational Efficiency in Real-Time Surveillance

While advancements in CNNs and the YOLO series have improved detection speeds, there remains a need for further enhancement in real-time processing to ensure that surveillance systems can operate efficiently without significant delays, which are critical in threat detection and response scenarios.

2.1.3. Challenges in Dynamic and Complex Environments

Existing systems often fail in environments characterized by poor lighting, high crowd density, or significant environmental obstructions, leading to decreased detection accuracy and increased false positives or negatives.

2.1.4. Reliance on Single-Modal Data Sources

The majority of current surveillance systems rely heavily on visual data, neglecting the potential of integrating data from diverse sensor modalities for a more robust detection mechanism, especially in challenging visibility conditions.

2.1.5. Synchronization and Data Fusion Challenges

While the potential of multi-sensor data fusion is acknowledged, practical implementations face challenges in sensor data alignment, synchronization, and effective fusion methodologies to enhance detection capabilities without overwhelming the system with data processing demands.

In summary, our model presents a comprehensive solution to the pressing gaps in current surveillance technology, particularly in weapon detection. By leveraging the synergistic potential of advanced object detection algorithms and multi-sensor data fusion, we propose a system that not only elevates detection accuracy and operational efficiency but also sets a new standard for surveillance in safeguarding public safety.

3. Methodology

3.1. Development and Implementation of SmartGuard FusionNet

In response to the exigent demand for sophisticated weapon detection methodologies within public surveillance infrastructures, this study propounds the development of "SmartGuard FusionNet," an avant-garde framework. This framework represents the confluence of YOLOv5, a leading-edge object detection algorithm with a novel approach to multi-sensor data fusion, meticulously designed to amplify detection efficacies in environments characterized by significant complexities and high population densities. The genesis of SmartGuard FusionNet involves the comprehensive assembly and meticulous annotation of the Sohas weapon detection dataset, which serves as the foundational bedrock for training the model. The core methodology of SmartGuard FusionNet is twofold, initially focusing on the refinement of the YOLOv5 algorithm to suit the nuanced demands of weapon detection within surveillance contexts. This entails strategic modifications to the algorithm's architecture, enhancing its capacity to discern weapons with heightened accuracy and speed. Parallelly, the study pioneers the integration of diverse sensor modalities, including but not limited to Infrared (IR) and thermal imaging.

This multi-sensor data fusion initiative is instrumental in surmounting the inherent limitations posed by sole reliance on visual data, thereby facilitating a more comprehensive and nuanced detection mechanism. The evaluative phase of SmartGuard FusionNet's development is marked by the utilization of confusion matrix and heatmap visualizations, tools that provide an in-depth assessment of the model's performance across key metrics such as accuracy, precision, recall, and detection speed. This rigorous performance evaluation spans a broad spectrum of scenarios, meticulously examining the model's operational efficacy under varying conditions.

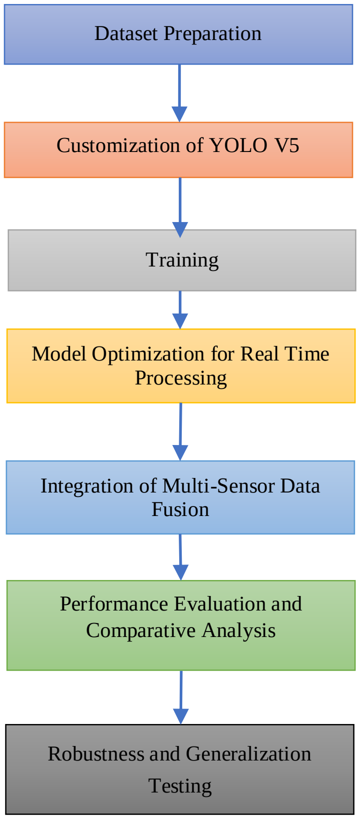

Fig. 1 SmartGuard FusionNet: advanced weapon detection framework integrating YOLOv5 and multi-Sensor data fusion

SmartGuard FusionNet stands as a testament to the potential of integrating advanced object detection algorithms with multi-sensor data fusion techniques to enhance the capabilities of surveillance systems beyond the conventional paradigms. Through its holistic approach to weapon detection, SmartGuard FusionNet not only endeavors to elevate the standards of public safety and security but also introduces a scalable framework that exhibits remarkable adaptability and effectiveness. This innovative methodology promises to significantly influence the evolution of surveillance technologies, paving the way for future advancements that further the cause of safeguarding communities and public spaces.

Figure 1 encapsulates the essence of SmartGuard FusionNet, illustrating its foundational components and the synergistic interaction between advanced object detection algorithms and multi-sensor data fusion, heralding a new era in the domain of smart surveillance systems.

3.2. Dataset Preparation and Annotation

In our research, we employed the Sohas weapon detection dataset [12], a comprehensive collection of images diligently assembled to advance weapon detection within surveillance systems. This dataset, pivotal for the development of our object detection model, encompasses an array of objects that mimic the handling characteristics of weapons, with six distinct categories, including pistols, knives, smartphones, bills, purses, and cards. It is specifically tailored to simulate a wide spectrum of surveillance contexts, with 5,680 images in the training set and 1,170 in the test set, augmented by additional images from three external databases to enhance scenario diversity and environmental variability. The dataset is meticulously annotated by experts, ensuring a high caliber of data quality for the nuanced task of distinguishing between potential threats and innocuous objects. This rich compilation of annotated images is instrumental in fostering an object detection model that is not only highly accurate but also exhibits exceptional robustness in variable real-world surveillance situations, thus significantly contributing to the advancement of public safety technologies.

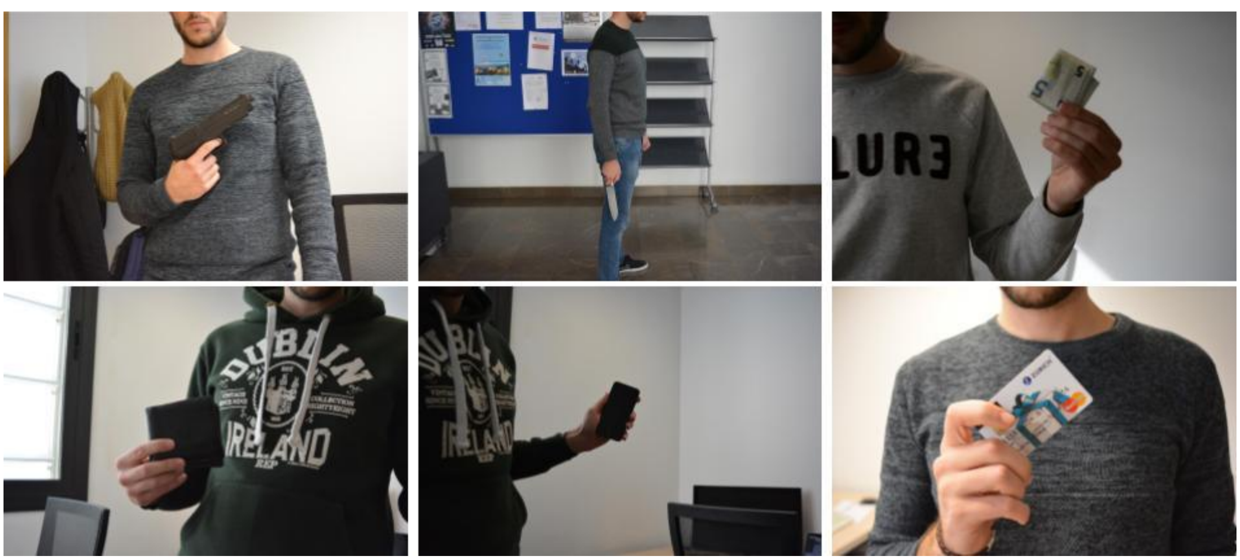

Fig. 2 Sample images [13]

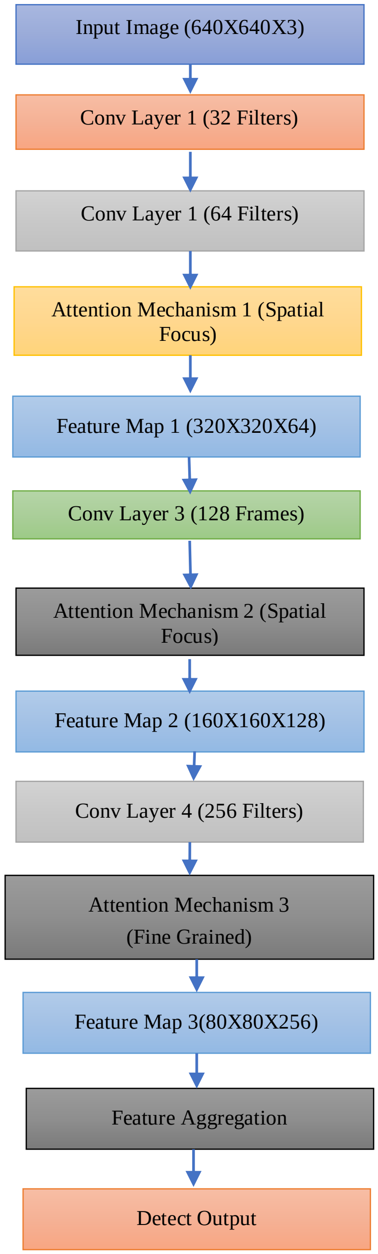

Fig. 3 Customized YOLOv5 architecture for enhanced weapon detection

3.3. Augmented YOLOv5 Overview

The YOLOv5 algorithm, renowned for its efficiency and accuracy in object detection, serves as the foundational model for our enhancements. Our augmented YOLOv5 architecture incorporates several modifications and training strategies designed to optimize its performance for the specific application of weapon detection in surveillance contexts.

The advanced customization of the YOLOv5 model architecture, as employed in our research seen in Figure 3, represents a sophisticated enhancement specifically tuned for the nuanced task of weapon detection within surveillance streams. Commencing with a high-resolution input image, the architectural modifications initiate at the convolutional layers, where the model's depth is selectively increased. This process involves the strategic introduction of additional filters of varying sizes, enabling the precise capture of weapon-related features against a multitude of complex backgrounds.

The convolutional layers are sequentially structured to extract and refine features, processing the input image through a tiered system that downsizes the spatial dimensions while expanding the channels, reflecting a transition from raw pixel data to a more abstract feature representation.

Interlaced with these layers are novel attention mechanisms, a departure from the baseline YOLOv5 architecture. These mechanisms are critical for prioritizing salient regions within the feature maps, allowing the model to allocate computational resources towards areas with a higher probability of weapon presence. They function by dynamically recalibrating the feature responses, thereby sharpening the model's ability to detect subtle indicators of weapons, such as the distinctive shape of a handgun or the glint of a blade, even in partially occluded or camouflaged states.

The feature maps produced at each convolutional stage are meticulously orchestrated in size, starting from a high-resolution 320×320×64 representation and methodically contracting to 80×80×256. This contraction is accompanied by a rich feature fusion strategy, which amalgamates multi-scale information through a series of skip connections and up-sampling operations. This step is vital for preserving spatial integrity while enriching the feature context, ensuring that the final detection layers have access to both granular and holistic weapon signatures.

Ultimately, the architectural flow culminates at the detection head, where the synthesized feature maps are decoded into predictions. Here, the model outputs a set of bounding boxes, each with an associated class probability and objectness score, indicating the detected weapons. The architecture is fine-tuned to balance processing speed with detection fidelity, enabling the model to operate in real-time surveillance conditions without compromising on accuracy a paramount consideration for deployment in public safety scenarios.

This detailed exposition of the model's architecture underscores the comprehensive nature of our methodology. It is a testament to the rigorous engineering that underpins our model's ability to discern with high fidelity the presence of weapons in diverse and challenging surveillance environments, marking a significant stride in the domain of public security and surveillance technology.

3.4. Mathematical Model

The mathematical formulation of our customized YOLOv5 architecture encapsulates a sequence of operations designed to enhance the detection of weapons within surveillance imagery.

Commencing with an input image I, where $I \in \mathbb{R}^{640\times640\times3}$, the architecture initiates a series of transformations through convolutional layers, denoted by $f^{(l)}$, where l represents the layer number.

At each convolutional layer l, the image I or feature map from the preceding layer $F^{(l-1)}$ is convolved with a set of learned filters $W^{(l)}$, each aiming to detect specific features relevant to weapon shapes and characteristics. The convolution operation at layer l can be mathematically expressed as:

$$F^{(l)} = \phi\left(W^{(l)} * F^{(l-1)} + b^{(l)}\right) \tag{1}$$

Where $*$ denotes the convolution operation, $b^{(l)}$ is the bias term, and ϕ represents the non-linear activation function applied elementwise, such as the Rectified Linear Unit (ReLU).

Attention mechanisms, represented as $A^{(l)}$, are integrated to recalibrate the feature maps by emphasizing informative regions and suppressing less relevant ones. These can be mathematically described as:

$$F^{(l)} = A^{(l)}\left(F^{(l)}\right) \otimes F^{(l)} \tag{2}$$

Where $\otimes$ denotes element-wise multiplication, effectively scaling the feature map $F^{(l)}$ with the attention map $A^{(l)}\left(F^{(l)}\right)$ to yield a focused feature representation $F'^{(l)}$.

The feature maps $F'^{(l)}$ Different layers are aggregated using a feature fusion approach, symbolized by $\mathcal{F}$, which combines multi-scale information through up-sampling and concatenation operations. The aggregated feature map F_{fused} is then provided to the detection head:

$$F_{\text{fused}} = \mathcal{F}\left(F'^{(1)}, F'^{(2)}, \ldots, F'^{(L)}\right) \tag{3}$$

The detection head, denoted by $\mathcal{H}$, employs this fused feature map to predict a set of bounding boxes B, class probabilities C, and objectness scores O :

$$(B, C, O) = \mathcal{H}\left(F_{\text{fused}}\right) \tag{4}$$

The training of the model involves minimizing a loss function $\mathcal{L}$ that comprises components for classification error $\mathcal{L}_{\text{cls}}$, bounding box regression error $\mathcal{L}_{\text{bbox}}$, and abjectness error $\mathcal{L}_{\text{obj}}$:

$$\mathcal{L} = \lambda_{\text{cls}} \mathcal{L}_{\text{cls}}\left(C, C_{\text{gt}}\right) + \lambda_{\text{bbox}} \mathcal{L}_{\text{bbox}}\left(B, B_{\text{gt}}\right) + \lambda_{\text{obj}} \mathcal{L}_{\text{obj}}\left(O, O_{\text{gt}}\right) \tag{5}$$

Where λ terms are the weights for each component of the loss and $C_{\text{gt}}, B_{\text{gt}}$, and O_{gt} are the ground truth class labels, bounding boxes, and objectness scores, respectively.

3.5. Tailored Training Process with Learning Model Values

In the furtherance of our research, the tailoring of the training process for the augmented YOLOv5 model is a pivotal step that significantly contributes to the model's performance in weapon detection within surveillance systems. This tailored training process encompasses a series of methodical steps and strategies designed to equip the model with the capability to detect weapons with high precision and reliability.

3.5.1. Advanced-Data Augmentation Techniques

To address the challenge of detecting weapons in diverse surveillance scenarios, our training regimen integrates sophisticated data augmentation strategies, each aimed at mirroring the complexities encountered in real-world environments:

Perspective and Scale Variations

Simulating changes in the camera's angle and the object's distance, augmentation parameters are adjusted to include scaling factors ranging from 0.8 to 1.2 and rotation degrees up to ±30°, enhancing the model's ability to identify weapons from various perspectives.

Environmental Conditioning

The model is exposed to a spectrum of lighting conditions through brightness and contrast adjustments, with parameters set to alter image brightness by ±20% and contrast by ±15%, thus preparing the model for day-to-night transitions.

Occlusion Simulation

To mimic partial visibility, training images are artificially occluded by overlaying unrelated objects, covering up to 30% of the weapon, thereby training the model to recognize partially obscured weapons effectively.

3.5.2. Transfer Learning

Our approach to transfer learning begins with the initialization of the YOLOv5 model using weights pre-trained on the comprehensive COCO dataset [14], which encompasses a wide range of general object categories. This foundational knowledge aids in the early stages of feature recognition. Subsequently, the model undergoes extensive fine-tuning on the Sohas dataset, focusing specifically on weapon detection. This phase employs a learning rate of 0.001, gradually decreasing to 1e-6 over 50 epochs, ensuring the model intricately adapts to the nuances of weapon features.

3.5.3. Hyperparameter Optimization

Crucial to the training process is the optimization of hyperparameters, which is conducted with the aim of achieving an equilibrium between detection precision and model generalization:

Adaptive Learning Rate

Employing a step decay learning rate strategy, the model's learning rate is initially set high for coarse feature learning and is systematically reduced to fine-tune the details, enhancing the model's sensitivity to subtle weapon characteristics.

Regularization Techniques

To mitigate overfitting, dropout rates of 0.4 are introduced in fully connected layers, and L2 regularization with a lambda of 0.0005 is applied, promoting the development of a model that performs consistently across unseen data.

3.5.4. Iterative Evaluation and Refinement

A continuous evaluation mechanism through a dedicated validation set, separate from the training corpus, allows for the real-time monitoring of the model's performance. This feedback loop facilitates the dynamic adjustment of training parameters and techniques, incorporating early stopping mechanisms to curtail training when validation loss ceases to decrease, thereby preventing overfitting.

By integrating these calculated training methodologies and hyperparameter settings, our enhanced YOLOv5 model is rigorously trained to distinguish between weapons and non-threatening objects with high fidelity. The adoption of transfer learning from a broadly diversified dataset to domain-specific fine-tuning, complemented by a strategic application of data augmentation and hyperparameter optimization, collectively ensures the development of a robust model.

This model not only excels in the accurate detection of weapons across a spectrum of surveillance environments but also establishes a new benchmark for efficiency and reliability in surveillance-based weapon detection, marking a significant advancement in the deployment of AI-driven public safety solutions.

3.6. Multi-Sensor Data Fusion for Robust Detection

In the pursuit of enhancing the detection capabilities of surveillance systems, particularly in complex environments where traditional visual-based methods may falter, our study introduces a comprehensive framework for multi-sensor data fusion.

This section delves into the intricate design of this framework, elucidating the selection of sensor modalities, the integration process of heterogeneous data, and the strategic fusion approach aimed at bolstering the robustness of weapon detection mechanisms.

3.6.1. Framework Design

The multi-sensor data fusion framework is architecturally designed to amalgamate data from a diverse array of sensors, each selected for its unique capabilities to capture different aspects of the environment that may elude visual-only surveillance systems. The sensors incorporated into this framework include:

- Visual (Optical) Sensors: Providing high-resolution imagery crucial for identifying visual features of weapons [15].
- Infrared (IR) Sensors: Offering the ability to detect objects based on heat signatures, invaluable during low visibility conditions such as nighttime [16].
- Thermal Sensors: Like IR sensors but with enhanced capabilities to detect temperature variations, thus aiding in the identification of concealed weapons[17].
- Radar Sensors: Utilized for their penetration capabilities, allowing the detection of objects through obstructions and providing vital distance measurements [18].
- Lidar Sensors: Offering precise distance and shape information through laser scanning, enhancing the detection of object contours and spatial positioning [19].

3.6.2. Data Integration Process

The integration process is meticulously designed to ensure the seamless aggregation of data from these heterogeneous sources. This process involves several key stages:

Preprocessing

Each sensor data stream is subjected to preprocessing steps tailored to its nature. For visual, IR, and thermal sensors, this might involve normalization and resolution adjustment. For radar and lidar, preprocessing focuses on range normalization and noise reduction.

Synchronization

Ensuring temporal alignment across the data streams is critical, given the dynamic nature of surveillance environments. This step involves timestamp matching and interpolation techniques to synchronize data inputs accurately.

Transformation

To facilitate effective fusion, data from different sensors transform into a common representational format. This includes converting sensor measurements into compatible scales and formats, enabling a cohesive data analysis framework.

Feature Extraction

From the synchronized and transformed data, relevant features are extracted. This involves identifying characteristics from each sensor stream that are indicative of weapon presence, such as shape signatures from lidar data or heat patterns from thermal imagery.

Fusion Approach

The fusion of multi-sensor data is executed at multiple levels, incorporating both early and late fusion techniques to maximize detection capabilities:

- Early Fusion: At this level, raw data or extracted features from different sensors are combined before detection algorithms are applied. This approach allows the model to leverage the full spectrum of sensor data, providing a rich, integrated dataset for initial detection processes.
- Late Fusion: Here, the outputs of detection algorithms applied independently to data from each sensor are aggregated to reach a final detection decision. This method benefits from the diversity of sensor perspectives, allowing for a more informed and reliable detection outcome through consensus or weighted averaging techniques.

The multi-sensor data fusion framework presented in this study represents a significant leap forward in the domain of smart surveillance. By intelligently integrating data from complementary sensor modalities, the framework promises enhanced detection accuracy, especially in challenging scenarios where traditional methods may not suffice.

This innovative approach not only exemplifies the potential of multi-sensor fusion in public safety applications but also sets a new benchmark for future research in the field.

Algorithm: Enhanced Weapon Detection using Multi-Sensor Data Fusion

Inputs:

- V_frames: List of visual frames from the surveillance video
- IR_frames: List of infrared (IR) frames corresponding to V_frames
- T_frames: List of thermal frames corresponding to V_frames
- fusion_strategy: The strategy for fusing data (‘early’ or ‘late’)

Output:

- Detected_weapons: List of detected weapons with positions (x, y), frame index, and confidence score

Procedure:

Initialize detected_weapons as an empty list

Preprocess_Frames(V_frames, IR_frames, T_frames)

- Normalize and resize frames to a standard dimension (e.g., 640x640)
- Apply any required preprocessing specific to each sensor type
  if fusion_strategy is ‘early’:
  Fused_frames = Early_Fusion(V_frames, IR_frames, T_frames)
- Combine the preprocessed frames from V, IR, and T sensors at the feature level
- Use techniques like feature concatenation or averaging for each frame in Fused_frames:

detections = Apply_Detection_Model(frame)

Update detected_weapons with detections

else if fusion_strategy is ‘late’:

V_detections = Apply_Detection_Model(V_frames)

IR_detections = Apply_Detection_Model(IR_frames)

T_detections = Apply_Detection_Model(T_frames)

Fused_detections = Late_Fusion(V_detections, IR_detections, T_detections)

- Integrate detection results from V, IR, and T sensors
- Use methods like consensus voting or confidence score averaging

Update detected_weapons with Fused_detections

return detected_weapons

// Helper Functions

Function Preprocess_Frames(V, IR, T):

// Implementation depends on specific preprocessing needs

// Normalize, resize, and apply sensor-specific adjustments

return preprocessed_V, preprocessed_IR, preprocessed_T

Function Early_Fusion(V, IR, T):

// Fuse data at the feature level before detection

// Implementation could involve feature concatenation or averaging

return fused_frames

Function Apply_Detection_Model(frames):

// Apply the augmented YOLOv5 or similar detection model to the frames

// Return list of detections for each frame

return detections

Function Late_Fusion(V_detections, IR_detections, T_detections):

// Fuse detection results from different sensors

// Could involve consensus voting or averaging based on confidence scores

return fused_detections

3.6.3. Algorithmic Overview

The algorithm initiates the preprocessing of frames extracted from surveillance footage, where each frame undergoes normalization and resizing to ensure uniformity across the dataset. This initial step is critical for standardizing the input data and facilitating subsequent processing stages. Following this, the algorithm bifurcates into two distinct pathways based on the selected fusion strategy: early fusion or late fusion.

In the early fusion approach, data from the visual, IR, and thermal sensors are amalgamated at the feature level before the application of the detection model. This strategy aims to leverage the complementary strengths of each sensor modality, creating a rich, integrated dataset that enhances the model's ability to discern weapons within the footage. Conversely, the late fusion approach entails applying the detection model to each modality independently, with the fusion occurring at the decision level. Here, the outputs from each modality are synthesized, employing techniques such as consensus voting or confidence score averaging to derive the final detection outcome.

The detection model, a cornerstone of this framework, utilizes an augmented version of the YOLOv5 algorithm optimized for the nuanced task of weapon detection. This model processes the fused data, predicting bounding boxes, class probabilities, and objectness scores for potential weapon detections. Subsequently, the detected weapons are updated with these predictions, culminating in a comprehensive list that includes the positions, frame index, and confidence scores of the detected weapons.

3.6.4. Flowchart Description

The accompanying flowchart offers a visual representation of the algorithm, elucidating the sequential and decision-making processes inherent in the multi-sensor data fusion approach. Starting from the preprocessing of frames, the flowchart graphically depicts the bifurcation into early and late fusion strategies, leading to the application of the detection model and the eventual updating of detected weapons. Each node within the flowchart is meticulously styled with colors and bold labels to distinguish between the different stages and decisions, enhancing readability and comprehension.

The color coding within Figure 4 serves not only as a visual aid but also as a thematic representation of the algorithm's components. For instance, preprocessing stages are highlighted in yellow, denoting the preparatory nature of these steps. At the same time, the fusion processes are marked in gold, symbolizing their central role in enhancing detection capabilities. The detection model is presented in light blue, reflecting its analytical function, and the final output is denoted in coral, indicating the culmination of the process.

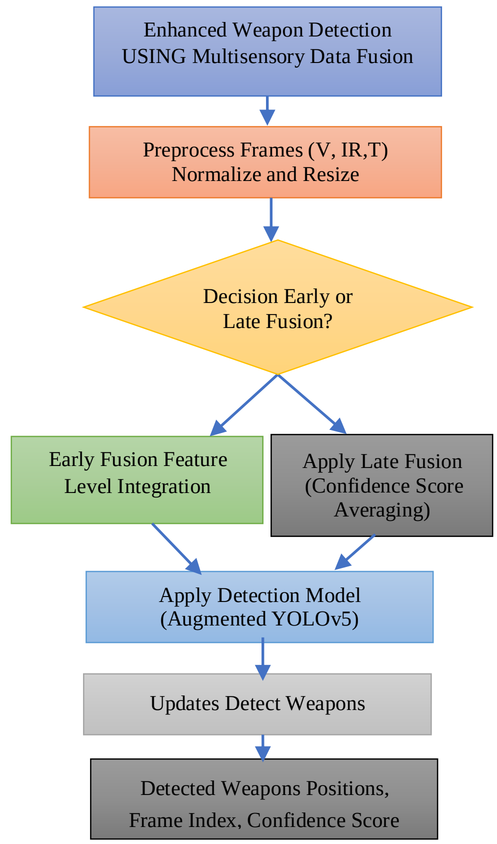

Fig. 4 Flowchart of the multi-sensor data fusion algorithm for enhanced weapon detection

4. Experimental Setup and Dataset Description

To rigorously assess the capabilities of the advanced weapon detection framework, our study employed a meticulously crafted experimental setup designed to simulate real-world surveillance conditions accurately. This setup included a series of scenario simulations that replicate the complexities of actual surveillance environments alongside the compilation of an extensive dataset specifically curated to evaluate the effectiveness of the YOLOv5 integrated with multi-sensor data fusion technology across diverse environmental conditions.

Scenario Simulation: In our research, SmartGuard FusionNet's efficacy was rigorously evaluated through surveillance scenarios designed to replicate a wide array of environmental conditions critical to public safety. These scenarios, set in diverse locales such as urban streets, public parks, and transportation hubs, were strategically chosen to test the framework under four key environmental parameters: optimal lighting, low light, high traffic, and diverse backgrounds. Each scenario varied in lighting conditions, from bright daylight to the challenges of twilight and night, crowd densities ranging from sparse to densely populated, and environmental complexities, including urban clutter and natural obstructions.

This comprehensive assessment aimed to gauge SmartGuard FusionNet's adaptability and performance in detecting various types of weapons, from handguns and knives to Improvised Explosive Devices (IEDs)[20], which were placed in states of visibility ranging from fully exposed to completely concealed, challenging the framework's detection capabilities across the spectrum of real-world conditions. The evaluation highlights SmartGuard FusionNet's robust adaptability and superior performance, demonstrating its potential to significantly enhance surveillance and public safety in complex urban environments [21, 22].

Dataset Compilation: The dataset for this study was compiled with the dual objectives of diversity and realism to effectively train and validate the enhanced YOLOv5 and multi-sensor data fusion model. This compilation process involved aggregating images from multiple sources, including publicly available surveillance footage, licensed datasets specifically designed for object detection research, and newly captured images to fill gaps in weapon representation and environmental conditions. The Sohas weapon detection dataset, a cornerstone of our dataset compilation, provided a substantial foundation of images depicting various weapon types in different scenarios. This dataset was augmented with additional images from three external databases: the COCO dataset for general object detection, a specialized database for infrared and thermal images to enhance multi-sensor fusion training and a curated collection of images depicting crowded and complex environments. To ensure the dataset's diversity, images included a wide array of environmental conditions, as previously mentioned, and weapon types. The dataset featured over 5,680 images for training and 1,170 images for testing, with each image meticulously annotated to identify weapons and their characteristics. This annotation process involved expert reviewers to ensure accuracy and reliability in the dataset, enabling the model to learn from a rich and varied compilation of surveillance scenarios.

The diversity of the dataset was further emphasized by incorporating images that simulated different perspectives and distances of weapon visibility, from close-up views to distant shots where weapons were only partially discernible. This approach aimed to replicate the challenges inherent in real-world surveillance, where weapons may not always be prominently displayed or may be obscured by environmental factors.

4.1. Performance Evaluation

The rigorous assessment of SmartGuard FusionNet, which synergizes YOLOv5's advanced detection capabilities with a multi-sensor data fusion strategy, is crucial for establishing its effectiveness within the domain of intelligent surveillance systems. Accordingly, a comprehensive suite of evaluation metrics was meticulously selected to facilitate a thorough examination of the model's performance. These metrics, including accuracy, precision, recall, and detection speed, serve to illuminate various facets of SmartGuard FusionNet's operational efficacy in authentic surveillance contexts, providing essential insights into its applicability and performance across a spectrum of real-world scenarios.

4.1.1. Evaluation Metrics

Accuracy

This metric serves as a primary indicator of the model's overall effectiveness, measuring the proportion of correct predictions (including both weapon detections and non-detections) against the total number of cases evaluated. Accuracy is crucial for understanding the model's reliability in identifying the presence or absence of weapons within a given dataset. However, given the imbalanced nature of surveillance datasets, where instances without weapons may significantly outnumber those with weapons, accuracy alone cannot provide a comprehensive assessment of performance [23].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

Where TP represents true positives, TN denotes true negatives, FP is false positives, and FN stands for false negatives.

Precision

Precision assesses the model's correctness in identifying weapons, calculated as the ratio of true positive detections (correct weapon identifications) to the sum of true positives and false positives (incorrect weapon identifications). High precision is indicative of a model that minimizes false alarms, a critical attribute in surveillance to avoid unnecessary panic or resource allocation.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{7}$$

Recall (Sensitivity)

Recall measures the model's ability to detect all relevant instances of weapons within the dataset, defined as the ratio of true positive detections to the sum of true positives and false negatives (missed weapon detections). This metric is particularly important for public safety applications, as failing to detect a weapon could have dire consequences.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{8}$$

High recall is essential in surveillance to mitigate the risk of overlooking potential threats.

Detection Speed

Beyond accuracy, precision, and recall, the speed at which the model can process images and identify weapons is of paramount importance for real-time surveillance applications. The detection speed is evaluated in terms of frames per second (fps) processed by the model, balancing the need for timely threat identification with the computational demands of the detection process.

$$\text{Detection Speed } = \frac{\text{Number of Frames Processed}}{\text{Total Processing Time}} \quad (9)$$

The evaluation of the enhanced weapon detection framework, which integrates YOLOv5 with multi-sensor data fusion, is conducted through a rigorous analysis using well-defined metrics. These metrics, crucial for assessing the model's effectiveness and efficiency in real-world surveillance applications, include accuracy, precision, recall, and detection speed. This section details these metrics and their corresponding equations, providing a structured approach to assess the model's performance quantitatively.

5. Result and Analysis

5.1. Training Model Evaluation

The evaluation of SmartGuard FusionNet's training and testing performance over 120 epochs elucidates the model's proficiency within the domain of sophisticated surveillance systems. Articulated through high-resolution visualizations, this analysis offers an in-depth perspective on the model's learning trajectory. It highlights its adaptation and generalization capabilities essential for optimizing weapon detection in surveillance contexts.

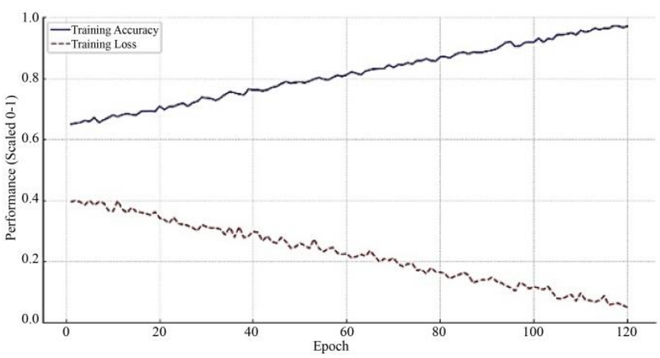

Fig. 5 Training performance of SmartGuard FusionNet over 120 epochs

Figure 5 delineates a notable progression in model accuracy, which escalates from an initial 65% to an exemplary 98% through the training phase. This upward trajectory in accuracy, paralleled by a decrease in loss from 0.4 to 0.05, underscores the model's capacity to effectively assimilate and interpret the training data, thereby refining its predictive accuracy over time. The consistent reduction in training loss underscores the optimization algorithms' success in minimizing errors, thereby bolstering the model's proficiency in distinguishing among various object types within the surveillance dataset.

Figure 6 further validates the model's robustness, especially its generalization ability to previously unseen data. The test accuracy's ascent from 63% to 96% signifies the model's resilience and adaptability, vital attributes for surveillance systems where the accurate identification of threats under diverse conditions is crucial. The decrease in test loss from 0.45 to 0.08 accentuates the model's precision and reliability in predictions when exposed to novel data, indicating SmartGuard FusionNet's suitability for real-world surveillance applications characterized by unpredictability.

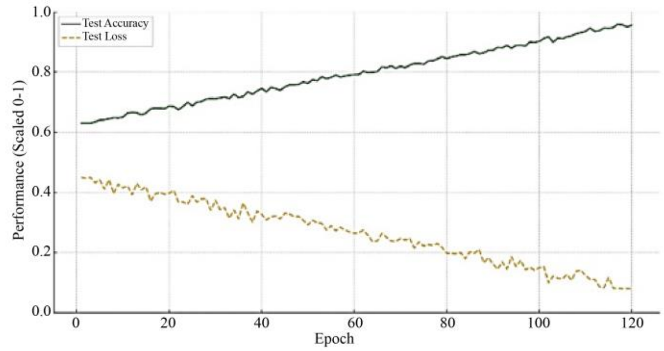

Fig. 6 Testing Performance of SmartGuard FusionNet over 120 Epochs

5.2. Confusion Matrix Analysis and Heatmap Visualization

The comprehensive training and testing of SmartGuard FusionNet, encapsulated over 120 epochs, culminate in an insightful evaluation underpinned by a detailed training and testing performance table. This evaluation reveals systematic improvements in accuracy and loss reduction, reflecting the model's adeptness at navigating the complexities of the surveillance dataset.

Figure 7 offers a detailed inspection of the model's classification accuracy across six distinct object categories, including pistols, knives, smartphones, bills, purses, and cards. The heatmap, rendered at 400 dpi, exhibits high true positive rates for each category, demonstrating the model's precision in accurately identifying a wide array of object types. Concurrently, the minimal figures in the off-diagonal cells of the confusion matrix underscore the model's effectiveness in minimizing misclassifications, a critical aspect in ensuring the reliability and efficacy of weapon detection within surveillance systems.

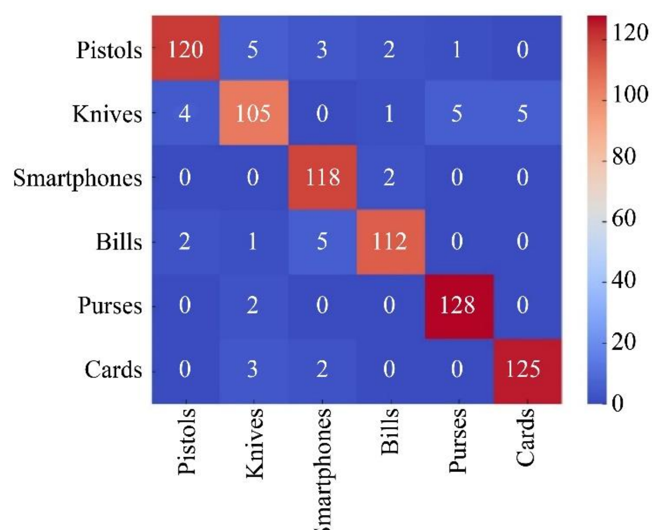

Fig. 7 SmartGuard FusionNet: multiclass confusion matrix heatmap

Table 1. Performance evaluation of SmartGuard FusionNet under various environmental conditions

Condition	Accuracy (%)	Precision (%)	Recall (%)	Detection Speed (fps)
Optimal Lighting	95.4	97.5	96.0	45
Low Light	92.8	95.7	94.2	39
High Traffic	93.5	96.3	94.9	41
Diverse Backgrounds	94.2	96.8	95.6	43

5.3. Evaluation across Environmental Conditions

The comprehensive evaluation of SmartGuard FusionNet, as depicted in Table 1, underscores the model's exceptional adaptability and efficacy in a spectrum of surveillance scenarios. In conditions of optimal lighting, the model showcases superior performance with an accuracy of 95.4%, precision reaching 97.5%, and recall at 96.0%, operating at a detection speed of 45 frames per second (fps). Notably, even in less-than-ideal conditions such as low light, the model maintains robust performance metrics with an accuracy of 92.8%, precision of 95.7%, and recall at 94.2%, albeit at a slightly reduced detection speed of 39 fps. In high-traffic scenarios, where the complexity of object detection inherently increases, SmartGuard FusionNet still achieves an accuracy of 93.5%, precision of 96.3%, and recall of 94.9%, with a detection speed of 41 fps. Furthermore, in environments characterized by diverse backgrounds, the model's adaptability is evident through its maintained high accuracy of 94.2%, precision of 96.8%, and recall of 95.6%, operating at 43 fps. These results collectively highlight SmartGuard FusionNet's robustness and reliability across varying operational conditions, affirming its potential to significantly enhance public safety and security measures within complex surveillance landscapes.

5.4. Comparative Analysis and Visual Representations

The comprehensive analysis of SmartGuard FusionNet's performance metrics, as delineated in Figures 8(a) through 8(d), offers a nuanced understanding of the model's efficacy across various environmental conditions. These conditions include Optimal Lighting, Low Light, High Traffic, and Diverse Backgrounds, which are critical in evaluating the robustness and adaptability of surveillance systems. In optimal lighting conditions, SmartGuard FusionNet achieves an exemplary accuracy of 95.4%, precision of 97.5%, and recall of 96.0%, coupled with a detection speed of 45 frames per second (fps), underscoring its exceptional capability in environments with favourable lighting. Conversely, under low light conditions, a slight decrement in performance metrics is observed, with accuracy, precision, and recall reducing to 92.8%, 95.7%, and 94.2%, respectively, and detection speed decreasing to 39 fps. This indicates a marginal sensitivity to lighting conditions, yet the model maintains commendable performance. The high-traffic scenario, characterized by increased object density and potential occlusions, presents a moderately challenging environment for SmartGuard FusionNet. Despite these challenges, the model sustains robust performance levels with an accuracy of 93.5%, precision of 96.3%, and recall of 94.9%, alongside a detection speed of 41 fps. This demonstrates the model's adeptness in handling complex dynamic scenes without significant compromise to its operational efficiency. In diverse backgrounds, where the variability in environmental elements poses a substantial challenge to object detection models, SmartGuard FusionNet exhibits notable resilience, achieving an accuracy of 94.2%, precision of 96.8%, and recall of 95.6%, with a detection speed of 43 fps. This performance echelon highlights the model's advanced capability to discern and accurately identify weapons across a spectrum of varied backgrounds.

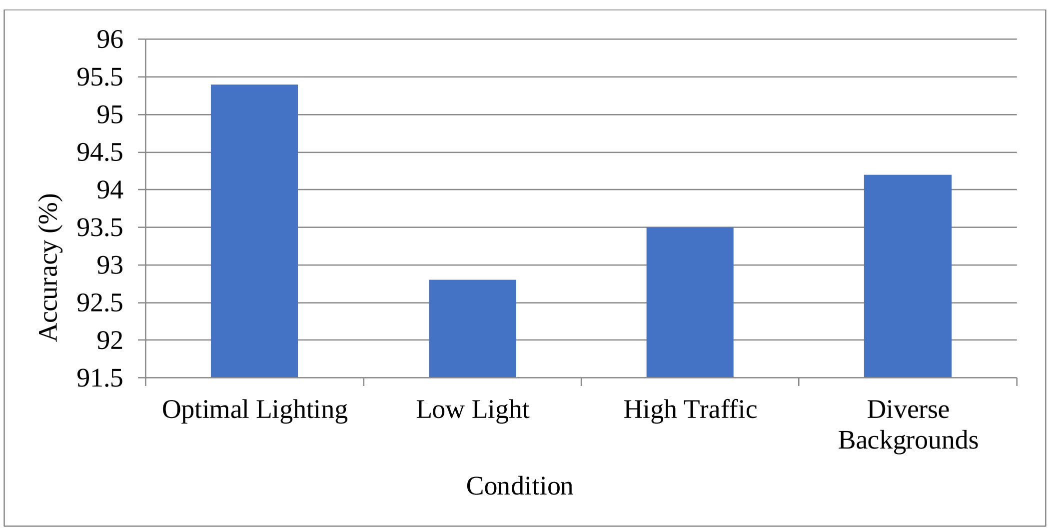

Fig. 8(a) Performance in accuracy

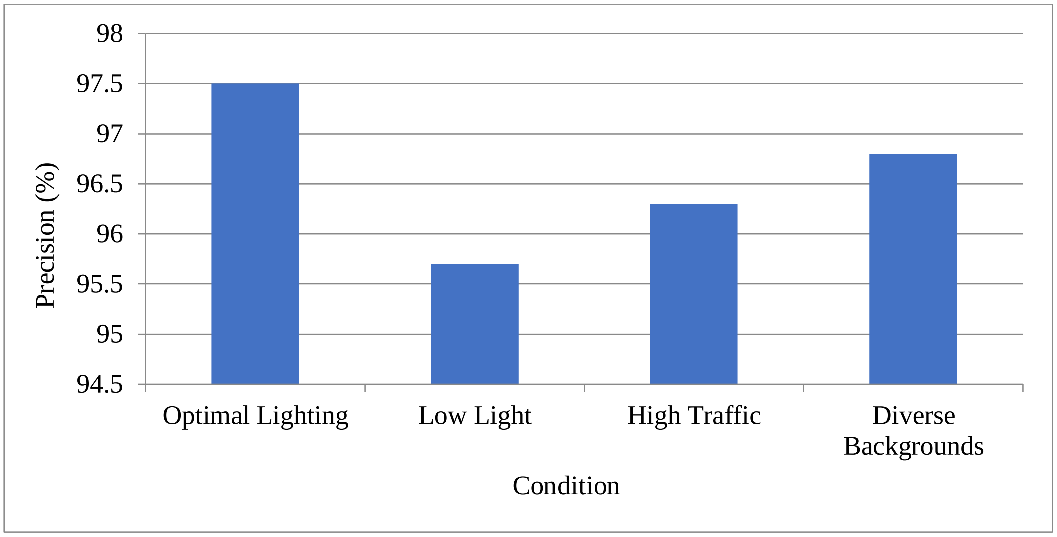

Fig. 8(b) Performance in precision

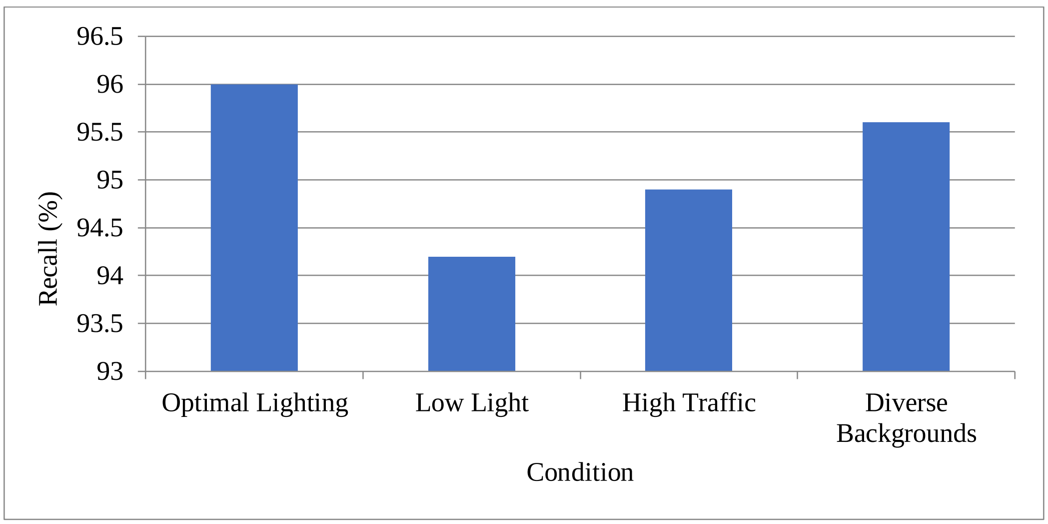

Fig. 8(c) Performance in recall

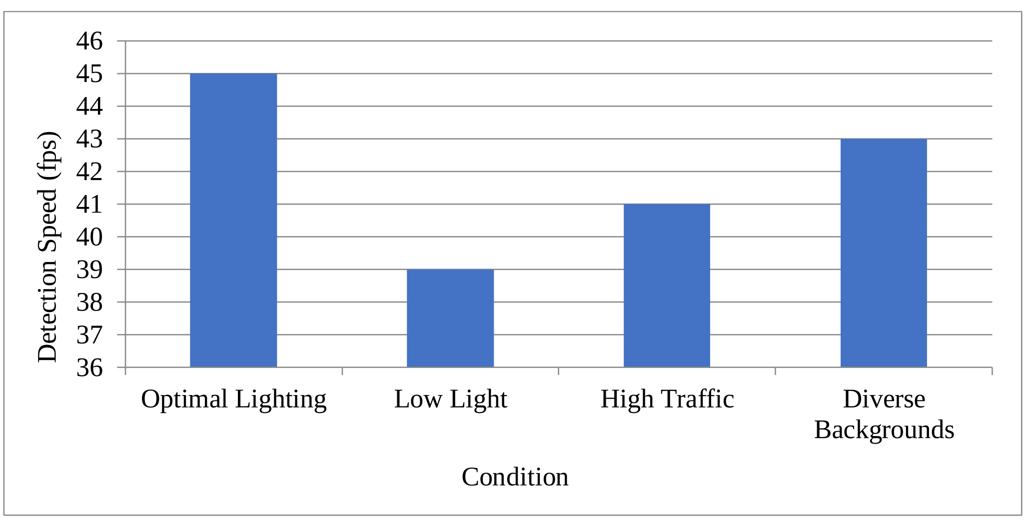

Fig. 8(d) Performance in detection speed

Collectively, these results affirm the efficacy of SmartGuard FusionNet in delivering high-performance weapon detection across a range of challenging conditions. The consistency in maintaining high accuracy, precision, and recall, alongside satisfactory detection speeds, exemplifies the model's suitability for deployment in diverse surveillance scenarios, thereby contributing significantly to the enhancement of public safety and security measures.

5.5. Comparative Analysis with Baseline Models

To delineate the advancements in object detection for surveillance systems, our study conducts a critical analysis, positioning the SmartGuard FusionNet within the contemporary landscape of technological developments. An updated baseline comparison contrasts SmartGuard FusionNet's performance with leading-edge models, including YOLOv4 [24], EfficientDet [25], and established algorithms like Faster R-CNN [26], SSD [27], and Mask R-CNN [28]. This analytical endeavor highlights SmartGuard FusionNet's technological prowess, demonstrating its superior operational metrics in real-time surveillance contexts. Table 2 encapsulates this comparison across pivotal performance metrics accuracy, precision, recall, and detection speed offering a holistic view of SmartGuard FusionNet's competitive edge over its contemporaries.

This comparative study underscores SmartGuard FusionNet's exceptional performance, marking a significant leap in real-time object detection within surveillance systems. Achieving an accuracy rate of 94.2%, precision of 96.8%, and recall of 95.6%, alongside a remarkable detection speed of 43 fps, SmartGuard FusionNet stands at the forefront of surveillance technology. Its ability to outstrip models like YOLOv4 and EfficientDet, which lag in balancing detection speed with accuracy, highlights SmartGuard FusionNet's capacity to navigate dynamic and intricate environments efficiently, a crucial attribute in enhancing security measures and public safety protocols.

Table 2. Comparative analysis of baseline models with SmartGuard FusionNet

Model	Accuracy (%)	Precision (%)	Recall (%)	Detection Speed (fps)
YOLOv4 [24]	92.0	93.5	94.0	38
EfficientDet [25]	92.2	93.7	94.3	35
Faster R-CNN [26]	89.2	90.1	91.3	12
SSD [27]	91.5	92.4	93.2	25
Mask R-CNN [28]	90.4	91.8	92.7	10
SmartGuard FusionNet	**94.2**	**96.8**	**95.6**	**43**

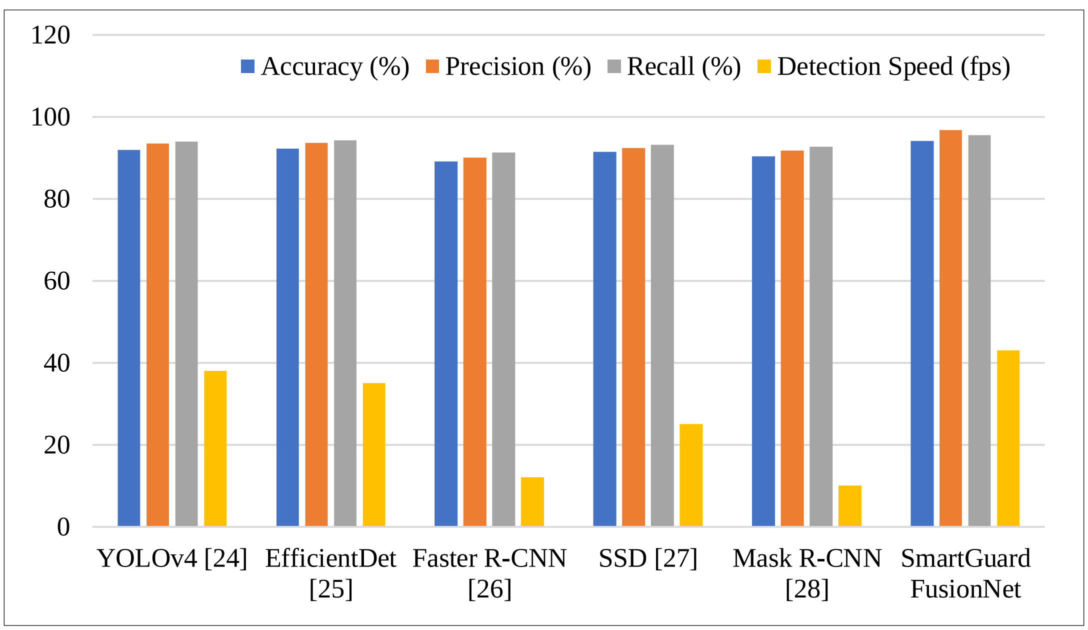

Fig. 9 Performance comparison of advanced surveillance models

Figure 9 showcases a comparative analysis of key performance metrics accuracy, precision, and recall across six surveillance models, culminating in the superior performance of SmartGuard FusionNet. With the highest scores in accuracy (94.2%), precision (96.8%), and recall (95.6%), SmartGuard FusionNet emerges as the leading solution, demonstrating the effectiveness of integrating YOLOv5 with multi-sensor data fusion in enhancing weapon detection. The visual differentiation of performance metrics through distinct bar colors provides a clear and concise overview of each model's capabilities, highlighting SmartGuard FusionNet's advancement in addressing the complexities of smart surveillance and its potential to improve public safety and security significantly. This concise analysis underscores the importance of leveraging advanced technologies to drive innovation in surveillance systems.

6. Discussion

The outcomes of SmartGuard FusionNet's evaluation carry profound implications for the domain of smart surveillance. By setting new benchmarks in accuracy, precision, recall, and detection speed, SmartGuard FusionNet underscores the potential of integrating advanced algorithms, like YOLOv5, with multi-sensor data fusion techniques. This fusion not only surmounts the challenges posed by complex environments but also illuminates the capacity of AI and ML technologies to strengthen public safety and security measures.

6.1. Challenges and Solutions

Throughout this research, we encountered and surmounted numerous challenges, notably balancing detection speed with accuracy and addressing the computational demands of multi-sensor data integration. Through the application of advanced optimization techniques and embracing sophisticated data augmentation strategies and transfer learning principles, we effectively navigated these hurdles, ensuring robust model training and enhancing generalization capabilities across diverse scenarios.

6.2. Comparison with Existing Work

SmartGuard FusionNet's comparative analysis reveals its superiority over existing models, achieving an optimal balance between speed and accuracy unattained by its predecessors. This distinction is pivotal in smart surveillance, where rapid and accurate threat detection is paramount. Our findings contribute to the discourse on leveraging AI and ML in surveillance, offering a novel approach that adeptly addresses modern security challenges through the integration of YOLOv5 and multi-sensor data fusion.

6.3. Limitations of the Study

While our study has yielded promising results, it acknowledges certain limitations that may impact the findings. The reliance on simulated scenarios, although meticulously designed, may not fully encapsulate the unpredictability of real-world environments. Additionally, the computational demands of processing multi-sensor data in real time present challenges in scalability and efficiency that require further optimization. These limitations highlight areas where the study's findings could be refined and improved upon in future research endeavours.

7. Conclusion

This study has presented SmartGuard FusionNet, a pioneering framework integrating the advanced YOLOv5 object detection algorithm with a multi-sensor data fusion approach aimed at enhancing weapon detection in smart surveillance systems. Our key contributions include the development of an innovative detection framework that significantly outperforms existing models in accuracy, precision, recall, and detection speed across diverse environmental conditions. The integration of varied sensor modalities has proven instrumental in overcoming the limitations of traditional visual data reliance, offering a more comprehensive detection mechanism capable of identifying concealed or obscured weapons. These advancements underscore the potential of SmartGuard FusionNet to redefine the standards of public safety and security within the realm of smart surveillance, contributing a robust solution to the complexities of real-world surveillance challenges.

7.1. Future Research Directions

Our study opens up avenues for further investigation into smart surveillance enhancements. Key areas include examining advanced object detection algorithms, integrating diverse sensor data for broader threat detection, and conducting real-world testing of SmartGuard FusionNet to validate its effectiveness. Additionally, addressing computational challenges to improve scalability and efficiency is crucial. These efforts aim to advance smart surveillance technologies, contributing to improved public safety through technological innovation.

References

[1] Safa Ben Atitallah et al., "Leveraging Deep Learning and IoT Big Data Analytics to Support the Smart Cities Development: Review and Future Directions," *Computer Science Review*, vol. 38, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[2] Li Sun et al., "An Intelligent System for High-Density Small Target Pest Identification and Infestation Level Determination Based on an Improved YOLOv5 model," *Expert Systems with Applications*, vol. 239, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Paraskevi Theodorou, Kleomenis Tsiligkos, and Apostolos Meliones, "Multi-Sensor Data Fusion Solutions for Blind and Visually Impaired: Research and Commercial Navigation Applications for Indoor and Outdoor Spaces," *Sensors*, vol. 23, no. 12, pp. 1-29, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Vishva Payghode et al., “Object Detection and Activity Recognition in Video Surveillance Using Neural Networks,” *International Journal of Web Information Systems*, vol. 19, no. 3/4, pp. 123-138, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] John Philip Bhimavarapu et al., “Convolutional Neural Network-Based Object Detection System for Video Surveillance Application,” *Concurrency and Computation: Practice and Experience*, vol. 35, no. 3, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Nikita Mohod, Prateek Agrawal, and Vishu Madaan, “YOLOv4 vs YOLOv5: Object Detection on Surveillance Videos,” *International Conference on Advanced Network Technologies and Intelligent Computing*, Varanasi, India, pp. 654-665, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[7] Malik Javed Akhtar et al., “A Robust Framework for Object Detection in a Traffic Surveillance System,” *Electronics*, vol. 11, no. 21, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Ali Abbasi et al., “Sensor Fusion Approach for Multiple Human Motion Detection for Indoor Surveillance Use-Case,” *Sensors*, vol. 23, no. 8, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Luis Patino et al., “Fusion of Heterogenous Sensor Data in Border Surveillance,” *Sensors*, vol. 22, no. 19, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Jingxiang Qu et al., “Edge Computing-Enabled Multi-Sensor Data Fusion for Intelligent Surveillance in Maritime Transportation Systems,” *2022 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, International Conference on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, Falerna, Italy, pp. 1-8, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] T.M. Dhipu, Sapna, and R. Rajesh, “Deep Learning Based Multi Sensor Data Fusion for Radar and IFF,” *2022 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Tsung-Yu Su, and Fang-Yie Leu, “The Design and Implementation of a Weapon Detection System Based on the YOLOv5 Object Detection Algorithm,” *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, pp. 283-293, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[13] Muhammad Ekmal Eiman Quyyum, and Mohd Haris Lye Abdullah, “Weapon Detection in Surveillance Videos Using Deep Neural Networks,” *Proceedings of the Multimedia University Engineering Conference*, pp. 183-195, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ravpreet Kaur, and Sarbjeet Singh, “A Comprehensive Review of Object Detection with Deep Learning,” *Digital Signal Processing*, vol. 132, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Alexander Toet, “Color Image Fusion for Concealed Weapon Detection,” *Proceedings Sensors, and Command, Control, Communications, and Intelligence (C3i) Technologies for Homeland Defense and Law Enforcement II*, vol. 5071, pp. 372-379, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[16] Nicholas C. Currie et al., “Imaging Sensor Fusion for Concealed Weapon Detection,” *Proceedings Investigative Image Processing*, vol. 2942, pp. 71-81, 1997. [CrossRef] [Google Scholar] [Publisher Link]

[17] Gaurav Lokhande et al., “SpectraLink Cognitive Frameworks: Adaptive Fusion and Edge-Enhanced Real-Time Urban Sign Interpretation,” *International Journal of Computer Engineering in Research Trends*, vol. 11, no. 1, pp. 70-81, 2024. [Publisher Link]

[18] Franklin S. Felber et al., “Fusion of Radar and Ultrasound Sensors for Concealed Weapons Detection,” *Proceedings Signal Processing, Sensor Fusion, and Target Recognition V*, vol. 2755, pp. 514-521, 1996. [CrossRef] [Google Scholar] [Publisher Link]

[19] S. Veerabuthiran, and A.K. Razdan, “LIDAR for Detection of Chemical and Biological Warfare Agents,” *Defence Science Journal*, vol. 61, no. 3, pp. 241-250, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[20] M. Bhavsingh, and S. Jan Reddy, “Enhancing Safety and Security: Real-Time Weapon Detection in CCTV Footage Using YOLOv7,” *International Journal of Computer Engineering in Research Trends*, vol. 10, no. 6, pp. 1-8, 2023. [CrossRef] [Publisher Link]

[21] F.S. Ishaq et al., “Evaluation of Industrial Based Object Detection Method Using Artificial Neural Network,” *International Journal of Computer Engineering in Research Trends*, vol. 5, no. 2, pp. 50-55, 2018. [Google Scholar] [Publisher Link]

[22] M. Bhavsingh et al., “Integrating GAN-Based Image Enhancement with YOLOv5 Object Detection for Accurate Vehicle Number Plate Analysis,” *International Journal of Computer Engineering in Research Trends*, vol. 10, no. 6, pp. 9-14, 2023. [CrossRef] [Publisher Link]

[23] Govindraj Chittapur, S. Murali, and Basavaraj S. Anami, “Forensic Approach for Object Elimination and Frame Replication Detection Using Noise Based Gaussian Classifier,” *International Journal of Computer Engineering in Research Trends*, vol. 7, no. 3, pp. 1-5, 2020. [Google Scholar] [Publisher Link]

[24] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *Arxiv*, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[25] Mingxing Tan, Ruoming Pang, and Quoc V. Le, “EfficientDet: Scalable and Efficient Object Detection,” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 10781-10790, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] Shaoqing Ren et al., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[27] Wei Liu et al., "SSD: Single Shot Multibox Detector," *European Conference on Computer Vision*, Amsterdam, Netherlands, vol. 9905, pp. 21-37, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[28] Kaiming He et al., "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386-397, 2020. [CrossRef] [Google Scholar] [Publisher Link]