Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Advancements in Automated Weapon Detection: A Review of Modern Techniques and Emerging Trends

Nisha Bhadauriya Agarwal
[0009-0003-2422-7250]
Ph.D (CSE) Scholar, Sage University
Indore, India

Dr. Deepak Kumar Yadav
[0009-0000-1226-487X]
HoD(CSE), Sage University
Indore, India

***Abstract*— The rise in weapon-related crimes has made it necessary to create smart surveillance systems that can spot threats in real time. Computer vision and deep learning-based automated weapon detection has become an important way to make the public safer. This paper offers an extensive examination of contemporary weapon detection methodologies, emphasizing traditional machine learning, deep learning-based object detection frameworks, and novel strategies. The performance, accuracy, and real-time usefulness of important architectures like YOLO, Faster R-CNN, and SSD are looked at. This paper also talks about some of the biggest problems, such as finding small objects, occlusion, and the limits of datasets. Finally, the paper talks about possible future research areas, such as edge deployment, transformer-based models, and multimodal systems.**

Keywords*— *Weapon Detection, Deep Learning, YOLO, Surveillance Systems, Computer Vision, Artificial Intelligence

I. INTRODUCTION

The rapid rise in weapon-related crimes has become a global issue, making it necessary to have advanced surveillance systems that can quickly find threats. Human monitoring is a big part of traditional surveillance systems, but people can get tired and make mistakes. AI-powered weapon detection systems, or automated weapon detection systems (Fig. 1), try to get around these problems by using computer vision and artificial intelligence.

Fig.1. The Future of Security: AI-Powered Weapon Detection Systems Explained

Recent studies show that automated solutions can speed up response times and make it less necessary to keep an eye on things manually. Convolutional neural networks (CNNs) and other deep learning techniques have been shown to be better at finding weapons in pictures and videos [1].

The rise in weapon-related crimes is a serious threat to public safety and security around the world. Most traditional surveillance systems depend on human operators to keep an eye on things, which is often not very effective, takes a lot of time, and is prone to mistakes because people get tired and lose focus. Because of this, there is a growing need for smart and automated systems that can quickly find these kinds of threats.

Automated weapon detection systems that can identify firearms and knives in images and videos have been made possible by recent advancements in computer vision and artificial intelligence (AI). These systems look at visual input and locate items with high accuracy using deep learning and machine learning methods. In real-time identification tasks, Convolutional Neural Networks (CNNs) and contemporary object detection frameworks such as YOLO, SSD, and Faster R-CNN have demonstrated remarkable performance.

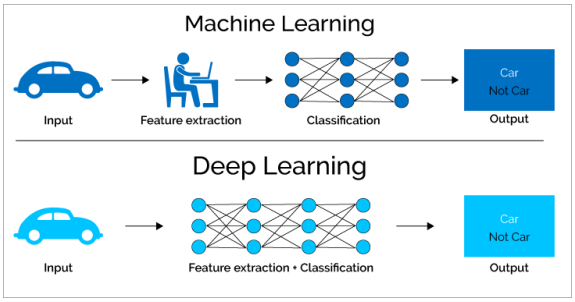

Fig. 2. Machine vs deep learning as artificial intelligence

The transition from conventional machine learning methods, which depend on manual feature extraction, to deep learning-based techniques has greatly enhanced detection precision, resilience, and scalability (Fig. 2). Nonetheless, notwithstanding these progressions, numerous challenges endure, such as the detection of small objects, occlusion, fluctuating lighting conditions, and the restricted availability of diverse datasets.

This paper provides a thorough examination of automated weapon detection methodologies, emphasizing contemporary deep learning techniques [10] and their utilization in surveillance systems. It also talks about important problems and looks at new trends like transformer-based architectures, multimodal systems, and edge computing for real-time deployment.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

II. BACKGROUND

Deep learning, machine learning, and computer vision have all made automated weapon detection a lot better in the last ten years. Initially, traditional machine learning methods were used, but deep learning models eventually took their place and improved them. To enhance detection, hybrid methodologies that integrate the optimal aspects of both paradigms have recently emerged. This section gives a brief overview of these methods.

A. Traditional Machine Learning Approaches

Conventional machine learning methods for weapon detection depend on manually crafted feature extraction methods and standard classifiers. Histogram of Oriented Gradients (HOG) [8], Scale-Invariant Feature Transform (SIFT), and Local Binary Patterns (LBP) are all common feature descriptors that take shape, texture, and edge information from images (Fig. 3). After that, these features are put into classifiers like Support Vector Machines (SVM) [8], k-Nearest Neighbors (k-NN), or Decision Trees to sort the objects.

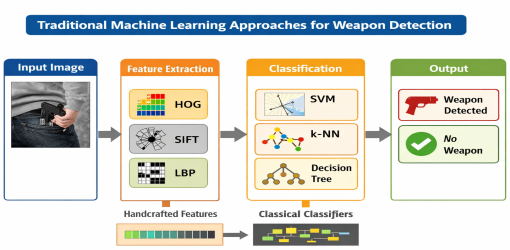

Fig. 3. Traditional machine learning approaches

These methods were the basis for early weapon detection systems, but they have some problems. These kinds of models are less flexible in complex and changing real-world situations because their performance depends mostly on how good the elements that people make are. They also have trouble with changes in lighting, scale, direction, and occlusion. Because of this, traditional methods often don't generalize well, are not very precise, and depend too much on feature engineering [1].

B. Deep Learning-Based Approaches

Deep learning has completely changed how weapons are identified by making it possible to learn from start to finish and automatically extract features. Deep learning-based methods have made weapon detection much better by making end-to-end learning and automatic feature extraction possible. As shown in Fig. 4, Convolutional Neural Networks (CNNs) build hierarchical representations directly from raw picture data, so they don't need to do manual feature engineering. This enables models to more effectively capture complex patterns and contextual information.

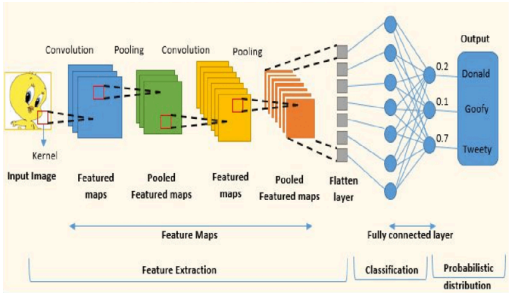

Fig. 4. A typical Convolutional Neural Network (CNN)

Modern object detection frameworks can be divided into two main groups: two-stage detectors and one-stage detectors (Fig. 5). Faster R-CNN and other two-stage detectors generate region proposals first and then classify them. This makes them more accurate, but it costs more to run. One-stage detectors like YOLO (You Only Look Once) [11] and Single Shot Detector (SSD) [17], on the other hand, do detection in one pass, which makes them much faster and better for real-time use.

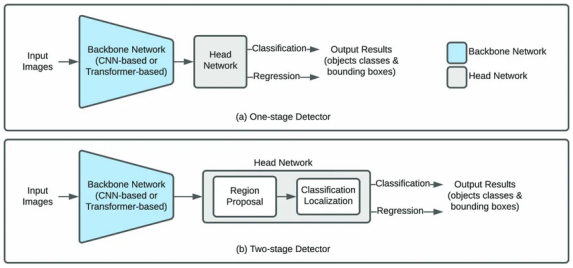

Fig. 5. A Comprehensive analysis of one stage and two stage object detector[23]

Deep learning models are better than traditional methods when it comes to accuracy, resilience, and scalability. But they need a lot of processing power and large datasets with notes. Even with these problems, deep learning is still the most common way to automatically detect weapons in modern surveillance systems.

1) Two-Stage Detectors

A type of deep learning-based object detection models known as "two-stage detectors" carry out detection in two consecutive steps: object categorization and region proposal. Because of their well-known high detection accuracy and accurate localization capabilities, these models are appropriate for situations where accuracy is more important than speed. The model creates a set of potential object regions, or "Region Proposals," in the first stage that are probably going to include objects. Algorithms like Selective Search (in R-CNN) or a Region Proposal Network (RPN) in more sophisticated designs like Faster R-CNN are commonly used to do this. Classifying these suggested locations into object types (such as weapon or non-weapon) and fine-tuning their bounding box coordinates constitute the second step.

R-CNN, Fast R-CNN [15], and Faster R-CNN are well-known two-stage detectors. Due to its increased efficiency and end-to-end training capacity, Faster R-CNN, as depicted in Fig. 6, is the most popular of them. In comparison to previous models, it greatly reduces computational cost by directly integrating the region suggestion mechanism into the network.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

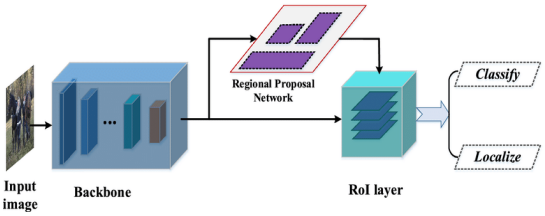

Fig. 6. Two stage object detectors (Faster R-CNN)

High precision, improved handling of small objects, and precise object localization are only a few benefits of two-stage detectors. These advantages, however, come at the expense of slower inference speed and more computational complexity, which makes them less appropriate for real-time applications like live surveillance systems. When it comes to weapon detection, two-stage detectors are frequently chosen in situations like forensic investigations or offline video processing that call for thorough analysis and great dependability. However, latency limits limit their use in real-time applications. Faster R-CNN and R-CNN family models offer superior localization and excellent accuracy. Nevertheless, they are slower and more costly to compute in real-time applications.

2) One-Stage Detectors

A "one-stage detector" is a deep learning-based object detection model that can find and classify objects in just one pass through the network. Unlike two-stage detectors, it doesn't make region proposals on its own. Instead, it uses the input image to quickly guess bounding boxes, class probabilities, and confidence scores. SSD [17] and YOLO (You Only Look Once) [11] are two examples of models that divide an image into grids and find multiple objects at once. This makes them very fast and useful for real-time tasks like surveillance. One-stage detectors are a good choice for activities that need to be done quickly, like finding weapons, because they are fast and accurate. However, they are sometimes not quite as accurate as two-stage detectors.

YOLO-based models are popular for the reasons listed below:

a) Fast speed

b) The capacity for real-time detection

c) Excellent mAP performance, sometimes reaching ~97%[2]

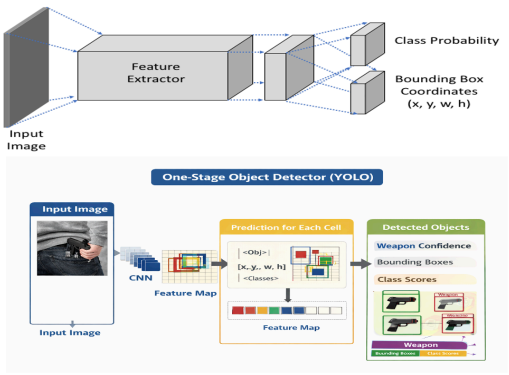

Fig. 7. One stage object detector

To enhance weapon identification performance, hybrid systems combine the advantages of deep learning and conventional machine learning methods [6]. In these systems, standard machine learning classifiers like Support Vector Machines (SVM) or Random Forests carry out the final classification, while deep learning models, usually Convolutional Neural Networks (CNNs), are employed as feature extractors. High-level characteristics are extracted from intermediate CNN layers and fed into a machine learning classifier in a typical hybrid pipeline. This method preserves strong feature representation while lowering computational cost. Furthermore, transfer learning is frequently used to refine pre-trained CNN models for weapon detection tasks (e.g., trained on ImageNet [22]).

Because hybrid approaches make use of pre-trained information and more straightforward classifiers, they are especially helpful in situations with less labeled data. Nevertheless, they might not fully utilize end-to-end optimization and might be less effective than detection frameworks that rely entirely on deep learning.

III. MODERN TECHNIQUES IN WEAPON DETECTION

Automated weapon identification systems now function much better thanks to recent developments in computer vision and artificial intelligence. Deep learning-based object identification frameworks, improved feature extraction techniques, and effective deployment tactics are the mainstays of contemporary methods. High accuracy, resilience, and real-time performance are the goals of these methods in intricate surveillance settings.

A. Advanced YOLO Variants

Recent versions such as YOLOv5, YOLOv7[14], and YOLOv8 [7] offer:

a) Improved detection accuracy

b) Faster inference

c) Better handling of small objects

YOLO-based systems dominate real-time surveillance applications due to their efficiency.

B. Attention Mechanisms

Attention mechanisms have become a significant improvement in deep learning models by allowing the network to concentrate on the most pertinent areas of an image. The Convolutional Block Attention Module (CBAM) [18] and Squeeze-and-Excitation (SE) blocks improve feature maps by highlighting important information in both the spatial and channel dimensions (Fig. 8).

Attention mechanisms help improve weapon detection accuracy, especially when the background is cluttered, there are obstructions, or the objects are small. By selectively emphasizing important features, these methods reduce noise and make the model easier to understand.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

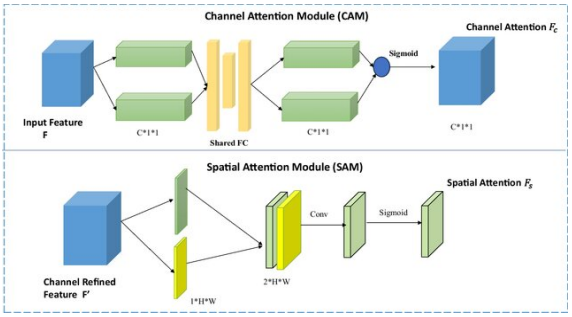

Fig. 8. A Typical Conventional Block Attention Module (CBAM)[24]

Attention modules like CBAM [18] (Convolutional Block Attention Module) and SE (Squeeze-and-Excitation) enhance performance by:

- Focusing on relevant features
- Reducing background noise

C. Data Augmentation and Synthetic Data

Data augmentation techniques play a crucial role in improving the generalization capability of deep learning models. Common augmentation methods include:

- Rotation and scaling
- Flipping and cropping
- Brightness and contrast adjustment
- Noise injection

These methods improve the diversity of datasets and assist models in learning resilient features in a variety of scenarios. Furthermore, the use of Generative Adversarial Networks (GANs) and simulation tools to generate synthetic data has grown in prominence as a solution to data scarcity and class imbalance in weapon detection datasets. Image augmentation and synthetic datasets are used to improve generalization techniques in order to get over dataset restrictions. Research shows that both synthetic and actual data greatly improve detection performance [1].

D. Data Augmentation and Synthetic Data

In order to take use of pre-trained models learned on massive datasets like ImageNet, transfer learning is frequently utilized in weapon detection [22]. Even with little labeled data, researchers can attain great accuracy by refining these models on weapon-specific datasets. This method improves performance while drastically cutting training time and computing cost, making it ideal for real-world applications.

E. Edge Computing and Real-Time Systems

The use of weapon detection models on edge devices has grown in significance because to the rising need for real-time surveillance. By enabling on-device processing, edge computing lowers latency and decreases reliance on cloud infrastructure. Deep learning models are optimized for deployment on resource-constrained devices, such as embedded systems and smart cameras, using methods including model compression, pruning, and quantization. This guarantees quick and effective detection in practical settings. Deployment on edge devices enables:

- Low latency
- Real-time monitoring
- Reduced cloud dependency

To attain high accuracy and real-time performance, contemporary weapon detection algorithms [9] make use of sophisticated deep learning models, attention mechanisms, data augmentation approaches, and edge computing. These strategies greatly outperform conventional techniques, although issues like data constraints and environmental unpredictability still need to be investigated.

IV. CHALLENGES IN WEAPON DETECTION

The creation of extremely dependable and resilient systems is still hampered by a number of issues, despite notable progress in automated weapon detection utilizing deep learning and computer vision techniques. Real-world complexity, environmental variability, and constraints in current datasets and models are the causes of these difficulties.

A. Small Object Detection

Accurately recognizing small objects, like knives or handguns, is one of the main hurdles in weapon identification, especially when they take up a relatively small percentage of the image. Weapons are frequently photographed at a distance or with low resolution in surveillance situations, which makes feature extraction challenging. Inadequate feature representation at deeper layers may make it difficult for deep learning models, especially one-stage detectors like YOLO, to identify such tiny objects. This frequently leads to false negatives, or missed detections, which can be crucial for security applications.

B. Occlusion and Cluttered Backgrounds

Weapons are often partially concealed or obscured by other items, such clothing, purses, or human body parts. Occlusion makes differentiating characteristics less visible, which makes it difficult for detection models to accurately identify the object. Furthermore, by adding noise and unnecessary features, complex and cluttered backgrounds in real-world settings significantly impair model performance. This raises the possibility of both false positives and false negatives.

C. Variations in Lighting and Environmental Conditions

The performance of vision-based weapon detection systems is greatly impacted by lighting conditions. Reduced detection accuracy might result from distorted visual features caused by poor illumination, shadows, glare, and nighttime surveillance circumstances. In low light or in unfavorable weather

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

conditions like fog, rain, or smoke, models trained on well-lit datasets frequently struggle to generalize. This emphasizes the necessity of multimodal strategies and reliable preprocessing methods (such as thermal imaging).

D. False Positives and False Negatives

In weapon detection systems, misclassification is still a serious problem. False positives happen when non-threatening items (such tools or cell phones) are mistakenly classified as weapons, resulting in pointless warnings. False negatives, on the other hand, occur when real weapons go unnoticed and present significant security threats. Finding a balance between recall and precision is crucial yet difficult, particularly in dynamic settings with a variety of item appearances.

E. Dataset Limitations and Imbalance

The caliber and variety of training data have a significant impact on how well deep learning models function. However, there are several issues with current weapon detection databases, including:

• There aren't enough labeled photos

• A lack of diversity in environments, weapon types, and orientations

• The disparity between classes (fewer photos of weapons than non-weapons)

These problems cause the model to behave biasedly and with low generalization. Additionally, access to real-world surveillance data is restricted due to privacy and security concerns, which makes it challenging to create comprehensive datasets.

F. Real-Time Processing Constraints

Systems for detecting weapons in surveillance settings must operate in real-time or almost real-time. However, complicated structures that need a large amount of processing power are frequently involved in high-accuracy models. This forces a trade-off between inference speed and detection accuracy, especially when deploying on edge devices with constrained computing resources.

G. Generalization and Robustness

Models that have been trained on certain datasets may not work well when they are put to use in new places because the domains have changed. Changes in the angle of the camera, the resolution, the background, and the way people act can all have a big impact on how well detection works. A big problem in research is making sure that systems are strong and flexible enough to work in a wide range of real-world situations.

Automated weapon detection systems are still not very effective because of problems like detecting small objects, occlusion, changing lighting, misclassification, limited datasets, real-time constraints, and generalization issues. To make sure that security solutions for the real world are reliable, scalable, and easy to use, we need to deal with these problems. Detecting small weapons is one of the biggest problems with current systems [1].

V. EMERGING TRENDS AND FUTURE SCOPE

Automated weapon identification systems are progressing beyond traditional deep learning frameworks due to the quick developments in computer vision and artificial intelligence. While addressing existing constraints, emerging technologies seek to improve detection accuracy, resilience, and real-time performance.

A. Transformer-Based Models

Recent advancements in Vision Transformers (ViTs) and Detection Transformers (DETR) [20] have instigated a transformative shift in object detection methodologies. Transformers [20] use self-attention mechanisms to find global contextual relationships in an image, which is different from Convolutional Neural Networks (CNNs). This makes it easier to find complicated objects and deal with occlusion and messy scenes. Future research is anticipated to concentrate on hybrid CNN-transformer models that integrate local feature extraction with global attention, resulting in enhanced and more efficient weapon detection systems. Vision Transformers (ViTs) [20] are becoming more popular because they can extract global features and help people understand context better.

B. Multimodal Detection Systems

RGB pictures are the main source of information used in traditional weapon detection systems, although these systems are sensitive to ambient factors. Multimodal methods incorporate several data sources, including:

- Thermal imaging
- The use of infrared sensors
- Sensors for depth

These technologies enhance detecting performance in conditions of poor light, fog, or darkness. The combination of several modalities improves their robustness and dependability, which qualifies them for practical surveillance uses. Future systems might integrate infrared sensors, thermal imaging, and RGB pictures. This enhances detection in difficult settings..

C. Edge Computing and Real-Time Deployment

The use of weapon detection models on edge devices, such as embedded systems and smart cameras, is becoming increasingly popular. Without depending on cloud infrastructure, edge computing lowers latency, uses less bandwidth, and permits real-time decision-making. In order to accomplish effective deployment on devices with limited resources, future systems will concentrate on lightweight model architectures, model compression methods, and hardware optimization.

D. Explainable AI (XAI)

It can be hard to understand why deep learning models make the decisions they do because they are often "black boxes." Explainable AI wants to make things clear by showing which parts of an image affect the model's prediction. Techniques like Grad-CAM and attention visualization help us understand how models work, which builds trust and

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

reliability in important areas like weapon detection. By doing the following, it makes things more open and trustworthy:

- Giving reasons for model choices
- Making black-box behavior less common

E. Federated and Edge Learning

Federated learning has become a viable solution as worries about data security and privacy have grown. It protects user privacy by allowing several devices to work together to train models without exchanging raw data. In surveillance systems where sensitive data cannot be transferred or stored centrally, this method is especially helpful. This knowledge allows for decentralized training and improves privacy.

F. Data Augmentation and Synthetic Data Generation

The absence of labeled and diverse datasets is one of the main obstacles to weapon detection. More sophisticated methods of augmenting data and creating artificial data with Generative Adversarial Networks (GANs) are becoming popular solutions. Particularly in situations with little real-world data, these techniques improve model generalization, diversify datasets, and lessen overfitting.

G. Real-Time IoT Surveillance Systems

Weapon detection will be better in the future if it works with smart city infrastructure and the Internet of Things (IoT) ecosystems. You can connect AI-powered surveillance systems to:

- Systems that send alerts automatically
- Databases for law enforcement
- Ways to respond to emergencies

This kind of integration makes it possible to find threats early and respond quickly, which greatly improves public safety.

VI. COMPARATIVE ANALYSIS

A comparative analysis of various weapon detection techniques [9] is essential to evaluate their effectiveness in terms of accuracy, speed, computational complexity, and suitability for real-world applications. A thorough comparison of contemporary deep learning-based object detection models, including both two-stage and one-stage detectors, and conventional machine learning techniques is provided in this section. The efficacy of traditional machine learning techniques is limited in complicated contexts since they rely on manually created features and traditional classifiers. On the other hand, hierarchical feature representations are automatically learned by deep learning models, which greatly increases detection robustness and accuracy.

Because of their sequential processing architecture, two-stage detectors like Faster R-CNN have slower inference speeds despite offering high precision and accurate localization. However, one-stage detectors, such as YOLO and SSD, are better appropriate for real-time surveillance systems since they provide quicker detection with marginally lower but competitive accuracy. Table I, summarizes the comparative analysis of Technologies with respect to various parameters like accuracy, speed, limitations etc.

TABLE I. COMPARATIVE ANALYSIS

Technology	Accuracy	Speed	Real-Time Suitability	Limitation
Traditional ML	Low	Medium	No	Poor generalization
Faster R-CNN[15][16]	High	Low	Limited	High computation
SSD [17]	Medium	High	Yes	Lower accuracy
YOLO	High	Very High	Yes	Struggles with small objects

The trade-off between inference speed and detection accuracy is a crucial component of weapon detection systems. While high-speed models like YOLO are favored for real-time surveillance, high-accuracy models like Faster R-CNN are appropriate for offline analysis and forensic applications.

- Two-stage detectors with high accuracy and low speed
- One-stage detectors: Moderate Accuracy + High Speed
- Poor Accuracy: Conventional techniques

When choosing a model based on the needs of the application, this trade-off is crucial. When compared to conventional techniques, deep learning models show more robustness, particularly when managing variances like:

- Conditions of lighting
- Clutter in the background
- The direction of objects

Nonetheless, the caliber and variety of training datasets have a significant impact on their effectiveness. Models that have been trained on small datasets might not be able to generalize well in practical situations. Different techniques are suitable for different deployment scenarios:

1. Traditional ML: Suitable for simple, controlled environments
2. Faster R-CNN: Ideal for high-accuracy requirements (e.g., forensic analysis)
3. SSD: Balanced performance for moderate real-time systems
4. YOLO: Best suited for real-time surveillance and edge deployment

In weapon detection tests, it is clearly demonstrated that deep learning-based methods perform noticeably better than conventional machine learning approaches. Among these, YOLO-based models provide the best speed-accuracy ratio, which makes them ideal for real-time applications. Faster R-CNN is still a good option, though, in situations when great precision is needed. The particular needs of accuracy, speed, and deployment environment ultimately determine which model is best.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

VII. CONCLUSION AND FUTURE WORK

One crucial use of computer vision and artificial intelligence to improve public safety and security is automated weapon detection. The current methods for detecting weapons are covered in this study, from sophisticated deep learning-based models to conventional machine learning techniques. Deep learning techniques, especially one-stage detectors like YOLO, are shown to offer the best possible balance between accuracy and real-time performance, which makes them ideal for surveillance applications.

This paper highlights the limitations of traditional methods, including their dependency on handcrafted features and poor generalization in complex environments. Deep learning models, on the other hand, proved to be more adept at managing changes in background, scale, and lighting. Additionally, detection performance has been greatly enhanced by the incorporation of transfer learning, data augmentation techniques, and attention processes. Small object detection, occlusion, dataset limits, and the trade-off between accuracy and computing efficiency are some of the issues that still need to be resolved despite these developments. Developing dependable and scalable weapon detection systems requires addressing these problems.

Future research directions emphasize the adoption of transformer-based architectures, multimodal detection systems, edge computing, and explainable AI to enhance system robustness and transparency. The integration of these emerging technologies with real-time surveillance and smart city infrastructure hold significant potential for proactive threat detection and rapid response.

In conclusion, even though automated weapon identification has advanced significantly, more study and creativity are required to create effective, precise, and deployable solutions that can function well in real-world settings.

ACKNOWLEDGMENT

I thank Almighty God for guidance and strength. I also thank my guide for valuable support and direction, and the faculty of Sage University for their continuous guidance.

REFERENCES

[1] A. Author et al., “Systematic review on weapon detection in surveillance footage through deep learning,” *Computer Science Review*, vol. 51, 2024.

[2] M. Talib and J. Saud, “A multi-weapon detection using deep learning,” *Iraqi Journal of Information and Communication Technology*, vol. 7, no. 1, 2024.

[3] S. Mane, “Weapon detection and classification using deep learning,” *ITEGAM Journal of Engineering*, 2024.

[4] S. Murugaiyan et al., “An enhanced weapon detection system using deep learning,” *Proc. ICNWC*, 2024.

[5] D. M. et al., “Weapon detection using deep learning,” *IJARCCE*, vol. 12, no. 4, 2023.

[6] S. Author et al., “A comprehensive study on weapon detection using machine learning and deep learning,” *Expert Systems with Applications*, 2023.

[7] A. Thakur et al., “Real-time weapon detection using YOLOv8,” *arXiv preprint*, 2024.

[8] N.B. Agarwal and Deepak Kumar Yadav, “ An Overview of Different Object Detection Algorithms and Libraries,” *IJREAM*, Vol-10, Issue-04, July 2024.

[9] K. Akhila and K. Ahmed, “Real-time deep learning weapon detection techniques,” *arXiv preprint*, 2024.

[10] Nisha Bhadauriya Agarwal, Dr. Deepak Kumar Yadav, 2024, A Comprehensive Analysis of Classical Machine Learning and Modern Deep Learning Methodologies, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 13, Issue 05 (May 2024),

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 779–788.

[12] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” *arXiv preprint arXiv:1804.02767*, 2018.

[13] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” *arXiv preprint arXiv:2004.10934*, 2020.

[14] C.Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors,” *arXiv preprint arXiv:2207.02696*, 2022.

[15] R. Girshick, “Fast R-CNN,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 1440–1448.

[16] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017.

[17] W. Liu et al., “SSD: Single Shot MultiBox Detector,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 21–37.

[18] S. Woo, J. Park, J.Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[19] A. Vaswani et al., “Attention Is All You Need,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 5998–6008.

[20] N. Carion et al., “End-to-End Object Detection with Transformers,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 213–229.

[21] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2012, pp. 1097–1105.

[23] https://www.researchgate.net/publication/358362847_A_Survey_of_Deep_Learning-Based_Object_Detection_Methods_and_Datasets_for_Overhead_Imagery/figures?lo=1

[24] https://www.researchgate.net/publication/390978920_Hybrid_attention-inflated_3D_architecture_for_human_action_recognition/figures?lo=1