Weapon Detection Using Deep Learning Methods

Dr. Geetha S
School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology
Chennai, India
geetha.s@vit.ac.in

Ashwin Parthasarathy
School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology
Chennai, India
ashwin.parthasarathy2021@vitstudent.ac.in

Harsh Rajesh Kaneria
School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology
Chennai, India
harshrajesh.kaneria2021@vitstudent.ac.in

Bhavya Jain
School of Computer Science and Engineering (SCOPE)
Vellore Institute of Technology
Chennai, India
bhavya.jain2021c@vitstudent.ac.in

***Abstract*— Detecting weapons effectively is crucial for improving public safety, especially in environments monitored by surveillance systems. This study compares the performance of three weapon detection methods: YOLOv11, Faster R-CNN, and a combination of both using ensemble techniques. YOLOv11 stands out for its speed and efficiency, making it ideal for real-time applications, while Faster R-CNN excels in precision but struggles with slower processing due to its complex region proposal step. By combining the strengths of both models in an ensemble, we aim to achieve faster, more accurate detection. Using datasets with real-world scenarios such as CCTV footage and weapon imagery, this analysis highlights the trade-offs between speed, accuracy, and computational complexity. The findings suggest that integrating YOLOv11 and Faster R-CNN through ensemble approaches can significantly enhance detection systems, offering practical solutions for modern security challenges.**

Keywords*— *Weapon Detection, YOLOv11, Faster R-CNN, Ensemble Techniques, Object Detection, Deep Learning, Real-Time Surveillance, Security Systems, Public Safety, Machine Learning.

I. INTRODUCTION

The demand for effective weapon detection has grown significantly in response to increasing security concerns and the need for public safety. Automated weapon detection systems are pivotal in preventing potential threats, particularly in sensitive environments such as airports, schools, and public venues. These systems rely heavily on advancements in deep learning and computer vision to ensure rapid and accurate identification of weapons. However, achieving a balance between speed, accuracy, and computational efficiency remains a pressing challenge in deploying these systems for real-time applications.

This study examines three weapon detection methodologies—YOLOv11, Faster R-CNN, and an ensemble approach combining both models—to address the limitations of existing systems. YOLOv11, the latest in the YOLO series, is designed for speed and efficiency, making it ideal for real-time applications. Its architecture incorporates advanced convolutional layers and attention mechanisms, enhancing its ability to detect small or obscured objects, such as concealed weapons. Despite its high speed, YOLOv11 occasionally compromises precision, especially in complex scenes or under poor lighting conditions.

Faster R-CNN, on the other hand, is renowned for its accuracy. It employs a Region Proposal Network (RPN) to generate candidate regions for object detection, which is particularly effective in ensuring precise localization. This multi-stage detection process is well-suited for identifying small or partially visible objects. However, its iterative processing pipeline results in slower inference times, posing challenges for applications requiring immediate responses.

To capitalize on the strengths of both models, this study proposes an ensemble approach that integrates YOLOv11's rapid detection capabilities with Faster R-CNN's precision. By merging predictions from the two models using a weighted voting mechanism, the ensemble method seeks to balance speed and accuracy effectively. The combination aims to address the trade-offs inherent in using either model independently, offering a robust solution for real-world weapon detection challenges.

The models were trained and evaluated using a comprehensive dataset that included CCTV footage and synthetic weapon imagery to mimic real-world scenarios. Performance metrics such as precision, recall, F1 score, and mean average precision (mAP) were used to assess detection accuracy, while inference time was measured to evaluate computational efficiency.

The findings of this study underscore the potential of integrating YOLOv11 and Faster R-CNN into a unified detection system. By addressing critical trade-offs between speed and precision, the proposed ensemble method offers a scalable and reliable solution for modern security applications. Future work could explore the integration of additional models and the use of multi-modal data to further enhance detection capabilities, paving the way for more advanced and effective public safety technologies.

II. RELATED WORK

Keerthana, S. M. & Sujitha, R. & Yazhini, P. in their work "Weapon Detection for Security Using the YOLO Algorithm with Email Alert Notification" [1] introduced a YOLOv8-based intelligent system capable of detecting weapons in real-time across images, videos, and live webcam streams. This system integrates user authentication and provides email alerts and audible notifications, achieving a 94% detection accuracy for predefined weapon classes.

Jadhav, Vedantika & Deshmukh, Rutuja in their study "Weapon Detection from Surveillance Footage in Real-Time Using Deep Learning" [2] analyzed YOLOv7, YOLOv8, CNN, and VGG models for weapon detection. The work highlights the comparative strengths of each model and implements an alert mechanism for dynamic monitoring.

Dugyala, Raman et al. in their research “Weapon Detection in Surveillance Videos Using YOLOv8 and PELSF-DCNN” [3] proposed a hybrid framework combining YOLOv8 with PELSF-DCNN for improved detection in complex scenarios involving motion and occlusion, enhancing reliability through sliding window and silhouette functions.

Khalid, Shehzad et al. in “Weapon Detection System for Surveillance and Security” [4] utilized YOLOv5 and Mask-RCNN to develop a robust weapon detection mechanism, achieving a high F1-score of 95.43% and enhancing segmentation accuracy with data augmentation techniques.

M. Pullakandam et al., in “Weapon Object Detection Using Quantized YOLOv8” [5], leveraged the YOLOv8 model with quantization techniques to develop an efficient weapon detection framework. Their approach enhanced real-time performance by reducing model size and computation requirements, achieving high accuracy and faster detection rates. A comparative analysis with YOLOv5 showed that YOLOv8 outperformed in precision and speed, demonstrating its suitability for resource-constrained environments like embedded systems and mobile devices. The study highlighted the importance of quantization in deploying advanced detection models for practical applications.

Jyothsna, V. et al. in their work “YOLOv8-Based Person Detection, Distance Monitoring, Speech Alerts, and Weapon Identification with Email Notifications” [6] integrated advanced features like distance estimation and audio-based detection into a YOLOv8-based system, emphasizing its real-world applications for comprehensive threat management.

Yadav, Pavinder & Gupta, Nidhi & Sharma, Pawan in their work “Robust Weapon Detection in Dark Environments Using YOLOv7-DarkVision” [7] introduce YOLOv7-DarkVision, optimized for low-light weapon detection. Combining image brightening algorithms and deep learning, the model achieves a precision of 95.50% and an F1-Score of 93.41%, addressing challenges of detecting weapons in nighttime CCTV footage.

Wang, Guanbo & Ding, Hongwei & Duan, Mingliang in their study “Fighting Against Terrorism: A Real-Time CCTV Autonomous Weapons Detection System” [8] propose an enhanced YOLOv4 with SCSP-ResNet and F-PaNet modules to improve detection accuracy and efficiency. Synthetic and real datasets were fused to enhance model robustness in complex CCTV environments.

Nale, Pranav & Gite, Shilpa & Dharrao, Deepak in their research “Real-Time Weapons Detection System Using Computer Vision” [9] develop a hybrid system combining YOLOv7 and Detectron2, emphasizing low-light detection and reduced false positives for real-time applications using a custom-compiled dataset.

Deshpande, Deepali & Kadam, Devika et al. in their work “Next-Gen Security: YOLOv8 for Real-Time Weapon Detection” [10] explore YOLOv8-Small for military applications, demonstrating its adaptability across terrains and lighting conditions, and its utility in enhancing situational awareness.

Varma, Ravi Kiran et al. in their research “VGG-SSD Model for Weapon Detection Using Image Processing” [11] integrate VGG-16 with SSD to detect weapons in surveillance footage. The model achieves an mAP of 87% and recall of 86.6%, showcasing efficiency in automated surveillance.

Luo, Chen et al. in their study “Image Recognition and Processing Algorithm of Power Grid Facilities Map Based on Deep Learning” [12] apply an enhanced YOLOv5 algorithm for power grid inspections. The study demonstrates a transfer to weapon detection applications due to improvements in accuracy and robustness in complex settings.

Sumi, Lucy & Dey, Shouvik in their work “A Deep Learning Approach to Gun Detection in Surveillance System using YOLOv9” [13] introduce YOLOv9 for gun detection, achieving superior precision (99.6%) and recall (99%) compared to prior YOLO versions. Their approach highlights significant improvements in firearm detection through the integration of Generalized Efficient Layer Aggregation Networks (GELAN).

Devasenapathy, Deepa et al. in their study “Artificial Neural Network Using Image Processing for Digital Forensics Crime Scene Object Detection” [14] applied artificial neural networks for weapon detection, emphasizing the importance of Histogram of Oriented Gradients (HOG) for feature extraction. The study achieved significant accuracy improvements in detecting guns and knives in forensic applications.

Akshaya, Poloju et al. in their research “Automatic Weapon Detection Using Deep Learning” [15] utilized YOLOv8 for detecting pistols, shotguns, and knives in real-time surveillance footage. Their implementation on custom datasets demonstrates the model's efficiency in reducing false positives while maintaining high detection speed and accuracy.

Reddy, Rahul et al. in their work “Multiclass Weapon Detection Using Multi Contrast Convolutional Neural Networks and Faster Region-Based Convolutional Neural Networks” [16] explored MC-CNN and Faster R-CNN for detecting concealed handheld weapons. Their comparative analysis underscores the industrial applications of thesetechniques in enhancing public security.

III. Dataset Description

A. Overview

The Weapon Detection Dataset is a comprehensive collection of annotated images designed to facilitate the development and evaluation of object detection models specifically tailored for detecting weapons. With a total of 9692 annotated images, this dataset covers a wide range of scenarios and environments, providing diverse and realistic examples of weapons commonly encountered in security, law enforcement, and surveillance applications.

B. Dataset Composition

The dataset consists of images containing instances of various weapon classes, including Rifles, Grenades, Handguns, Knives, and Swords. Each image is meticulously annotated with bounding boxes, accurately delineating the location and extent of each weapon instance within the scene.

C. Dataset Split

The Weapon Detection Dataset has been meticulously divided into three distinct subsets: the training set, the validation set, and the test set. These subsets are carefully crafted to serve specific purposes in the development and evaluation of object detection models tailored for detecting weapons.

The training set comprises 6787 annotated images meticulously curated to provide a diverse and representative sample of weapon instances. These images are used to train the object detection model, to recognize various types of weapons, including automatic rifles, handguns, knives, and more. Through the iterative process of training, the models learn to identify weapons accurately, leveraging the rich visual information contained within the training images.

Consisting of 2062 annotated images, the validation set serves as a critical component in the model development pipeline. These images are held out from the training process and used to evaluate the performance of the trained models iteratively. By assessing the model's performance on unseen data from the validation set, researchers and developers can fine-tune model hyperparameters, mitigate overfitting, and ensure that the model generalizes well to new, unseen weapon instances.

The test set, comprising 843 images, represents the ultimate benchmark for assessing the efficacy and generalization ability of the trained object detection models. These images are entirely independent of both the training and validation sets, ensuring an unbiased evaluation of the model's performance. By subjecting the trained models to the test set, researchers and practitioners can obtain robust insights into the model's real-world applicability, its ability to detect weapons accurately across diverse scenarios and environments.

Class	***Instances***
Grenade	1416
Handgun	4923
Knife	2899
Rifle	3395
Sword	93

Table 1: Structure of the Dataset Used

IV. Proposed Architecture

YOLO (You Only Look Once) is recognized for its fast object detection, particularly valued for its ability to process images in real-time. Unlike other methods that analyse images segment by segment, YOLO examines the whole image simultaneously. It employs a CNN (Convolutional Neural Network) backbone to extract features from the image, segments the image into a grid, and within each grid cell, it predicts bounding boxes and class probabilities for objects in just one pass through the network. This method greatly accelerates the detection process, making it ideal for real-time applications, although it may sacrifice a bit of accuracy when compared to multi-stage detectors.

The methodology involves systematic data preprocessing to ensure input consistency, tailored training of each model to maximize performance, and the application of ensemble techniques to enhance detection accuracy and generalization. This structured framework is designed to harness the collective strengths of the models, providing a reliable and scalable solution for weapon detection.

YOLOv11 represents the latest advancement in the YOLO series, building on its predecessors' strengths while introducing cutting-edge innovations to further enhance object detection performance. It features a hybrid architecture combining convolutional neural networks (CNNs) with vision transformers (ViTs), enabling the model to efficiently extract both local and global features from an image. YOLOv11 incorporates advanced multi-scale feature fusion and dynamic attention mechanisms, allowing it to handle complex scenarios with overlapping objects and varying object scales. Its streamlined one-pass detection pipeline delivers exceptional speed and accuracy, making it ideal for real-time applications across diverse domains, including autonomous systems, surveillance, and underwater exploration. For this project, YOLOv11 is trained and fine-tuned on annotated sonar images to detect and classify objects across varying conditions, including noise and low visibility.

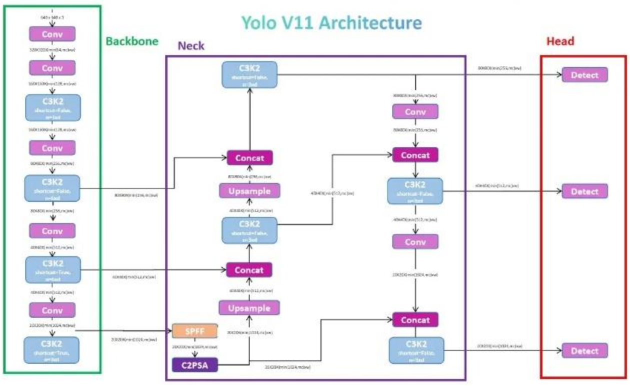

Fig. 1: YOLOv11 Model Architecture

Faster R-CNN combines precision with advanced region proposal capabilities, making it ideal for detecting objects with intricate boundaries in sonar images. The model operates in two stages: a Region Proposal Network (RPN) identifies areas likely to contain objects, and a CNN refines these proposals while classifying the objects. This approach allows Faster R-CNN to excel in detecting overlapping and partially occluded objects, which are common in underwater scenarios. The model's ability to balance accuracy and computational efficiency is utilized to enhance detection reliability in diverse sonar environments.

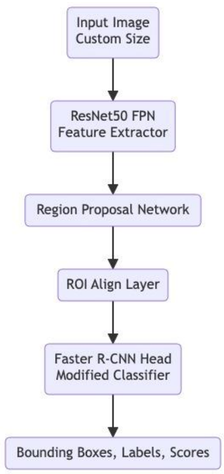

Fig. 2: Faster R-CNN Architecture

The ensemble model leverages the complementary strengths of YOLO and Faster R-CNN architectures to enhance object detection and localization. YOLO's real-time detection capabilities are combined with Faster R-CNN's precision in extracting detailed spatial relationships, enabling the model to perform well across diverse scenarios. By applying Weighted Box Fusion (WBF) to aggregate predictions from both architectures, the ensemble reduces individual model biases and improves robustness, especially in cases of overlapping detections and varying confidence levels. This hybrid approach is particularly effective in harmonizing results, achieving a balanced trade-off between speed and accuracy, and addressing challenges like class imbalances and visually similar object classes in complex environments.

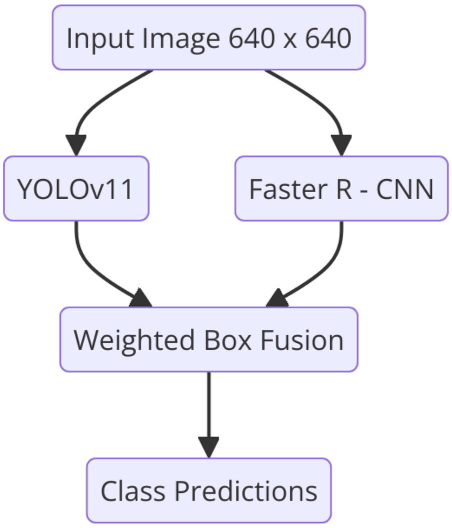

Fig. 3: Ensemble Model Architecture

V. Implementation and Results

The implementation phase of this project focuses on transforming the proposed methodology into actionable systems capable of real-time weapon detection and classification. Leveraging state-of-the-art deep learning models, such as YOLOv11, Faster R-CNN, and a YOLO-Faster R-CNN ensemble, the implementation combines individual model strengths to address the unique challenges of sonar-based marine monitoring. Each model is meticulously trained, optimized, and evaluated on a diverse dataset of annotated sonar images to achieve high accuracy and robust performance.

A. Setting up the Environment

For the practical implementation of our proposed solution, we selected Google Colab as the development environment due to its user-friendly and interactive nature, conducive to developing machine learning models. Google Colab offers a cloud-based platform that facilitates seamless experimentation and development without the need for extensive setup or infrastructure. Its integration with popular data visualization and manipulation libraries significantly streamlines the model development process.

B. Algorithms Used

1) YOLOv11

YOLOv11 is a state-of-the-art object detection model designed for real-time performance. For this project, YOLOv11 was trained on the sonar dataset using bounding box annotations to detect fish, debris, and background objects. The data was preprocessed using resizing (e.g., 640x640 resolution), normalization, and data augmentation techniques such as flipping and brightness adjustments to simulate various underwater conditions.

2) Faster R-CNN

Faster R-CNN is a two-stage detector combining a Region Proposal Network (RPN) with a CNN for object classification and bounding box regression. This model was fine-tuned using ResNet50 as the backbone for feature extraction and an ROI pooling layer to refine region proposals.

3) Ensemble Method

The ensemble model combines YOLO's real-time detection with Faster R-CNN's precise localization, enhancing accuracy and robustness. Using Weighted Box Fusion, it harmonizes predictions, reducing biases and improving performance on overlapping detections and challenging class distinctions, achieving a balance between speed and precision.

C. Results

The results generated from each algorithm were rigorously analysed and compared based on performance metrics such as Accuracy, Precision, Recall, F1 – Score, and Mean Average Precision (mAP). Through this comprehensive analysis, our study provides important insights on the use of computer vision techniques for marine life detection, emphasizing the role of data preprocessing, data augmentation and, use of various object detection algorithms in achieving robust predictive performance.

The performance evaluation highlights the strengths and trade-offs of the YOLOv11 and Faster R-CNN models. YOLOv11 demonstrated high precision at 0.779 and a strong mAP@50 of 0.691, indicating its effectiveness in correctly identifying objects with fewer false positives. However, its mAP@50-95 of 0.486 reflects challenges in handling detections across varying IoU thresholds. Conversely, Faster R-CNN excelled in recall at 0.675, showcasing its ability to detect a broader range of objects, with an F1 score of 0.619 balancing precision and recall. While its accuracy was 0.537, indicating room for improvement in overall prediction correctness, the complementary strengths of both models make them suitable for integration in an ensemble approach to maximize detection performance.

Given below are the predictions done by the models.

Fig. 4: Sample Prediction Image – YOLOv11

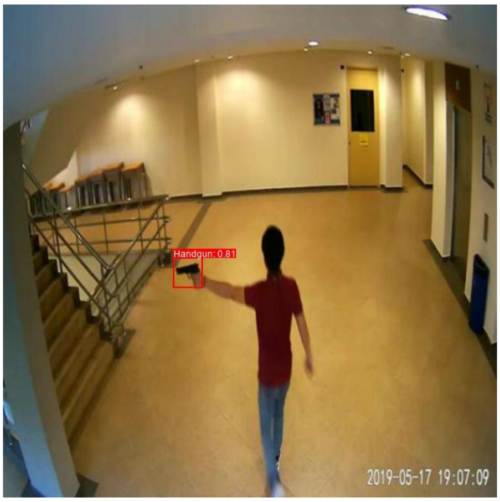

Fig. 5: Sample Prediction Image – Ensemble Model

Fig. 6: Sample Prediction Image – Faster R-CNN

VI. Conclusion And Future Work

In conclusion, this project demonstrated the strengths and limitations of several machine learning models—YOLOv11, Faster R-CNN, and a YOLOv11–Faster R – CNN Ensemble—in the context of weapon object detection and classification. YOLOv11 proved particularly valuable for real-time detection due to its speed, while Faster R-CNN excelled in precision and recall, making it ideal for offline applications that require highly accurate object localization. The ensemble approach successfully integrated the strengths of both models, achieving a balanced solution that improved accuracy and reduced inference time, making it well-suited for real-world weapon detection challenges. The findings highlight the importance of selecting and combining models based on application-specific needs to address critical trade-offs between speed and precision.

Future research could explore advanced techniques to enhance the detection system's performance and efficiency. Employing data augmentation strategies and Generative Adversarial Networks (GANs) to generate synthetic datasets may improve model robustness and generalizability. Integrating YOLOv11's real-time detection capabilities with the precision of ensemble approaches might offer a balance between speed and accuracy. Investigating adaptive ensemble methodologies could enable dynamic prioritization of model contributions based on detection scenario complexity or environmental factors. Additionally, applying optimization techniques such as pruning and quantization may facilitate faster real-time deployment with reduced computational demands.

References

[1] S. M. Keerthana, R. Sujitha, and P. Yazhini, "Weapon Detection for Security Using the YOLO Algorithm with Email Alert Notification," 2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET), Chennai, India, 2024, pp. 1–7.

[2] V. Jadhav, R. Deshmukh, P. Gupta, S. Ghogale, and M. Bodireddy, "Weapon Detection from Surveillance Footage in Real-Time Using Deep Learning," 7th International Conference On Computing, Communication, Control And Automation (ICCUBEA), Pune, India, 2023, pp. 1–8.

[3] R. Dugyala, M. V. V. Reddy, C. T. Reddy, and G. Vijendar, "Weapon Detection in Surveillance Videos Using YOLOv8 and PELSF-DCNN," E3S Web of Conferences, vol. 391, 2023, Art. no. 01071.

[4] S. Khalid, A. Waqar, H. U. A. Tahir, and I. T. Tenebe, "Weapon Detection System for Surveillance and Security," 2022 IEEE International Conference on IT Innovation and Knowledge Discovery (ITIKD), Islamabad, Pakistan, 2022, pp. 1–7.

[5] M. Pullakandam, K. Loya, P. Salota, R. M. R. Yaoarnala, and P. K. Javvaji, "Weapon Object Detection Using Quantized YOLOv8," 2023 5th International Conference on Energy, Power and Environment (ICEPE), Warangal, India, 2023, pp. 1–6.

[6] V. Jyothsna, C. Alle, R. Kurnutala, G. K. N., and K. K. R. Kushal, "YOLOv8-Based Person Detection, Distance Monitoring, Speech Alerts, and Weapon Identification with Email Notifications," 2024 International Conference on Expert Clouds and Applications (ICOECA), Tirupathi, India, 2024, pp. 1–8.

[7] P. Yadav, N. Gupta, and P. K. Sharma, "Robust Weapon Detection in Dark Environments Using YOLOv7-DarkVision," Digital Signal Processing, vol. 145, 2024, Art. no. 104342.

[8] G. Wang, H. Ding, and M. Duan, "Fighting Against Terrorism: A Real-Time CCTV Autonomous Weapons Detection System," Digital Signal Processing, vol. 132, 2023, Art. no. 103790.

[9] P. Nale, S. Gite, and D. Dharrao, "Real-Time Weapons Detection System Using Computer Vision," 2023 Third International Conference on Smart Technologies, Communication and Robotics (STCR), Pune, India, 2023, pp. 1–6.

[10] D. Deshpande, D. Kadam, and M. Jain, "Next-Gen Security: YOLOv8 for Real-Time Weapon Detection," 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud), Pune, India, 2023, pp. 1055–1061.

[11] R. K. Varma, S. G. N. V. G., and K. Raju, "VGG-SSD Model for Weapon Detection Using Image Processing," 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC-2024), Bhimavaram, India, 2024, pp. 890–897.

[12] C. Luo, Y. Liu, and L. Chen, "Image Recognition and Processing Algorithm of Power Grid Facilities Map Based on Deep Learning," 2024 Asia-Pacific Conference on Software Engineering, Social Network Analysis and Intelligent Computing (SSAIC), Guizhou, China, 2024, pp. 17–22.

[13] L. Sumi and S. Dey, "A Deep Learning Approach to Gun Detection in Surveillance System using YOLOv9," Control Instrumentation System Conference (CISCON), Nagaland, India, 2024, pp. 1–7.

[14] D. Devasenapathy, V. N, R. M, T. Jayasudha, R. K. Dwibedi, and D. Ganesh, "Artificial Neural Network Using Image Processing for Digital Forensics Crime Scene Object Detection," Second International Conference on Edge Computing and Applications (ICECAA 2023), Chennai, India, 2023, pp. 651–655.

[15] P. Akshaya, P. B. Reddy, P. Panuganti, P. Gurusai, and A. Subhahan, "Automatic Weapon Detection Using Deep Learning," International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence, and Telecommunication Engineering (RMKMATE), Narsapur, India, 2023, pp. 501–507.

[16] R. Reddy, G. K. Vallabh, and S. Sharan, "Multiclass Weapon Detection Using Multi Contrast Convolutional Neural Networks and Faster Region-Based Convolutional Neural Networks," 2nd International Conference for Emerging Technology (INCET), Belgaum, India, 2021, pp. 1–8.