Detection and Recognition of Handheld Armaments based on video surveillance using Deep Learning Technique

Neha Gupta[1,2,a)] and Bharat Bhushan Agarwal[3,b)]

[1]*Computer Science and Engineering Department, IFTM University, Moradabad, India*

[2]*Computer Science and Engineering Department, Moradabad Institute of Technology, Moradabad, India*

[3]*Computer Science and Engineering Department, School of Computer Science and Applications, IFTM University, Moradabad, India*

Corresponding author: [a)] *discoverneha@gmail.com*
[b)] *bharat_agarwal@iftmuniversity.ac.in*

Abstract: This research tackles a crucial component of security in the modern world by introducing a revolutionary weapon detection method based on YOLOv8. The surveillance footage that can either be recorded for live viewing while the activity is happening or replayed after it has been captured and tested on the YOLOv8 model. Every aspect of life is impacted by the current trend toward "automation," and video analytics is no exception. To guarantee system reliability and lower the possibility of needless alarms, a strong emphasis on high accuracy with few false positives is focussed. It is significant that the focus on helping to create intelligent security solutions is in line with the continuous endeavours to utilize cutting-edge technologies for the benefit of society. Encouraging safer communities is a common objective, and this research helps to meet the pressing demand for reliable threat detection systems in the dynamic security environment of today. Emphasizing the system's adaptability to different security needs is a way to highlight its versatility in deployment in a variety of environments, including stadiums, schools, airports, and urban centres. Preventing possible security breaches and guaranteeing people's safety in public areas require proactive security measures.

Keywords: YOLOv8, RCNN,FPN,mAP, cls loss, dfl loss.

INTRODUCTION

To protect public areas in a time when security worries are on the rise, it is critical to build effective and precise weapon detection systems. Modern threats can often be defeated by traditional security measures, which emphasizes the necessity for cutting-edge technology like computer vision and deep learning[1]. Using YOLOv8 (You Only Look Once version 8), a cutting-edge object identification model well-known for its excellent accuracy and real-time performance, this research presents a novel approach to weapon detection.

The main goal of this research is to create a reliable weapon detection system that can quickly and reliably identify several kinds of weapons, such as knives, grenades, pistols, rifles, and missiles, in a variety of environmental settings. In order to do this, we have considered a number of deep learning models based on object detection[2][6], such as You Only Look Once (YOLO), Mask Region-Based Convolutional Neural Networks, Region-Based Convolutional Neural Networks (R-CNN), and Faster Region-Based Convolutional Neural Networks (Faster R-CNN). Ultimately, we have decided to use YOLOv8 based on the project's requirements for accuracy and prediction time. The issue of the prompt and precise detection of firearms in crowded situations is supported by YOLOv8, a faster method with good detection accuracy.

A weapon detection system is a piece of equipment intended to detect the presence of firearms or other potentially harmful weapons in a specific area and notify law enforcement or security officials. To improve security and safety measures, these systems are frequently utilized in a variety of settings, including airports, public transportation hubs,

schools, and other public areas. The following are some typical methods and technology found in weapon detecting systems:

Metal Detectors: To identify metal objects, such as knives and firearms, traditional metal detectors are frequently utilized. Airports, governmental structures, and high-security locations frequently use them.

X-ray Scanners: These devices are used to provide finely detailed images of items that are contained within packages or bags. They are frequently employed to identify hidden threats at security checkpoints in airports and other sensitive areas.
Acoustic Gunshot Detection: A few technologies employ microphones to pick up gunshot sounds. These devices assist security professionals in taking swift action by locating the firing by analysing the acoustic signature.

Video Analytics: To analyse video feeds in real-time, sophisticated video surveillance systems can integrate machine learning and artificial intelligence algorithms. Based solely on visual cues, these systems can recognize weapons and notify security staff of their presence.
Intelligent Sensors: Intelligent sensors, such as Internet of Things (IoT) devices, can be used to track and evaluate environmental data. Sensors could, for instance, identify the existence of substances linked to explosives.

It is noteworthy that the efficiency of weapon detection systems is contingent upon a few elements, such as the technology employed, the surroundings, and the operator's level of experience. When putting such systems in place in public areas, privacy and ethical issues also need to be taken into mind.

By utilizing artificial intelligence (AI) methods to evaluate visual data, such as photos or video feeds, weapon detection through deep learning algorithms entails determining the existence of weapons. Applications involving the identification of weapons can benefit from deep learning, a branch of machine learning that has shown promise in image recognition tests.

SYSTEM OVERVIEW

The main goal of this research is to create a reliable weapon detection system that can quickly and reliably identify several kinds of weapons, such as knives, grenades, pistols, rifles, and missiles, in a variety of environmental settings. In order to do this, we have considered a number of deep learning models based on object detection [2][6], such as You Only Look Once (YOLO), RNN,LSTM, Cubic SVM [9], Mask Region-Based Convolutional Neural Networks [10], Region-Based Convolutional Neural Networks (R-CNN), and Faster Region-Based Convolutional Neural Networks (Faster R-CNN) [11]. Ultimately, we have decided to use YOLOv8 based on the project's requirements for accuracy and prediction time. The issue of the prompt and precise detection of firearms in crowded situations is supported by YOLOv8, a faster method with good detection accuracy. The fig.1 shows the existing solution approaches to solve the similar category task by using other algorithms and models.

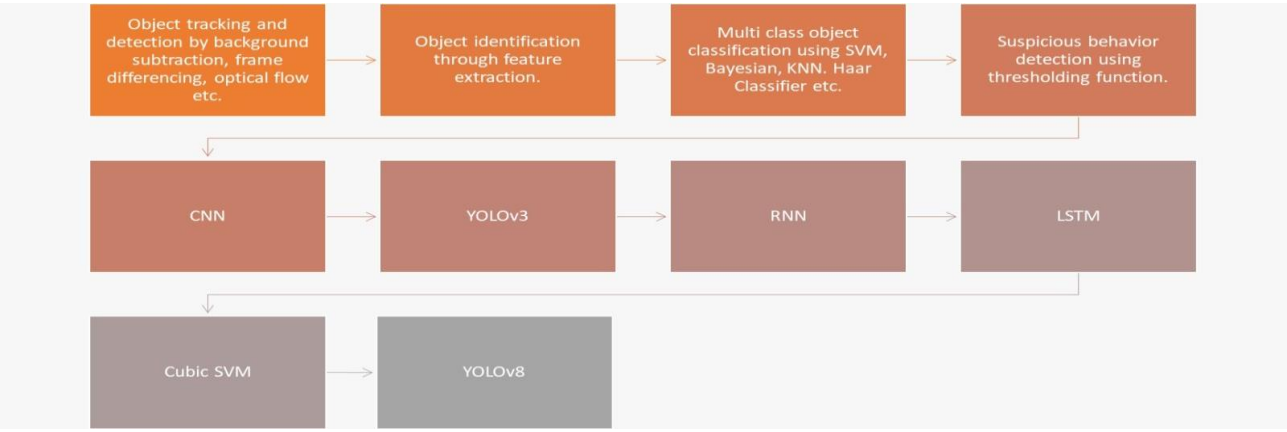

FIGURE 1: Existing Solution Approaches

We are introducing a Real-time Weapon Detection System that uses an Internet Protocol (IP) address as input to host the video stream array and detect and recognize different types of weapons, including grenades, missiles, pistols, rifles, and knives, in a live video broadcast stream. The system's goal is to increase security. Using the weapons dataset, the YOLOv8 model is refined by the system through the use of a You Only Look Once (YOLO) technique for object recognition. The system promptly notifies the registered staff of the potential threat to security when a weapon with a frame of the object and a security alert is discovered.

YOLOv8's design smoothly incorporates a deep CNN to guarantee effective and precise object recognition. YOLOv8, which stands out for its simple design, uses a single neural network to immediately infer class probabilities and bounding boxes from input images, allowing for real-time processing. In this design, predictions are refined by detecting heads functioning at several scales, while feature extraction is handled by the backbone network. By utilizing innovations such as feature pyramid networks (FPN) and optimization approaches, YOLOv8 achieves an optimal speed-precision ratio, making it a top choice for a variety of applications, most notably weapon detection. Because of its flexible architecture, threat detection capabilities are significantly improved and it can be easily integrated into security frameworks and surveillance systems.

The effectiveness of the suggested weapon detecting system is evaluated through extensive testing. The assessment process entails a meticulous analysis of various reference datasets and actual surveillance video obtained from a range of settings. Metrics like as precision, recall, and mean average precision (mAP) are carefully used to assess system robustness and detection accuracy. Furthermore, a qualitative analysis is carried out to examine how well the system handles intricate scenarios such as occlusions, crowded backdrops, and different viewing angles. The system's performance under demanding situations is carefully examined through methodical experimentation and comprehensive assessment, guaranteeing its fitness for confident real-world deployment.

An outline of the proposed approach that can be used to detect weapons is provided in fig.2 below:

FIGURE 2: Workflow of Proposed Approach

Data Retrieval: Compiling a varied and well-labeled dataset is the initial stage in creating a deep learning-based weapon detection system. Images or video frames showing instances of weapons and non-weapon objects in a range of settings and circumstances should be included in this dataset.
Custom Dataset Preparation: To test the model, custom videos are recorded through holding guns of various types, knife and other category objects were shown through google images on phone to test the model.

Model Architecture: Using Google Collab, we ensured the computational power of CUDA-enabled devices and imported the YOLOv8 model using the Ultralytics package [5]. We then used the model's command line training mode to fine-tune, starting with the YOLOv8X.pt weights.

Model Training: Make use of the gathered dataset to train the deep learning model. The model picks up patterns and characteristics linked to weapons throughout training. To reduce the discrepancy between the model's predictions and the ground truth labels, the parameters of the model are adjusted during the training phase.

Object Detection and Recognition: Integrate a framework for object detection into the deep learning model to find and identify things in pictures or videos that are of interest, like weapons. Typical object detection frameworks are Faster R-CNN, You Only Look Once (YOLO), and Single Shot Multibox Detector (SSD).

IMPLEMENTATION DETAILS

Software developer Ultralytics is well-known for creating solutions for deep learning and computer vision, especially when it comes to object identification and picture segmentation. A platform called Roboflow makes it easier to train and use computer vision models. It enables users to organize and prepare their image collections for a range of computer vision applications, including classification, object recognition, and image segmentation.
The dataset is fetched is from roboflow dataset API. The You Only Look Once - Object Bounding Box (YOLO-OBB) format is used in this collection.There are 14,981 annotations in all, with an average of 1.6 per image spread across 5 classes [0: Pistol, 1: Rifle, 2: Grenade, 3: Missile, 4: Knife]. The dataset contains 24 Null samples and 0 missing annotations. A picture's typical size is 0.41 million pixels (mp) on average[3]. The YOLOv8 model was fine-tuned using the following dataset with the classification, as shown in fig.3and 4 and the outcome is what we obtained in result section.

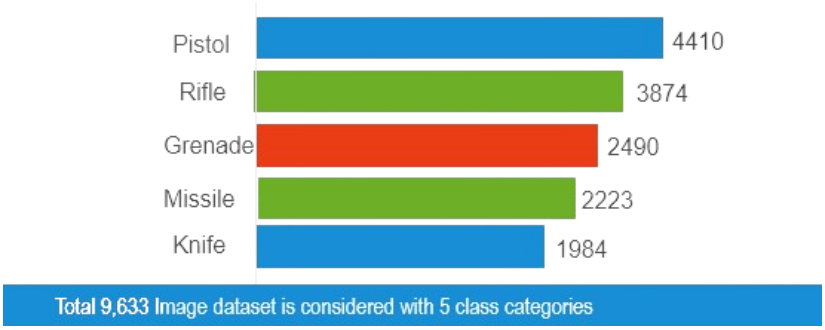

FIGURE 3: Dataset Description

FIGURE 4: Training dataset used

We imported the YOLOv8 model using the Ultralytics package [5], making use of its command line training mode and fine-tuning starting with the YOLOv8X.pt weights. We started the training with the mode set to "train" using Google Collab, making sure to utilize the processing power of CUDA-enabled devices. Divided into intervals of 20, 30, and 40 epochs, the fine-tuning procedure lasted 400 epochs in total. We repeatedly used the "last.pt" weights file from the previous training session to help with continuity. By taking a methodical approach, we were able to gradually

improve the model's performance and move toward lower false detection rates and higher detection rates. Both detection accuracy and confidence are increased through fine-tuning.
Users can enter the internet protocol (IP) address of a camera module that feeds live video broadcasts using the user interface, which displays a front-end view. The data is accessed through this Internet Protocol (IP) address. A "Start" button on the UI also initiates the complete object detection procedure on the incoming video feed. The back-end module processes the video stream in the background, analyzing and identifying objects in the live feeds by running object detection algorithms.
After receiving the input Internet Protocol address, the back-end integration to the front-end uses the open-cv script to start receiving the continuous live feed. Next, it loads the Best.pt file for YOLOv8 best set weights and uses it for object detection, which identifies objects per video frame. Finally, it uses the Python SMTP module to return the object detected frame and an alert signal to the authorized email address.

We have incorporated the sequential development processes of testing and debugging into our project, which was built utilizing an agile development methodology. The effort has been built upon the continuous creation and application of the various datasets in conjunction with customized training possibilities. To evaluate the detection accuracy, three validation checks were carried out: image-based detection, video-based detection, and, in the last step, live video stream-based detection.

After the video stream device has been placed at the desired position, the project deployment entails connecting the surveillance equipment and the device over WiFi. uploading the "Best.pt" file, which contains the best training weights, setting up the project's environment, providing the IP address of the video stream device, changing the authority emails to the necessary email addresses, and finally doing the detection.

PERFORMANCE METRICS

The YOLOv8 model was fine-tuned using the dataset as shown in fig. , with varying the number of epochs, the results were measured in the form of confusion matrix, box loss, cls loss and dfl loss.

Predicting precise bounding boxes is a crucial component of object detection, which is the process of identifying and categorizing items within an image. The "box_loss" most likely refers to the loss function used in the regression task of predicting bounding box coordinates, which are commonly expressed as (x, y, w, h).For measuring the box loss, a model detects objects in an image and its goal is to forecast the bounding box's dimensions and coordinates, which encloses each object firmly. After then, the ground truth bounding box—which is marked in the training data—and the anticipated bounding box are contrasted. The model is more accurate at localizing items in the picture when the "box_loss" is smaller. It is noteworthy that "box_loss" is frequently merged with other elements, including objectness loss and classification loss, in the overall loss function of object detection models. In order to enhance the model's overall performance in object localization and classification tasks, the training method entails minimizing this combined loss.

"cls loss" refers to the classification loss, one of the elements of the overall loss function that is employed in the object detection model training process. Object detection includes both the localization of objects using bounding boxes and their categorization. The model's ability to correctly anticipate the class label for each bounding box that is identified is gauged by the classification loss. Typically, YOLOv8 bases its class predictions on a softmax activation function. The class scores are normalized by the softmax function, which converts them into probabilities.
A variety of components, including localization loss (coord loss), confidence loss (obj loss), and classification loss (cls loss), make up the total loss function employed in YOLOv8. By modifying its parameters, the model is trained to minimize this combined loss and enhance object localization and class prediction accuracy.

The purpose of distributional focal loss (dfl) is to mitigate the issue of class imbalance that arises during object detection training. The model is able to concentrate more on difficult, incorrectly categorized examples by introducing a modifying component that down weights the loss attributed to well-classified examples.

Some other typical assessment metrics that are employed to evaluate the model's performance are precision, recall and mAP.

Precision gauges how well the model predicts the good outcomes. It is the proportion of all positively anticipated instances to all accurately predicted instances as shown in equation-1. A high precision means that there are fewer false positive predictions being made by the model.

$$\text{Precision} = \frac{TP}{(TP + FP)}$$ .....Eq. 1

Recall gauges how well the model can identify and include all pertinent cases. It is the proportion of all real positive instances to all correctly projected positive instances as shown in equation-2. A high recall means that a significant proportion of real positive cases are being captured by the model.

$$\text{Recall} = \frac{TP}{(TP + FN)}$$ .....Eq. 2

A composite statistic called mAP takes precision and recall into account at various confidence levels. It is frequently used to assess a model's performance over a range of precision-recall trade-offs in object detection tasks. Each class's Average Precision (AP) is determined, and the mean of all the classes is then computed to obtain the mean precision (mAP). Better overall performance is indicated by higher mAP levels.

The precision and recall at a specified intersection over union (IoU) criterion of 50% are assessed using a particular statistic called mAP 50. IoU calculates the overlap between the bounding boxes that are anticipated and those that are based on ground truth.
The area under the precision-recall curve is computed to yield mAP 50, and the curve is generated at the 50% IoU threshold.
A more thorough statistic called mAP 50-95 takes precision and recall into account over a range of IoU thresholds, usually from 50% to 95% in increments of 5%. This offers a more comprehensive comprehension of the model's functionality at various IoU levels.
By averaging the AP (Average Precision) values acquired at each IoU threshold, the mAP 50-95 is determined. The effectiveness of a model at various bounding box overlap levels can be evaluated using both metrics. While mAP 50 focuses exclusively on the 50% IoU threshold, mAP 50-95 provides a more thorough examination of performance across a variety of IoU thresholds

TESTING AND RESULT ANALYSIS

The surveillance footage that may be viewed live while the activity is being recorded, or it can be played back after it has been captured. The contemporary trend toward "automation" has an impact on all aspects of life, including video analytics.

Two distinct situations are used to test the build model: offline video and online video stream via webcam.A model's robustness and efficacy can be ensured by testing it under various scenarios. Below is a quick synopsis of the two test cases:

1. Video Testing Offline:
This scenario involved testing the weapon detection model using offline or pre-recorded video material as shown in fig.5. Following benefits were attached in this scenario:

- Since the video data is pre-recorded, regulated testing settings are made possible.
- Permits recurrent testing on the same dataset for evaluation that is consistent.
- Helpful in evaluating how well the model performs on particular video scenes.

FIGURE 5: Representative samples of the simulated dataset used in our experiment.

2. Feed Testing Online:
This scenario involved testing the weapon detection model on a real-time or live video feed. Following benefits were attached in this scenario:

- Reflects how well the model performs in dynamic, real-world circumstances.
- Mimics the circumstances that the model is expected to face during implementation.
- Provide information about how well the model can handle changing and continuous input.

One of the best-known features of YOLOv8 (You Only Look Once version 8) is its real-time object identification. It immediately predicts bounding boxes and class probabilities after dividing the input image into a grid. YOLOv8 works well in situations requiring prompt and precise detection as shown in fig 7.

A synthetic dataset is used in the experimental evaluation of our suggested methodology for offline video testing. The artificial dataset is intended to mimic actual situations where one or more weapons are detected in a video. The graphic displays a few exemplary samples from the synthetic dataset with 75 percent confidence setting as shown in fig 6.

FIGURE 6: Detection and Recognition performed by the model in offline video testing

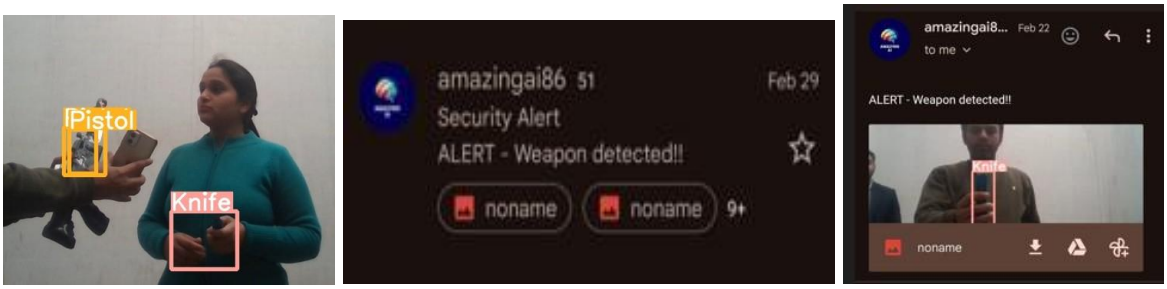

FIGURE 7: Detection ,Recognition and Alert performed by the model during online feed testing

In both testing scenarios, assess the model's mAP (mean average precision), accuracy, precision, and recall. Take into account any difficulties or restrictions encountered while testing, such as the lighting, the occlusion of objects, or the camera angles. To enhance the model's performance in various settings, make adjustments based on the testing results. Maintaining the weapon detection system's dependability and efficacy requires ongoing testing and improvement, particularly when it is used in a variety of contexts.

After completion of 790 epoch following performance metrics were obtained as show in fig 8 and 9.

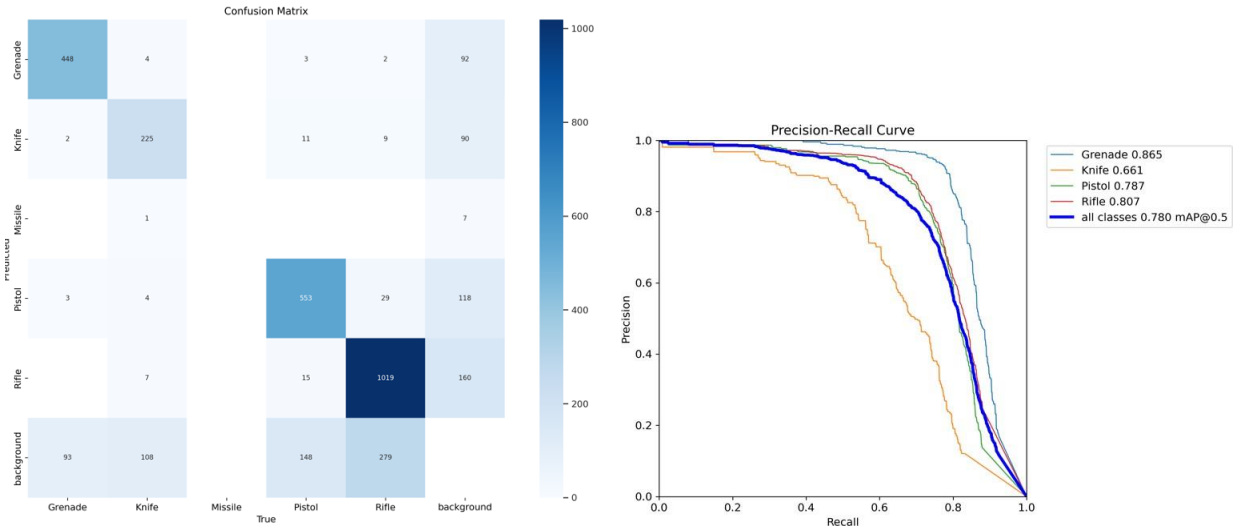

FIGURE 8: Confusion Matrix and PR Curve

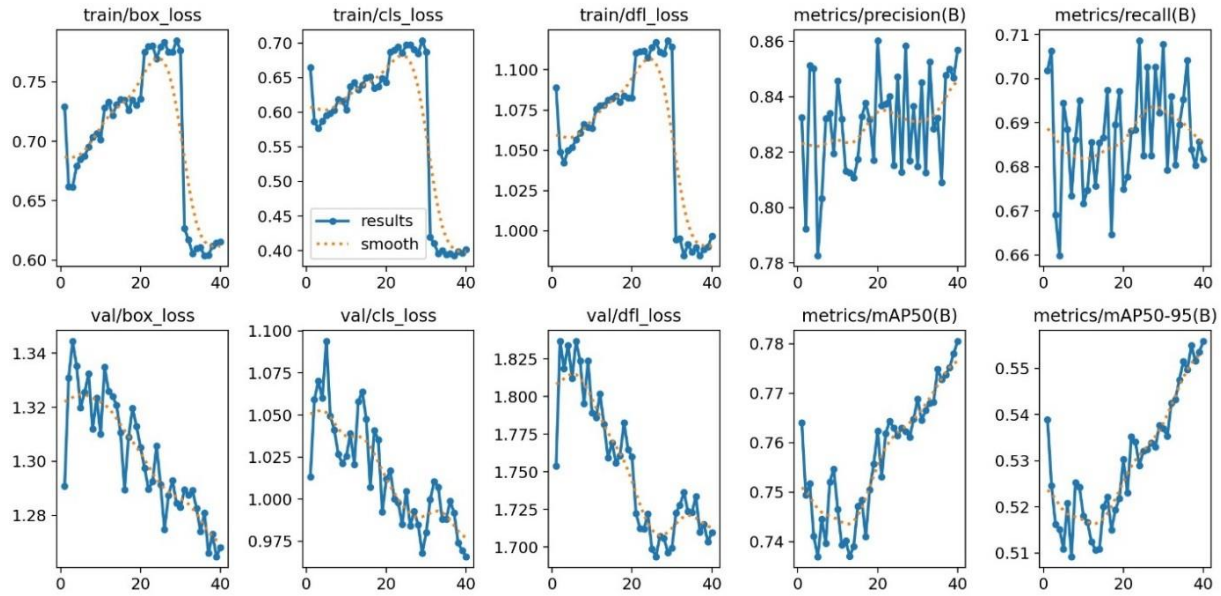

FIGURE 9: Resulting graphs against box loss, class loss and dfl loss

CONCLUSION

The suggested weapon detection method may have more impact and legitimacy if it includes specific difficulties, technical information, or validation findings. A thorough grasp of the system's implementation would also benefit from taking privacy issues, ethical issues, and other system constraints into account in various scenarios. For integration with Surveillance Systems, either use the trained deep learning model to integrate with already-in-use surveillance systems or use it to build a weapon detection system on its own. When weapons are spotted, the system can assess real-time video feeds, examine individual frames, and sound an alarm or sent a message to notify registered

user. It is recommended to incorporate systems for ongoing adaptation and improvement. To maintain the model current and useful, add new data on a regular basis. Consider things like evolving threats, potential false positives, and changes in the appearance of weaponry. All things considered, it looks like your research makes a substantial contribution to the subject of security and has the potential to greatly improve threat detection capabilities, which will ultimately help to create safer settings.

FUTURE SCOPE

The model's performance in real-world scenarios can be improved by fine-tuning it using specific data that is relevant to the deployment context. In this step, the model gets assistance in adjusting to the features and changes found in the target environment. When implementing weapon detection technologies, be sure to adhere to ethical and regulatory requirements as well as privacy guidelines. Furthermore, test the system extensively and make sure it's ready for real-world situations. Deploying deep learning-based weapon detection systems requires careful consideration of privacy issues, ethical issues, and legal ramifications. To guarantee the system's dependability and reduce false positives and negatives, extensive testing and assessment should also be carried out. Sustaining the system's efficacy over time requires frequent updates and continuous monitoring. Weapon detection technology can be applied to IP-based detection CCTV cameras in the future that can handle input feeds from multiple video sources simultaneously. Once the models have demonstrated their effectiveness, the approach can be further expanded by incorporating later iterations of the Yolo models (e.g., YOLOv9)[7].To increase the support for commodity hardware, more input must be accepted, and a distinct training environment[8] with graphics different from those in the corresponding video environment must be created.

REFERENCES

1. M. Monika, Udutha Rajender, A. Tamizhselvi, Aniruddha S Rumale, “Real-time object detection in videos using deep learning models”, ICTACT Journal on image and video processing, 2023, Volume. 14
2. Chang Ho Kang, Sun Young Kim, “Real-time object detection and segmentation technology: an analysis of the YOLO algorithm”, JMST Adv. 5,69-76, 2023
3. Roboflow Universe, weapon detection Computer Vision Project, https://universe.roboflow.com/test-7awfy/weapon-detection-f1lih/
4. Roboflow Universe, Explore the Roboflow Universe, https://universe.roboflow.com/
5. Github, ultralytics, Ultralytics, https://github.com/ultralytics/ultralytics
6. Delong Qu, Weijun Tan, Zhufu Liu, Q Yao, Jingfeng Liu, “A Dataset and system for real-time gun detection in surveillance video using deep learning“ in arXiv preprint arXiv: 2105.01058, 2021
7. Chien-Yao Wang, I-Hau Yeh, Hong-Yuan Mark Liao, “YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information” in arXiv preprint arXiv: 2402.13616, 2024
8. Yuan Zhu, Yanqiang Wang, Yadon An, Hong Yang, Yiming Pan, “ Real-Time Vehicle Detection and Urban Traffic Behaviour Analysis Based on UAV Traffic Videos on Mobile Devices” in arXiv preprint arXiv: 2105.01058, 2024
9. N. Bordoloi, A. K. Talukdar, and K. K. Sarma, "Suspicious Activity Detection from Videos using YOLOv3," in 17th India Council International Conference, New Delhi, India, Dec. 2020, pp. 1–5
10. K. K. Verma, B. M. Singh, and A. Dixit, "A review of supervised and unsupervised machine learning techniques for suspicious behavior recognition in intelligent surveillance system," International Journal of Information Technology, vol. 14, no. 1, pp. 397–410, Feb. 2022, https://doi.org/10.1007/s41870-019-00364-0.
11. S. Nandyal and S. Angadi, "Recognition of Suspicious Human Activities using KLT and Kalman Filter for ATM Surveillance System," in International Conference on Innovative Practices in Technology and Management, Noida, India, Feb. 2021, pp. 174–179, https://doi.org/ 10.1109/ICIPTM52218.2021.9388322.