International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 2
***Article Received:** 05 January 2021 **Revised:** 30 January 2021 **Accepted:** 25 February 2021*

Firearm Detection Using YOLOv8 Deep Learning Model

Sugavaneshwari P[1], Priadarsini M[2], Narmatha R[3]

[1] sugavaneshwari@gcedpi.edu.in, [2] priadarsini.m@gcesaelm.edu.in, [3] narmatharaj@gcedpi.edu.in

[1,3] Department of Computer Science and Engineering, Government College of Engineering, Dharmapuri- 636704, India.

[2] Department of Computer Science and Engineering, Government College of Engineering, Salem- 636011, India.

ABSTRACT

Firearm detection in surveillance imagery is an important computer vision task for automated security monitoring. This study presents a deep learning-based firearm detection system using the YOLOv8 (You Only Look Once, version 8) object detection framework. A dataset of 1,600 firearm images was prepared and annotated to train and evaluate the detection model under varying visual conditions, including changes in illumination, viewing angle, and partial occlusion. The proposed system performs real-time firearm detection from images and video streams and identifies the location of detected firearms using bounding boxes. Post-processing is employed to improve the interpretation of detection results and support alert generation. The model is designed with real-time deployment in mind and can be adapted for surveillance applications where computational efficiency is required. Experimental evaluation on the prepared dataset achieved an mAP@0.50 of 99% indicating effective firearm recognition under the evaluated conditions. The results demonstrate the potential of YOLOv8-based object detection for automated firearm detection in surveillance environments.

Keywords: Firearm Detection, YOLOv8, Deep Learning, Object Detection, Computer Vision, Surveillance, Real-Time Detection.

I. INTRODUCTION

Firearm-related incidents present significant challenges for public safety and surveillance systems. Conventional CCTV systems depend heavily on continuous human monitoring, which can become difficult when operators are required to observe multiple video streams simultaneously. Automated object detection can assist surveillance personnel by identifying potentially dangerous objects and directing attention to relevant regions of an image or video frame.

Computer vision techniques have been increasingly applied to automated firearm and weapon detection. Early approaches used handcrafted visual features and conventional image-processing techniques for identifying firearms. Tiwari and Verma proposed a computer-vision-based framework for visual gun detection using the Harris interest point detector [1]. In another study, Tiwari and Verma investigated SURF-based features for visual gun detection [2]. These approaches demonstrated the applicability of image features for firearm recognition but relied on manually designed feature representations.

Concealed weapon detection has also been investigated using imaging technologies. Sheen et al. proposed three-dimensional millimeter-wave imaging for concealed weapon detection [3]. Xue et al. investigated the fusion of infrared and visual images for concealed weapon detection [4], while Blum et al. proposed a multisensor concealed weapon detection approach using a multiresolution mosaic technique [5]. Upadhyay and Rana further investigated exposure fusion for concealed weapon detection [6]. These approaches demonstrate the importance of combining suitable image representations with detection techniques for identifying weapons under challenging imaging conditions.

With the development of deep learning, convolutional neural networks and object-detection methods have provided an alternative to handcrafted feature-based approaches. Deep learning models can automatically learn discriminative visual representations from training data and can perform object localization and classification within a unified framework. He et al. demonstrated the effectiveness of deep residual networks for visual recognition tasks [7], while Tan et al. presented EfficientDet as a scalable object-detection

***Article Received:** 05 January 2021 **Revised:** 30 January 2021 **Accepted:** 25 February 2021*

framework that considers both detection accuracy and computational efficiency [8].

Object-detection performance is commonly evaluated using metrics such as precision, recall, F1-score, and mean average precision. Padilla et al. provided a systematic review of performance metrics used for object-detection algorithms and highlighted the importance of selecting appropriate evaluation measures for reliable comparison [9].

In this study, YOLOv8 is evaluated for firearm detection using an annotated dataset containing 1,600 images. The objective is to investigate the capability of the trained model to identify and localize firearms under variations in orientation, scale, illumination, background, and partial occlusion. The dataset is divided into training, validation, and testing subsets, and the model performance is analyzed using detection results, confusion matrices, precision-confidence curves, recall curves, and mean average precision.

The main contributions of this study are as follows:

1. A dataset containing 1,600 firearm images was prepared and annotated with bounding boxes.
2. The dataset was divided into training, validation, and testing subsets using a 60:20:20 split.
3. A YOLOv8-based firearm detection model was trained using the annotated dataset.
4. The detection performance was evaluated using precision, mAP, confusion-matrix analysis, and confidence-based evaluation curves.
5. The experimental results and limitations of the trained model are discussed in the context of automated firearm detection.

II. RELATED WORK

Early firearm-detection systems primarily relied on handcrafted visual features. Tiwari and Verma investigated gun detection using Harris interest points and FREAK descriptors [1], while another study used SURF-based features for visual gun detection [2]. Grega et al. developed an automated firearm and knife detection approach for CCTV images, demonstrating the feasibility of detecting weapons directly from surveillance imagery [3]. Concealed weapon detection has also been investigated using millimeter-wave imaging and multimodal image fusion techniques [4–7].

The development of deep learning enabled automatic extraction of visual features for object detection. Girshick et al. introduced R-CNN for object detection using CNN-based features and region proposals [8], followed by Fast R-CNN [9] and Faster R-CNN [10], which improved detection efficiency. Redmon et al. introduced YOLO as a single-stage object detector designed for real-time detection [11], followed by YOLO9000 with improvements in speed and detection performance [12]. SSD provided another single-stage detection framework using predictions from multiple feature-map scales [13].

Multi-scale feature representation has become particularly important for detecting small objects. Feature Pyramid Networks (FPN) improved object detection through multi-scale feature representations [14]. ResNet provided deeper feature extraction through residual learning [15], while MobileNet introduced a lightweight architecture suitable for computationally constrained applications [16]. RetinaNet addressed foreground-background imbalance using focal loss [17]. EfficientDet subsequently combined efficient feature extraction and multi-scale feature fusion to improve the accuracy-efficiency trade-off [18].

Deep-learning approaches have been specifically applied to firearm detection. Olmos et al. proposed automatic handgun detection in video using deep learning and Faster R-CNN [19]. A subsequent study investigated binocular image fusion to reduce false positives in handgun detection [20]. Singleton et al. investigated handgun identification using TensorFlow and MobileNet for real-time surveillance [21]. Lim et al. evaluated deep neural networks for gun detection in surveillance videos using a large CCTV dataset [22].

Other studies have addressed specific challenges in firearm detection. Egiazarov et al. investigated firearm detection and segmentation using an ensemble of semantic neural networks [23]. Vallez et al. studied deep-autoencoder-based false-positive reduction for handgun detection [24]. Salazar-González et al. investigated real-time gun detection in CCTV and highlighted small-object detection as an important challenge [25]. Brightness-guided preprocessing has also been investigated for weapon detection under varying illumination conditions [26].

The literature indicates that firearm detection remains challenging because firearms can occupy a small portion of surveillance frames and may be affected by occlusion, illumination changes, background

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 2
***Article Received:** 05 January 2021 **Revised:** 30 January 2021 **Accepted:** 25 February 2021*

complexity, and image quality. False positives and computational efficiency are also important considerations for real-time surveillance systems. Object-detection evaluation should therefore consider multiple metrics rather than a single accuracy value. Padilla et al. reviewed commonly used object-detection metrics, including precision, recall, F1-score, and mean average precision [26].

Based on these observations, the present study evaluates a YOLOv8-based firearm detection model using a dataset of 1,600 annotated images. The dataset is divided into training, validation, and testing subsets, and the model is evaluated using precision, mean average precision, confusion-matrix analysis, and confidence-based precision and recall curves. The study focuses on evaluating the model on the prepared dataset and examining its applicability to automated firearm detection.

III.METHODOLOGY

Deep learning is at the core of this firearm detection system, enabling quick and accurate identification of weapons in real-time. It is a specialized branch of machine learning that uses artificial neural networks to analyze complex visual patterns and data. This project specifically utilizes YOLOv8 (You Only Look Once, version 8), a highly optimized deep learning model designed for fast and efficient object detection. YOLOv8 relies on convolutional neural networks (CNNs) to process images and video frames, allowing it to detect firearms with exceptional speed and precision.

To achieve reliable detection, the system undergoes rigorous training using a diverse dataset of firearm images, covering variations in size, orientation, lighting, and occlusions. During this process, YOLOv8's deep neural network progressively learns essential features, beginning with basic visual elements such as edges and textures, and advancing to more complex characteristics unique to firearms. The model divides images into grids, predicting bounding boxes and class probabilities for each grid cell, enabling the system to detect and localize firearms efficiently within a single pass.

This approach ensures real-time performance, making it a valuable asset for security surveillance, law enforcement, and public safety applications. By incorporating AI-powered detection into CCTV cameras, drones, and monitoring systems, this project enhances security infrastructure by reducing response times, preventing threats, and improving emergency preparedness. YOLOv8's ability to process visual data swiftly and accurately makes it a powerful tool for modern security systems, allowing authorities to respond effectively to potential firearm-related incidents before they escalate.

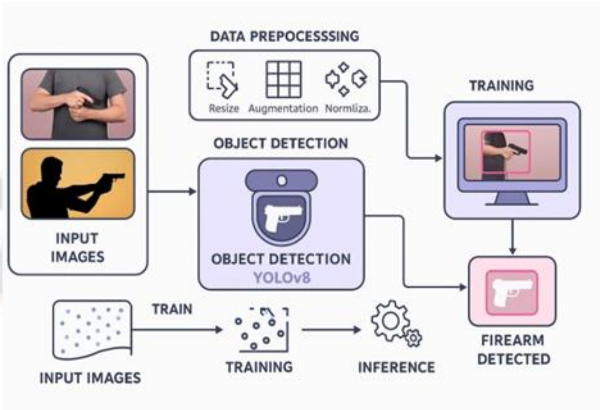


Figure 1. Proposed model

1.DATASET COLLECTION

The dataset collection process was performed to provide sufficient visual variation for training the firearm detection model. A total of 1,600 images were collected from publicly available firearm images, CCTV/video sources, Internet-based repositories, and manually captured images. The collected images contain firearms under different orientations, scales, backgrounds, and illumination conditions.

The dataset includes different firearm categories, including pistols, rifles, and shotguns. Each image was manually annotated using bounding boxes to identify the location of firearm instances. The annotations contain the class identifier and the corresponding bounding-box coordinates in a format compatible with the YOLO object-detection framework.

The complete dataset was divided into training, validation, and testing subsets. A total of 960 images were used for training, 320 images for validation, and 320 images for testing. Thus, the dataset distribution was 60% training, 20% validation, and 20% testing.

Table 1. Dataset Distribution

Dataset	Number of Images	Percentage
Training	960	60%
Validation	320	20%
Testing	320	20%
Total	**1600**	**100%**

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 2
Article Received: 05 January 2021 **Revised:** 30 January 2021 **Accepted:** 25 February 2021

Figure2.Training images

2. PREPROCESSING

Data preprocessing was performed before training to make the collected images suitable for object detection. The images obtained from different sources varied in resolution, dimensions, lighting conditions, and background characteristics. The images were therefore prepared according to the input requirements of the YOLOv8 training pipeline. Data augmentation was applied to increase the visual variability of the training samples. Augmentation operations included transformations such as flipping, rotation, scaling, brightness adjustment, and contrast modification. These transformations were intended to expose the model to variations in object orientation, scale, and illumination.

The firearm instances were manually annotated using bounding boxes. Each annotation specifies the class of the firearm and its location within the image. The dataset was subsequently divided into training, validation, and testing subsets as shown in Table 1.The testing subset was kept separate from the training process and was used for evaluating the detection performance of the trained model.

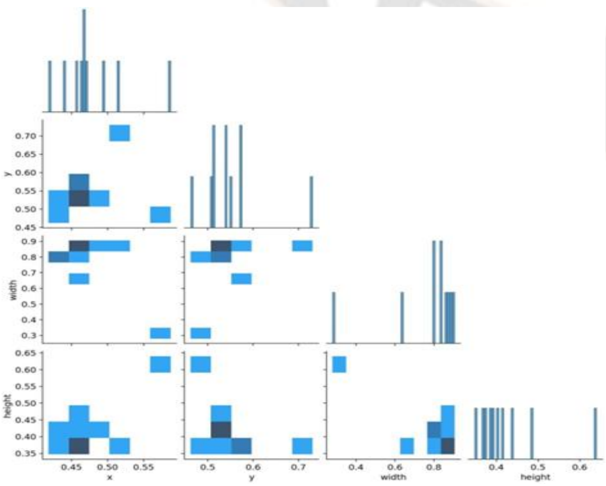


Figure3. Labels correlogram

3. YOLOv8 MODEL

YOLOv8 is a single-stage object-detection model designed to perform object localization and classification efficiently. In the present study, YOLOv8 was trained to identify firearm instances in annotated images.

The model processes an input image and generates object detections containing bounding-box coordinates, class information, and confidence scores. The detection framework is suitable for applications in which both detection accuracy and computational efficiency are important.

The model was trained using the prepared firearm dataset. During training, the annotated bounding boxes and corresponding class labels were used to learn the visual characteristics of firearm instances. Data augmentation was used to introduce variations in the training samples and improve the model's ability to handle changes in orientation, scale, illumination, and background.

After training, the model was applied to previously unseen test images. For each detected firearm, the model produces a bounding box and an associated confidence score. Non-maximum suppression is used during the detection process to reduce redundant overlapping detections.

The performance of the trained model was evaluated using the available detection results, including precision, mean average precision, confusion-matrix analysis, and confidence-based precision and recall curves.

4.FIREARM DETECTION

Firearm detection refers to the process of identifying and locating firearms within digital images or video streams using advanced machine learning and deep learning techniques. This system is designed to automatically detect firearms in real-time by analyzing surveillance footage, typically from CCTV cameras, to ensure public safety in sensitive areas such as airports, schools, or public buildings. The core of this system relies on deep learning models, particularly object detection algorithms like YOLOv8 (You Only Look Once version 8), which are optimized to identify objects (in this case, firearms) in images or videos.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 2
***Article Received:** 05 January 2021 **Revised:** 30 January 2021 **Accepted:** 25 February 2021*

Figure4. Labeled images

The firearm detection system operates by processing video frames or static images captured by surveillance cameras andclassifyingthemintodifferentobjectcategories. Themodelis trainedtoidentifyvarious types of firearms, including pistols, rifles, and shotguns, by recognizing key features, shapes, andpatterns unique to these objects. This detection process involves detecting the objects' location (bounding boxes) within the frame and classifying them with high confidence to determine whether they are firearms.

Key components of firearm detection include data collection and preprocessing, model training, and real-time inference. The data collection process involves gathering a diverse range of firearm images, including different weapon types and scenarios, to train the detection model. Preprocessing techniques, such as image resizing, normalization, and data augmentation, are applied to ensure the model cangeneralizewellacrossvariousenvironments, lighting conditions, and object orientations. The model is then trained using annotated datasets to recognize firearms by their visual features, and after training, it is capable of detecting firearms in new, unseen images.

In a live system, once the firearm detection model is deployed, it analyzes the incoming video feed, detecting firearms within milliseconds. Upon detecting a firearm, the system can generate alerts, notifying security personnel of a potential threat, thereby reducing responsetimesandenhancingsecuritymeasures. The system's real-time capability is essential for preventing incidents before they escalate,making it an invaluable tool in modern security infrastructures.

The efficiency and accuracy of firearmdetection systems continue to improve with advancements in deep learning, particularlywith models like YOLOv8, which offer high detection speeds and precision. This system can be integrated with existing surveillance infrastructures, ensuring a seamless addition to the security ecosystem without requiring extensivehardwareupgrades.Byautomating firearm detection, security personnel can focus on critical decision-making, enhancing overall safety in public spaces.

5.POST-PROCESSING

Post-processing is a critical phase in the firearm detection system, occurring after the initial detection by the YOLOv8 model. Once the model identifies potential firearms in the video frames or images, post-processing techniques are appliedto refine and enhance the results, ensuring accurate detection andreducing false positives or negatives. The primary function of post-processing in this context involves filtering out irrelevant detections and improving the accuracy of the bounding boxes. This is achieved through techniques like Non-Maximum Suppression (NMS), which helps eliminate redundant or overlapping bounding boxes by retaining only the most confident prediction for each firearm. NMS works by selecting the bounding box with the highest confidence score and removingother boxes that significantly overlap, ensuring that each firearm is detected only once.

Another important post-processing step is classifying the detected objects. After identifying potential firearms, the systemchecks their corresponding classes (such as pistol, rifle, shotgun, etc.). If multiple objects are detected, the classification algorithmensures that the correct firearm type is assigned to each bounding box. Additionally, thresholding is applied to filter out detections with low confidence scores, ensuring that only themostreliable predictions are considered.If a firearm is detected but the confidence level is too low, the system may discard that detectionto avoid false alarms. This helps in ensuring the system's accuracy and reliability, particularly in complex environments where

***Article Received:** 05 January 2021 **Revised:** 30 January 2021 **Accepted:** 25 February 2021*

objects might resemble firearms but are not actual weapons.

In some cases, post-processing can involve tracking detected firearms acrossmultiple video frames. This allows the system to monitor the movement of the firearm within the camera's view, providing security personnel with continuous updates. Object tracking can be done using algorithms like Kalman Filtering or optical flow techniques, ensuring that the firearm's position is consistently tracked over time. Post-processing also ensures that alerts are triggered only when a legitimate firearm is detected. Once the system is confident that a firearm is present with a high level of accuracy, an alert is sent to the security personnel, allowing them to take appropriate action.

Overall, post-processing improves the quality of the initial detections made by YOLOv8, ensures more accurate results, and enhances the overall efficiency of the firearm detection system by reducing the likelihood of false alerts. It plays a vital role in making the system robust and reliable for real-time applications in securityand surveillance.

6.ALERTSYSTEM

The alert system in a firearm detection application plays a crucial role in notifying security personnel in real time about potential threats detected by the model. After the YOLOv8 model processes video frames and detects firearms, the alert system comes into play to notify the user about the presence of a firearm. Once a firearm is detected and the confidence level meets a certain threshold, an alert is triggered. This alert is displayed as a simple text message on the system interface, informing the security personnel of the type of firearm detected (e.g., pistol, rifle) and its location within the video frame. The location can be indicated in terms of coordinates (e.g., bounding box positions) or by specifying the section of the image (e.g., left, center, or right) to help the security team identify the firearm's position quickly.

The alert system may also include additional information such as the time the firearm was detected, which camera captured the footage, and even the confidence level of the detection. By providing this relevant information, the alert system enables security personnel to make informed decisions on the next course of action. Sincethealertisdisplayedastext,itishighlyaccessible, and the security team can respond promptly to the alert without needing tointerpret complex visual data. This feature enhances the real-time responsiveness of the security system and ensures that any potential threat can be addressed as soon as possible.

Alerts are typically designed to be clear, concise, and unobtrusive, ensuring that securitypersonnel can focus on the task at hand while being immediately notified when a firearm is detected. Depending on the system's configuration, multiple alerts can be displayed simultaneously if multiple firearms are detected in different parts of the monitored area.

IV. RESULTS AND DISCUSSION

The trained YOLOv8 model was evaluated using the prepared firearm dataset. The dataset consisted of 1,600 images, with 960 images used for training, 320 images for validation, and 320 images for testing. The test set was used to examine the ability of the trained model to identify and localize firearm instances in previously unseen images.

A. Detection Results

Figure 5 presents representative detection results obtained from the trained model. The detected firearm regions are represented using bounding boxes together with their corresponding prediction confidence values.

Figure5.Labeling with prediction percentage

The detection examples demonstrate that the trained model can identify visible firearm instances in the evaluated images. The model is able to localize the detected firearm within the image and provide the corresponding prediction confidence. Variations in firearm orientation, scale, background, and illumination are present in the dataset and provide different visual conditions for evaluating the detector.

However, detection performance can be affected when firearms occupy a small region of an image, are partially occluded, or are captured under poor

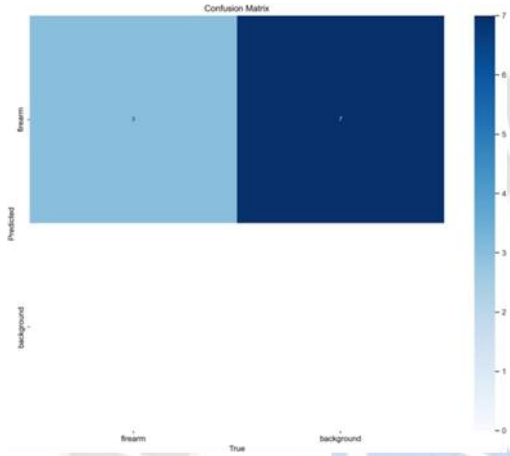
International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 2
***Article Received:** 05 January 2021 **Revised:** 30 January 2021 **Accepted:** 25 February 2021*

illumination. Such conditions can reduce the amount of visual information available to the detector and may result in missed or low-confidence detections.

B. Confusion Matrix Analysis

Figure 6 presents the confusion matrix obtained during model evaluation, while Figure 7 presents its normalized form.

The confusion matrix provides information about the classification behavior of the detector. It can be used to examine the distribution of correct and incorrect predictions among the evaluated classes. The normalized confusion matrix provides the corresponding proportions and facilitates interpretation of class-wise prediction behavior.



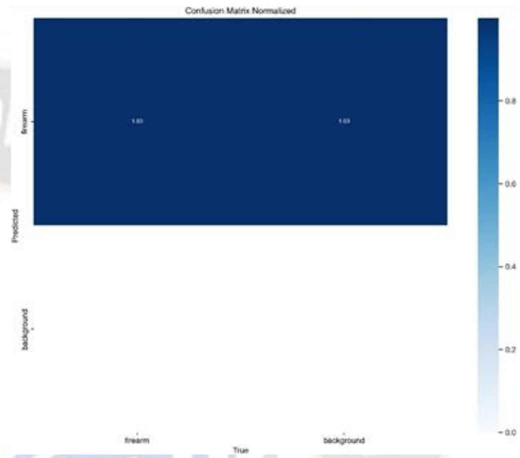


Figure 6. Confusion matrix Figure 7. Normalized confusion matrix with accuracy 1.0

Since the proposed task is an object-detection problem, the confusion matrix is considered together with detection-specific metrics such as precision and mean average precision. Therefore, the normalized confusion matrix is not treated as an independent measure of overall detection accuracy.

C. Precision and Recall Analysis

Figure 8 presents the precision-confidence curve obtained from the trained model. The curve illustrates how the precision of the detections changes with the confidence threshold.

Figure 9 presents the recall curve, while Figure 10 presents the recall-confidence relationship. These curves provide information about the relationship between confidence threshold selection and detection performance.

Increasing the confidence threshold can remove low-confidence detections and consequently reduce false detections. However, an excessively high threshold may also eliminate valid firearm detections. Therefore, the confidence threshold should be selected according to the desired balance between precision and recall.

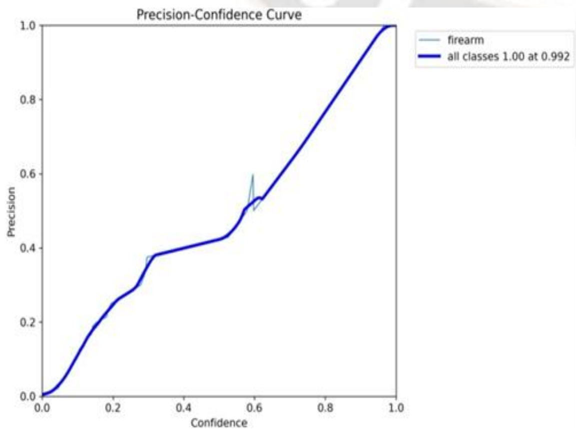


Figure8.Precision-Confidencecurve

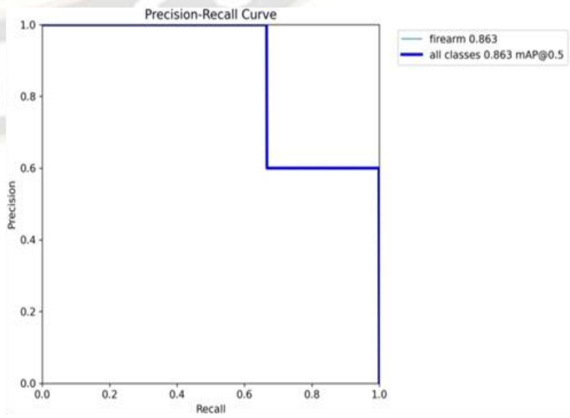


Figure9.Recallcurve

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 9 Issue: 2
***Article Received:** 05 January 2021 **Revised:** 30 January 2021 **Accepted:** 25 February 2021*

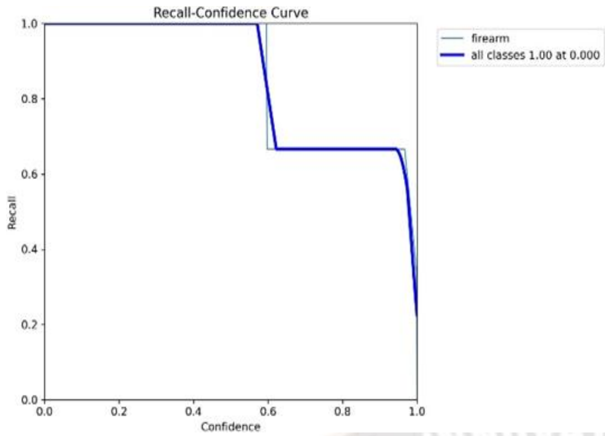


Figure10.Recall-confidencecurve

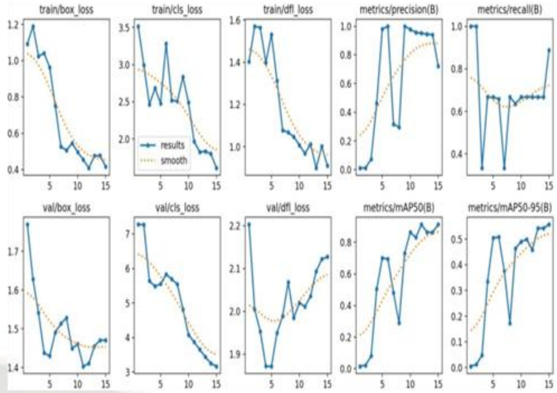


Figure11.Overall results

D. Overall Detection Performance

The experimental results reported in this study show a precision of 99% and a reported mAP value of 99% for the evaluated model. These results indicate that the trained YOLOv8 model achieved high detection performance on the prepared dataset.

The reported values should, however, be interpreted in relation to the characteristics of the dataset and experimental configuration. The dataset contains 1,600 images collected from multiple sources, and factors such as image diversity, annotation quality, class distribution, and similarity between training and testing samples can influence the measured performance.

Therefore, the reported results demonstrate the performance of the model on the prepared evaluation dataset rather than establishing that the same performance will necessarily be obtained on all real-world CCTV environments.

E. Comparison with Existing Studies

Previous research has investigated firearm detection using handcrafted image features, concealed-weapon imaging, image fusion, and deep-learning-based object-detection techniques [1–9].

Tiwari and Verma investigated visual gun detection using Harris interest point and SURF-based features [1,2]. Grega et al. investigated automated firearm and knife detection in CCTV images [10]. Other studies investigated concealed weapon detection using millimeter-wave imaging and multimodal image fusion [3–6]. Deep-learning-based object-detection approaches have subsequently provided improved mechanisms for learning visual representations directly from training data [7,8].

The present study evaluates YOLOv8 on the prepared firearm dataset. The reported precision and mAP values demonstrate the performance obtained under the experimental conditions used in this study.

Direct numerical comparison with previously published studies should be interpreted carefully because different studies may use different datasets, firearm categories, annotation procedures, train-test splits, evaluation metrics, and experimental hardware. Consequently, performance values obtained from different datasets cannot be considered directly equivalent.

LIMITATIONS AND DISCUSSION

Although the trained model produced high detection performance on the evaluated dataset, several limitations should be considered.

First, the dataset contains a relatively limited number of images compared with the diversity of scenes encountered in large-scale surveillance environments. Additional images representing different camera viewpoints, resolutions, illumination conditions, backgrounds, and firearm appearances would provide a broader basis for evaluating generalization.

Second, images obtained from Internet sources may differ substantially from CCTV imagery. Internet images frequently provide clearer views of firearms, whereas surveillance cameras may produce low-resolution images, motion blur, unusual viewing angles, and partial occlusion. This difference may result in a domain gap between the training data and operational surveillance footage.

Third, the present evaluation focuses on image-level object detection. Further evaluation using continuous CCTV video sequences would be useful for determining the model's inference speed and temporal stability. Measures such as frames per second, inference latency, and detection consistency across consecutive frames would provide additional information about real-time deployment.

Finally, the reported performance should not be interpreted as evidence that the system can independently determine whether a detected object represents an immediate threat. The model identifies visual patterns corresponding to firearm instances in the training data. Human review and appropriate operational procedures would still be required when using automated detections in practical surveillance settings.

Overall, the experimental results demonstrate the feasibility of applying YOLOv8 to firearm detection on the prepared dataset. Further evaluation using larger and independently collected CCTV datasets would be necessary to establish the robustness and generalization of the approach across diverse surveillance environments.

REFERENCES:

[1] R. K. Tiwari and G. K. Verma, “A computer vision based framework for visual gun detection using Harris interest point detector,” *Procedia Computer Science*, vol. 54, pp. 703–712, 2015.

[2] R. K. Tiwari and G. K. Verma, “A computer vision based framework for visual gun detection using SURF,” *International Journal of Computer Applications*, 2015.

[3] M. Grega, A. Matiolański, P. Guzik, and M. Leszczuk, “Automated detection of firearms and knives in a CCTV image,” *Sensors*, vol. 16, no. 1, 2016.

[4] G. Sheen, D. McMakin, and T. Hall, “Three-dimensional millimeter-wave imaging for concealed weapon detection,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 49, no. 9, pp. 1581–1592, 2001.

[5] M. Xue, D. M. G. M. de Oliveira, and others, “Fusion of visual and infrared images for concealed weapon detection,” 2002.

[6] J. Blum, J. Aggarwal, and others, “Multisensor concealed weapon detection,” 2004.

[7] S. Upadhyay and P. Rana, “Exposure fusion for concealed weapon detection,” 2014.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 580–587, 2014.

[9] R. Girshick, “Fast R-CNN,” in *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, 2015.

[10] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 28, 2015.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[12] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7263–7271, 2017.

[13] W. Liu et al., “SSD: Single shot multibox detector,” in *European Conference on Computer Vision (ECCV)*, pp. 21–37, 2016.

[14] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2117–2125, 2017.

[15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[16] A. G. Howard et al., “MobileNets: Efficient convolutional neural networks for mobile vision applications,” *arXiv preprint arXiv:1704.04861*, 2017.

[17] T.-Y. Lin et al., “Focal loss for dense object detection,” in *Proc. IEEE International Conference on Computer Vision (ICCV)*, pp. 2980–2988, 2017.

[18] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10781–10790, 2020.

[19] R. Olmos, S. Tabik, and F. Herrera, "Automatic handgun detection alarm in videos using deep learning," *Neurocomputing*, vol. 275, pp. 66–72, 2018.

[20] R. Olmos, S. Tabik, and F. Herrera, "Automatic handgun detection alarm in videos using deep learning," *arXiv preprint arXiv:1702.05147*, 2017.

[21] M. Singleton, D. D. S. et al., "Gun identification using TensorFlow and MobileNet," 2018.

[22] S. Lim et al., "Gun detection in surveillance videos using deep learning," 2019.

[23] V. Egiazarov et al., "Firearm detection and segmentation using an ensemble of semantic neural networks," 2020.

[24] J. Vallez et al., "Deep autoencoder for false positive reduction in handgun detection," 2020.

[25] J. L. Salazar-González, C. Zaccaro, J. A. Álvarez-García, L. M. Soria-Morillo, and F. Sancho-Caparrini, "Real-time gun detection in CCTV: An open problem," *Neural Networks*, vol. 132, pp. 297–308, 2020, doi: 10.1016/j.neunet.2020.09.013.

[26] R. Padilla, S. L. Netto, and E. A. B. da Silva, "A survey on performance metrics for object-detection algorithms," in *Proc. International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 237–242, 2020, doi: 10.1109/IWSSIP48289.2020.9145130.