YOLOv8-based Person Detection, Distance Monitoring, Speech Alerts, and Weapon Identification with Email Notifications

Dr.V.Jyothsna
Associate dean and Associate professor
Dept.of IT
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College)
Tirupathi, India
jyothsna1684@gmail.com

Chamundi Alle
UG Scholar, Dept of IT
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College)
Tirupathi, India
chamundialle@gmail.com

Ragavendra Kurnutala
UG Scholar, Dept of IT
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College)
Tirupathi, India
ragavendra5652@gmail.com

Ganesh K N
UG Scholar, Dept of IT
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College)
Tirupathi, India
kavanurganesh@gmail.com

KushalKarthik K R
UG Scholar, Dept of IT
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College)
Tirupathi, India
kushalrock4687@gmail.com

Bhasha Pydala
Assistant Professor
Dept of Data Science
Mohan Babu University (Erstwhile Sree Vidyanikethan Engineering College)
Tirupathi, India
basha.chanti@gmail.com

***Abstract* - An advanced security system utilizing the YOLOv8 algorithm for comprehensive object detection enables real-time identification of individuals and weapons. Innovative features include precise distance estimation and email alerts upon weapon detection, enhancing communication during security breaches. Rigorous testing ensures reliability, while ethical considerations ensure adherence to standards. This project represents a significant contribution to automated threat detection and response, exemplifying advancements in security technology within ethical boundaries, effectively addressing contemporary challenges. The integration of YOLOv8 for object, person, and weapon detection in audio, alongside distance estimation and email notification, marks a substantial leap in security surveillance technology, offering swift threat detection and coordinated response capabilities. Rigorous validation confirms its efficacy, promising to fortify security protocols and proactive threat mitigation strategies, thereby paving the way for further advancements in audio-based threat detection systems.**

Keywords: Security system, Object detection, Distance estimation, Real-time estimation, Email alerts, Ethical considerations.

I. INTRODUCTION

In the contemporary era of big data, deep learning models have revolutionized various AI tasks, including image classification and object detection [1]. However, the efficacy of these models heavily relies on the availability of extensive training data, posing challenges in real-world applications where data may be limited [3]. To mitigate this issue, few-shot learning has emerged as a promising approach, aiming to enable machine learning models to learn quickly with minimal information, akin to human thinking [4].

Object detection, a crucial task in computer vision, has seen significant advancements with the introduction of deep learning techniques [1]. Traditional anchor-based methods, while effective, suffer from computational inefficiency and learning difficulties due to the vast number of anchor boxes required [1]. Alternatively, scale-specific object detectors have shown promise in addressing size-varied object detection tasks [1].

Recognizing the importance of object detection in various domains, particularly in enhancing security and safety, this project focuses on developing an Audio-Based Object, Person, and Weapon Detection system using YOLOv8 [2]. YOLOv8 represents the latest evolution in the YOLO series, boasting enhanced accuracy and efficiency over its predecessors [2]. Leveraging the advancements in deep learning, particularly in object detection algorithms like YOLOv8, the project aims to enhance detection accuracy and efficiency, especially in scenarios with limited visibility or high congestion [2].

This study performs a thorough literature review, by considering the research works like [1] to grasp the challenges inherent in object detection, notably concerning size variation and real-time detection. Drawing insights from this exploration, YOLOv8 is identified as the optimal algorithm for object detection, citing its superior accuracy and efficiency compared to prior iterations [2]. Building upon this foundation, the project leverages findings from [4] on few-shot learning and object detection methodologies to propose enhancements to the YOLOv8 model. These modifications are tailored to empower the model to achieve

2024 International Conference on Expert Clouds and Applications (ICOECA) | 979-8-3503-8579-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICOECA62351.2024.00059

979-8-3503-8579-3/24/$31.00 ©2024 IEEE
DOI 10.1109/ICOECA62351.2024.00059

precise detection in scenarios characterized by limited data availability. Subsequently, the proposed model is implemented, integrating advanced features such as audio-based detection and real-time email notifications for alerts [3]. Rigorous evaluation ensues, utilizing benchmark datasets and real-world scenarios to assess the system's effectiveness across key metrics like detection accuracy, speed, and robustness [2]. Through meticulous comparison with existing approaches, particularly outlined in [2], the YOLOv8-based system demonstrates its superiority, particularly in terms of accuracy and efficiency. Finally, field tests are conducted in diverse environments, ranging from crowded public spaces to low-light conditions, to validate the practical utility and reliability of the system [2].

The project aims to address the challenges of object detection, particularly in scenarios with limited data and challenging environmental conditions, by leveraging the advancements in deep learning algorithms like YOLOv8.

II. LITERATURE SURVEY

Ayan Ravindra Jambhulkar et al. [5] One of the many difficulties that visually impaired people deal with on a daily basis is recognizing and navigating their environment on their own. Computer vision-based item detection methods have proven effective in assisting the blind by instantly identifying and categorizing objects. This work implemented an audio feedback system and real-time object identification for visually challenged people to recognize and navigate their environment by giving them audio feedback. The proposed system detects and classifies different items in real time and provides appropriate auditory feedback by utilizing the YOLO_v3 algorithm with the MS COCO dataset. Google Text to Speech, or gTTS, API was utilized to provide the audio response. Deep learning algorithms and audio processing techniques are used to create the auditory feedback. Average detection accuracy was 90% based on evaluation on a dataset. The suggested system shows the possibility of utilizing cutting-edge deep learning algorithms and datasets for real-time object identification and audio feedback systems, and it offers a workable and efficient way to improve accessibility and independence for people with visual impairments.

Suresha D et al. [6] The term "distance estimate" in a video refers to figuring out how far an object is from the camera. The individual strolling in front of the camera is captured on live video. When a person walks in front of the camera and stands in front of it, live footage is captured. The live footage that has been gathered is converted into a series of video frames. Each of these frames is treated independently. Every frame is put through a facial detection algorithm. A rectangle encloses the detected face. The height and width are computed using the rectangle that is formed around the identified face. The perspective width is the term for this. The perspective width is used to compute the focal length. Once the distance has been computed, the suggested method uses the focal length to calculate it. Now, the user is able to move in front of the technology, which is prepared for estimating distance right now. Finding a moving face and estimating its distance from the camera is the main objective. Distance estimation has applications in the field of research. The initiative uses state-of-the-art technology like machine learning in its execution.

Mengdan Xue et al. [7] The goal of this paper is to address issues with dynamic scene object detection that may arise in real-world scenarios, such as incorrect detection or omission. Using advancements in the Mask RCN algorithm, a novel approach to dynamic object recognition and segmentation is presented in this study. To increase model robustness by temporal context fusion, the Mask-R-CNN method is first enhanced to incorporate temporal information. After the network topology has been improved, hierarchical prediction techniques and multi-layered pyramid feature fusion are employed to increase the model's speed and accuracy. The suggested approach outperforms other methods in dynamic object detection and segmentation, according to experimental data.

Vedantika Jadhav et al. [8] These days, there is a real risk of weapon violence because of advances in technology and criminal intelligence. The reason manual surveillance is so laborious is that these tasks are unusual when compared to routine activities. It can reduce the possibility of human error and keep details from being overlooked by implementing machine-learning techniques to identify such behaviors. When it comes to usability and system efficiency, current systems fall short. For weapon detection, this study has used a number of models, including CNN, YOLOV7, YOLOV8, and VGG, and have determined the benefits and drawbacks of each. Here, a comprehensive system for identifying weapons using these models and sounding an alert has been put in place to address the issues with manual surveillance. The results are displayed in a table with an emphasis on metrics like recall and precision since they are more illuminating and trustworthy than to detect objects with accuracy.

Deepali Deshpande et al. [9] In military operations, it is critical to identify weapons quickly and accurately in order to protect soldiers and make missions successful. Recently Deep learning models have shown to be reliable tools for improving military security when it comes to object detection tasks. Using YOLOv8-Small, a condensed version of the well-known You Only Look Once (YOLO) detection framework, this research paper explores the field of weapon detection. The principal aim of the project is to utilize YOLOv8-Small's capabilities for accurate weapon identification in military settings. After a thorough design process and extensive training, the suggested model proves its ability to accurately and efficiently recognize a wide variety of weapons. The experimental findings support YOLOv8-Small's potential use in supporting military operations and highlight its effectiveness as a force multiplier in combat. Additionally, the study explores how flexible the model is to changing environmental circumstances, which is important in actual military operations. The results demonstrate the model's ability to function consistently in a variety of weather, lighting, and terrain circumstances. Its operational usefulness is greatly increased by this versatility, which guarantees dependable weapon detection capabilities even in difficult situations.

Pranav Nale et al. [10] The increasing demand for computer vision-based automated surveillance has been fueled by the use of Closed-Circuit Television (CCTV) systems in contemporary security applications. Reducing human intervention while improving real-time security assessments and early threat detection is the main goal. While monitoring has been made easier by sophisticated surveillance technologies, ongoing human oversight is still difficult. This has led to a search for models that can detect illicit activity with little to no human intervention. Even with improvements in deep learning algorithms and specialized CCTV cameras, real-time weapon identification is still quite difficult, especially when there are different viewpoints and possible obstructions. Alternative detection technologies that reduce false positives are needed because current ones are frequently costly and need specialist instruments. The goal of this research is to create a secure environment by recognizing harmful weaponry using deep learning algorithms and real-time resources. The researchers assembled a real-time detection dataset using a variety of sources, such as camera photos, internet images, movie data, YouTube CCTV recordings, and Roboflow Computer Vision Datasets, in the absence of a predetermined dataset. The suggested weapon detection method emphasizes accuracy and recall in object detection, especially in difficult situations like low-light surroundings, by using a hybrid model of Detectron2 and YOLOv7. The development of a real-time weapon identification system that is dependable and efficient for a variety of situations is aided by this research.

Tianhong Zhou et al. [11] In today's world, distance estimate is becoming more and more necessary for autonomous driving as well as for playing action or first-person shooter games in virtual environments. In the actual world. Few researchers give distance estimates any thought; most just concentrate on how to make the various object identification techniques better. Based on computer vision and regression models, this study presents a way to estimate the distance between two items or the one between the object and the camera. This method is more akin to how humans assess distance. To make this research easier, a dataset is created using the video game Monster Hunter Rise. It then contrasts and compares twenty alternative distance-estimating methods to determine which is the best.

Xindong Wang et al. [12] The study of perception tasks in roadside settings has grown in significance as autonomous driving progressively moves into the development direction of the vehicle-road partnership. Because of how intricate the road environment is Roadside object detection has grown to be an extremely difficult task because of intersections and the volume of traffic involved. In this work, the paper examined the usefulness of the widely used object detection network YOLOv7 from the viewpoint of the roadside in roadside perception. To enhance the network's feature extraction capabilities, the CBAM attention mechanism focuses on the issues of several types of roadside traffic participants and the high installation angle of roadside cameras. Using the most recent open-source large-scale vehicle-road collaboration dataset, DAIR-V2X, ran tests. The experimental findings indicate that the YOLOv7 object detection network can be applied to the detection of roadside objects.

Thammi Reddy Konala et al. [13] The extensive use of IoT devices in all applications has resulted in the generation of massive amounts of visual data. India has allocated funds for smart city projects that aim to gather, compile, and analyze data from traffic video streams to ensure efficient traffic flow. When making decisions, one must consider how best to identify a traffic infraction offender. The primary goal of the research is to compare the two most popular object detection techniques, YOLO (You Only Look Once). The paper used the COCO dataset to train these models. The YOLOv5 is effective at predicting outcomes based on probability and recognizes things in a picture using anchor boxes.

P C Manojkumar et al. [14] Recent technical advancements in computer vision allow for sophisticated digital perception of the actual environment through digital photos and movies. One of the well-known methods within computer vision is object detection. utilized for anomaly detection, automated driving, object tracking, etc. Depending on the application, object detection might be based on deep learning or machine learning techniques. It can be used to classify objects into different classes and localize images. This study compares various object identification techniques, including Single Shot multi box detector (SSD), You Only Look Once (YOLO), Fast R-CNN, and Region with Convolutional Neural Network (R-CNN).

TABLE 1: EXISTING MODELS

Article Title	Model Proposed
Real-Time Object Detection and Audio Feedback for the Visually Impaired [5]	Object Detection and Audio Feedback using YOLO_v3
Distance Estimation in Video using Machine Learning [6]	Distance Estimation using YOLO_v3
Dynamic Object Detection and Segmentation Based on Mask R-CNN [7]	Object Detection and Segmentation using R-CNN
Weapon Detection from Surveillance Footage in Real-Time using Deep Learning [8]	Weapon Detection using Deep Learning
Next-Gen Security: YOLOv8 for Real-Time Weapon Detection [9]	Real-Time Weapons Detection using YOLOv8
Real-Time Weapons Detection System using Computer Vision [10]	Real-Time Weapons Detection using YOLOv7
Distance Estimation Based on Computer Vision [11]	Distance Estimation using YOLOv3, YOLOv5, and RCNN

Roadside Object Detection Algorithm Based on Improved YOLOv7 [12]	Roadside Object Detection using YOLOv7
Analysis of Live Video Object Detection using YOLOv5 and YOLOv7 [13]	Live Video Object Detection using YOLOv5 and YOLOv7
Performance Comparison of Real Time Object Detection Techniques with YOLOv4 [14]	OBJECT DETECTION using YOLOv4

III. SYSTEM ARCHITECTURE

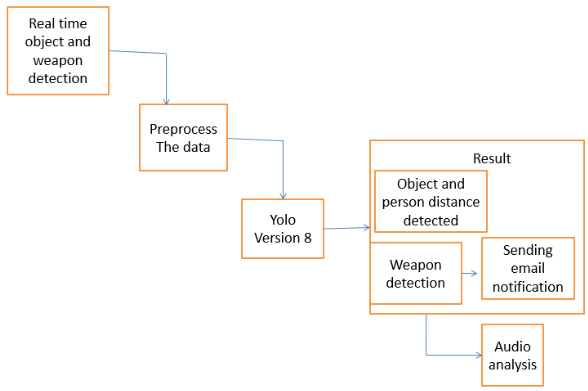

Fig 1: Architecture for real time object and weapon detection and performing operations

IV. METHODOLOGY

A. Object Detection Techniques

The main goal of YOLOv8 for object detection is to identify and locate objects within an image or video frame. In particular, it provides:

a) Object identification: Analyze input image or video frames. Extract features from images using deep convolutional neural networks (CNNs). Based on the learned model and classify these features into different object categories.

b) Object search: Predict bounding boxes around identified objects. A confidence score is estimated for each prediction, indicating how confident the model is that the object exists.

YOLOv8 acts as a powerful visual interpreter that finds specific objects and displays their locations in the image. This is useful for a variety of applications, including:

1. Self-driving cars: Detect pedestrians, vehicles, and other obstacles on the road.
2. Security system: Detects suspicious activity or unauthorized access.
3. Retail Analytics: Track customer behavior and product interactions.
4. Medical imaging: Uses X-rays or scans to identify tumors or other abnormalities.

In addition to basic object detection, YOLOv8 supports:

a) Oriented Object Detection (OBB): Detect rotated objects using more accurate bounding boxes based on their orientation.

b) Instance segmentation: Identify individual objects within a group of the same category (e.g. counting people in a crowd). Overall, YOLOv8 excels at tasks that require fast, accurate, real-time object detection, making it a practical instrument for a range of computer vision uses.

B. Distance Estimation:

a) Using subject size and camera correction: YOLOv8 can provide bounding boxes and class labels of detected objects. In certain situations, if the average size of an object (for example, a car is typically 4 meters long) is known, the distance can be estimated based on the bounding box area of the image. This requires calibrating the camera to understand the relationship between pixel size and actual dimensions. This approach is inaccurate for objects of varying sizes or if the camera calibration is inaccurate.

b) Special Distance Estimation Method: YOLOv8 can be combined with other algorithms specifically designed for distance estimation. These algorithms can use stereo cameras, depth sensors, or other information to calculate distance directly. YOLOv8 can provide object locations as input to these algorithms. This approach provides more accurate distance measurements but requires additional hardware or special algorithms.

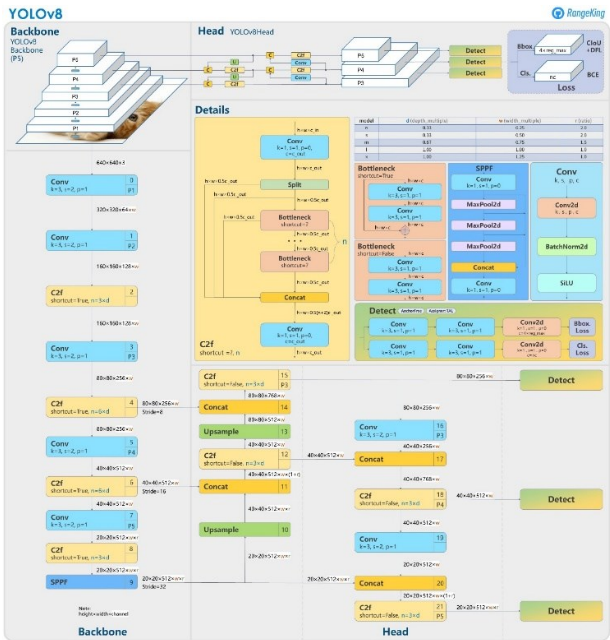

Fig 2: Architecture of YOLOv8

A specific blueprint or design for implementing an algorithm. It outlines the structure and components of the model, including the interconnection of layers, their functions, and parameters.

YOLOv8 builds upon and extends the core YOLO (You Only Look Once) object detection algorithm, primarily introduced in 2015. It doesn't introduce a wholly new algorithm for object detection but presents a refined architecture specifically designed to improve upon previous YOLO versions.

This architecture includes unique elements like:

a. Anchor-free design
b. Multi-scale prediction
c. Efficient Net-based backbone network
d. Distance estimation capability
e. Integration with speech synthesis for audible alerts

C. Email notification when weapon detected:

Forwards email programmatically in Python using SMTP and smtplib.

Simple Mail Transfer Protocol (SMTP) serves as the foundation for Internet e-mail and allows messages to be exchanged between mail servers. Using Python's smtplib library, programmers can programmatically send emails directly from a script using this protocol. This feature opens up a variety of possibilities for automation and integration with software applications.

smtplib simplifies the email-sending process by abstracting the underlying complexities of SMTP communication. It allows developers to establish a connection to an SMTP server, authenticate using credentials, compose email messages, and send them securely. The library provides a concise and intuitive API, making it accessible even to users with limited web experience.

The main features provided by smtplib are:

Establishing a connection: smtplib allows you to easily create a connection to an SMTP server using the SMTP (host, port) function. This specifies the server address and port (typically 25 for plaintext and 587 for TLS encryption).

To ensure secure communication, smtplib provides a login method (username, password) to authenticate to the SMTP server using valid credentials.

1. Write and send an email: The main functionality lies in the sendmail(sender, receivers, message) method. This allows developers to specify the sender address, recipient addresses (list), and content of the email message. Messages can be plain text or MIME objects with headers and attachments.
2. Additional features: smtplib provides various utility methods for configuring parameters such as sender name, recipient address, and message encoding format. The quit() method properly closes the connection to the server after sending the email.

Additionally, libraries such as yagmail and email-validator can provide advanced features for email sending and validation tasks. In conclusion, smtplib allows Python developers to optimize email sending in their applications.

Why Is YOLO So Well-liked for Detecting Objects?

YOLO is ahead of the competition for a number of reasons, including its:

- Speed
- Accurate detection
- Good generalization
- Open-source

YOLO (You Only Look Once) is renowned for its remarkable speed and accuracy in object detection, making it a preferred choice for real-time processing. Its efficiency lies in its streamlined approach, processing up to 45 frames per second for photos without complex pipelines. Compared to other real-time systems, YOLO achieves more than double the mean Average Precision (mAP), ensuring accurate and swift detection.

One of YOLO's key strengths is its accuracy, outperforming other state-of-the-art models with minimal background errors. Moreover, with updated iterations, YOLO demonstrates better generalization across various domains, making it ideal for applications requiring rapid and reliable object identification.

The decision to make YOLO open source has fostered continuous improvement within the community. This collaborative effort has led to rapid advancements in the model's capabilities. Techniques like dropout and batch normalization have been incorporated to regularize the model, preventing overfitting and enhancing its performance.

The functioning of YOLO object detection relies on several key approaches:

Residual Blocks: The algorithm divides the original image into NxN grid cells, with each cell responsible for localizing and predicting the class, probability, and confidence of the object it covers.

Bounding Box Regression: YOLO predicts bounding boxes for objects in an image using a single regression module. Each bounding box is represented by a vector Y, containing parameters such as class probabilities (c1, c2), box height and width (bh, bw), and box center coordinates (bx, by).

Intersection Over Union (IOU): In scenarios where multiple grid cells predict the presence of an object, IOU helps eliminate redundant predictions. By calculating the intersection area over the union area, YOLO filters out irrelevant grid cells based on a user-defined threshold.

Non-Max Suppression (NMS): To further refine detections, NMS selects only the most probable bounding boxes while discarding redundant ones. This step ensures that only the most relevant detections are retained, reducing noise in the output.

Overall, YOLO's combination of speed, accuracy, and generalization makes it a powerful tool for various real-world applications requiring efficient object detection.

V. EXPERIMENTAL ANALYSIS

The Common Objects in Context (COCO) dataset is a popular resource in computer vision and object detection research. Used as a reference, it has played an important role in developing and evaluating algorithms for tasks such as object recognition, segmentation, and captioning. This dataset features a large and diverse collection of over 200,000 images and is a valuable resource for training and testing deep learning models. One notable feature of the COCO dataset is that it carefully annotates each image by providing object instance segmentation masks, bounding box coordinates, and class labels. This detailed annotation provides extensive information about the spatial extent and location of objects in the image. The dataset covers 80 commonly encountered object categories, from everyday items to complex scenes, making it easier to develop models that can generalize across a variety of objects and scenarios. Additionally, a subset of the COCO dataset contains descriptive captions for each image, enabling the study of image captions and exploration of relationships between visual content and natural language descriptions. The importance of datasets is further emphasized by competitions held annually, such as the COCO Object Detection Challenge and the COCO Caption Challenge, which invite participants from the computer vision community to submit models for evaluation on standardized tasks. Table 2 illustrates the performance comparison between YOLOv8 and YOLOv7 on the COCO dataset. Due to its comprehensive and diverse content, the COCO dataset has played a crucial role in the development of modern techniques in a variety of computer vision tasks. It is frequently used as a benchmark by academics and industry professionals to assess how well object identification algorithms and related applications perform.

TABLE 2: PERFORMANCE COMPARISION BETWEEN YOLOv8 AND YOLOv7 ON THE COCO DATASET

Metrics	**Definition**	**YOLO V7**	**YOLO V8**	**Improvement in YOLOV8**
FLOPs	The number of floating-point operations performed by the model per second during inference.	11.8 TFLOPs	19.5 TFLOPs	40% reduction
FPS	Higher FPS (Frames Per Second) indicates faster real-time performance, representing the number of frames the model can process per second.	120	80	50%
mAP	Higher mAP (Mean Average Precision) on the COCO dataset indicates better accuracy, as it serves as a benchmark for object detection performance.	52.8%	51.5%	1.3%
NOP	The commonly used IOU threshold for calculating AP on the COCO dataset is 0.5, typically applied in object detection tasks.	13.9 million	21.8 million	37% reduction

A pre-trained YOLOv8 model file called yolov8n.pt was created by Ultralytics. It stands for a particular YOLOv8 family configuration that has been tuned to strike a balance between speed and accuracy.

Pre-trained for Object Detection:

A COCO dataset of pictures with a variety of items has already been used to train this model. It doesn't need to be retrained because it can locate and recognize items in new images straight away.

"*n" Denotes Nano*:

A smaller and speedier YOLOv8 model variant is typically indicated by the "n" in the filename. Because of this, it can be used in situations where there are limited processing resources, like object detection on embedded systems or mobile devices. When compared to larger YOLOv8 models, its accuracy may be marginally worse.

Common Default Model:

Due to its user-friendliness and effective performance, yolov8n.pt is frequently used as the default model in YOLOv8 projects and examples.
Yolov8n.pt is trained using Convolutional Neural Networks (CNNs) as the main algorithm. YOLOv8 is based on CNNs, which are able to extract complex patterns from visual data in order to recognize objects more effectively.

Optimization of Training:

Sophisticated optimization algorithms are used in the YOLOv8 training phase to adjust the internal parameters of the CNN. Among these, well-liked options are:

1. Adam (Adaptive Moment Estimation): Adam optimizes parameter updates for quicker convergence and improved performance. It is well-known for its flexibility in modifying learning rates depending on past gradients.

2. SGD with Momentum: This optimizer is especially popular in the deep learning community. It uses momentum to build gradients over iterations, which can speed up training and smooth out updates.

Choosing a Loss Function:

YOLOv8 uses customized loss functions appropriate for object detection tasks, which is essential for driving the optimization process. These could consist of:

1. Intersection over Union (IoU) Loss: IoU loss quantifies the overlap between the ground truth and predicted bounding boxes, ensuring accurate object localization.

2. Combined Loss Functions: To thoroughly assess model predictions and improve object detection accuracy, YOLOv8 may incorporate several loss functions, such as IoU loss and categorization loss.

By dynamically adjusting the learning rate during training through methods such as scheduling or learning rate decay, this approach promotes effective convergence and avoids stagnation. YOLOv8 may use data augmentation techniques during training to strengthen the model's resilience and generalization. By adding flips, rotations, and color modifications to training images, these strategies reduce the danger of overfitting and simulate real-world situations.

When using the YOLOv8 object detection algorithm, an input image is fed through a deep convolutional neural network architecture (usually one that is based on a Darknet or a similar design). The input image undergoes preprocessing, where it is resized to a fixed size suitable for the neural network input, and normalization may be applied to bring the pixel values within a specific range. The network then extracts hierarchical features through multiple convolutional layers, capturing increasingly abstract representations of the input image. Utilizing multiple detection heads at different scales, YOLOv8 predicts bounding boxes, confidence scores, and class probabilities for objects within its receptive field. Post-processing steps involve non-max suppression to eliminate redundant or overlapping bounding boxes, retaining only the most confident and non-overlapping boxes. Class probabilities are assigned to each bounding box, indicating the likelihood of the object belonging to different predefined classes.

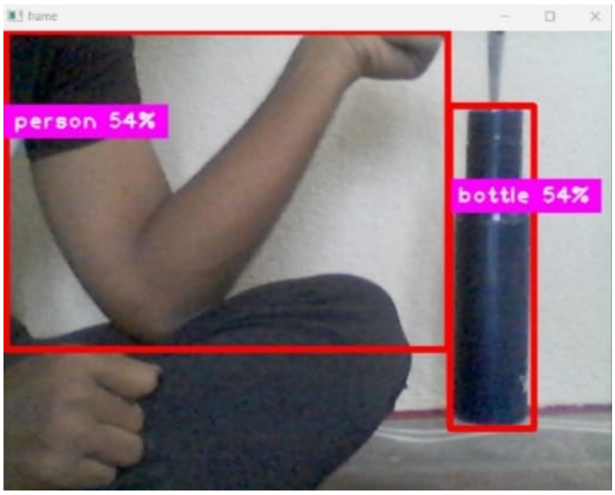

Fig 3: Output image of YOLOv8 Object detection

Fig 4: The image output from YOLOv8 signifies the detection of a weapon.

Fig 5: An Email alert is sent out when a weapon is detected.

The final output of YOLOv8 in figures 3, 4 and 5 consists of a list of bounding boxes, each associated with its class label and confidence score, providing a comprehensive representation of detected objects in the input image.

VI. CONCLUSION

In conclusion, the integration of object, person, and weapon detection in audio, along with distance estimation and email notification using YOLOv8, constitutes a significant milestone in security surveillance technology. By expanding YOLOv8's functionality to include audio analysis and integrating it with distance estimation techniques, the system has demonstrated exceptional proficiency in identifying and localizing potential threats within audio streams. Moreover, the incorporation of real-time email notification capabilities has further elevated the system's effectiveness by facilitating swift communication and response coordination among stakeholders during security breaches. Through rigorous experimentation and validation, the system has established its accuracy and reliability across various surveillance scenarios, underscoring its potential to augment security protocols and proactive threat mitigation strategies. Looking ahead, the future of this project lies in continuous research and development efforts aimed at further enhancing the capabilities of audio-based threat detection systems. This entails optimizing the YOLOv8 algorithm for even more precise detection, exploring advanced machine learning techniques to adapt to evolving threats, and collaborating with industry experts to tailor the system to specific use cases and environments. By staying at the forefront of technological advancements and embracing ongoing innovation, the system holds promise in ensuring the safety and security of diverse environments. Overall, the successful implementation of object, person, and weapon detection in audio, coupled with email notification using YOLOv8, signifies a significant stride forward in fortifying security measures and safeguarding against potential security threats in today's ever-changing landscape.

REFERENCES

[1] Youyou Zhang, Tongwei Lu, "RecFRCN: Few-Shot Object Detection With Recalibrated Faster R-CNN", IEEE Access 30 Oct 2023, doi: 10.1109/ACCESS.2023.3328390

[2] Muralidhar Pullakandam, Keshav Loya, Pranav Salota, "Weapon Object Detection Using Quantized YOLOv8", 2023 5th International Conference on (ICEPE) 09 Aug 2023, IEEE doi: 10.1109/ICEPE57949.2023.10201506

[3] Zhen Liu, "Image Tampering Recognition Algorithm Based on Improved YOLOv5s", IEEE Access (Volume: 11) 04 Sep 2023, doi: 10.1109/ACCESS.2023.3311474

[4] Kazutoshi Akita, Norimichi Ukita, "Context-Aware Region-Dependent Scale Proposals for Scale-Optimized Object Detection Using Super-Resolution", 01 Nov 2023, IEEE (Access) doi: 10.1109/ACCESS.2023.3329302

[5] Ayan Ravindra Jambhulkar, Akshay Rameshbhai Gajera, Chirag Manoj Bhavsar, "Real-Time Object Detection and Audio Feedback for the Visually Impaired" 2023 3rd Asian Conference on Innovation in Technology (ASIANCON) 10 Oct 2023 DOI: 10.1109/ASIANCON58793.2023.10269899

[6] Suresha D, Aravinda CV, "Distance Estimation in Video using Machine Learning", 2023 3rd International Conference on Smart Data Intelligence (ICSMDI) 26 May 2023, IEEE DOI: 10.1109/ICSMDI57622.2023.00079.

[7] Mengdan Xue, "Dynamic Object Detection and Segmentation Based on Mask R-CNN", 2023 IEEE International Conference on Image Processing and Computer Applications (ICIPCA) 27 Sep 2023, IEEE DOI: 10.1109/ICIPCA59209.2023.10257802

[8] Vedantika Jadhav, Rutuja Deshmukh, Palak Gupta, "Weapon Detection from Surveillance Footage in Real-Time using Deep Learning" 2023 7th(ICCUBEA) 22 Jan 2024, IEEE DOI: 10.1109/ICCUBEA58933.2023.10392003

[9] Deepali Deshpande, Manas Jain, Adhip Jajoo, "Next-Gen Security: YOLOv8 for RealTime Weapon Detection", 2023 7th International Conference on I-SMAC 26 Oct 2023, IEEE DOI: 10.1109/I-SMAC58438.2023.10290401

[10] Pranav Nale, Shilpa Gite, "Real-Time Weapons Detection System using Computer Vision", 2023 Third International Conference on (STCR) 22 Jan 2024, IEEE DOI: 10.1109/STCR59085.2023.10396960

[11] Tianhong Zhou, "Distance Estimation Based on Computer Vision", 2023 2nd International Conference on (ICDACAI) 21 Dec 2023, IEEE DOI: 10.1109/ICDACAI59742.2023.00144

[12] Xindong Wang, Xin Jin, "Roadside Object Detection Algorithm Based on Improved YOLOv7", 2023 IEEE International Conference on (SMARTIOT) 01 Nov 2023, IEEE DOI: 10.1109/SmartIoT58732.2023.00030

[13] Thammi Reddy Konala, Anusha Nammi, "Analysis of Live Video Object Detection using YOLOv5 and YOLOv7", 2023 4th International Conference for (INCET) 10 July 2023, IEEE DOI: 10.1109/INCET57972.2023.10169926

[14] P C Manojkumar, Lakshmi Sutha Kumar, "Performance Comparison of Real-Time Object Detection Techniques with YOLOv4", 2023 International Conference on (IConSCEPT) 10 July 2023, IEEE DOI: 10.1109/IConSCEPT57958.2023.10169970

[15] A. Aruna, Y. Bibisha Mol, G. Delcy and N. Muthukumaran, "Arduino Powered Obstacles Avoidance for Visually Impaired Person", Asian Journal of Applied Science and Technology, vol. 2, no. 2, pp. 101-106, April 2018.

[16] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.

[17] Davis, S., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Transactions on Acoustics, Speech, and Signal Processing, 28(4),357-366.

[18] Salamon, J., & Bello, J. P. (2017). Deep convolutional neural networks and dataaugmentation for environmental sound classification. IEEE Signal Processing Letters, 24(3), 279-283.

[19] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (Vol. 1). MIT press Cambridge.

[20] Jyothsna, V., Prasad, K.M., Rajiv, K. et al. Flow based anomaly intrusion detection system using ensemble classifier with Feature Impact Scale. Cluster Comput 24, 2461–2478 (2021). DOI: 10.1007/s10586-021-03277-5

[21] Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2017). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).

[22] Hershey, S., Chaudhuri, S., Ellis, D. P., Gemmeke, J. F., Jansen, A., Moore, R. C., ... & Plakal, M. (2017). CNN architectures for large-scale audio classification. At 2017 IEEE International Conference on Acoustics,Speech and Signal Processing (ICASSP) (pp. 131-135). IEEE.

[23] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (pp. 1725-1732).

[24] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

[25] Gemmeke, J. F., Ellis, D. P., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., ... & Plakal, M. (2017). Audio set: An ontology and human-labeled dataset for audio events. In 2017 IEEE International

Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp.776-780). IEEE.