WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Weapon Detection and Person Tracking

Pratik Kamble[a], Aditya Jadhav[a], Gayatri Bagal[a], Monali Jamdade[a], Prof. Nita Dimble [b]*

[a] Student, NGI Faculty of Engineering, Naigoan, India
[b] Project Guide, NGI Faculty of Engineering, Naigoan, India

A B S T R A C T:

This paper presents a review of weapon detection and person tracking systems using deep learning techniques. The proposed system utilizes YOLOv8, a state-of the-art model in object detection, to detect firearms and other weapons, integrated with the DeepSORT algorithm for object tracking. After extensive testing and evaluation, YOLOv8 was found to offer superior accuracy and performance, outperforming alternative models such as YOLOv5 and YOLOv11. Additionally, this system is optimized to detect weapons using a standard webcam, making it adaptable for real-time surveillance. This paper provides a comprehensive overview of the research, objectives, methodology, and current progress, along with future directions for this technology.

Keywords: Weapon Detection, Person Tracking, Public Safety Deep Learning, YOLOv8, DeepSORT, Transfer Learning.

Introduction

Advanced surveillance systems that can identify threats and follow people in real time are desperately needed as public safety concerns in places like airports, schools, and shopping malls develop. Automated solutions are crucial for enhancing security and situational awareness since human error, weariness, and slow response times are common drawbacks of traditional surveillance techniques that depend on human operators. Using deep learning techniques, this study investigates the creation of a automated weapon detection and person tracking system that offers a dependable, real-time solution. The suggested method makes use of DeepSORT for person tracking and YOLOv8 for weapon detection to continuously monitor those carrying identified weapons.

In comparison to models like YOLOv5, YOLOv7, and experimental versions like YOLOv11, YOLOv8, a state-of-the-art object detection model, showed superior accuracy in identifying weapons. DeepSORT, which uses both motion and appearance-based tracking, effectively handles occlusions, crowded scenes, and dynamic movements, reducing identity switches and maintaining consistent tracking across frames. This paper outlines the goals, difficulties, and techniques involved in creating a reliable, real-time surveillance system that can be used in a variety of settings, such as low light levels and complex, densely populated areas.

Additionally, by optimizing YOLOv8 and DeepSORT to work well with standard webcams, the system improves accessibility and deployability for both public and private security applications. The technology gives security staff real-time information through this connection, enabling proactive threat identification and quicker reaction times, which eventually increases the effectiveness of monitoring as a whole.

Methodology

This section delves into the core modules of the system, including data collection, the weapon detection model using YOLOv8, and person tracking via DeepSORT.

Data Collection and Preprocessing

Data collection plays a critical role in ensuring the robustness and accuracy of the weapon detection model. For this project, a dataset comprising thousands of images of various weapons, including firearms and knives, was gathered from public databases such as the Custom and Roboflow dataset, as well as specialized collections for security research. The dataset includes diverse scenarios and lighting conditions, to ensure that the model performs well in various environments, from well-lit areas to low visibility situations. Each image was carefully pre-processed, resized, and annotated to define the exact location of weapons within frames.

To further enhance the model's generalization capabilities, data augmentation techniques, such as rotation, scaling, flipping, and brightness adjustments, were applied. These techniques help simulate real-world conditions and improve the model's robustness, ensuring accurate detection even in crowded and complex scenes. This comprehensive approach to data preparation is crucial to achieving high detection accuracy with YOLOv8 in real-time applications.

Hardware Requirement

- **Processor**
- **Video card**
- **Memory**
- **Webcam**

1) Processor :

A processor is a tiny electrical device that executes input/output commands, logical processes, and other fundamental tasks needed by computers. The operating system transmits these directives to the processor. Although the terms "processor" and "CPU" are frequently used synonymously, the CPU is only one type of processor that can be found in a computer. The **Graphics Processing Unit (GPU)** and some types of hard drives that can conduct some processing are examples of additional processors found in computers.

Video Card :

A video card, also known as a graphics processing unit (GPU), is a specialized hardware component that is designed to process and output graphical data to a display device such as a monitor. It functions as a coprocessor to the computer's central processing unit (CPU), specifically dedicated to handling complex graphical computations such as rendering 3D images, video editing, and gaming. The GPU contains hundreds or thousands of processing cores that work in parallel to quickly execute large amounts of data in real-time. In addition to providing enhanced graphics capabilities, a powerful GPU can also improve the overall performance of a computer system, especially in tasks that require heavy graphic processing.

Memory :

Memory, also known as RAM (Random Access Memory), is a type of computer hardware that is used to temporarily store data that is currently being accessed or processed by the CPU (Central Processing Unit). It allows for quick and easy access to data, improving the overall performance and speed of a computer system. RAM is typically measured in gigabytes (GB) and can be upgraded or expanded to meet the needs of a computer user. However, it is important to note that RAM is a volatile memory, which means that it only stores data temporarily and will lose that data if the computer is shut down or power outages.

Webcam :

A webcam is a type of digital camera that captures video and audio and is designed to be connected to a computer or laptop, allowing for live streaming or video conferencing. It usually features a small, built-in microphone and lens, and can be adjusted to capture different angles and perspectives. The video captured by the webcam can be stored or transmitted in various file formats, such as AVI or MP4, and can be used for a wide range of purposes, including video calls, online meetings, remote learning, and live streaming events. Overall, webcams have become an essential tool for communication and collaboration in today's digital world.

Software Requirement

Google Colab

- VS Code
- Open CV
- Python
- Nvidia CUDA and CNN
- PyTorch and YOLOv8
- DeepSORT

Google Colab :

Google Colab (short for Colaboratory) is a cloud-based platform provided by Google that allows users to write and execute Python code through a web browser. It supports free access to computational resources, including GPUs and TPUs, making it ideal for machine learning, deep learning, and data analysis tasks. Colab is widely used in research and education due to its collaborative features, integration with Google Drive, and support for Jupyter Notebook environments without any setup.

VS Code :

VS Code, short for Visual Studio Code, is a free source-code editor developed by Microsoft. It supports a wide range of programming languages and has features such as debugging, syntax highlighting, code completion, and Git integration. It also has a customizable user interface, allowing users to personalize the editor according to their preferences. VS Code's popularity stems from its lightweight nature, fast performance, and extensive range of extensions available, making it a popular choice among developers of all levels. Its cross-platform compatibility allows it to be used on Windows, macOS, and Linux operating systems.

OpenCV :

OpenCV (Open Source Computer Vision) is a widely used open-source computer vision and machine learning software library. It provides various tools and algorithms for image and video analysis, processing, and manipulation. OpenCV supports multiple programming languages such as C++, Python, and Java, and runs on various platforms like Windows, Linux, macOS, and Android. With its easy-to-use functions and powerful

capabilities, OpenCV is popularly used in areas such as object detection, facial recognition, and gesture recognition. It also offers support for hardware acceleration with GPUs and supports deep learning frameworks like PyTorch.

Python :

Python is a high-level, interpreted programming language known for its simplicity, readability, and ease of use. It is designed to be beginner-friendly, making it a popular choice for beginners in programming. Python is widely used for a variety of tasks such as web development, data analysis, machine learning, and artificial intelligence. One of the key features of Python is its extensive library, which makes it easy to accomplish complex tasks with just a few lines of code. Additionally, Python's syntax is clear and concise, making it easy to understand and debug. Overall, Python is a versatile and powerful programming language suitable for a wide range of applications.

NVIDIA CUDA and CNN :

NVIDIA CNN refers to a type of convolutional neural network (CNN) architecture that has been optimized for use on NVIDIA GPUs (graphics processing units) using the CUDA parallel computing platform. NVIDIA has developed a range of deep learning software libraries and tools, such as the cuDNN library, which is specifically designed to optimize deep learning computations on NVIDIA GPUs. This makes it possible to train and deploy complex neural networks, including CNNs, with high efficiency and performance.

PyTorch and YOLOv8 :

PyTorch is a popular opensource machine learning library based on the Torch library, primarily developed by Facebook's AI Research (FAIR) team. PyTorch provides a wide range of tools and functions for building and training deep neural networks, including support for GPU acceleration, automatic differentiation, and dynamic computational graphs. PyTorch is commonly used for computer vision, natural language processing, and other machine learning applications. YOLOv8 is the latest version in the You Only Look Once (YOLO) family of real-time object detection models. Developed by Ultralytics and built on PyTorch, YOLOv8 offers improvements in accuracy, speed, and flexibility compared to its predecessors. It features a modular design, better anchor-free detection architecture, and supports tasks beyond object detection, including instance segmentation and image classification. YOLOv8 is widely used for real-time applications in areas such as surveillance, autonomous driving, and smart retail.

7) DeepDORT :

DeepSORT (Deep Simple Online and Realtime Tracking) is an advanced multi-object tracking algorithm that builds upon the traditional SORT (Simple Online and Realtime Tracking) framework by incorporating deep learning-based appearance descriptors. While SORT relies primarily on Kalman Filters and Hungarian matching for tracking based on motion, DeepSORT enhances accuracy by using a deep convolutional neural network to extract robust appearance features from detected objects. This enables better handling of occlusions, re-identification after temporary disappearance, and reduces identity switches during tracking. DeepSORT is widely used in real-time surveillance, pedestrian tracking, and autonomous systems due to its balance of speed and accuracy.

Weapon Detection Model using YOLOv8

The weapon detection module relies on the YOLOv8 model, chosen after extensive experimentation with various architectures. YOLOv8's architecture allows for highspeed processing and accurate detection, which is critical for real-time surveillance. Earlier iterations with models like YOLOv5, YOLOv7, and even an experimental YOLOv11 version were tested, but none matched the accuracy and efficiency of YOLOv8. This model was finetuned using transfer learning on the specialized dataset, significantly enhancing detection precision for diverse weapon types.

A key feature of YOLOv8 is its ability to handle complex backgrounds and crowded environments, where false positives can often arise. The model's high accuracy in detecting even small or partially occluded weapons is particularly valuable for applications in public spaces where fast, reliable identification is essential. Additionally, the system is configured to work with standard webcams, enabling adaptability for both indoor and outdoor scenarios. By leveraging YOLOv8's strengths in real-time detection and integrating it with live video feeds, the model has proven capable of providing accurate weapon detection across varied settings, establishing its utility for enhanced surveillance.

Another advantage of using YOLOv8 lies in its modular and extensible architecture, which allows for easy integration with other systems such as object tracking algorithms and alert generation mechanisms. In this project, YOLOv8 is coupled with a lightweight tracking module to monitor the movement of individuals carrying detected weapons across video frames. This not only ensures continuous surveillance but also aids in generating behavioral analytics for security personnel. The integration supports scalability, enabling the system to be deployed in various environments including schools, transportation hubs, and public venues without requiring significant computational resources.

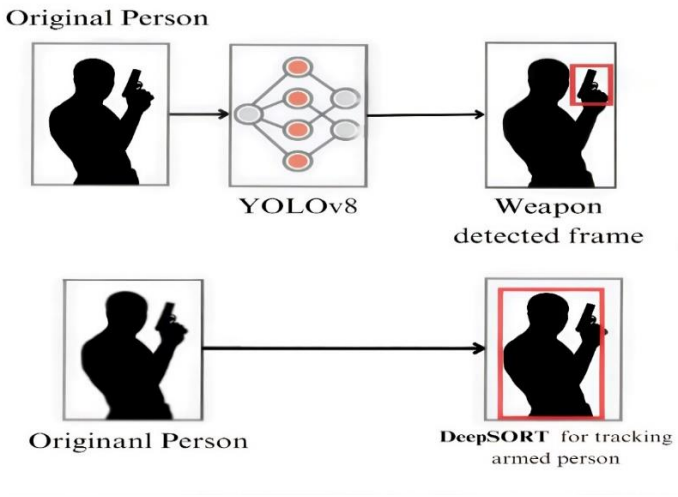
Person Tracking DeepSORT

There are several ways to monitor people in security surveillance, and one of the best techniques for real-time multi-object tracking is DeepSORT. Because traditional tracking techniques only use mobility data, they are vulnerable to identity swaps, particularly in congested settings. By combining motion and appearance information, DeepSORT improves tracking and makes it possible to identify people more reliably even when they momentarily go out of frame or change course.

Once a possible danger, like a weapon, is identified, tracking specific persons becomes essential in security applications. YOLOv8 effectively recognizes firearms in pictures and videos by enclosing them in exact bounding boxes. However, tracing the individual carrying the weapon guarantees ongoing surveillance and assists security professionals in taking the appropriate action. Weapon detection alone is insufficient .DeepSORT enables this by associating the detected weapon with a specific individual and maintaining their identity across multiple frames. Even in complicated situations, the technology makes sure the suspect is not lost by handling occlusions, fast motions, and several persons in the area. DeepSORT considerably lowers false tracking errors compared to other straightforward tracking techniques because of its capacity to adjust to variations in appearance and movement patterns.

In order to enable prompt and well-informed actions, this guarantees that security personnel receive real-time alerts on questionable locations. The device improves situational awareness by utilizing sophisticated tracking algorithms, which makes monitoring more effective and proactive. In addition to keeping an eye on suspects, tracking offers useful movement data that enable security professionals foresee possible threats and take action before a situation gets out of hand.

System Diagram



Result

The system effectively detects weapons and tracks them in real-time. Results:

Weapon Detection :

The system detects a variety of weapons, such as knives, guns, and other types of guns in real-time by using YOLOv8, a cutting-edge deep learning model. The model minimizes false positives and false negatives while achieving high detection accuracy by utilizing a rich dataset that includes a variety of weapon kinds, lighting situations, and occlusions. Using precise bounding boxes, it locates firearms with accuracy while processing pictures and video feeds in an efficient manner. Cutting-edge feature extraction methods improve detection accuracy, guaranteeing that guns are recognized even in difficult situations including crowded backdrops, partial occlusions, dim lighting, and different object sizes. The system is very useful for security applications in surveillance, law enforcement, and public safety monitoring since it continually adjusts to new threats by fine-tuning and retraining on updated datasets.

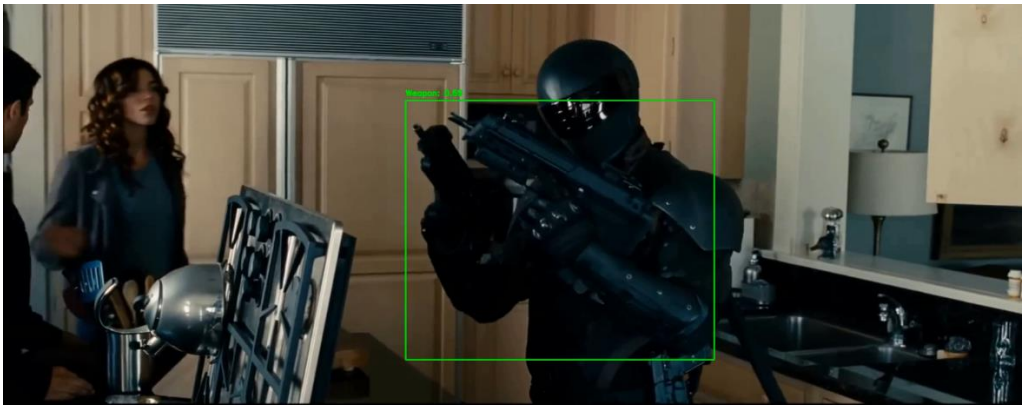


Fig. 4.1 Weapon Detection

B. Person Tracking :

Even in congested or dynamic surroundings, the system continually tracks people carrying suspected weapons using DeepSORT, a powerful multi-object tracking algorithm. It reduces identity shifts and tracking failures by linking detected weapons to particular individuals, ensuring constant tracking across video frames. To provide precise tracking, the model effectively manages occlusions, quick motions, and changing illumination conditions. It differentiates between several people using motion prediction and deep feature extraction, guaranteeing that the weapon carrier is recognized even in challenging situations. Security monitoring is improved by this real-time tracking feature, which enables authorities to react quickly to any threats and stop dangerous situations from getting worse.

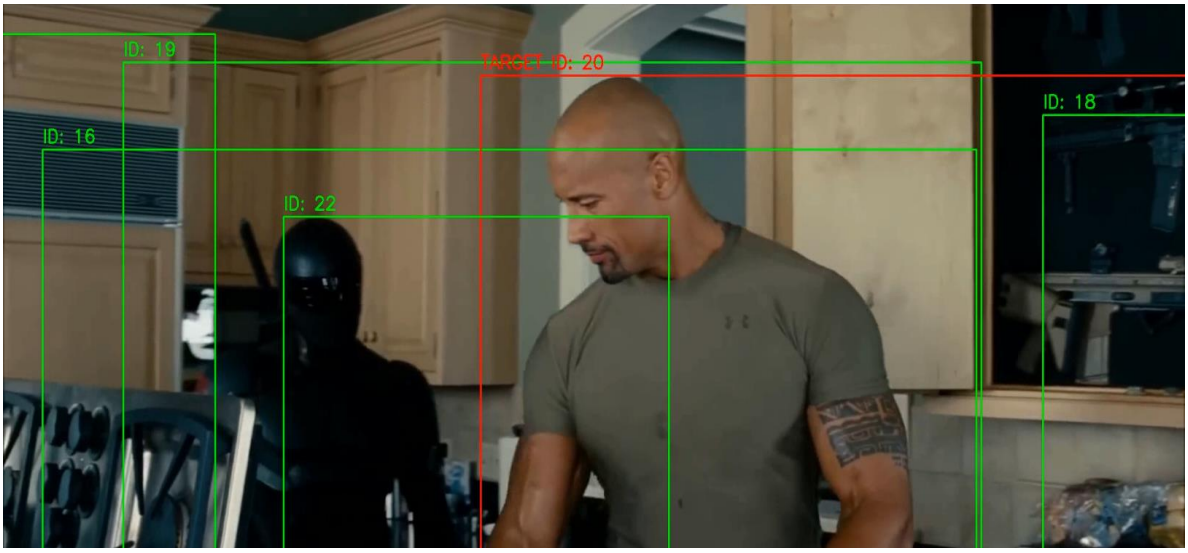


Fig. 4.2 Person Tracking

Challenges and Future Scope

Although the YOLOv8 and DeepSORT integration has facing issue but the system faces ongoing challenges with environmental variability. Lighting conditions, occlusions, and the presence of non-standard weapon appearances can impact detection accuracy. Future research aims to address these issues by integrating multimodal data, such as thermal and audio inputs, to enhance the detection accuracy in challenging settings.
Additionally, optimizing the system for edge device deployment could further reduce latency and power consumption, making it suitable for deployment on security cameras and other constrained hardware. Another focus for improvement is reducing the model's computational load, which would enable faster processing without compromising accuracy. These enhancements could significantly improve the scalability and effectiveness of the system, particularly in high density public areas and resource-limited environments

- **Biometric Fusion for Stronger Verification**
  Integrating multiple biometrics like facial recognition, voice, and iris scanning will enhance accuracy and reduce spoofing or impersonation risks.
- **Mobile App Integration for Remote and Hybrid Learning**
  A secure facial recognition mobile app can allow students to mark attendance remotely, supporting online classes and blended learning models.
- **Smart Classroom and IoT Integration**
  Connecting the system with IoT devices (e.g., smart doors, seating sensors) can provide context-aware attendance verification and automate entry-based tracking.
- **Liveness Detection with 3D Face or Depth Estimation**
  Enhancing the system with depth-sensing or anti-spoofing techniques (e.g., blinking detection, 3D face maps) can prevent fraud using printed photos or videos.
- **Cloud-Based Real-Time Analytics Dashboard**
  A centralized dashboard for teachers and administrators to monitor live attendance, flag proxies, and analyze student behavior across sessions or classes.

Conclusion

In this research, an effective and real-time solution for weapon detection and person tracking has been developed by integrating the YOLOv8 object detection model with the DeepSORT tracking algorithm. YOLOv8 was chosen for its high accuracy and processing speed, making it well-suited for surveillance and security applications. Its ability to perform webcam-based detection enhances the system's adaptability across diverse environments. By incorporating DeepSORT, the system ensures continuous tracking of individuals carrying weapons as long as they remain within the camera frame.

This integrated approach demonstrates strong potential for real-world deployment in smart surveillance systems to enhance public safety and threat detection.

REFERENCES

1. T. Diwan, G. Anirudh, and J. V. Tembhurne, ”Object detection using YOLO: challenges, architectural successors, datasets and applications,” 2021. DOI: 10.1007/s11042-022-13644-y
2. M. T. Bhatti, M. G. Khan, M. Aslam, and M. J. Fiaz, ”Weapon detection in real-time CCTV videos using deep learning,” IEEE Access, 2021. DOI: 10.1109/ACCESS.2021.XXXXX.
3. N. Wojke, A. Bewley, and D. Paulus, ”Simple online and realtime tracking with a deep association metric,” in Proc. IEEE Int. Conf. Image Processing (ICIP), 2017. DOI: 10.48550/arXiv.1703.07402.
4. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ”You Only Look Once: Unified, real-time object detection,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016. DOI: 10.1109/CVPR.2016.91.
5. A. Demidovskij et al., ”OpenVINO deep learning workbench: Towards analytical platform for neural networks inference optimization,” J. Phys.: Conf. Ser., vol. 1828, 2021.DOI: 0.1088/1742-6596/1828/1/012012.
6. S. Kumar et al., ”Automatic weapon detection in surveillance videos: State-of-the-art, challenges, and future directions,” ACM Computing Surveys, 2021. DOI: 10.4018/IJSKD.2020010103.