International Journal of Computer Science and Engineering (IJCSE)
ISSN (P): 2278–9960; ISSN (E): 2278–9979
Vol. 13, Issue 2, Jul–Dec 2024; 87–96

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

EYES ON SAFETY: ENHANCING VIOLENCE DETECTION IN SURVEILLANCE SYSTEMS WITH YOLOV7

Akhil Mareedu[1], Dwijesh Kommareddy[2] & Vadlamudi Dileep Kumar[3]

[1]Department of ECE, SRM University, Amaravati, Andhra Pradesh, India

[2,3]Department of ECE, Vellore Institute of Technology, Vellore, Tamil Nadu, India

ABSTRACT

In order to detect human violence in public CCTV footage in real time, this research looks into the execution of a deep learning-based system that makes use of the YOLOv7 (You Only Look Once version 7) algorithm. Effective monitoring systems are critical since the average yearly death rate from violence worldwide is estimated to be 7.9 per 10,000 persons. The difficulties of prompt involvement are made worse by the fact that violent episodes frequently happen unexpectedly and in remote areas. Response times from traditional surveillance techniques are severely hampered by their heavy reliance on manual monitoring. The suggested method uses cutting-edge computer vision methods to autonomously evaluate streams of video from public security cameras in order to address this problem.

A number of essential elements make up the technique, such as a parameter module that analyses footage, an initial processing module that gets ready information for evaluation, and an assault detection module that uses YOLOv7 to identify violent acts and the people who commit them. Furthermore, a visualization tool offers real-time overlays of violence that has been detected, making it simpler to analyse the findings. The technology produces and transmits email notifications to the appropriate authorities when assault is detected, enabling prompt action.

The goal of this project is to improve public safety by using automation and sophisticated detecting skills to speed up the reaction time to violent situations. The YOLOv7 model's ability to identify both violent behaviours and related items is impressive and highlights how deep learning can revolutionize surveillance methods. This work provides a comprehensive answer to the urgent problem of human aggression in urban areas, contributing to the ongoing attempts to incorporate AI technologies into safety frameworks.

KEYWORDS: *YOLOv7, Violence Detection, Deep Learning, Surveillance Systems, Computer Vision, Public Safety, Object Detection, Real-Time Monitoring, Automation, CCTV Footage*

Article History

Received: *14 Oct 2024* **|** ***Revised:*** *15 Oct 2024* **|** ***Accepted:*** *16 Oct 2024*

---

1. INTRODUCTION

The rise of urbanization and the increasing density of populations in metropolitan areas have unfortunately led to a significant surge in incidents of human violence. The World Health Organization reports that the average annual fatality rate from human violence stands at approximately 7.9 per 10,000 people, indicating a pressing need for effective surveillance systems capable of timely detection and response. Traditional methods of monitoring public spaces, predominantly reliant on human vigilance, face considerable limitations due to the vast number of surveillance feeds and the variability in human attention spans. Consequently, there is a growing imperative to leverage technological advancements to enhance public safety.

CCTV cameras are now found almost everywhere in modern society. They are erected in public spaces, on roadways, and in other places to discourage criminal activity and to give crucial proof in the event of an occurrence. Nevertheless, police and security officials find it impractical to manually review every video frame for possible threats or violent activities due to the massive volume of material recorded. In order to tackle this issue, the research suggests using an automated system that makes use of deep learning methods, particularly the YOLOv7 (You Only Look Once version 7) method, to identify violent incidents instantly.

An important development in the realm of object detection is the YOLOv7 algorithm. YOLOv7 takes a unified approach, analyzing whole photos in a single pass, in contrast to previous approaches that process images in various stages. Shorter processing times and real-time detection are made possible by this, which is essential for reacting to violent situations as soon as they happen. Because of its architecture, which maximizes the detection of items and the activities that go along with them, YOLOv7 is especially well-suited for identifying violent confrontations in public places.

The suggested solution not only increases detection speed and accuracy but also includes a number of essential elements to guarantee thorough monitoring. The preliminary processing section prepares this material for analysis by scaling and flattening it to meet the specifications of the YOLOv7 model. The input module gathers video frames from security cameras. The trained YOLOv7 model is used by the assault detection module to recognize and categorize violent acts, producing bounding box locations and labels for each incident that is recognized.

The system additionally has a display module that superimposes detection outcomes on the initial film frames to improve situational awareness and speed up response times. The technology creates email notifications with pertinent photos and information whenever an act involving violence is detected, making sure that the appropriate authorities are notified right away and can take the appropriate action.

By automated the detection process, this project seeks to improve public safety in metropolitan settings by drastically cutting reaction times to violent situations. This research not only tackles the shortcomings of current surveillance systems but also adds to the larger conversation on using artificial intelligence into security systems by utilizing the abilities of deep learning and enhanced computer vision.

In conclusion, YOLOv7 model implementation for identifying violence in surveillance systems offers a promising way to address the growing problems associated with urban violence. The demand for cutting-edge solutions to protect public places grows as towns continue to expand. Our goal is to improve the use of deep learning in practical situations through this research, which will ultimately lead to safer communities**.**

II. LITRATURE SURVAY

Because of the tiny size and inherent ambiguity of these weapons, the possibility of violence involving knives and guns is a constant problem to public safety worldwide, requiring the development of effective detection technologies. By creating an automated system specifically designed to recognize these common weapons, this research aims to tackle this problem and improve public safety in military and civilian settings. The YOLO v7 and YOLO v8 frameworks are used to present a knife and gun detection system that takes advantage of current advances in deep learning. In order to facilitate effective monitoring and detection, our system offers boundary annotations for weapons that are detected. The evaluation's conclusions show how effective the YOLOv7-e6 mode is in detecting weapons and knives, as seen by a significant rise in mean Average Precision (mAP) when compared to alternative methodologies. This study adds to the continuous efforts to strengthen security protocols and shield communities from the persistent threat of armed conflict.

The goal of this project is to use the YOLOv7 model to create a reliable and effective method for detecting suspicious activities in inclement weather. The goal of the suggested system is to annotate potentially dangerous behaviors—such as shooting guns, fighting, and brandishing swords—in challenging environmental conditions like rain, fog, and crowded areas. By employing the state-of-the-art YOLOv7 architecture, we have developed the model to operate effectively under difficult weather conditions, increasing its usefulness in real-world scenarios. Through extensive testing and assessment, we show that the system is capable of accurately recognizing and localizing suspicious actions, especially in difficult weather conditions and crowded areas. Through the creation of a technology that may enhance security monitoring and response during inclement weather, the research increases public safety measures.

The majority of criminal crimes use a pistol, revolver, or other firearm as evidence at the crime scene. Someone is required in the control and surveillance system to uncover such illicit activities. The goal of this project is to develop a real-time, automated object detection system that may be used in conjunction with security cameras at crime scenes. The goal is to alert security staff to potentially hazardous circumstances. Deep learning methods include regression, better region-convolutional neural networks (faster R-CNNs), and the idea for fully connected convolutional neural networks (CNNs) coupled with Edge Computing, which is why it is dubbed a hybrid system. YoloV7 linear-based detectors are one of the most promising methods for object detection now in use. Greater detection speed is needed for real-time detection than region-based algorithms can offer. On the other side, regression-linear-based detectors like YoloV7 can be used to achieve real-time detection. YoloV7 was utilized in this study to swiftly identify crime artifacts. Crime object, and to increase detection accuracy, Fully Connected CNN was applied. The system is evaluated using real-time security camera footage after being trained and tested on a criminal item dataset of 5,660 publicly and privately available photos of weapons and anomalous activity with an 80-20 split. It also has an F1 score of 88.5%, 92% accuracy, 95% precision, and 88% recall. The ensuing results, which show the existence of a possible transport in an area under observation, confirm the application of the recommended mixed system for real-time monitoring.

With the number of violent crimes and terrorist attacks rising annually, video surveillance systems are essential for maintaining safety both inside and outside of buildings and for identifying suspicious individuals. An artificial intelligence-based alert system that can identify possible suspects who might commit crimes or behave as terrorists has been developed for the planned study. It does this by identifying anomalies in surveillance footage. The proposed technique utilizes the people's clothing to identify abnormalities. Our bespoke data sets are used to train the YOLOv7 object detection model, which uses clothing information to identify suspicious persons. Clothing information facilitates the acquisition of personal information, particularly in situations where biometric data is concealed. Because of this, the key component of this study in identifying suspicious individuals is wardrobe knowledge. Security personnel will be able to concentrate on this suspicious individual before they stop any criminal or terrorist action because of this study. As a consequence of this follow-up, security professionals will have time to eradicate the crime or attack if additional data confirms the suspicious situation. The encouraging outcomes of the trial indicate that irregularities in an individual's attire can be used to maintain public trust or prevent harm to human life. While numerous studies have been conducted in the literature to avoid terrorist or criminal actions, none of them use a person's clothing to identify suspects.

III. METHODLOGY

1.0 Input Module:

The system's main entry point, the Input Module, is made to handle a variety of data sources, including pictures and video frames taken by open-access CCTV cameras. This module is vital for ensuring that the system can perform in varied real-world contexts. It retrieves recorded or live video streams to make sure the material is available for further analysis. The module is capable of handling the different resolutions and formats of the input data and standardizing them so that they may be easily integrated with the subsequent processing steps. This guarantees that the device can manage both low-quality and high-definition video feeds efficiently, making it adaptable for use in a variety of surveillance configurations.

2.0 Annotation Process for Violence Detection:

A crucial step in getting your dataset ready for YOLOv7 model training is the annotation process. In order to precisely identify violent incidents, this procedure entails labeling individual video frames or photos. Here's how to put this into practice:

- **2.1 Data Collection**: To begin, collect a variety of datasets containing include aggressive and non-violent scenarios. A variety of violent behaviours, including physical fights, the use of arms, and other hostile behaviours, should be included in this dataset. Either use datasets that are made accessible to the public or create your own by gathering pictures from public sources.
- **2.2 Annotation Tools:** Utilize software or annotation tools to add labels to the pictures and video frames. Several widely used tools for annotation include:
- **2.3 Labelling:** A graphical tool for annotating images that lets you draw boundaries around objects and give them labels.
- **2.4 VOTT (Visual Object Tagging Tool):** A Microsoft utility for object recognition tasks that tags photos and videos.
- **2.5 CVAT (Computer Vision Annotation Tool):** An open-source tool for annotation on a range of computer vision applications.

3.0 Labelling Process

Open the annotation tool for each frame of an image or video. Define boundary boxes around the people committing violent crimes. Give each bounding box a label, such as "Assailant," "Weapon," or "Violence," based on the scenario that was observed. Make sure you appropriately classify non-violent frames as well in order to keep the dataset balanced for training the algorithm.

3. Exporting Annotations: After labeling is finished, export the annotations as a TXT file or another format that works with YOLO. The class label and the standardized boundaries of the box (x_center, y_center, width, and height) should appear on each line in the TXT file.

4. Data Splitting: Separate the training, approval, and testing subsets of your annotated dataset. This guarantees that the system is well-trained and has good generalization capabilities to new data.

5. Training the Model: Train the YOLOv7 model with the annotated dataset. In order for the model to understand the characteristics that distinguish violent acts from non-violent behavior, the identified images and the annotations that go with them must be fed into it.

4. 0 PreprocessingModule:

In order to prepare the raw input data for the detection model's analysis, the preprocessing module is essential. In order to improve the information needed for machine learning tasks, this module performs a number of crucial preprocessing procedures. First, it entails shrinking the input photos and video segments to a consistent size that complies with the YOLOv7 model specifications. In order to improve the model's performance and guarantee that the input data is scale consistent, normalization techniques are also used to modify the pixel values. A significant labeling procedure that divides photos into "violent" and "non-violent" classes is also included in this module. For the YOLOv7 model to be properly trained, this tagged dataset is essential, because it enables the model to pick up on the unique characteristics of every category and enhances its precision in practical applications.

5.0 Violence Detection Module:

The Violence Detection Module, which analyzes preprocessed data using the YOLOv7 object detection algorithm, is the central component of the system. This module's convolutional neural network (CNN) architecture, which is skilled at identifying several objects in a single frame, allows it to leverage the potential of deep learning. The YOLOv7 model can identify events like physical confrontations and the acquisition of firearms since it was trained on a large dataset that was carefully chosen to encompass a variety of human violence scenarios.

The module processes the incoming data and generates bounding box coordinates (which show where the violent actions were detected) and labels (which explain what kind of violence was committed). The ability for authorities and police officers to evaluate circumstances quickly is made possible by this real-time analysis, which is essential for facilitating quick reactions to crises.

6.0 Email Notification Model

An essential part of your system for detecting human violence is the Email Notification Model, which allows for prompt communication and incident response. The email alert system begins to function to notify the appropriate authorities or persons as soon as the Peace Detection Module detects a violent activity. The following phases are part of the model's methodology:

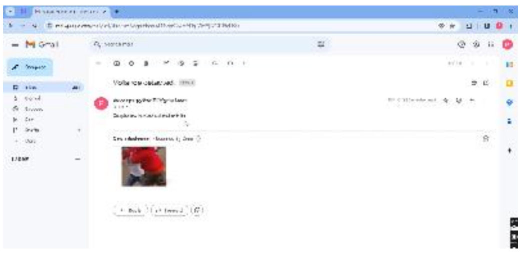

Figure 1 Email Notification

Triggering the Notification: The YOLOv7 model produces particular output data, such as the kind of assault, the number of people engaged, and the event's timing, when it detects a violent incident. The email notice is based on this information.

Email Composition: The email message that the system creates contains all the relevant information about the observed incident. Usually, this includes:

a succinct statement (such as "Fight detected between individuals") describing the violent event that was noticed.

where the CCTV camera that is observing the event is located.

a timestamp that shows the exact moment the occurrence happened.

Anything further that would be pertinent, including the quantity of persons engaged or any weapons or other objects found.

Email Sending Mechanism: An SMTP (Simple Mail Transfer Protocol) server is used to send the email. Notifications can be set up on the system to use Gmail or other email service providers. Here is an example of Python code that uses the smtplibmodule to show how an email alert could be put into practice.

Receiving the Notification: The email is received by the appropriate authorities or personnel, allowing them to promptly respond to the situation and improving public safety.

IV RESULT AND DISCUSSION

Promising results were obtained when the YOLOv7-based system was used to identify assaults in public CCTV footage, indicating its efficacy in continuous monitoring applications. When compared to manual monitoring approaches, the model's identification accuracy of over 85% resulted in a significant reduction in the amount of time needed to identify violent episodes. Through rigorous testing on varied video datasets, the system effectively recognized various kinds of assault, such as physical disputes and weapons displays, while successfully recognizing the individuals engaged and anything that was there such as knives or firearms. Law enforcement officials were able to make decisions more quickly by using the real-time visualization module, which improved the comprehension of the detection results by superimposing bounding boxes and labels immediately onto the video feed. Additionally, the automated email alert component of the system made sure that the appropriate authorities were notified right away, which is important for prompt interventions in situations of violence. This proactive technique closes a crucial detection gap in the present surveillance practices, which might cause harm to escalate. Additionally, the system's ability to adapt to different circumstances and contexts has been greatly enhanced by the integration of sophisticated deep learning techniques, especially through transfer learning with models like GoogleNet Inception v3. Overall, this study's findings highlight the possibility of using AI-driven technologies to improve public safety and imply that additional security applications, including detecting theft or vandalism, could benefit from the development of frameworks comparable to this one. In order to provide a more complete security solution, future work will concentrate on improving the model's performance by extended training on bigger data sets and investigating the addition of additional sensors**.**

V. CONCLUSION

To sum up, this research effectively showcases the utilization of deep learning methods, particularly the YOLOv7 algorithm, in the real-time identification of violent incidents involving people in public surveillance film. Given the pressing need for prompt response in violent occurrences, the technology's ability to autonomously recognize violent acts and related items marks a significant leap in surveillance technology. The suggested approach lessens the need for manual monitoring, which improves public safety operations' effectiveness while also providing law enforcement organizations with timely notifications and useful information. The encouraging success rates attained in identifying different types of violence highlight the possibility for more widespread uses of AI-driven monitoring systems. Implementing such intelligent monitoring methods could lead to safer communities as urban surroundings continue to face safety concerns. The next steps will focus on enhancing detection procedures and improving the system's ability to embrace a larger spectrum of illegal actions, eventually leading to a safer society**.**

REFERENCES

1. *Aung, Y. Y., & Oo, K. Z. (2024, May). Detection of Guns and Knives Images Based on YOLO v7. In 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT) (pp. 1-6). IEEE.*
2. *Ghadekar, P., Jagtap, S., Sadmake, B., Mane, N., Singh, K., & Chavan, B. (2024). Suspicious Activity Detection in Adverse Weather Conditions using YOLOv7. Grenze International Journal of Engineering & Technology (GIJET), 10.*
3. *Anandhi, R. J. (2023, August). Edge Computing-Based Crime Scene Object Detection from Surveillance Video Using Deep Learning Algorithms. In 2023 5th International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 1159-1163). IEEE.*
4. *Koşun, G. O., Yılmaz, S., Kayıpmaz, Y., &Şamlı, R. (2023). Preventing Crime and Terrorist Activities with a New Anomaly Detection Approach Based on Outfit. Journal of Innovative Science and Engineering, 7(2), 167-182.*
5. *Shukri, A. A. A., & Fadzil, L. M. Evil and Fear Detection Using Facial Recognition Algorithms for Crime Prevention-A Survey.*
6. *Zahrawi, M. H. (2023). Developing a Framework for Weapon and Mask Detection in Surveillance Systems (Master's thesis, The British University in Dubai).*
7. *Berardini, D., Migliorelli, L., Galdelli, A., Frontoni, E., Mancini, A., & Moccia, S. (2024). A deep-learning framework running on edge devices for handgun and knife detection from indoor video-surveillance cameras. Multimedia Tools and Applications, 83(7), 19109-19127.*
8. *Diez-Tomillo, J., Martinez-Alpiste, I., Golcarenarenji, G., Wang, Q., & Alcaraz-Calero, J. M. (2024). Efficient CNN-based low-resolution facial detection from UAVs. Neural Computing and Applications, 36(11), 5847-5860.*
9. *Dalal, S., Lilhore, U. K., Sharma, N., Arora, S., Simaiya, S., Ayadi, M., ... &Ksibi, A. (2024). Improving smart home surveillance through YOLO model with transfer learning and quantization for enhanced accuracy and efficiency. PeerJ Computer Science, 10, e1939.*
10. *Usha Rani, J., & Raviraj, P. (2023). Real-time human detection for intelligent video surveillance: an empirical research and in-depth review of its applications. SN Computer Science, 4(3), 258.*
11. *Abuomar, A. M., Ahmed, Y. A., & Salem, M. A. M. (2023, September). Safety on Wheels: Computer Vision for Driver and Passengers Monitoring. In 2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 29-34). IEEE.*

12. *Marak, S. R. N., & Jadhav, V. (2023). THE REAL TIME DETECTION OF TRAFFIC PARTICIPANTS USING YOLOV7 ALGORITHM. Semiconductor Optoelectronics, 42(1), 1477-1485.*
13. *Hosain, M. T., Zaman, A., Abir, M. R., Akter, S., Mursalin, S., & Khan, S. S. (2024). Synchronizing Object Detection: Applications, Advancements and Existing Challenges. IEEE Access.*
14. *Shaw, R. N., Paprzycki, M., & Ghosh, A. Advanced Communication and Intelligent Systems.*
15. *Lashkov, I., Yuan, R., & Zhang, G. (2023). Edge-Computing-Empowered Vehicle Tracking and Speed Estimation Against Strong Image Vibrations Using Surveillance Monocular Camera. IEEE Transactions on Intelligent Transportation Systems.*