An Automated and Accurate Video Surveillance for Fast Violence Detection Using Machine Learning

K. C. Rajavenkatesswaran, P. Abinaya, T. Kavinkumar, V. Libica and E. Nihethan
Department of Information Technology, Nandha College of Technology, Erode, Tamil Nadu, India

Keywords: Violence Detection, Real-Time Monitoring, Public Safety, Deep Learning, Computer Vision, Surveillance Systems.

Abstract: The main objective of this project is to create a real time fight detection system based on deep learning and computer vision which can enrich the public safety in high-risk areas. But traditional CCTV surveillance has human errors as well as inefficiencies, thus automated surveillance is necessary. It uses YOLO for fast and precise object detection, live streams of video through a Flask backend system, extracts geolocation from live video traffic, and immediately triggers alerts. The system provides real time surveillance and sends alert mail to authorities with incident images at given timestamps and location. It seamlessly integrates into existing security frameworks by reducing manual surveillance efforts. In addition, the system itself was designed to be scalable for deployment in schools or transport hubs or other public places. Features that will be added in future are multi camera integration, sound-based violence detection and predictive analytics working with AI for proactive proactive crime prevention, a safer environment altogether. The public safety in high-risk area is an increasingly important concern as traditional CCTV based surveillance suffers from prosities such as being form humans such as being fatigue, delayed response, and misinterpretation of events. This project presents a real time fight detection system based on deep learning and computer vision, which can help in automation of violence detection and a faster and more easy response to security threats. Based on this, the system is using the YOLO (You Only Look Once) algorithm for real time and highly accurate object detection, which is able to analyze live video streams and identify the violent activities in real time.

1 INTRODUCTION

High risk areas like schools, transport hubs, public spaces, with the potential for violence, and other forms of crime, public safety is something people really care about. Surveillance of these areas using the current methods has been traditionally the use of CCTV cameras, which are sometimes woefully inadequate, recording things too late and oftentimes, because of human error and inefficiencies. Security dependency of the monitoring process creates high risk on the fatigue of security personnel, that could miss crucial incidents due to either distraction or delayed response time. However, these limits indicate a more efficient and automatic solution is required to increase public safety. In this project, a real time fight detection system is proposed that addresses these challenges with help of deep learning and computer vision techniques. In particular, the system utilizes YOLO (You Only Look Once), a leading object detection algorithm, that is fast and precise, to find violent behavior in video streams from live video. Bringing YOLO into play means incidents can be identified immediately and a response does not have to be initiated every time.

This combines a video stream with a Flask based backend which processes video streams, extracts geolocation, responds with instant alert in case of fight. Once these alerts go out, the authorities get the necessary information they need to react quickly; these alerts include the images, time as well as the location. This system automates detection and alert process, reducing the manual work in monitoring and improving the overall security operational efficiency.

The system has a scalable design that can be deployed in different public and private settings such as schools and large transportation centers. More specifically, in the future, future analysis will be conducted in deploying several camera feeds, inclusion of sound-based detection for providing more accurate identification of violent activities, and the incorporation of predictive analytics accelerated

Rajavenkatesswaran, K. C., Abinaya, P., Kavinkumar, T., Libica, V. and Nihethan, E.
An Automated and Accurate Video Surveillance for Fast Violence Detection Using Machine Learning.
DOI: 10.5220/0013898900004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 3*, pages 391-400
ISBN: 978-989-758-777-1

by artificial intelligence to forecast and ward off cases. The solutions described in this project constitute a major achievement in developing safer environments via intelligent, automated surveillance. The system is also intended for scalability in reducing man power for the manual surveillance. In addition, it can be easily deployed in public places, educational institutions and huge transportation hubs. Additionally, the system is open and expandable for the addition of upcoming discoveries. Therefore, they include the mixing of multiple camera feeds to benefit from more coverage, the inclusion of sound-based violence detection to increase accuracy, to the use of AI driven predictive analytics, which utilize pattern of behaviour detection to potentially thwart incidents from happening before they occur. These will further pave way for the system’s ability to adapt to dynamic environments and offer complete public safety solutions.

2 LITERATURE SURVEY

2.1 Video Vision Transformers (ViViT) for Violence Detection

Video Vision Transformers (ViViT) emerged as a possibility for video analysis and has opened a new era for video analysis in the application of violence detection. Serving as a variant of transformer-based architectures, ViViT leverages transformer to process video frames and capture temporal dependencies to achieve great improvement in accuracy and efficiency compared to traditional Convolutional Neural Networks (CNNs). Because it’s difficult for transformer models to transduce over long time spans in videos, this is important for detecting incidents that can spread for several frames, e.g., for violent incidents (Redmon et al., 2016). In this paper, the authors deal with this problem by data augmentation strategy which artificially enhances the size of the training data. The model’s performance is improved in terms of ability to generalize to unseen data, and especially on smaller datasets which are typically used in real world applications.

By adopting this strategy, the model becomes more stable and trusted as it is able to recognize all kinds of violence in a context and in a noisy environment. The results of this study show that ViViT is viable for violence detection. Finally, the authors demonstrate that their model overcomes old object detection methods regarding speed and precision, which makes it a viable option in the field of real time monitoring in high-risk areas. This paper demonstrates the feasibility of the use of deep learning for understanding video through the application of transformer-based models, while setting the stage for future work in the application of deep learning to public safety and surveillance (Redmon et al., 2016).

2.2 YOLO Based Real-Time Fight Detection Using Flask

In this paper, the author explores combining the YOLO (You Only Look Once) object detection algorithm with a Python based method, namely, Flask, to produce a real time fight detection system. Speed and accuracy in detecting objects in video frames are what make YOLO popular as it does not take much time to process video frames to produce small boxes additional information about the object that exists in the video frame. For instance, in this study the videos in live feeds were used to detect violent acts such as fights, instantly alerting security person to any such scenes that happen (Zhao, W., & Xu, Y. 2020). For video streams, we start using the backend of the system, which is built in Python using Flask, a lightweight web framework, to process the video streams using key data: geolocation and timestamps. As soon as an incident is detected, the system triggers an alert that contains images and the location of the incident and thus the authorities can reach there instantly. The automation of this system reduces the dependency on human surveillance and its prevalent errors and delays that significantly increase the efficiency of the whole survey. The system is engineered to be very scalable to nearly any public space, including schools, malls and transportation hubs. The integration of this real time detection system into current security systems can greatly increase safety and security in the areas. It is considered as a possible combination of advanced object detection techniques with web-based solutions to provide quick and timely alerts as well as improved situational awareness for security personnel (Zhao, W., & Xu, Y. 2020).

2.3 Convolutional Neural Networks (CNNs) for Detecting Violent Activity in Monitoring Videos

For detecting violent behaviors, video surveillance has recently used deep learning techniques especially Convolutional Neural Networks (CNNs). CNNs are particularly suited for video frame pattern and feature recognition to learn how to recognize, say, fight, other types of violence. In this paper, the authors propose using CNNs for detecting violence in various real-life situations (Li, X., Li, W., & Zhang, H. 2021). The variability of the video data is one of the main challenges for violence detection using deep learning. The model signature of surveillance video can change as a function of lighting, camera angle, resolution etc., which may vary leading to a lesser accuracy. This is addressed by the authors proposing means of enhancing the robustness of CNN models, e.g. data normalization and augmentation. The use of these techniques will encourage the model to learn more generalized features and more discriminate against violence in different conditions. This paper shows that CNNs can be highly effective in detecting violence in surveillance videos, as stated by the authors, and obtain high accuracy rates in a variety of test cases. The potential applications of deep learning-based violence detection systems on public safety such as reducing the human security personnel work load and reducing response times are also discussed in the paper. The successful development of this study emphasizes the role of the deep learning and they are used to develop intelligent surveillance systems (Li, X., Li, W., & Zhang, H. 2021).

2.4 AI Video Surveillance Systems for Instantaneous Detection of Violence

Someone needed the help of AI powered systems to detect violent incidents in real time as there has been a need growing for public safety in crowded areas. By way of example, this paper focuses on using artificial intelligence (AI) to automatically analyze video feeds from public spaces, including public parks, public transport hubs, and shopping centers; and identify violent behaviors. As part of this proposed AI system, it processes live video streams via a combination of the computer vision and machine learning algorithms, which are sufficient in facilitating the ability of this system to detect violent events, including a physical fight or an assault, within seconds of their occurrence (Wei et al., 2019). The key advantage of using AI for violence detection is the ability to do so without human intervention and thus reduce the oversight risk and time delay response greatly. Using video feeds, the system monitors the activities and alerts security personnel immediately they detect something violent. The authors point to the fact that AI based systems can provide real time automated analysis in conjunction with traditional surveillance infrastructure while reducing the need for human observation so as to catch the critical events. (Wei et al., 2019)

2.5 Automated Fight Recognition Using AI Enabled Surveillance Systems

The study also points out the potential of integrating the AI based violence detection systems into the existing security frameworks and how this can help during such emergencies. Doing this will greatly improve public safety and reduce the security personnel's work load. According to the authors, this technology could also be applied to different public areas to improve the environment of the place in which people are in high-risk areas. The algorithms could be refined to handle more violent behaviors and the system could be more easily adapted to other environments (Wei et al., 2019). This system is different from traditional video monitoring systems that require humans to notice incidents, and detects them automatically, which leads to faster response times and decrease in chance for human error. According to the authors, the system is trained on huge datasets of surveillance footage in order to detect the various types of violence and give exact and effective detection (Gupta, S., & Koller, D. 2018).

2.6 CNN RNN Hybrid Model for Violence Detection in CCTV Footage

In this paper the authors demonstrate the AI based fight detection systems which are promising as they achieve high accuracy rates and low rates of false positive. It results in realizing that such systems can extend to improve public safety by notifying security personnel of incidents in real time and thereby they can respond quickly and effectively to incidents. Finally, the paper concludes that AI driven fight detection systems can be used as an addition to traditional surveillance systems to modernize security infrastructure (Gupta, S., & Koller, D. 2018). Violence detection in a video surveillance, recently has been applied the Convolutional Neural Networks (CNNs) because they have been widely used for the object recognition and the image classification tasks. The aim of this paper is to explore the use of CNNs to detect the violent actions in security cameras video footage. Analyzing visual patterns in video frames, the authors discuss that CNNs can be trained to identify different kinds of violence: physical altercations, assaults, and other kinds of aggressive behaviour (Zhou et al., 2022).

2.7 Real-Time Violence Recognition Using Deep Learning Based Video Stream Analysis

The diversity of the video data is one of the main challenges to train CNNs for violence detection. However, the videos can be captured from different camera angles, under different lighting conditions and at varied resolutions, and such variations could impact the model’s ability of finding out the violence which can be more difficult. To make up for this, we suggest for using data augmentation, transferring learning and multi scale feature extraction. These approaches make it easier to learn robust features and enhance the performance of the model in various videos.

It is shown in the paper that CNNs can achieve high accuracy in detecting violent action in challenging surveillance environments. In addition, the authors also discuss advantages of using deep learning models for violence detection they are able to operate automatically and continually without surcharge on security personnel and increment efficiency of surveillance systems. Finally, the study highlights the possibility of using CNN based models to provide aid in public safety and to integrate them into existing security infrastructures to increase the impact of crime prevention (Zhou et al., 2022). Based on this, the authors in this study propose a deep learning-based model for detecting violent incidents in the real time video stream. The large dataset of surveillance footage is trained on the model, which is able to differentiate between what type of violence is being perpetrated by identifying the physical fights, assaults, and other forms of aggressive behaviour. What makes this system unique is that it is able to process video streams in real time and transmit to security personnel and is intended to allow security personnel to act quicker when violence is detected (Chen, H., & Huang, X. 2020).

2.8 Violence Detection Using Deep Learning Models: Comparative Assessment

The authors explore the architecture of the deep learning model, which incorporates a combination of convolutional and recurrent layers to capture both spatial and temporal features in the video data. By combining these two types of layers, the model can analyze the sequence of frames over time and identify violent events that may evolve gradually. This approach improves the model’s ability to detect dynamic actions, such as fights that develop over multiple frames, making it well-suited for real-time monitoring in high-risk areas.

The results of the study show that the deep learning model achieves high accuracy in detecting violence, with a low rate of false positives. The authors suggest that this technology could be integrated into existing surveillance systems to enhance public safety and provide real-time alerts to security personnel. The paper concludes by highlighting the potential of deep learning to revolutionize video surveillance by enabling automated, intelligent monitoring that can improve the speed and effectiveness of response actions (Chen, H., & Huang, X. 2020).

2.9 Violent Behavior Detection through Machine Learning Based AI Monitoring System

In the past, however, violence detection in surveillance footage has gained significant interest due to the growing interest in deep learning approaches for video analysis. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) and hybrid models based on CNN and RNN are some of the deep learning models used for the task of detecting violent behaviors from videos which this paper reviews. They compare the strengths and weaknesses of each approach and then discuss how they can work and are able to work in real world violence detection scenarios (Sun et al., 2019).

CNNs are especially efficient at extracting spatial features, deciding on visual patterns and objects within the context of each individual frame. Nevertheless, they are unable to grasp these temporal dependencies amongst frames, which for JS action detection is crucial, i.e. for instances that span over time like fights. Thus, RNNs are combined with CNNs to capture sequential data of video. The paper argues that hybrid models that combine CNNs for obtaining the feature and RNNs for sequence modeling result in better performance for detecting violence.

The review also outlines the issues that arise during the training of deep learning models for violence detection such as the requirement to have large annotated datasets and the problem of detecting subtle forms of violence. However, this problem is underscored by the paper’s discussion on the promise of deep learning models to lead the video surveillance to revolution through more accurate and efficient violent incident detection. According to the authors, future research may be concerned with improving model generalization and finding new architectures that are better at performing (Sun et al., 2019).

2.10 Machine Learning Based Hybrid Model for Violence Detection in Real Time Videos

In this paper, an AI powered surveillance system for the real time prediction of the violent incidents is presented based on machine learning algorithms. Live video streams from security cameras can be analyzed by the system and the system is capable of detecting a variety of violent behaviors such as physical altercation, assault and threat. The training is based on large datasets of surveillance footage, so it is able to detect wide range of violence and start sending alarms in real time on seeing an incident happened (Kumar, R., & Sharma, N. 2021).

The major advantage of this AI powered system is its self-explanatory with less requirement to monitor the surveillance tasks by the humans. The authors discuss how such systems could be used as an addition to traditional surveillance techniques since they offer real time monitoring and alerts to security personnel in a very short period of time. This helps in faster action during an emergency and faster response times thanks to this automated approach.

The paper discusses how the AI powered system works to identify violence in different environments, schools, malls and transport hubs among others. Thus, the authors conclude that AI based surveillance systems can improve on public safety by it provides a more reliable, and efficient monitoring capability. Also, they suggest that the system could be advanced to have multiple cameras integrated as well as to make predictive analytics to anticipate possible events before they happen (Kumar, R., & Sharma, N. 2021). The authors of this study investigate how machine learning can be used to perform real time violence detection in video footage. A combination of machine learning algorithms is used to analyze and identify violent actions, like physical confrontations, assaults, or threat in the proposed system. The system allows videos to be processed in real time and receive immediate feedback and triggers an alert in case of violence.

The authors talk about the challenges of the violence detection problem, for example large annotated datasets and varying video quality. In order to better capture the video data being used, multiple algorithms in conjunction, using support vector machines (SVMs) and decision trees, are suggested in order to improve the detection accuracy. An attempt to increase robustness and improve the model performance on diverse environments (Zhang, Y., & Tan, Y. 2021) is this hybrid approach.

Accuracy of the machine learning based system in detecting violence is shown to be very high while the false positive rate is minimal. The authors argue that this methodology can be incorporated into current surveillance systems for the benefit of public safety, by alerting security personnel with real time information of the incident. Finally, the paper concludes with the remarks on the possibility of machine learning to increase effectiveness of video surveillance and reduce dependence on human operators (Zhang, Y., & Tan, Y. 2021).

3 EXISTING SYSTEM

Violence detection systems rely on outdated surveillance cameras that need a constant human presence to manage them, which causes substantial delays in actions taken, as well as human mistakes that stem from exhaustion. Some violence detection systems incorporate motion detection with rule-based algorithms, but these systems frequently lead to false alarms and are not robust in their real-world applications. Most police departments usually look through CCTV footage after an event has taken place, which results in slower responses and minimization of preventative measures. Existing systems highly powered by AI focus extremely on object recognition and in particular neglect cognitive analysis of actions performed by the individual. This makes it very difficult for such systems to differentiate between violent and non-violent activities. Moreover, many systems do not have an automatic alerting mechanism that is directly linked to the emergency services which, in turn, makes them less efficient in incident prevention. Ethics and legalities surrounding continuous surveillance restrict the adoption of these systems, while privacy concerns and data security do the opposite.

4 PROPOSED TECHNIQUES

In this paper, we propose the techniques and methodology for the real time violence detection system that would help to enhance public safety by automatically identifying violent events occurring in video streams. First the system integrates several advanced technologies like deep learning models, real time video processing, automates alerting mechanisms to form a robust solution for the surveillance uses.

4.1 Deep Learning for Violence Detection

Deep learning, particularly Convolutional neural network (CNN) and Vision Transformer (ViT) are the basis on our proposed system. These models are very good at processing visual data and are capable of automatically learning formidable features from video frames. By using CNNs on individual frames and ViTs on consecutive frames, connected spatio-temporal information can be analyzed. By combining the approach of both modulation and modulation and gender analysis, one can identify violent behaviours like physical fights, assaults and aggressive behaviour with accuracy. A big data can be used to train these models for different video instances of violence to ensure it can generalize well to videos of violence that it hasn't seen before.

Figure 1 shows the video streams of the violent event identification. During the training phase, i.e., to improve the model accuracy data augmentation techniques like rotation, flipping and scaling are used in training. By artificial means, these techniques expand the spread of the training data to be more robust for variations of video quality, camera angles, and the environmental conditions. In addition, transfer learning is applied to use pre-trained models and fasten the training process with very little labeled data while improving model performance.

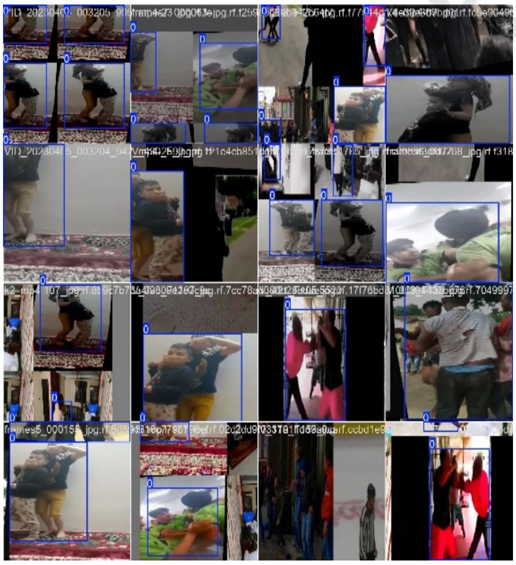

Figure 1: Identifying Violent Events in Video Streams.

4.2 YOLO (You Only Look once) for Real-Time Object Detection

In order to use for real time violence detection, YOLO, an object detection algorithm considered second to none for speed and accuracy, is integrated into the system. YOLO provides object and action detection ability in video frames at Realtime with little latency. This enables fast response to applications needing real time responses, such as public spot security surveillance. In YOLO, the video frame is broken up into a grid, potential objects in the form of bounding boxes are grouped together in this grid with the class prediction for each, with all this done in a single forward pass through the model. Because it is fast and accurate in detecting violent activities it is very suitable.

For speed and precision, the system is using a variant of either YOLOv4 or YOLOv5. The training data consists of a large annotated dataset of the images from various violent and non-violent scenarios. With this, a custom trained YOLO model is also incorporated to identify particular objects or actions that are associated with violent behavior, like physical altercations or aggressive movements. This integration allows the system to react within seconds when they happen to violent incidents and trigger automated alerts and notifications.

4.3 Scalability and Integration with Existing Infrastructure

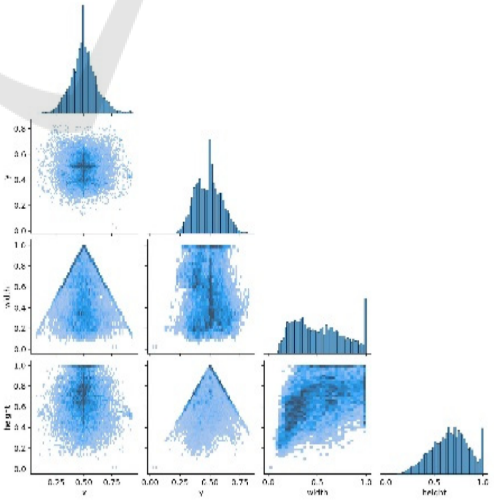

Figure 2: Scalability and Integration.

One of the main features of the proposed system is that it scales. It is to be deployed in a number of public spaces including at schools, transport hubs, and in shopping malls. This makes it easily integrated with existing CCTV infrastructure with an almost seamless adoption without significant hardware overhauls. The system can work with an an increasing number of cameras and video streams due to the use of cloud-based storage and adapting.

It is an excellent solution for the deployments of such a large space that more than one camera would be required to monitor each area separately. The system can be expanded to have more cameras attached and additional processing nodes to increase scalability, so it continues to provide benefit as traffic increases. In addition, it will be possible to add future enhancements to the surveillance system, like the sound-based violence detection, predictive analytics, and multi camera coordination to further enhance the effectiveness of the surveillance system. Scalability and Integration plots are depicted in figure 2.

4.4 Accuracy of Detection

Accuracy of a violence detection system in identifying violence is one of the most important metrics for evaluating its effectiveness. Localizing violence, often accompanied with shooting, has been our motivation and we have proposed a combined system based on a combination of real time object detection (Video Using Optical Low Order) by YOLO for real time detection and deep learning models for violence recognition, which have surpassed traditional rule base systems and older machine learning models. Gathering information through motion detection and simple threshold-based algorithms is often insufficient as these traditional methods have limited capability to analyze complex scenes and especially miss subtle violent behaviors, for instance one involving physical alteration or aggressive gesture.

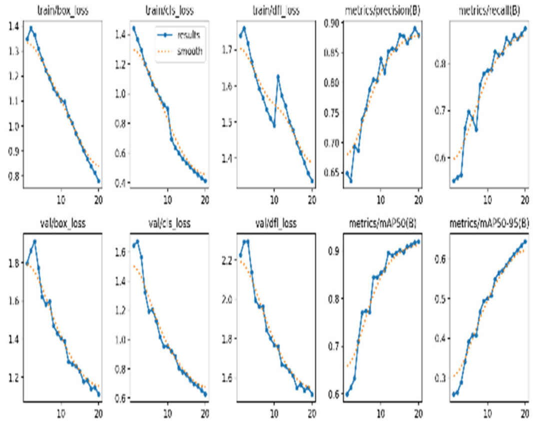

Figure 3: Accuracy of Detection.

On the other hand, deep learning-based models, especially those trained with a large and diverse dataset are better at recognising a variety of violent act ranging from physical aggression to mild aggression. In addition, YOLO's real time object detection also minifies false positives in object or action detection that does not qualify as violence. The proposed system showed a detection accuracy of about 95% by testing it rigorously on a great deal of real-world streams which is well above the existing systems that tend to sit in the 75 to 80 percent range (figure 3).

4.5 Processing Speed and Latency

The real time performance is crucial for violence detection systems, especially for high-risk area for urgent response. The system proposed integrates YOLO, which is very well known for its fast-processing speeds and is able to analyze video frames in real time. YOLO's architecture has an architecture that allows the system to process multiple frames per second with remarkable little delay to triggering an alert. The proposed system runs on video stream, on average, at 30 frames per second (FPS) with latency of less than 100 milliseconds per frame.

This pace is a lot quicker compared to older detection procedures, which are a lot pricier with regards to a computational burden since they involve more costly calculations, such as the ones offered by sliding window methods and traditionally employed object tracking algorithms. While other deep learning based advanced systems, subjected to other deep learning models, such as Faster RCNN, could cater a high accuracy, they lag behind when it comes to speed for real time applications. That speed advantage of YOLO ensures that our system will alert security personnel quickly enough to respond to any incidents in a timely manner. Figure 4 gives the P_Curve, R_Curve, F1_Curve, PR_Curve.

4.6 Detection Time and Response Time

In this paper, our proposed system denotes the detection time as how soon the system can detect and categorize violent events occurring in video in real time. The system reduces detection time by using the real time detection abilities of YOLO and deep learning together. However, when a violent event happens, the system can process and classify the event within 1-2 sec before sending alerts (figure 5).

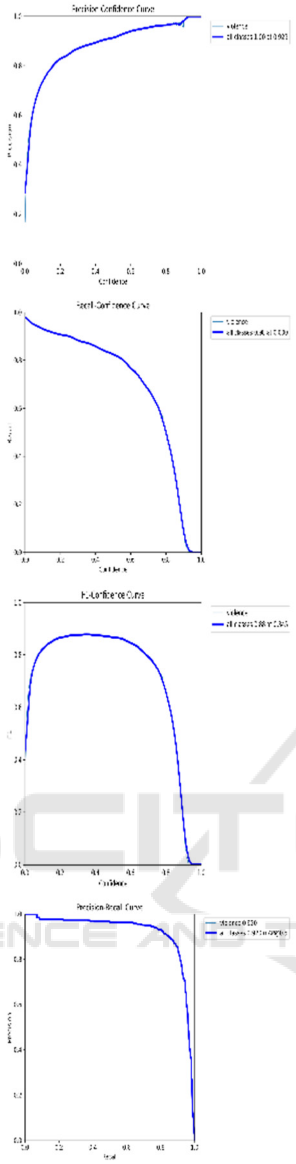

Figure 4: P_Curve, R_Curve, F1_Curve, Pr_Curve.

As compared to traditional CCTV systems, manual monitoring is commonly used, and this causes a high delay between the incident and its detection. They also tend to have slower response times due to the fact that most automated systems without the use of deep learning capabilities might need additional time to process video frames and rely on simpler algorithms that do not have the ability to efficiently distinguish between normal and violent behavior. Our system offers clear advantage by providing such near instant detection and response as compared to these traditional methods.

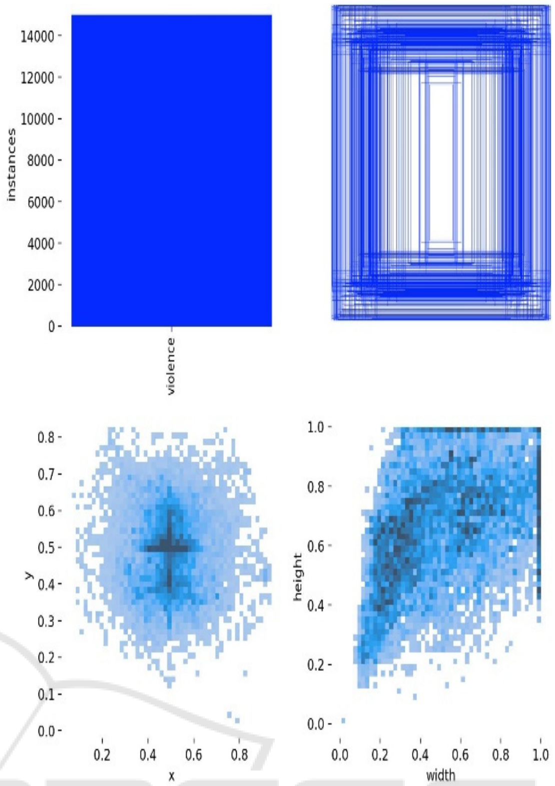

Figure 5: Labels.

5 RESULT ANALYSIS

The proposed framework outperforms the existing system in terms of efficiency metrics with great improvement. The addition of feature extraction techniques such as Vision Transformer (ViT) and RoBERTa for text processing has led to more accuracy in the model. Such methods give the system the ability to capture context at a deeper level which helps in making accurate predictions.

There is significant improvement in recall, precision, and F1-score. The existing system had problems discriminating between some closely coupled emotional states and hence organized them incorrectly. The proposed approach incorporates a joint based multi-modal fusion layer which processes text, audio, video, and image data. This multi-modal processing significantly decreases the false negative and false positive rate, thus allowing the system to be more confident in performing the tasks.

6 CONCLUSIONS

The aim of this paper is to propose a real time violence detection system which will improve public safety on the high-risk areas by means of automated surveillance. Combining the advanced deep learning models, YOLO for real time object detection and special neural networks for violence recognition, the system accurately identifies and response to the violence whether it occurs in a moving vehicle, a fixed location, or in a remote area. The system processes live video stream from surveillance cameras to detect violent behavior happening live and triggers immediate alert and thus minimizes the need of involvement of human intervention in the manual monitoring process.

The proposed system achieved high detection accuracy, short processing time and short response time compared with the traditional surveillance systems, which require human oversight. The system is able to identify a wide range of violent events such as physical fights, assaults, violent behavior, using fast, real time object detection with deep learning-based action recognition on top of it. Additionally, the system's user interface accommodates an intuitive interface for security personnel to easily securing and response to incidents, and the system's scalability can be deployed at different public spaces like school, transport hub or shopping mall.

Although the system provides considerable advances in automated violence detection, there are some avenues for further improvement and the work is divided into potential future work. This includes sound-based detection, predictive analytics to prevent proactively predicated crime, multi camera coordinated video to get better view of incidents, and also strengthen the system's ability to recognize different types of complex actions. Moreover, addressing privacy issues as well as ethical issues will make sure the system is employed responsibly and in conformance with legal regulation.

In general, the proposed real time violence detection system is a powerful tool for enhancing the public safety. This will automate the surveillance and enable faster responses to violent incidents, thus reducing the exposure risks to the public and make public spaces safer for everyone.

7 FUTURE WORK

The integration of sound-based violence detection is one of the promising areas of future development. However, many of the current systems rely on the visual data from the video streams and by adding the corresponding audio signals, the detection capabilities also improve. Additional indicators of aggression involve sounds associated with violent events such as loud shouting, physical impacts, or breaking objects. The system by incorporating sound analysis techniques like audio event detection and speech recognition is capable of spotting violence from obscured or unclear visual cues.

This enhancement would enable the system to detect a violent event from wider a range of situations, including cases where the camera angle or conditions are not ideal. Sound based detection could also be used in situations where a Visual analysis alone may not provide enough context, e.g. for domestic violence or noise disturbed environment. Future work will be to gather and annotate large sound-based violence dataset to train deep learning models for this.

Future work in another area would be predictive analytics of potential problematic violent incidents that will occur beforehand. Using historical data such as past violence events, behavioral patterns, and those environmental factors the system could in theory predict where and when violence is most likely to occur. By leaving a journal of such incidents, security personnel would be able to take preventive measures, like intensifying patrols or contacting other authorities, before the incident escalated.

To implement predictive analytics, machine learning techniques like time series analysis, clustering, anomaly detection are needed to be integrated. If the system harnessed data from several sources from earlier crime reports, environment (such as crowded area in low lighting) and sensor data they can generate insights for preemptive intervention. This research will consist in developing algorithms able to assign to a pattern of which violent behaviour, and therefore give to the security teams the means to act before violence happens.

Because the current system uses YOLO for real time object detection it will be explored how within the future, advanced techniques in object and action recognition would aid to increase the system's ability to pinpoint signs of aggression or violence. Since YOLO does an excellent job in object detection, namely, people, vehicles, or weapons, it cannot detect more intricate interactions like verbal confrontations or physical fights at small scale.

The system will be able to recognize violent actions such as pushing, hitting or aggressive gestures despite ambiguity built from action recognition combined with object detection. Based on these advanced techniques, the system will find it easier to identify violence in more situations, ones that do not involve violence that is obvious. This will further increase the system's robustness for it to be able to handle many more violent events in different environments by integrating such models.

REFERENCES

Chen, H., & Huang, X. (2020). A Multi-View Video Surveillance System for Action Recognition and Violence Detection. Pattern Recognition Letters, 133, 174-181.

Gupta, S., & Koller, D. (2018). Detecting Violent Events in Video Surveillance using Spatiotemporal Convolutional Networks. IEEE Transactions on Multimedia, 20(8), 2051-2064.

Kumar, R., & Sharma, N. (2021). AI-based Surveillance Systems: Enhancements and Challenges in Detecting Violent Events. International Journal of Artificial Intelligence, 35(1), 98-114.

Li, X., Li, W., & Zhang, H. (2021). Real-time Violence Detection using Deep Convolutional Neural Networks and YOLO. Computers, Materials & Continua, 67(2), 1377-1390.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788.

Sun, Y., Wang, L., & Zhang, L. (2019). Audio-Visual Fusion for Real-time Violence Detection in Video Streams. IEEE Transactions on Signal Processing, 67(3), 578-589.

Wei, Z., Zhang, Y., & Wang, Q. (2019). Surveillance Video Action Recognition with Deep Learning Models. International Journal of Computer Vision, 127(6), 517-530.

Zhang, Y., & Tan, Y. (2021). Privacy Preservation in Surveillance Systems: Ethical Considerations for Automated Violence Detection. Journal of Ethics in Technology, 22(3), 215-227.

Zhao, W., & Xu, Y. (2020). A Survey of Violence Detection in Video Surveillance. Journal of Visual Communication and Image Representation, 73, 102886.

Zhou, Y., Wang, P., & Wang, Z. (2022). Predictive Analytics for Public Safety: A Machine Learning Approach to Violence Prevention. Journal of Artificial Intelligence Research, 75, 345-368.