SentinelEyes: Violence Detection System

Dhruv Mistry [1], Sahil Deshmukh [2], Shubh Joshi [3] and Chitra Bhole [4]

[1, 2, 3, 4] K. J. Somaiya Institute of Technology, Sion, Mumbai, Maharashtra

Abstract: Automated surveillance systems play a critical role in detecting and responding to violent conduct in order to maintain social stability and public safety. Security protocols can be improved by utilizing technological innovations like deep learning algorithms and surveillance systems.This research introduces an innovative system focused on enhancing public safety by deploying a real-time violence detection mechanism. Using the MobileNetV2 architecture, our system efficiently analyzes live CCTV feeds, simultaneously identifying potential instances of violence. Integral to our approach is the immediate dissemination of alerts through a dedicated Telegram bot, providing authorized entities with essential details, including date, location, and a corresponding image. This system aims to contribute significantly to the improvement of public safety infrastructure by facilitating rapid responses to security incidents.

Keywords: MobileNetV2 ,Violence Detection ,Alert System ,Yolov8 ,Telegram bot

1. Introduction

It is critical to address aggressive behavior in public spaces because it not only endangers the safety of the community but also threatens societal cohesiveness and economic growth. Beyond just causing immediate physical harm, violence has a negative impact on property values, productivity, and the provision of critical social services.
Recent advances in deep learning techniques have led to increased interest in automating the identification of violence in video footage. These techniques are particularly good at extracting spatiotemporal data from video streams, capturing motion dynamics between consecutive frames as well as spatial information from individual frames in real-time situations. Real-time violence detection systems reduce the need for laborious manual review by quickly identifying violent activity by utilizing deep learning.
Because security cameras are so common, this solution integrates seamlessly with existing CCTV infrastructure, harnessing the power of the MobileNetV2 architecture to elevate surveillance capabilities.
In this paper, we apply MobileNetv2, a cutting-edge deep learning architecture known for its accuracy and efficiency, to the design of a Real-Time violence alarm system.
OpenCV is used to convert the real-time video frames into images in the first stage .The recognized frames are then subjected to enhancing procedures to increase visual sharpness and detail, guaranteeing precise interpretation by authorities.MobilenetV2 analyzes the video frames to predict any violent behavior through the updated parameters and sends an alert on prediction of a violent frame above the threshold.

Initially, we utilized the YOLOv8 model for real-time violence detection due to its high accuracy. However, recognizing the need for a more lightweight solution, we transitioned to the more simpler MobileNetV2 architecture. This shift enhances computational efficiency while effectively identifying violent activities in real-time, aligning with our goal of seamlessly integrating with existing CCTV infrastructure with great accuracy.

A noteworthy facet of our system is the incorporation of a Telegram bot, functioning as an instantaneous alert conduit. This feature ensures that when a violent occurrence is detected authorized entities responsible for public safety receive prompt notifications containing crucial details – the precise date, location, and a visual snapshot encapsulating the nature of the identified incident. This rapid information dissemination equips security personnel to respond swiftly and decisively, minimizing potential risks and cultivating an environment of enhanced safety.

Our Real-Time Violence Detection System isn't just a regular deep learning model, it's about making our public spaces safer. As we dive into how it works and why it matters for everyone's safety, we're excited to share a system that goes beyond just being smart. It's all about keeping our communities secure and starting important conversations about how we can make the world a safer place for everyone.

2. Literature Review

Rising rates of violence are posing an increasing threat to public safety in urban areas across the globe. To tackle this issue, researchers are looking into deep learning methods, which are a branch of artificial intelligence, in sophisticated surveillance systems. By allowing the real-time identification and categorization of violent occurrences, these methods have the potential to strengthen law enforcement and enable quicker and more efficient response.

Numerous deep learning applications for violence detection have been studied. A ResNet-based method for video streams was proposed by Shripriya [1]. In this method, the video is first supplied concurrently to a single-shot detector and the ResNet model. One thousand videotapes of hockey games made up the hockey fighting dataset that was employed.

Additionally, Narenthirakumar Appavu et al. [2] suggested a multi-sensor technique for early violence detection in response to the urgent problem of bullying in primary schools. The study used two motion sensors and Fusion and improved Relief-F algorithms to gather information on seven different categories of violence-related actions. The relief-F approach was used to filter features, and Random Forest and a radial basis function neural network were used to build a two-level classifier. Using a decision layer approach, the recognition results from two sensors were combined, yielding detection accuracy of

97.3% in real-world settings and 84.4% in elementary school violence. This creative method greatly aids in the early identification of violence in elementary classrooms.

According to research by V. D. Huszár [3], the quick development of digital video technology has led to a global increase in the use of surveillance cameras. However, the increase of video data makes it difficult for humans to analyze it in real time, which is why automatic violence detection in surveillance tape is becoming more popular. This paper addresses the intricacy of dynamic interactions in video footage by exploring the use of smart networks with 3D convolutions to capture spatial and temporal features. Using action recognition models that have already been trained, the research focuses on effective and precise violence detection. Testing on several publicly available datasets demonstrates the effectiveness of the method, which improves accuracy by 2% while requiring fewer model parameters. The tests also demonstrate how well the approach works with typical compression artifacts that are present in applications using remote server operations.

Effective surveillance technologies are necessary to combat the ubiquitous problem of violence, as human-supervised solutions are no longer sufficient. With the limitations of automated violence detection technology acknowledged, recent developments in deep learning present a viable remedy. Despite encouraging experimental results, this literature study highlights the dearth of real-time implementations in monitoring systems [4]. The solution put forth advocates for the implementation of a deep learning-based real-time violence detection system on unmanned aerial vehicles (UAVs). It has demonstrated outstanding performance metrics, including 93.69% mAP(0.5), 0.114 FLOPS(B), and a small weights size of 564.3 KB, all of which are compatible with real-time requirements.

Particularly in the fields of public safety, violence detection systems have greatly advanced in recent years thanks to advances in computer vision and machine learning. The difficulties of identifying violence in crowded environments are discussed in this paper, which also suggests an intelligent system that makes use of cutting-edge deep learning models. The study compares and contrasts ResNet50 with YOLOv8 in detail, assessing both models' strengths and weaknesses in terms of violence detection [5].

With an astounding 98.63% accuracy rate, the study by [6] offers a revolutionary method for separating violent and peaceful images using a combination Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) model. 11,043 416x416 original pictures made up the dataset, which was downscaled to 256x256 for efficient model training. The recommended hybrid model performed better when compared to cutting-edge Deep Learning systems and conventional Machine Learning techniques.

The problem of domestic abuse is widespread and leaves victims feeling helpless. The method suggested in [7] uses cutting-edge technology, such as computer vision, natural language processing, and speech recognition, to address this problem. With the help of a smartphone app that allows victims to gather evidence and access support services, this multimodal method seeks to identify and address both verbal and physical abuse. When abusive activity is recognized, machine learning algorithms improve system accuracy and allow law enforcement to receive automatic alerts.

The literature emphasizes how important intelligent surveillance systems are for managing video data on multimodal violence. To overcome the difficulties in weakly supervised learning for violence detection, the research by [8] presents a neural network with mutual learning branches, a post-processing technique, and a co-attention module. On the XD-Violence dataset, the suggested model beats current state-of-the-art models by 1.92% in AP (Average Precision).

In [9], a novel algorithm for detecting violent behavior is introduced. It makes use of optical flow and local spatio-temporal variables. In order to overcome computational difficulties, the method combines optical flow with the Harris 3D spatio-temporal interest point detector to create a physical contact detection technique. The suggested method shows excellent efficacy in identifying aggressive behaviors in a variety of situations and with varying numbers of participants [9].

Modern techniques for violent crime early detection are required due to the rising rate of crime worldwide. In this paper, [10] suggests a machine learning approach for video stream violence detection that makes use of 3D convolutional neural networks. Supervised learning is utilized for binary and multi-class violence classification, utilizing Inception-v3 and Gated Recurrent Units. Transfer Learning improves model performance on a variety of datasets, such as YouTube, human recordings, movies, and surveillance footage [10].

In today's world, protecting human lives is vital, and video surveillance is frequently used for security in a variety of contexts. But depending just on CCTV cameras makes it difficult to monitor an area effectively, which could result in mistakes being made while detecting crimes. To improve real-world accuracy, [11] has developed a violence detection system that makes use of optical flow. By converting regular CCTVs into smart cameras, the solution will enable law enforcement to react to situations more quickly and pro-actively [11].

The increasing use of CCTV surveillance is a reflection of the increased necessity to deal with the aftermath of violent incidents in public areas. A thorough real-time video analysis framework using CNN for feature extraction and LSTM for chronological interpretation is suggested in [12] in order to improve prompt response. Through the combination of a mobile app and Telegram bot, this technique seeks to detect and swiftly report cases of violence to relevant authorities, achieving effective monitoring and intervention [12].

Algorithms for detecting violence in videos have been developed in response to citizen insecurity, specifically with regard to physical violence between individuals. Three convolutional neural network models—Xception, InceptionV3, and VGG16—coupled with recurrent LSTM networks were the subject of a comparative investigation by [13]. According to the analysis, the Real Life Violence Situations dataset's InceptionV3 model had the best accuracy (94%) in identifying violent and non-violent situations. This implies that the algorithm can effectively identify instances of physical aggression between people in citizen security footage [13].

Via violence event detection, safety in the context of smart cities must be guaranteed. Prior research has mostly examined the application of 2D-Convolutional Neural Networks (2d-CNN) in conjunction with Recurrent Neural Networks (RNN) to extract spatial and temporal features. Transformer networks have shown successful in several areas, but large datasets are necessary for them to function well. In order to address this, [14] presented the data-efficient video transformer (DeVTr), which uses an embedding layer that is a pre-trained 2d-CNN. Experiments conducted on the Real-life violence dataset (RLVS) produced an astounding 96.25% accuracy rate, outperforming earlier techniques [14].

Smart surveillance has drawn a lot of interest, especially when it comes to violence detection. Current methods for detecting violence frequently depend on interval sampling, which might lead to the omission of important action information. Using ResNet50 as the image encoder, the literature put forth by [15] presents a novel key framing technique based on encoding and clustering for frame extraction. The Mix dataset, which combines the violent flows, movies, and hockey battles datasets, is used to train the model. Even with 360 fewer frames in a one-minute video clip, the key framed model beats state-of-the-art algorithms according to a comparative evaluation on a manually annotated test dataset from YouTube [15].

In order to lessen the impact of violent situations, a violence detection and warning system that uses human body pose identification from public CCTVs is proposed in [16]. Using OpenPose for real-time skeleton recognition, the system analyzes atypical limb movement angles and speeds to detect potentially aggressive behavior. The system detects anything and sounds an alarm to alert the appropriate authorities so they can act quickly. This method helps overcome barriers to victims receiving help on their own and helps provide victims with rapid assistance. Target aggressive actions can be successfully detected, according to experimental data [16].

Because manual surveillance has its limits, computerized violence detection is essential for resolving city security problems [17]. Current methods make use of conventional machine learning or CNN and LSTM networks together. To improve accuracy and efficiency, a clustering-based keyframe extraction technique is used in the proposed Efficient ConvLSTM based Violence Detection System (ECLVDS) by [17]. The model performed admirably, outperforming some previous models with an accuracy of 98.90% when tested on the Hockey Fight Dataset [17].

The research shows a troubling trend of 7% annual growth in human violence, which frequently happens abruptly in remote locations. Research indicates that one of the main challenges in resolving these occurrences is the difficulty of delayed information communication. In order to get around this, automatic detection methods based on computer vision and CCTV footage were put into place by [18]. A deep learning model that combined the YOLO-5 and Convolutional Neural Network (CNN) was particularly effective and achieved an astounding accuracy rate of 98.63% [18].

When it comes to smart surveillance systems, the ability to identify violent situations is essential for efficient monitoring. Transformer networks, albeit requiring large amounts of data, have demonstrated impressive results in video action identification. In order to increase accuracy with minimal datasets, the Data-efficient video transformer (DEVTr) family incorporates pre-trained convolutional neural networks [19]. In order to beat earlier methods in violence event recognition under resource restrictions, this study by [19] expands on DEVTr, introduces unique data augmentation approaches, and achieves exceptional accuracy on Real-life violence, NTU CCTV-fight, and UBI-Fight datasets.

While violent action recognition, especially in public areas, is still difficult, new developments in deep neural networks and transfer learning have shown promise in identifying violent behavior. A unique deep NeuralNet system for violence detection using motion feature extraction from RGB dynamic images is presented in the literature by [20]. The method accurately predicts violent content by fine-tuning the previously trained Inception-Resnet-V2 model. Validation using widely used benchmarks, such as the Real Life Violence Dataset, the Hockey Fight dataset, and the movie dataset, shows how much better and more successful the model performs [20].

3. Methodology

After experimenting with various pipelines for our project, including different CNN(Convolutional Neural Network) architecture and YOLO models, we settled on a methodology that relies on the MobileNetV2 for frame processing generated through CCTV feed.

We compared the YOLOv8 model, known for its sophisticated object detection capabilities, with the MobileNetV2 architecture. YOLOv8 adopts a complex deep neural network structure, incorporating bounding boxes to enable efficient detection of multiple objects in a single pass. Despite its rich capabilities, YOLOv8 achieved an accuracy of approximately 88% in real-time analysis, lagging behind the MobileNetV2 model.

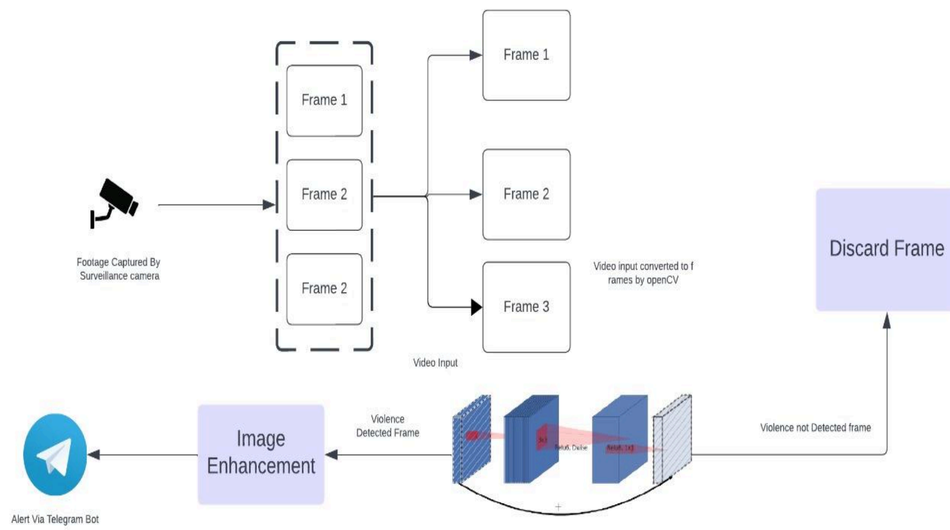


Fig 3.1: Project Flow

MobileNetV2, with its lightweight architecture, excelled with a remarkable 96% accuracy, making it the preferred choice for our specific violence detection task.

As seen in Fig 3.1 ,Frames are extracted from the CCTV feed at predefined intervals to predict violent activities. These frames undergo processing to enhance the model's predictive capabilities, and they are then fed into the MobileNetV2 model, trained on historical data, to predict violent activity in real-time.

Upon detection of violent frames, notifications are dispatched to authorized personnel or entities registered in the system through a Telegram bot. The notification includes the violent frame image, along with location and timestamp details. These alarming images are stored in Firebase, a choice made begrudgingly, as it offers a scalable and secure Cloud Storage solution. Additionally, Firebase allows the deployment of serverless functions with real-time analytics, features that, while beneficial, are overshadowed by the compromises in the chosen methodology.

3.1 Data Collection and Preprocessing

A historical dataset with videos categorized as 0 and 1 that included both violent and non-violent scenes was used to train the algorithm. To create image data, the video footage was then divided into frames at predefined intervals. After that, augmentation techniques were applied to these images utilizing libraries like imgaug, imageio, and OpenCV, improving their quality and getting them ready for further analysis.

IMGAUG , IMAGEIO , OPENCV:

The libraries imgaug, imageio, and OpenCV were all very important to our project. Diverse picture augmentation was made possible by Imgaug, and flawless input and output procedures for several formats were guaranteed by Imageio. The flexible computer vision library OpenCV offered the necessary features for tasks like brightness modification and scaling. These libraries worked together to streamline our image processing pipeline, improving our model's overall accuracy and performance for both training and real-time analysis.

3.2 Activity Detection with MobileNetV2

After extracting and processing frames, they are forwarded to the MobileNetV2 architecture for real-time violence detection. MobileNetV2, designed for image classification, object detection, and segmentation on devices with limited resources, proves efficient in real-time scenarios due to its simplicity. The MobileNetV2 base as seen in Fig 3.2 is initialized with average pooling, excluding the top classification layer. Subsequently, a new Dense layer with a sigmoid activation function is appended to enable binary classification, yielding a complete model suitable for tasks like binary image classification. The model undergoes fine-tuning on a curated dataset, optimizing parameters to achieve a validation accuracy of around 96%.

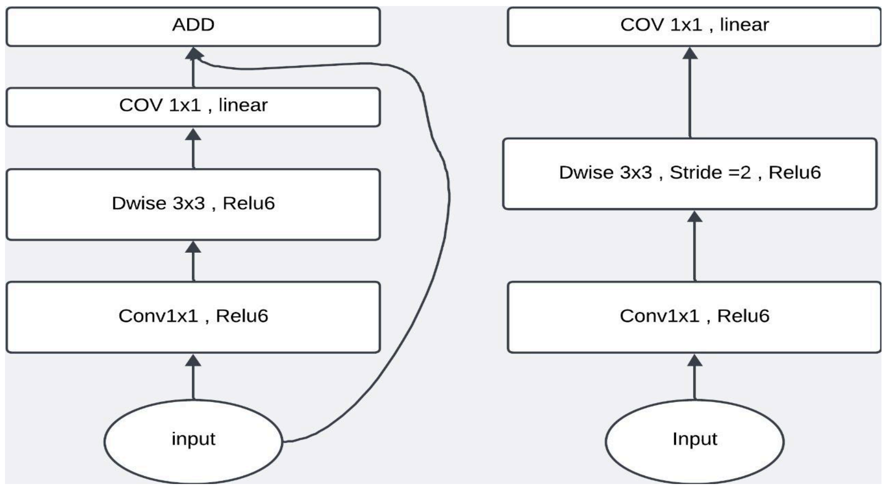


Fig 3.2: MobileNetv2 Architecture

CALLBACKS:

We utilized TensorFlow Keras callbacks for efficient training and performance monitoring. ModelCheckpoint saved optimal weights, LearningRateScheduler adjusted learning rates dynamically, and TensorBoard provided insightful visualizations. EarlyStopping prevented overfitting, and ReduceLROnPlateau adapted the learning rate for improved model performance. These callbacks collectively streamlined training, ensuring the development of a robust and accurate model.

3.3 Real-time CCTV Feed Integration

Even while CCTV has historically been a major tool for detecting crimes, it frequently required constant manual observation that took place around-the-clock. By automating the process of gathering and detecting illegal activity through CCTV feeds, our approach presents a practical answer. This novel approach captures video frames and uses the MobileNetV2 architecture to automatically detect activity. It interacts smoothly with current real-time CCTV systems. By lowering the dependency on continual human interaction, this method not only improves the effectiveness of surveillance but also simplifies the crime detection process.

3.4 Telegram Bot Integration

A Telegram bot was created using the Telegram Bot API to facilitate real-time alerts. The bot was integrated with the activity detection system, enabling it to send alerts containing pertinent information, such as timestamps and images capturing the detected violent activity, to a designated system or entity.The system stores information about the concerned entity, and as soon as violence is detected in a frame as seen in Fig 3.3, an alert is promptly sent. The alert not only includes a clear image of the incident as evidence but also provides the location information.

Telepot:

Telepot is a Python library that simplifies the development of Telegram bots, streamlining the interaction between our project and the Telegram messaging platform. With its intuitive API and user-friendly design, Telepot enables seamless integration of bot functionalities, facilitating the prompt delivery of alerts and notifications generated by our violence detection system to authorized personnel or entities. Its simplicity and robust features make Telepot a valuable tool in our project, enhancing communication and alerting capabilities through the Telegram platform.

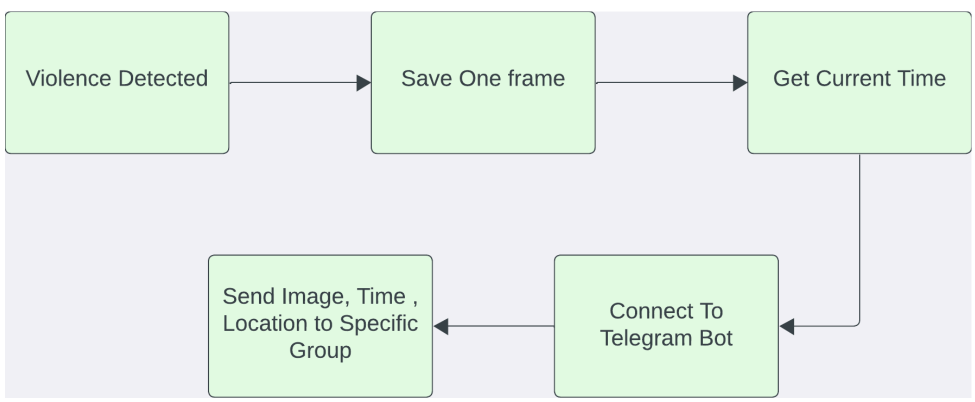


Fig 3.3: Sending Alert

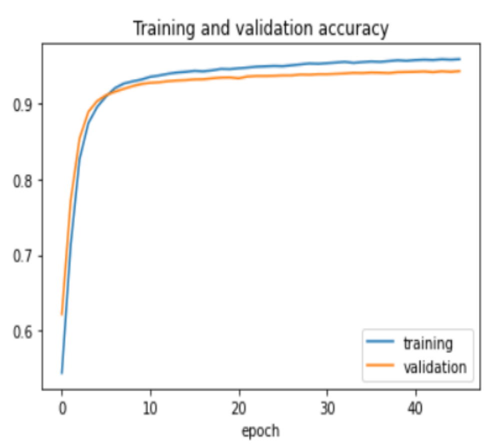
.

4. Result And Analysis

Our research has been focused on developing a precise violence detection classifier capable of real-time operation, promptly alerting authorized entities with minimal latency. After considering various object detection models, we have narrowed down our choices to two widely used networks: YOLOv8 and MobileNetV2.

MobileNetV2 emerges as a preferable option for real-time violence detection due to its simpler architecture. It achieves higher accuracy of 95 percent, primarily owing to its lightweight design that facilitates quicker processing. MobileNetV2 is also quicker and easier to train, making it suitable for scenarios with limited training data and crucial when model size and computational complexity are key considerations

We achieved outstanding results with an accuracy of 95 percent and a minimal loss of 0.14 as seen in Fig 4.1 and 4.2. This outcome signifies the model's exceptional capability in accurately classifying images as violent and non-violent. The high accuracy is a proof of the network's ability to extract intricate patterns and features within the data, showcasing its robustness in image recognition tasks. The low loss value further underscores the efficiency of MobileNetV2 in minimizing prediction errors during the training process



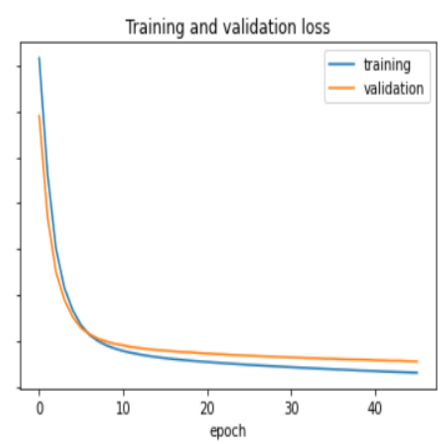


Fig 4.1 : Training and validation accuracy of MobileNetV2

Fig 4.2 : Training and validation loss for for MobileNetV2

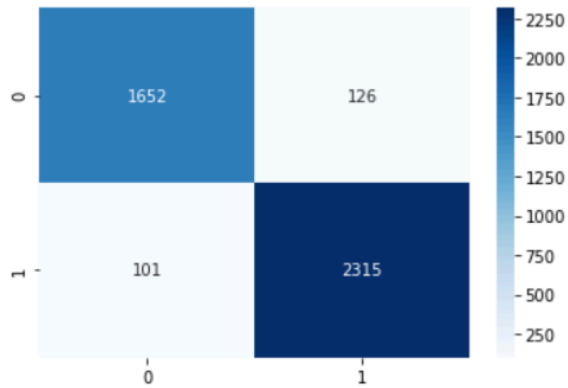


Fig 4.3: Confusion matrix for MobileNetV2

As seen in Fig 4.3 , MobileNetV2 classifies 3975 instances correctly compared to the 226 instances that were classified incorrectly which led to a high accuracy. Compared to the YOLOV8 model that could predict just 88 percent of the instances correctly as seen in Fig 4.4 , our proposed model is more efficient in real-time situations despite being a smaller And lightweight architecture.

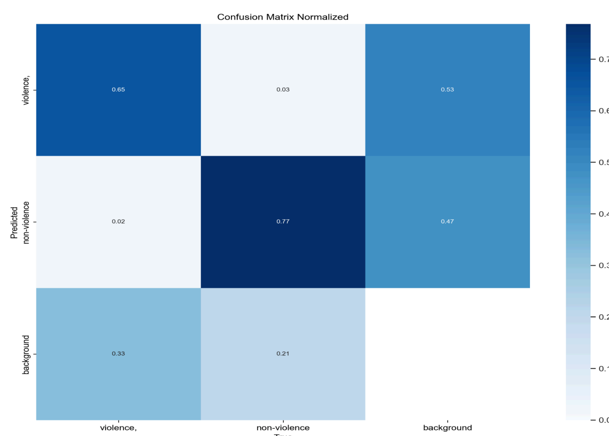


Fig 4.4: Confusion matrix for YOLOV8

Table 1: Comparison between the models used

Metrics	MOBILENETV2	YOLOV8
Accuracy	95%	89%
Validation Loss	0.12	2.0
Model Size	Compact and lightweight, suitable for edge devices	Larger model size, may require more storage space
Speed	Faster inference, especially on resource-constrained devices	Slightly slower due to its comprehensive detection

YOLOv8 employs a single neural network to predict bounding boxes and class probabilities for multiple objects in an image. While it achieves high accuracy in video classification, the training process may be time-consuming due to the inclusion of bounding boxes for each object. The complexity of its architecture poses challenges in real-time predictions.

YOLOv8 is pretty good at spotting violence with an 88 percent accuracy rate, but it experiences a bit more confusion, indicated by a loss of 2.0. This means it could make some errors in predicting violence.

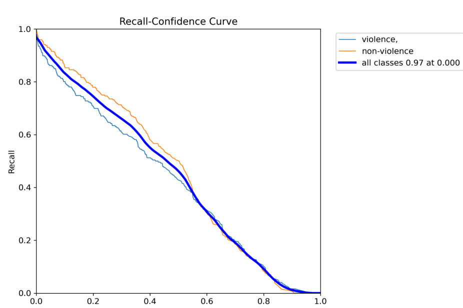


Fig 4.5 : Recall curve for Yolov8

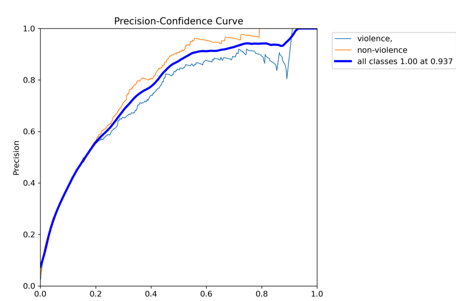


Fig 4.6 : Precision curve for Yolov8

5. Conclusion

The advancement of real-time savagery discovery frameworks powered by deep learning innovations is evident in the interest of improving urban security and safety. Fundamentally noteworthy are the combinations of multi-person 2D posture estimate, fast individual localization via yolov8, and Convolutional Neural Arrange (CNN) and Long Short-Term Memory (LSTM) for savagery categorization. OpenPose's precise identification of human postures provides a strong foundation for further discovery phases, while Yolov8's productivity ensures timely hazard recognition. CNN and LSTM work together to leverage spatiotemporal highlights, completing the circle of progressing accuracy by tracking designs and settings over time. Reconnaissance demands accuracy and efficiency, and the clustering-based keyframe extraction method reduces false warnings, maximizes processing, and decreases repetitive outlining. Using the practical analysis of LSTM strengthens the system's ability to distinguish between benign and aggressive activity, providing law enforcement with a proactive tool for safer city environments. As this thorough writing audit outlines, the mix of cutting-edge technology shows significant promise in mitigating the problems posed by urban brutality and misbehavior. The course for future investigations, which will focus on increasing efficiency, enhancing system performance, and exploring innovative deep learning applications within urban security, is made clear by this audit.

6. Future Scope

Looking ahead, the future scope of our Real-Time Violence Detection System holds exciting possibilities for further refinement and expansion. We are planning to integrate firebase into our project that will serve as a database to store the recognised images that have been sent to the entities through our bot.
This will help to store the face and details or the suspect and can be used to detect the culprit in other CCTV feeds through facial recognition.

6.1. Firebase Integration

Continuing with our project, the subsequent phase involved facial detection of the suspected individual across different CCTV feeds, with the storage of facial extractions and associated metadata in Firebase.

FireBase:
Our project is not possible without Firebase, which has unmatched benefits over alternative databases. Efficient picture storage is ensured by its safe and scalable cloud storage, and data synchronization is made effortless by its real-time database. An

adaptable and resource-efficient processing and analytics solution is offered by the implementation of serverless functions. With capabilities like real-time updates and serverless function deployment, Firebase optimizes our violence detection system and is a better option than other databases due to its variety and ease of integration.

6.2. Facial Recognition

To identify the face of the person engaged in the violent activity, we employed facial recognition technology. This involved utilizing a CNN architecture to assess the similarity between frames extracted from CCTV footage and the facial features stored in Firebase. The implementation leveraged the face_recognition library, which is constructed on Dlib and OpenCV, providing an expedited approach for detecting stored faces in new frames with a more straightforward and less intricate model.

FACE RECOGNITION:

The face_recognition library stands out as a valuable asset in our project, leveraging Dlib and OpenCV to provide efficient facial recognition capabilities. Built on top of these powerful frameworks, face_recognition offers a streamlined and less complex model for detecting stored faces in new frames. Its accelerated face detection and recognition make it a key component in our facial recognition technology implementation. With its simplicity and speed, the face_recognition library significantly contributes to the accuracy and real-time efficiency of identifying individuals involved in violent activities across various CCTV feeds in our project.

References

1. C. Shripriya, J. Akshaya, R. Sowmya and M. Poonkodi, "Violence Detection System Using Resnet," 2021 5th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2021, pp. 1069-1072, doi: 10.1109/ICECA52323.2021.9675868. keywords: {Systematics;Surveillance;Streaming media;Feature extraction;Cameras;Motion detection;Spatiotemporal phenomena;resnet-violence detection;surveillance;object detection},
2. N. Appavu and C. N. Kennedy Babu, "Multisensor fusion sensor and improved Relief-F algorithms Based violence detection in schools," 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2023, pp. 1-6, doi: 10.1109/ICONSTEM56934.2023.10142726. keywords: {Time-frequency analysis;Redundancy;Radial basis function networks;Sensor fusion;Feature extraction;Motion detection;Sensors;Primary School Violence;Artificial Intelligence;Enhanced Relief-F;Activity Recognition;System Recognition},
3. V. D. Huszár, V. K. Adhikarla, I. Négyesi and C. Krasznay, "Toward Fast and Accurate Violence Detection for Automated Video Surveillance Applications," in IEEE Access, vol. 11, pp. 18772-18793, 2023, doi: 10.1109/ACCESS.2023.3245521.

4. H. H. Nguyen, Q. Trung Le, V. Q. Nghiem, M. Son Hoang and D. A. Pham, "A novel violence detection for drone surveillance system," 2023 International Conference on Communication, Circuits, and Systems (IC3S), BHUBANESWAR, India, 2023, pp. 1-6, doi: 10.1109/IC3S57698.2023.10169405. keywords: {Deep learning;Training;Measurement;Surveillance;Sociology;Real-time systems;Object tracking;violence detection;deep learning;UAVs},
5. L. Sachan, P. Katiyar, Y. Kumbhawat, G. K. Rajput and T. Mehrotra, "Comparative Analysis on Violence Detection Using Yolo and ResNet," 2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2023, pp. 89-92, doi: 10.1109/SMART59791.2023.10428272. keywords: {Deep learning;Surveillance;Employment;Market research;Public security;Intelligent systems;Residual neural networks;Intelligent Surveillance System;Violence Detection;Deep Learning Models},
6. R. G. Tiwari, H. Maheshwari, A. K. Agarwal and V. Jain, "Hybrid CNN-LSTM Model for Automated Violence Detection and Classification in Surveillance Systems," 2023 12th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2023, pp. 169-175, doi: 10.1109/SMART59791.2023.10428538. keywords: {Deep learning;Adaptation models;Surveillance;Computational modeling;Computer architecture;Convolutional neural networks;Long short term memory;convolutional neural network;long short-term memory;image classification;violence detection;non-violence images;deep learning;hybrid model},
7. F. K. S, R. P. Haroon, F. Fathima, T. Thajudeen and M. Fudin, "Domestic Violence Detection System Using Natural Language Processing," 2023 International Conference on Innovations in Engineering and Technology (ICIET), Muvattupuzha, India, 2023, pp. 1-6, doi: 10.1109/ICIET57285.2023.10220859. keywords: {Visualization;Technological innovation;Computer vision;Machine learning algorithms;Law enforcement;Speech recognition;Object detection;Natural Language Processing(NLP);Object detection;Image processing;Speech recognition},
8. J. Cheng, C. Sun, J. Chen and P. Lu, "Audio-visual mutual learning for Weakly Supervised Violence Detection," 2023 8th International Conference on Information Systems Engineering (ICISE), Dalian, China, 2023, pp. 5-8, doi: 10.1109/ICISE60366.2023.00008. keywords: {Visualization;Fuses;Surveillance;Soft sensors;Supervised learning;Neural networks;Feature extraction;Video violence detection;multi-modal fusion;weak supervision learning},
9. Y. Lyu and Y. Yang, "Violence Detection Algorithm Based on Local Spatio-temporal Features and Optical Flow," 2015 International Conference on Industrial Informatics - Computing Technology, Intelligent Technology, Industrial Information Integration, Wuhan, China, 2015, pp. 307-311, doi: 10.1109/ICIICII.2015.157. keywords: {Image motion analysis;Computer vision;Optical imaging;Three-dimensional displays;Optical detectors;Detection algorithms;Monitoring;Violence Detection;Harris 3D Spatio-temporal Interest Points Detector;Pyramid Lucas-Kanade Optical Flow Method;Motion Coefficient},
10. M. Gadelkarim, M. Khodier and W. Gomaa, "Violence Detection and Recognition from Diverse Video Sources," 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022, pp. 1-8, doi: 10.1109/IJCNN55064.2022.9892660. keywords: {Video on demand;Three-dimensional displays;Annotations;Computational modeling;Surveillance;Neural networks;Transfer learning;Violence Detection;Action Recognition;Multi-class Violence Classification;Transfer Learning;Supervised Learning;3D Convolutional Neural Networks;Inception-v3;Gated Recurrent Units (GRUs);Recurrent Neural Networks (RNNs);Frame-Level Annotations},

11. K. Basu and A. M. Parimi, "Optical Flow based method for Violence Detection," 2023 First International Conference on Cyber Physical Systems, Power Electronics and Electric Vehicles (ICPEEV), Hyderabad, India, 2023, pp. 1-6, doi: 10.1109/ICPEEV58650.2023.10391925. keywords: {Law enforcement;Semantic segmentation;Smart cameras;Video surveillance;Real-time systems;Rail transportation;Safety;Optical flow;Gunnar-Farneback algorithm;Image segmentation;Violence detection},
12. A. N. Appavu and C. N. K. Babu, "An Xception Model Based Real-time Violence Detection," 2023 IEEE International Conference on Advanced Systems and Emergent Technologies (IC_ASET), Hammamet, Tunisia, 2023, pp. 1-6, doi: 10.1109/IC_ASET58101.2023.10151034. keywords: {Representation learning;Law enforcement;Surveillance;Streaming media;Feature extraction;Chatbots;Real-time systems;CCTV;CNN;LSTM;Violence;Non-Violence;Xception},
13. H. Calderon-Vilca, K. C. Ramos, E. D. Quiroz, J. A. Rojas, R. C. Vilca and A. A. Tarqui, "The Best Model of Convolutional Neural Networks Combined with LSTM for the Detection of Interpersonal Physical Violence in Videos," 2021 29th Conference of Open Innovations Association (FRUCT), Tampere, Finland, 2021, pp. 81-86, doi: 10.23919/FRUCT52173.2021.9435563. keywords: {Technological innovation;Convolutional neural networks;Security;Videos},
14. A. R. Abdali, "Data Efficient Video Transformer for Violence Detection," 2021 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT), Purwokerto, Indonesia, 2021, pp. 195-199, doi: 10.1109/COMNETSAT53002.2021.9530829. keywords: {Learning systems;Recurrent neural networks;Satellites;Event detection;Smart cities;Memory management;Feature extraction;Transformer;CNN;Deep Learning;violence Detection;RLVS;Spatio-temporal;Video Classification;Smart Cities},
15. B. J, D. R. A, V. K. B and C. G, "Physical Violence Detection in Videos Using Keyframing," 2023 International Conference on Intelligent Systems for Communication, IoT and Security (ICISCoIS), Coimbatore, India, 2023, pp. 275-280, doi: 10.1109/ICISCoIS56541.2023.10100407. keywords: {Video on demand;Computational modeling;Surveillance;Sampling methods;Motion pictures;Encoding;Security;Violence detection;Keyframe extraction;Spatial-temporal CNN},
16. K. -C. Chang and Y. -C. Liao, "Design of Violence Event Detection System Based on CCTVs by Human Body Pose Recognition," 2023 International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), PingTung, Taiwan, 2023, pp. 695-696, doi: 10.1109/ICCE-Taiwan58799.2023.10226669. keywords: {Image recognition;Law enforcement;Event detection;Weapons;Biological system modeling;Feature extraction;Skeleton},
17. S. K. Parui, S. K. Biswas, S. Das, M. Chakraborty and B. Purkayastha, "An Efficient Violence Detection System from Video Clips using ConvLSTM and Keyframe Extraction," 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), Jaipur, India, 2023, pp. 1-5, doi: 10.1109/IEMECON56962.2023.10092302. keywords: {Urban areas;Neural networks;Manuals;Feature extraction;Video surveillance;Data models;Spatiotemporal phenomena;Violence Detection;Keyframe Extraction;Hockey Videos;CNN;LSTM;ConvLSTM},

18. V. Gautam, H. Maheshwari, R. G. Tiwari, A. K. Agarwal and N. K. Trivedi, "Automated Detection of Violence in Detached Areas using Hybrid Deep Learning Models: A YOLO-5 and CNN Approach," 2023 2nd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2023, pp. 1276-1282, doi: 10.1109/ICACRS58579.2023.10404840. keywords: {Deep learning;YOLO;Computer vision;Terminology;Convolutional neural networks;Object recognition;Context modeling;Closed-Circuit Television (CCTV);Human Violence;Deep Learning;Machine Learning;Object Detection;Computer Vision;Deep Neural Networks},
19. A. R. Abdali and A. A. Aggar, "DEVTrV2: Enhanced Data-Efficient Video Transformer For Violence Detection," 2022 7th International Conference on Image, Vision and Computing (ICIVC), Xi'an, China, 2022, pp. 69-74, doi: 10.1109/ICIVC55077.2022.9886172. keywords: {Event detection;Convolution;Surveillance;Urban areas;Neural networks;Memory management;Transformers;Transformer;CNN;Deep Learning;violence Detection;RLVS;Spatio-temporal;Video Classification;Smart Cities},
20. A. Jain and D. K. Vishwakarma, "Deep NeuralNet For Violence Detection Using Motion Features From Dynamic Images," 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2020, pp. 826-831, doi: 10.1109/ICSSIT48917.2020.9214153. keywords: {Feature extraction;Training;Dynamics;Motion pictures;Computer architecture;Task analysis;Violence Detection;Deep NeuralNet;Transfer Learning;Dynamic Images;Inception-Resnet-V2;CNN},