Detection of Criminal Activities Through CCTV Surveillance

Submitted in partial fulfillment of the requirements of the degree

BACHELOR OF ENGINEERING

IN

COMPUTER ENGINEERING

By

Richita Karira D12C-30

Tanmay Chaudhary D12C-12

Manav Keswani D12C-34

Soumil Tawde D12C-66

Name of the Mentor

Dr. Rohini Temkar

Department of Computer Engineering

Vivekanand Education Society's Institute of Technology,

An Autonomous Institute affiliated to University of Mumbai

HAMC, Collector's Colony, Chembur,

Mumbai-400074

University of Mumbai (AY 2023-24)

CERTIFICATE

This is to certify that the Mini Project entitled **“Detection of Criminal Activities Through CCTV Surveillance”** is a bonafide work of **Richita Karira(D12C-30), Tanmay Chaudhary(D12C-12), Manav Keswani(D12C-34), Soumil Tawde(D12C-66)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **“Bachelor of Engineering”** in **“Computer Engineering” .**

(Prof. Rohini Temkar)
Mentor

(Prof.____________________________)
Head of Department

(Prof.____________________________)
Principal

Mini Project Approval

This Mini Project entitled "Detection of Criminal Activities Through CCTV Surveillance" by **Richita Karira(D12C-30), Tanmay Chaudhary(D12C-12), Manav Keswani(D12C-34), Soumil Tawde(D12C-66)** is approved for the degree of **Bachelor of Engineering** in **Computer Engineering.**

Examiners

1……………………………………….
(Internal Examiner Name & Sign)

2……………………………………….
(External Examiner name & Sign)

Date: October 21, 2023

Place:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society’s Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Dr. (Mrs.) Rohini Temkar** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to the Head of the Computer Department **Dr.(Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J.M. Nair** , for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

ABSTRACT

Hearing about the violent activities that occur on a daily basis around the world is quite overwhelming. Personal safety and social stability are seriously threatened by the violent activities. A variety of methods have been tried to curb the violent activities which includes installing surveillance systems. It will be of great significance if the surveillance systems can automatically detect violent activities and give warning or alert signals.

The whole system can be implemented with a sequence of procedures. Firstly, the system has to identify the presence of human beings in a video frame. Then, the frames which are predicted to contain violent activities have to be extracted. The irrelevant frames are to be dropped at this stage. Finally, the trained model detects violent behavior and these frames are separately saved as images. These images are enhanced to detect faces of people involved in the activity, if possible. The enhanced images along with other necessary details such as time and location is sent as an alert to the concerned authority.

The proposed method is a deep learning based automatic detection approach that uses Convolutional Neural Network to detect violence present in a video. But, the disadvantage of using just CNN is that it requires a lot of time for computation and is less accurate. Hence, a pre-trained model, MobileNet, which provides higher accuracy and acts as a starting point for the building of the entire model. An alert message is given to the concerned authorities using a telegram application.

List of figures

Table of Contents

Chapter 1 : Introduction

Chapter 2 : Literature Survey

Chapter 3 : Proposed System

Chapter 1: Introduction

This chapter includes the introduction to the topic, motivation behind choosing the topic, a concrete problem definition, objectives and the organization of the proposed project.

1.1 Introduction to the project

Violent behavior in public places is a pressing issue that necessitates urgent attention. Such acts of violence not only disrupt social harmony but also erode communities, leading to reduced productivity, diminished property values, and disruptions in essential social services. Moreover, violence poses a significant public health concern, affecting individuals across all age groups, from infants to the elderly. Recognizing and addressing violence in a timely and efficient manner is a complex challenge, particularly because it must be detected in real-time videos captured by a multitude of surveillance cameras situated in diverse locations. The primary objective of this project is to develop a robust real-time violence alert system that can reliably detect and promptly alert the relevant authorities when violent activities occur.

1.2 Motivation for the project

The motivation behind this project stems from the urgent need to enhance public safety and social stability by addressing the menace of violence in public spaces. Across the globe, violence has far-reaching consequences, and it is imperative to harness the power of technology to mitigate its impact. Existing public video surveillance systems, while valuable, are limited by the human effort required to monitor hours of footage to identify fleeting moments of violence. This project is inspired by the potential of deep learning techniques, particularly Convolutional Neural Networks (CNN), in automating the detection of violence in video streams, thus enabling swift and accurate response from law enforcement agencies. By leveraging the capabilities of pre-trained models like MobileNetv2, we aim to create an efficient and accurate violence detection system that can reduce the burden on human surveillance.

1.3 Problem Statement & Objectives

Problem Statement: The problem at hand is the real-time detection of violent activities within video footage obtained from public surveillance cameras. The challenge lies in differentiating between regular activities and violent behavior, promptly alerting the authorities, and ensuring that the system functions with high accuracy and efficiency. This problem addresses the critical need for automated violence detection to alleviate the burden on human surveillance, allowing for faster response times and enhanced public safety.

Objectives:

- Develop an automatic violence detection system using deep learning techniques, specifically MobileNetv2, to analyze video frames in real-time.
- Identify human presence in video frames and extract frames indicative of violent activities while filtering out irrelevant ones.
- Enhance the extracted frames to improve the identification of individuals involved in the violent behavior.
- Collect essential information, including the timestamp and location of the incident.
- Establish a real-time alert system that sends notifications to the concerned authorities via the alert module of the proposed system.
- Achieve high accuracy and efficiency in violence detection to improve public safety and reduce the need for manual video monitoring.
- To help curtail cases of Women Harassment in future scope of development.

1.4 Organization of project

The project is organized systematically to provide a comprehensive understanding of its scope and development. It commences with the requisite administrative elements, including approval certificates and acknowledgments, followed by an abstract. Lists of abbreviations, figures, tables, and symbols are also included for reference. The main body of the report is divided into chapters, with Chapter 1 introducing the project's context, motivation, problem definition, objectives, and the overall organization of the report. Chapter 2 conducts a literature survey, highlighting existing systems, research gaps, and the mini project's contributions. Chapter 3 explores the proposed system, encompassing an introduction, architectural details, algorithm and process design, methodology, hardware and software requirements, experiments, result analysis, and concluding remarks, along with future work. A comprehensive list of references and, if applicable, annexes containing published papers, camera-ready documents, business pitches, or proof of concepts conclude the document, ensuring a structured and informative presentation of the project.

Chapter 2: Literature Survey

2.1 Survey of Existing System

The field of violence detection has witnessed a variety of approaches, broadly classified into three categories: visual-based, audio-based, and hybrid approaches. Visual-based methods focus on extracting relevant features from visual information, encompassing local and global features. Audio-based approaches employ hierarchical techniques, including Gaussian mixture models and Hidden Markov models, to detect violence-related sounds such as gunshots and explosions [2]. Hybrid methods combine both visual and audio characteristics to identify violent incidents. The CASSANDRA system, for instance, detects aggression in surveillance videos by analyzing motion features and audio cues associated with violent events [4].

Several methods have been explored for spatio-temporal modeling, including approaches utilizing 2D CNNs with motion saliency maps and temporal squeeze and excitation modules, which emphasize frame grouping. Space-Time Interest Points (STIP) extract features by analyzing spatial and temporal differences, using 3D volumes to capture how image segments evolve over time. Low-level features, such as LHOG and LHOF descriptors, are used in violence detection. Additionally, 3D CNNs are employed for spatio-temporal feature extraction, although these methods often come with high computational costs. Sensor-network approaches extract images from video streams and employ deep neural networks for violence frame recognition [3].

Paper Name	Methodology	Data set	Limitations
Detection of Real-world Fights in Surveillance Videos Authors: Mauricio Perez, Alex C. Kot Published in: ICASSP 2019	The paper proposes a pipeline for fight detection in surveillance videos.The authors evaluate different feature extraction methods, including Deep Learning and Local Interest Points.They also explore different classifiers, such as end-to-end CNN, LSTM, and SVM.The Two-Stream approach, combining spatial and temporal features, shows the best performance.	The authors introduce a novel and challenging dataset called CCTV-Fights.The dataset contains 1,000 videos of real fights, with over 8 hours of annotated CCTV footage.The videos cover various locations, including public places, schools, shopping malls, and residential or commercial areas.The dataset includes both CCTV footage and videos recorded from mobile cameras, car cameras (dash-cams), and drones or helicopters.	The effectiveness of existing surveillance systems in detecting fights is questionable, as they rely on continuous human supervision. Previous work on fight detection either focuses on short clips or unrealistic scenarios (e.g., movies, sports, fake fights).The lack of available datasets comprising real-world fights from surveillance cameras has been a research gap.The paper acknowledges the need for further improvements in spatial features, leveraging sequential information, and designing early detection methods.

Paper Name	Methodology	Data set	Limitations
Vision-based Fight Detection from Surveillance Cameras Authors: Gözde Ayşe Tataroğlu, Şeymanur Hazım Kemal Ekenel Published in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019	The proposed method utilizes a combination of a Bidirectional Long Short-Term Memory (Bi-LSTM) network and a self-attention layer.The Bi-LSTM network is used for sequence learning, while the attention layer improves the performance of the model.The authors also employ Convolutional Neural Networks (CNNs) for feature extraction, specifically VGG16 and Xception architectures.	The authors collected a new surveillance camera fight dataset for their study.The dataset consists of fight scenes from surveillance camera videos available on YouTube, as well as security camera footages from various locations such as cafes, bars, streets, buses, and shops.The dataset includes various types of fight scenarios, such as kicks, fists, hitting with objects, and wrestling. The collected surveillance camera fight dataset is publicly available .	1. The results for the surveillance camera dataset were not as good as the other datasets, indicating the challenge of generalizing to diverse and real-world surveillance camera footage. 2. The Peliculas dataset used in the experiments had a limited number of fight scene samples, which affected the accuracy of the model on this dataset.

Paper Name	Methodology	Data set	Limitations
Fight Detection in Surveillance Videos Authors: Ersin Esen, Mehmet Ali Arabaci, Medeni Soysal Published in: 11th International Workshop on Content-Based Multimedia Indexing (CBMI), 2013	The paper proposes a novel method for fight detection in surveillance videos using a motion feature called Motion Co-occurrence Feature (MCF). The method involves motion estimation, motion vector quantization, and MCF extraction. Motion vectors are obtained using a block matching algorithm, and the MCF is calculated based on the co-occurrence distributions of motion vectors in magnitude and direction domains. The algorithm utilizes a k-Nearest Neighbor (kNN) classifier for fight detection.	The authors evaluated their algorithm using a dataset that includes samples from existing databases (BEHAVE and CAVIAR) and videos collected from the Internet. The dataset consists of labeled test data with various actions, including walking, running, fight, and non-fight scenes. The authors emphasize the need for more diverse datasets to improve fight detection algorithms.	The paper acknowledges the limited number of studies and datasets specifically focused on fight detection. Existing datasets from movies or sports events may not be suitable for fight detection in surveillance cameras. The authors highlight the importance of more diverse and challenging datasets to enhance the performance of fight detection algorithms.

Paper Name	Methodology	Data set	Limitations
Automatic fight detection in surveillance videos Authors: Eugene Yujun Fu, Hong Va Leong, Grace Ngai, Stephen C.F. Chan Published in: International Journal of Pervasive Computing and Communications, Vol. 13 Issue: 2, 2017	The paper proposes an approach to detect fights in surveillance videos based on low-level visual features and natural language processing techniques.The authors focus on detecting fights as a special kind of social event and interaction in surveillance scenarios.They propose using high-level features such as behaviors, actions, or visual words to detect fights or aggression.The approach aims to achieve real-time detection with low computation overhead.	The authors collected a new surveillance camera fight dataset for their study. The dataset consists of fight scenes from surveillance camera videos available on YouTube, as well as security camera footages from various locations such as cafes, bars, streets, buses, and shops. It includes various types of fight scenarios, such as kicks, fists, hitting with objects, and wrestling.The collected surveillance camera fight dataset is publicly available and can be accessed through the provided GitHub link.	1. The paper proposes an approach to detect fights in surveillance videos based on low-level visual features and natural language processing techniques. 2. The authors focus on detecting fights as a special kind of social event and interaction in surveillance scenarios.

2.2 Limitations of Existing Systems

While these approaches have made significant contributions to violence detection, they exhibit several limitations. Many of these methods involve high computational costs, making them unsuitable for real-time applications. Additionally, they may suffer from low detection rates and high false alarms in crowded scenes or occlusion-prone environments. The computational demands of feature extraction, as seen in the STIP method, may be impractical for use in surveillance and media rating systems. Moreover, sensor-network approaches require substantial computational power, which can be a hindrance to real-time deployment.

2.3 Mini Project Contribution

In the context of the aforementioned limitations and challenges, this mini project aims to contribute to violence detection using an efficient and accurate approach. Our approach leverages 2D CNNs with frame grouping and a Temporal Squeeze and Excitation Block to effectively capture spatio-temporal features. By emphasizing the identification of moving objects and motion boundaries, this method can highlight regions associated with violence in videos. This approach is computationally efficient, although it may require a fixed camera position, and it minimizes the need for extensive calculations. Our project's focus is to offer a practical and reliable solution for violence detection, particularly in real-time applications and scenarios characterized by occlusion and crowded scenes. We aim to enhance the accuracy and efficiency of violence detection in a resource-efficient manner, thus addressing the limitations of existing systems.

Chapter 3: Proposed System

This chapter presents a brief description about our proposed system, its framework, algorithm and process design, methodology applied, details about hardware and software used in our system, all the experiments and results, result analysis and discussions, conclusion of the project and the future work that will be done.

3.1 Introduction to Proposed System

The proposed system aims to develop a real-time surveillance system with the capability to recognize violent activities and promptly alert the relevant authorities. This system will analyze video footage obtained from surveillance cameras, identify human presence, and apply deep learning techniques to detect violence in real-time. Upon detection, the system will enhance the relevant frames for clarity and dispatch an alert to the nearby police station, ensuring a swift response to violent incidents. This project leverages the MobileNetV2 model to efficiently detect violent behavior within video sequences

3.2 Architectural Framework / Conceptual Design

The architecture of the system involves the processing of surveillance camera footage, including frame extraction, violence detection, image enhancement, and alert notification. The core component is the MobileNetV2 classifier for violence detection, and the system employs image enhancement using the Python Imaging Library (PIL). Additionally, an alert module is incorporated to send notifications to the authorities upon detecting violent activities in the video stream.

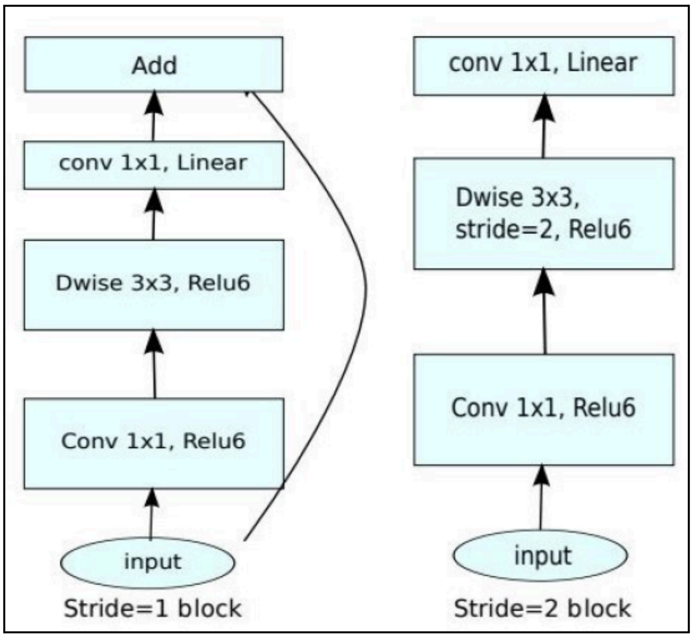


Figure 3.1 Architecture

3.3 Algorithm and Process Design

The system involves multiple stages, including the extraction of frames from surveillance camera footage, the application of the MobileNetV2 classifier for violence detection, image enhancement using PIL, and the alert notification process. The violence detection component is based on deep learning techniques, which analyze video frames for signs of violence. The alert module uses consecutive frame analysis to trigger alerts, ensuring that continuous violent behavior is detected.

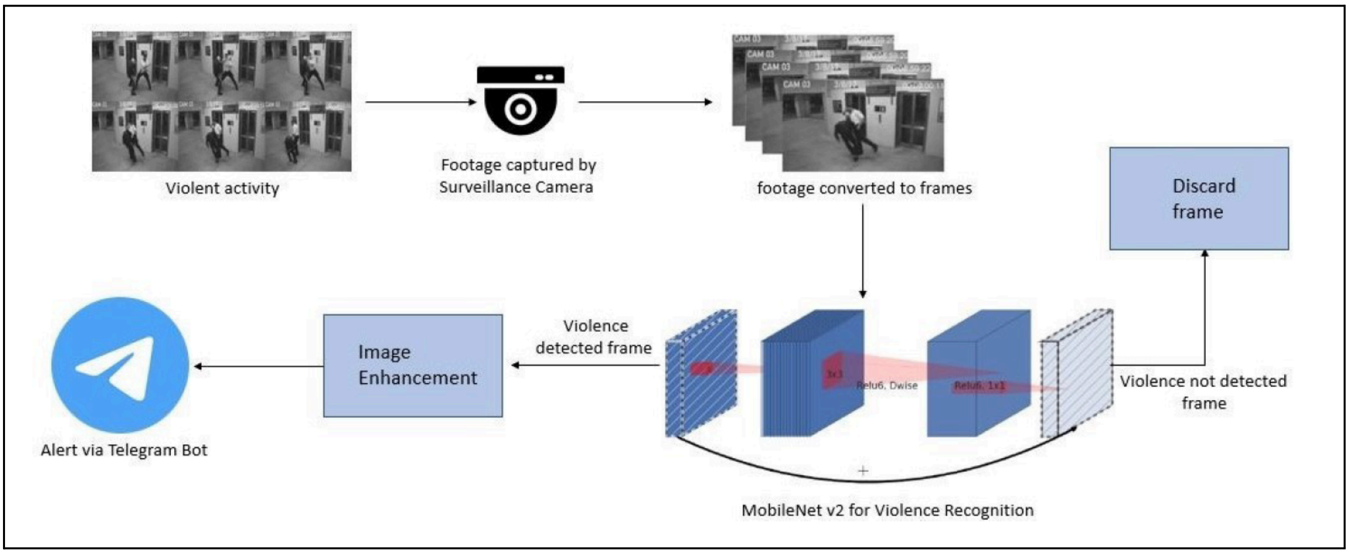

Violent activity
Footage captured by Surveillance Camera
footage converted to frames
Discard frame
Violence detected frame
Image Enhancement
Violence not detected frame
Alert via Telegram Bot
3x3
Relu6, Dwise
Relu6, 1x1
+
MobileNet v2 for Violence Recognition

Figure 3.2 Algorithm/Process Design

3.4 Methodology Applied

The methodology applied in this project relies on deep learning, specifically the use of the MobileNetV2 architecture for violence detection. This method takes advantage of spatio-temporal features to identify violent activities in real-time video streams. Additionally, image enhancement is performed using PIL to improve the quality of the frames related to violent incidents.

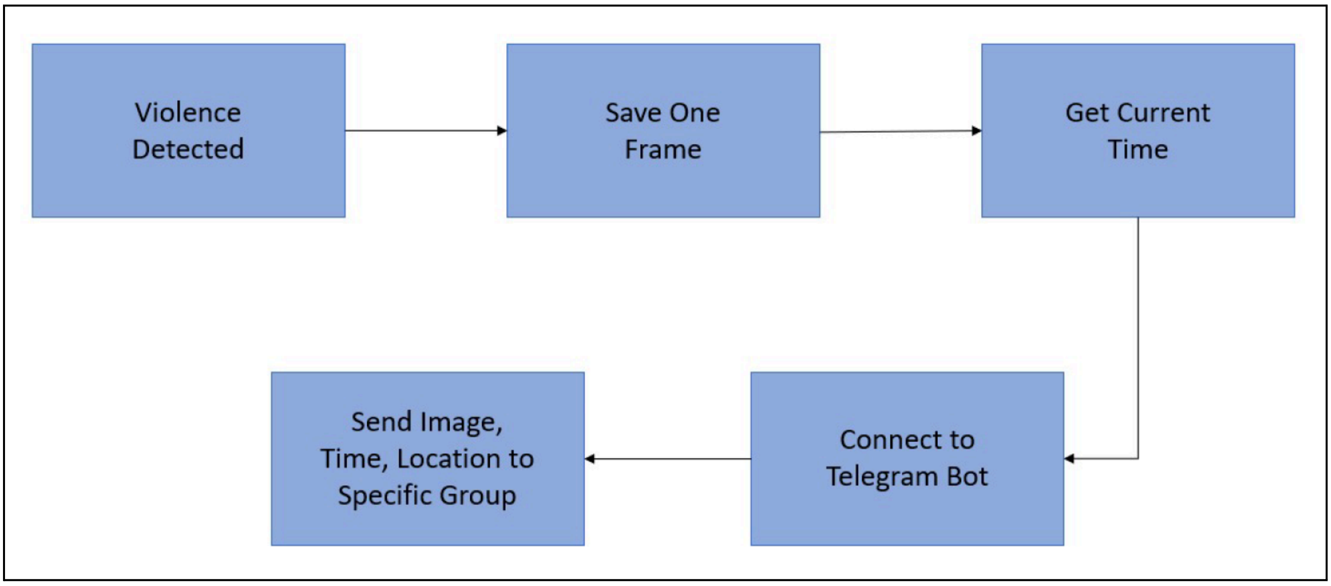


Figure 3.3 Methodology

3.5 Hardware & Software Requirements

Hardware: The hardware requirements for the project comprise a high-performance computer equipped with a robust CPU, such as an Intel Core i7 or an equivalent processor, accompanied by ample RAM for efficient processing. Additionally, sufficient storage capacity is necessary to accommodate and manage the video data effectively. While not mandatory, the inclusion of a GPU (Graphics Processing Unit) is highly recommended for accelerated deep learning tasks, particularly for faster model training.

Software:On the software side, the project necessitates Python as the primary programming environment and relies on deep learning frameworks like TensorFlow, alongside essential libraries for image processing and machine learning, including OpenCV, to support the development and execution of the violence detection system.

3.6 Experiment and Results

The project uses a dataset containing 1000 video clips, divided into two classes: violence and non-violence. The average duration of these video clips is 5 seconds, with the majority sourced from CCTV footage. Experiments involve training the MobileNetV2 classifier on this dataset to detect violence within video sequences.

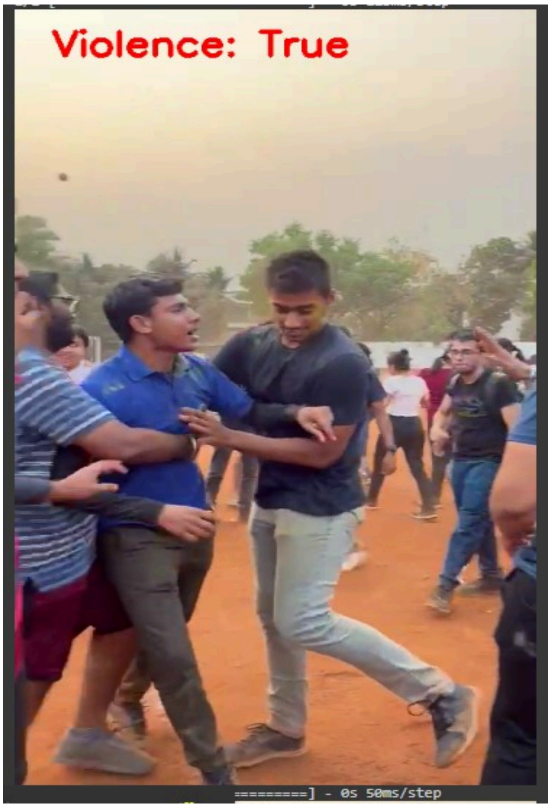
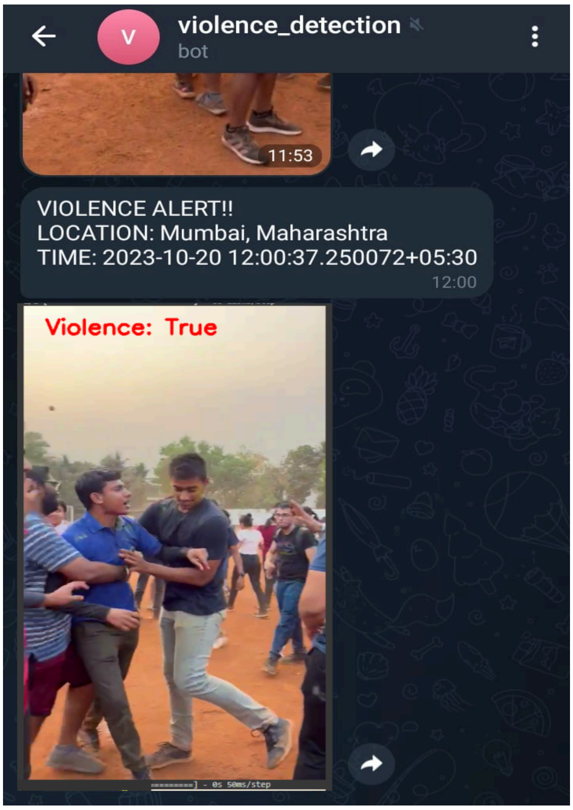
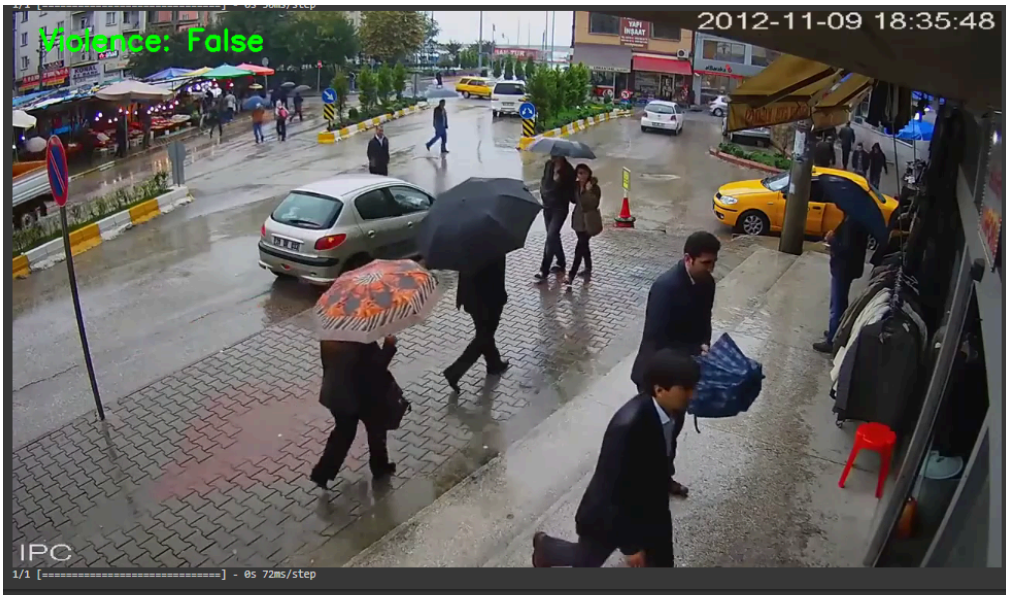


Fig 3.3 Video Frame Detected Violence



VIOLENCE ALERT!!
LOCATION: Mumbai, Maharashtra
TIME: 2023-10-20 12:00:37.250072+05:30
12:00

Fig 3.4 Alert System



```
1/1 [==============================] - 0s 72ms/step
```

Fig 3.4 Non-Violent Frame Detection

3.7 Result Analysis and Discussions

The results of the experiments, including the accuracy and efficiency of violence detection, are analyzed and discussed. The project evaluates the system's performance in real-time scenarios, highlighting its strengths and limitations.

3.8 Conclusion

In conclusion, the project aims to contribute to the development of an efficient and reliable real-time surveillance system for violence detection. The combination of deep learning techniques, image enhancement, and alert mechanisms ensures a timely response to violent incidents, laying a solid foundation for improved public safety. For future work, further enhancements in algorithm efficiency, real-time processing, and expanding the system's adaptability to various surveillance scenarios are potential areas of focus, as these developments can advance the system's effectiveness in mitigating violence and fostering social stability.

References

[1] J. C. Vieira, A. Sartori, S. F. Stefenon, F. L. Perez, G. S. de Jesus and V. R. Q. Leithardt, "Low-Cost CNN for Automatic Violence Recognition on Embedded System," in IEEE Access, vol. 10, pp. 25190-25202, 2022, doi: 10.1109/ACCESS. 2022.3155123.

[2] P. Sernani, N. Falcionelli, S. Tomassini, P. Contardo and A. F. Dragoni, "Deep Learning for Automatic Violence Detection: Tests on the AIRTLab Dataset," in IEEE Access, vol. 9, pp. 160580-160595, 2021, doi: 10.1109/ACCESS. 2021.3131315.

[3] M. -S. Kang, R. -H. Park and H. -M. Park, "Efficient Spatio-Temporal Modeling Methods for Real-Time Violence Recognition," in IEEE Access, vol. 9, pp. 76270-76285, 2021, doi: 10.1109/ACCESS.2021.3083273.

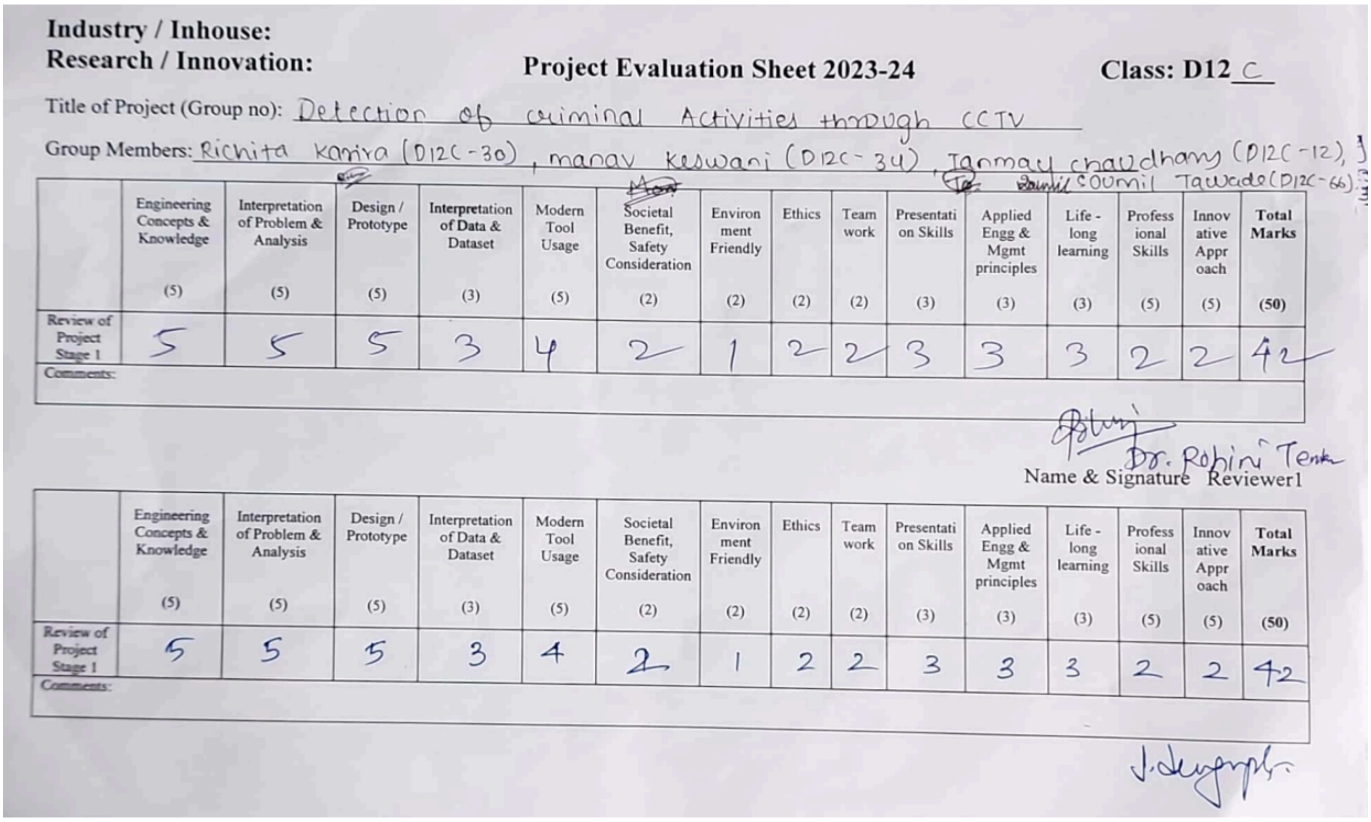
[4] Ullah FUM, Ullah A, Muhammad K, Haq IU, Baik SW. Violence Detection UsingSpatiotemporal Features with 3D Convolutional Neural Network. Sensors (Basel). 2019 May 30;19(11):2472. doi: 10.3390/s19112472.

[5] Khan SU, Haq IU, Rho S, Baik SW, Lee MY. Cover the Violence: A Novel Deep-Learning-Based Approach Towards Violence-Detection in Movies. Applied Sciences.2019; 9(22):4963. https://doi.org/10.3390/app9224963J. Clerk Maxwell,A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[6] M. Ramzan et al., "A Review on State-of-the-Art Violence Detection Techniques," in IEEE Access, vol. 7, pp. 107560-107575, 2019, doi: 10.1109/ACCESS. 2019.2932114.

[7] Sandler, Mark Howard, Andrew Zhu, Menglong Zhmoginov, Andrey Chen, Liang-Chieh. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. 4510-4520. 10.1109/CVPR.2018.00474.

Annexure



Industry / Inhouse:
Research / Innovation:

Project Evaluation Sheet 2023-24

Class: D12 C

Title of Project (Group no): Detection of criminal Activities through CCTV

Group Members: Richita Karira (D12C-30), manav keswani (D12C-34), Tanmay chaudhary (D12C-12), Soumil Tawade(D12C-66).

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg & Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	**Total Marks (50)**
Review of Project Stage 1	5	5	5	3	4	2	1	2	2	3	3	3	2	2	42
Comments:															

Dr. Rohini Temk
Name & Signature Reviewer1

	Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (3)	Applied Engg & Mgmt principles (3)	Life - long learning (3)	Profess ional Skills (5)	Innov ative Appr oach (5)	**Total Marks (50)**
Review of Project Stage 1	5	5	5	3	4	2	1	2	2	3	3	3	2	2	42
Comments:															