SOMAIYA VIDYAVIHAR

K J Somaiya Institute of Technology
(Formerly known as K J Somaiya Institute of Engineering and Information Technology)
An Autonomous Institute Permanently Affiliated to University of Mumbai.

Somaiya TRUST

Department of Computer Engineering

Academic Year 2022-23

LY Major Project A

# SentinelEyes: Real-time Violence Detection using CCTV Feed

**Group Members:**
Sahil Deshmukh
Shubh Joshi
Dhruv Mistry

**Supervisor:**
Prof. Chitra Bhole

# Outline of Project

- Introduction
- Problem statement
- Project Objective
- Literature Survey
- Summary/ Finding of Literature Survey
- Proposed system / idea
- System Architecture
- Algorithm used for proposed system
- Conclusion
- References

# INTRODUCTION

The project aims to develop "SentinelEyes," an intelligent violence detection system using advanced deep learning techniques. It combines YOLO v3 for person detection, OpenPose for pose estimation, and CNN-LSTM models for sequence classification. The system will automatically identify and classify instances of violence in real-time by analyzing CCTV feed. The main challenges include data collection, model architecture design, training, and optimization. Successful implementation will contribute to public safety efforts by enabling real-time monitoring and response to violent activities in various settings, such as security systems and social media platforms.

# PROBLEM STATEMENT

Develop an automated technological solution utilizing live CCTV feeds to detect street crime, violence, burglary, theft, infiltration, and unauthorized access in real-time. The system will employ computer vision and machine learning algorithms for incident detection, and upon identification, it will generate alerts to the nearest Police Station. The solution aims to enhance public safety, reduce response time, and deter criminal activities for improved security in urban environments. It will be scalable, prioritize data privacy, and provide an intuitive user interface for seamless integration and comprehensive monitoring.

# PROJECT OBJECTIVES

1. Develop an automated system capable of detecting incidents related to street crime, violence, burglary, theft, infiltration, and unauthorized access in real-time using live CCTV feeds.
2. Implement a robust and accurate mechanism for detecting firearms and dangerous weapons in video frames to enhance security measures.
3. Design the system to work effectively in crowded areas, considering challenges such as occlusions, overlapping individuals, and complex interactions.
4. Ensure the system operates seamlessly with live footage from CCTV cameras, processing the streams in real-time without significant delays or latency.
5. Integrate the solution with a notification system to alert the nearest Police Station immediately upon the detection of a potential violent incident.
6. Employ advanced techniques to minimize false alarms and increase the system's precision, reducing unnecessary interventions by law enforcement.
7. Develop a user-friendly interface for law enforcement personnel to access and monitor the detected incidents, facilitating quick responses and interventions.

# LITERATURE REVIEW

The literature review discusses various approaches for violence detection in CCTV feeds using neural networks. Researchers have explored different methods to extract video features and implement classifiers for violence recognition. Some methods focus on detecting specific violent elements like blood, gunshots, or explosions, but suffer from high false alarm rates. Others use spatio-temporal descriptors or hybrid deep learning networks for human detection and pose estimation. The use of CNN+LSTM and 3D CNNs has also been explored for feature extraction and action recognition.

Vision-based Fight Detection from Surveillance Cameras (Ş. Aktı et al.) introduced a dataset and used a CNN+LSTM network with a self-attention layer for fight detection. Liu et al. proposed a Global Context-Aware Attention LSTM for 3D action recognition from skeleton data. Various deep ensemble learning strategies have been employed to improve prediction accuracy. Additionally, good frame selection is crucial for reducing computational costs and enhancing model accuracy.

# LITERATURE REVIEW

The proposed system pipeline combines OpenPose for multi-person pose estimation and YoloV3 for person detection, followed by a small Li-Net CNN for violence action classification. This approach allows for real-time violence detection as OpenPose's bottom-up approach reduces computation time, and YoloV3 has fast processing time for person detection.

In summary, the literature review showcases the diverse range of neural network-based methods for violence detection in CCTV feeds, with a focus on real-time performance and improved accuracy through feature extraction, pose estimation, and action recognition techniques.

# FINDINGS OF LITERATURE SURVEY

1. Combining Different Learning Methods: Some studies use a mix of different learning methods to detect people and understand their poses. This helps in identifying aggressive behaviors based on how people are positioned.

2. Recognizing Patterns Over Time: Using a combination of CNN and LSTM, researchers are able to recognize patterns that happen over time in the videos. This helps in understanding if a situation is becoming violent or not.

3. Improving Accuracy with Many Models: By combining multiple models, researchers improve the accuracy of violence detection. It's like having different experts look at the same problem and give their opinions to make a better decision.

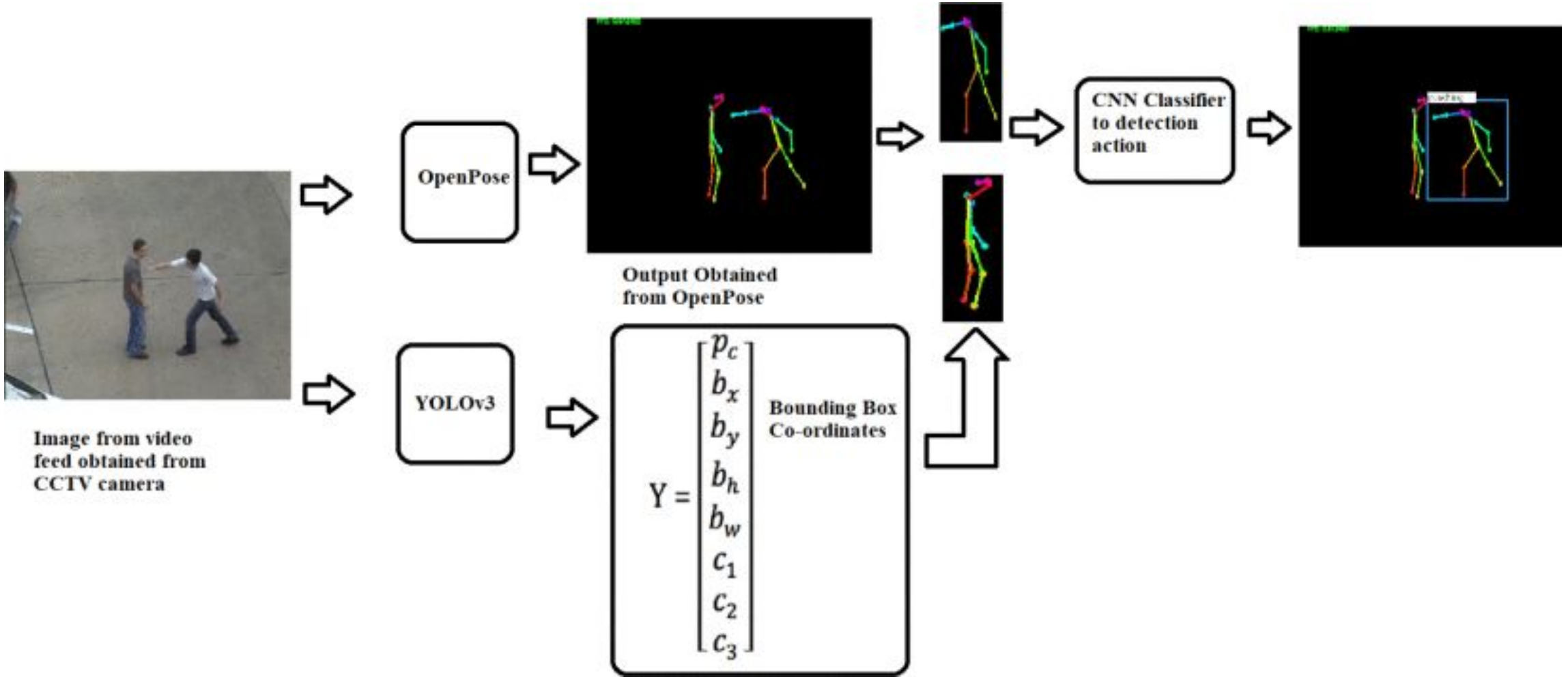
4. Detecting Violence in Real-Time: Detecting violence quickly is challenging, especially with many people in the video. However, using YOLO and OpenPose, researchers can process videos quickly to spot violence as it happens.

5. Useful Datasets for Training: Special datasets have been created that contain videos with violent activities. These datasets help researchers teach their models to recognize violence more effectively.

# PROPOSED SYSTEM

The proposed final system is an intelligent CCTV-based violence detection, alert, and security solution designed to automatically detect street crime, violence, burglary, theft, infiltration, and unauthorized access using live CCTV feeds. It integrates advanced neural network models for object detection, temporal analysis, and pose estimation to achieve real-time and accurate violence detection. By deploying strategically placed CCTV cameras in public areas, the system continuously captures and preprocesses live video feeds to enhance data quality. The YOLOv3 object detection model identifies violent objects, while a CNN-LSTM architecture analyzes frame and pose sequences, capturing temporal dependencies for improved accuracy. OpenPose estimates body positions for analyzing body language during potential incidents. Audio analysis detects gunshots, and facial recognition aids suspect identification. Predictive analytics identifies crime hotspots for proactive security measures. In emergencies, the system generates alerts to Police Stations and emergency units for swift action. Continuous model improvement refines detection capabilities, ensuring a safer environment with utmost consideration for ethical implications, data privacy, and compliance with local laws and regulations.

# System Architecture



# ALGORITHMS USED

1. **YOLOv3 (You Only Look Once):**
YOLOv3 is a real-time object detection model that efficiently identifies and localizes multiple objects in images or video frames. It is known for its speed and accuracy, making it well-suited for violence detection in live video feeds.

2. **CNN-LSTM (Convolutional Neural Network - Long Short-Term Memory):**
CNN-LSTM is a combination of CNNs and LSTM networks. It captures temporal dependencies in sequential data and excels at analyzing actions and behaviors over time. In violence detection, it improves accuracy by recognizing complex patterns associated with violent actions.

3. **OpenPose:**
OpenPose is a pose estimation framework that estimates body joint positions from images or videos. It provides a skeletal representation of human poses, enabling the analysis of body language and gestures. In the violence detection system, it helps detect aggressive poses and actions, enhancing the system's ability to identify violent behavior.

# CONCLUSION

The integration of YOLOv3, CNN-LSTM, and OpenPose in violence detection systems has shown promising results in accurately identifying and analyzing violent behavior. YOLOv3 provides efficient object detection, CNN-LSTM captures temporal dependencies, and OpenPose offers comprehensive pose estimation. By combining these techniques, researchers have achieved enhanced accuracy, real-time performance, and the ability to capture complex spatial and temporal patterns associated with violence. Further research in this domain should focus on refining these integrated models, addressing biases, and exploring new architectures to advance the field of violence detection using neural networks.

# REFERENCE


1-https://ieeexplore.ieee.org.library.somaiya.edu/document/9712793
2-https://ieeexplore.ieee.org.library.somaiya.edu/document/9712676
3-https://ieeexplore.ieee.org.library.somaiya.edu/document/9886172
4-https://ieeexplore.ieee.org.library.somaiya.edu/document/9675868
5-https://ieeexplore.ieee.org.library.somaiya.edu/document/9461921
6-https://ieeexplore.ieee.org.library.somaiya.edu/document/10092302
7-https://ieeexplore.ieee.org.library.somaiya.edu/document/9483558
8-https://ieeexplore.ieee.org.library.somaiya.edu/document/9904115
9-https://ieeexplore.ieee.org.library.somaiya.edu/document/9530829
10-https://ieeexplore.ieee.org.library.somaiya.edu/document/10097219

# Thank You…!