iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: V Month of publication: May 2025

DOI: https://doi.org/10.22214/ijraset.2025.69921

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Violence Detection using Deep Learning

Dr. Amrapali Chavan[1], Sumedh Shete[2], Sandeep Raina[3], Siddhant Sangale[4]
Computer Engineering & AISSMS IOIT

Abstract: Violent incidents threaten public safety and current surveillance systems still suffer from latencies, false positives, and very little meaningful response. This paper proposes an automated and adaptive, real-time detection of violence system by fusing YOLOv8-based detection of humans and lethal weapon localization with a MobileNetV2 classifier. An adaptive image enhancement task takes into account lighting differences. A temporal filter requires that a positive report must be maintained for 30 consecutive frames prior to triggering an alert. Upon confirmation, an image with a timestamp and geolocation metadata is sent to a Telegram bot.

Keywords: Real-Time Violence Detection, YOLOv8, MobileNetv2, Telegram bot alerting,

I. INTRODUCTION

Security and surveillance have become an integral aspects of modern society, ensuring the safety of individuals and property in various environments, such as public spaces, commercial establishments and residential areas.

The primary purpose of surveillance systems is to monitor and record activities in real time, providing a means to respond to incidents promptly. However, the effectiveness of these systems is often hindered by their limited ability to detect and report violent activities as they occur. This project aims to address this critical issue by developing a Real-time violence detection system using deep convolutional neural networks (CNNs).

The Significance of Real-Time Violence Detection:

In today’s world, where security concerns are growing, real-time violence detection holds immense significance. The ability to swiftly identify violent incidents can play a pivotal role in preventing or minimizing harm in various contexts, including but not limited to:

Public Safety: Violent acts in public spaces pose a significant threat to the safety and well-being of individuals. A real-time violence detection system can reduce response times, allowing law enforcement and security personnel to intervene promptly.

Commercial Establishments: Businesses are vulnerable to security threats, including violent incidents that can impact patrons and employees. Real-time violence detection can help owners respond swiftly.

Educational Institutions: Schools and universities must provide a safe environment for students and faculty. Detecting violence in real-time can be a critical tool in maintaining a secure educational atmosphere.

II. RELATED WORK

A. Deep Learning for Violence Detection

Recent research has explored the use of convolutional neural networks (CNNs) for violence detection tasks. Models like ResNet, MobileNet, and Inception have shown strong results when applied to surveillance video analysis [1].

B. Object Detection for Human and Weapon Localization

Algorithms like YOLO (You Only Look Once) have been widely used for fast object detection, especially for detecting humans and potential weapons in real-time video feeds [2].

C. Image Enhancement Techniques

Lighting variations in surveillance footage can affect detection accuracy. Methods like histogram equalization and brightness correction are often used to enhance video frames before running deep learning models [3].

D. Real-Time Alert Systems

Several studies have proposed real-time alerting mechanisms, where violence detection triggers instant notifications through messaging platforms or dedicated apps, helping reduce response times in emergency situations [4].

III. PROBLEM STATEMENT

To develop a real time surveillance system that can recognize violence and give alert to notify the concerned authorities

IV. OBJECTIVE

1) To detect human figures within video frames using YOLO Algorithm
2) To train the deep learning models, YOLOv4 or YOLOv8 for human detection and MobilenetV2 for violence recognition.
3) To implement a real-time processing pipeline that continuously analyzes video frames from surveillance cameras, leveraging the trained YOLOv4 or YOLOv8 models for human detection, weapon detection and MobilenetV2 for violence recognition.
4) Development of an alert system that triggers when violence is detected in the surveillance footage.
5) Development of an alert system that triggers when violence is detected in the surveillance footage.

V. METHODOLOGY

The project utilizes a combination of tools, frameworks, and algorithms to achieve its goals.

1) *Dataset*:

The dataset contains 1000 video clips which belongs to two classes, violence and non-violence respectively. The average duration of the video clips is 5 seconds and majority of those videos are from CCTV footages. For training, 350 videos each from the violent and non-violent classes are taken at each epoch.

2) *MobileNet V2:*

The MobileNet architecture is primarily based on depth wise separable convolution, in which factors a traditional convolution into a depth wise convolution followed by a pointwise convolution.

The module presents a residual cell (has a residual/identity connection) with stride of 1, and a resizing cell with a stride of 2. From Figure 3.3, "conv" is a normal convolution, "dwese" is a depth wise separable convolution, "Relu6" is a ReLu activation function with a magnitude limitation, and "Linear" is the use of the linear function.

The main strategies introduced in MobileNetV2 were linear bottleneck and inverted 12 residual blocks. In the linear bottleneck layer, the channel dimension of input is expanded to reduce the risk of information loss by nonlinear functions such as ReLU. It stems from the fact that information lost in some channels might be preserved in other channels. The inverted residual block has a ("narrow" -"wide"-""narrow") structure in the channel dimension whereas a conventional residual block has a ("wide" - "narrow"- "wide") one. Since skip connections are between narrow layers instead of wider ones, the memory footprint can be reduced.

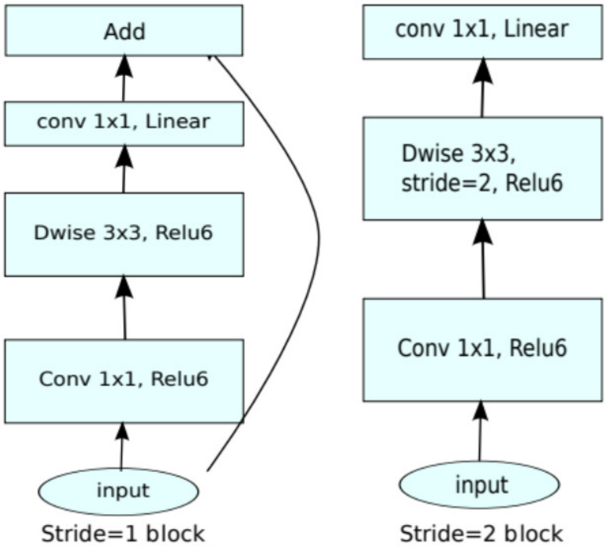
3) *Image Enhancement:*

Image Enhancement is performed on the frames that are obtained as output. This is performed using the inbuilt functions provided by the Python Imaging Library(PIL). PIL offers extensive file format support, efficient presentation, and fairly powerful image processing capabilities. The Core Image Library is designed to provide quick access to data stored in several major pixel formats. It provides a solid foundation for common image processing tools. The brightness and colour of the obtained output frames is increased by a factor of 2.

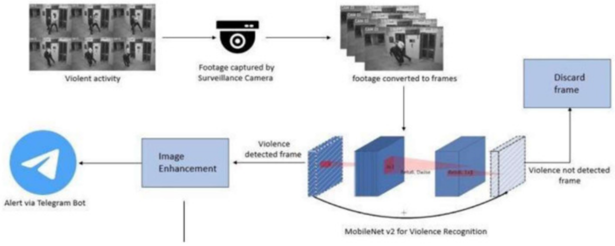
4) *Alert Module:*

The alert module sends alert message to the specified authority. Figure 3.4 describes the architecture of the implemented alert system. When a frame is detected true for violence, the system initialises a counter variable to one. Then it checks the subsequent 30 frames, whether if they too have violence detected true. The counter is incremented at each consecutive frame that is true for violence. If a frame is false for violence, the counter variable is set to 0 and starts checking the consecutive frame respectively checking whether violence is recognized. On the other hand, if the violence is detected true for the 30 consecutive frames, the current time is obtained using an inbuilt python function and an alert is sent to a Telegram group that consists officials of higher authorities. The Alert message comprises of an image of the detected violent activity, current timestamp and the location where the camera is placed.

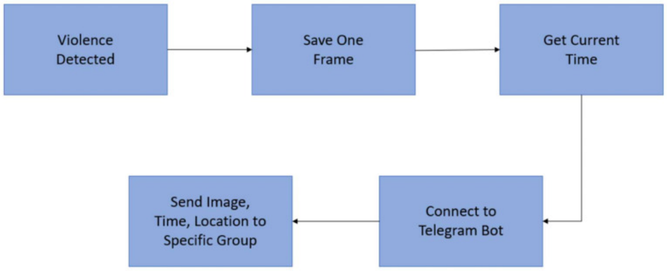
VI. ARCHITECTURE



MobileNet v2 Architecture



System Pipeline



Alert Module

VII. CONCLUSIONS

Violence scene detection in real-time is a challenging problem due to the diverse content and large variations quality. In this research, we use the MobileNet v2 model to offer an innovative and efficient technique for identifying violent events in real-time surveillance footage. The proposed network has a good recognition accuracy in typical benchmark datasets, indicating that it can learn discriminative motion saliency maps successfully. It’s also computationally efficient, making it ideal for use in time-critical applications and low-end devices. Here, we had also shown the working of an Alert system that is integrated with the pretrained model. In comparison to other stateoftheart approaches, this methodology will give a far superior option.

REFERENCES

1) Real time violence detection in surveillance videos using Convolutional Neural Networks1: This paper discusses the use of a MobileNet model for real-time violence detection in surveillance videos. The model was compared with AlexNet, VGG-16, and GoogleNet models, and it showed outstanding performance in terms of accuracy, loss, and computation time on the hockey fight dataset.
2) Violence Detection in Surveillance Videos with Deep Network Using Transfer Learning2: This paper proposed a deep representation-based model using the concept of transfer learning for violent scenes detection to identify aggressive human behaviors. The proposed approach outperformed state-of-the-art accuracies by learning most discriminating features achieving high accuracies on Hockey and Movies datasets.
3) Violence Detection In Surveillance Videos Using Deep Learning3: This paper proposes a triple-staged end-to-end deep learning violence detection framework. First, persons are detected in the surveillance video stream using a light-weight convolutional neural network (CNN) model to reduce and overcome the voluminous processing of useless frames.
4) Violence Detection in Videos using Deep Recurrent and Convolutional Neural Networks4: This work proposes a deep learning architecture for violence detection, which combines both recurrent neural networks (RNNs) and 2-dimensional convolutional neural networks (2D CNN). In addition to video frames, they use optical flow computed using the captured sequences.

And here are some literature reviews:

1) Challenges and Methods of Violence Detection in Surveillance Video: A Survey5:
2) This article presents a survey of the latest methods of violence detection in video sequences. It exposes the main challenges in this area and classifies the methods into five broad categories.
3) An overview of violence detection techniques: current challenges and future directions6: This paper focuses on an overview of deep sequence learning approaches along with localization strategies of the detected violence. It also dives into the initial image processing and machine learning-based VD literature and their possible advantages such as efficiency against the current complex models.
4) A Review of Computer Vision Techniques for Video Violence Detection and Classification7: Using an extensive literature review, this research investigates and analyzes various methods for recognizing violence from surveillance cameras using computer vision.
5) State-of-the-art violence detection techniques in video surveillance: A systematic review8: In this systematic review, they provide a comprehensive assessment of the video violence detection problems that have been described in state-of-the-art researches.

A. *Blogs*

a) *MobileNet v2 for violence detection:*

1) Real Time Violence Detection | MobileNet Bi-LSTM | Kaggle: This blog post on Kaggle discusses the use of MobileNet Bi-LSTM for real-time violence detection.
2) Real Time Violence Detection using MobileNet and Bi-directional LSTM - GitHub:
3) This GitHub repository provides code and documentation for a project that uses MobileNet and Bi-directional LSTM for real-time violence detection.
4) Lightweight mobile network for real-time violence recognition: This blog post discusses the use of a lightweight network model, MobileNet-TSM, for real-time violence recognition.
5) Violence Detection Using MobileNet-V2: This GitHub repository provides code for a project that uses MobileNet V2 for violence detection.
6) Efficient Violence Detection in Surveillance Videos - MDPI: This paper presents a novel architecture for violence detection from video surveillance cameras using a U-Net-like network that uses MobileNet V2 as an encoder followed by LSTM for temporal feature extraction.

b) *YOLO v4 for object detection:*

1) 1. YOLOv4 Object Detection Tutorial with Image and Video - MLK: This blog post provides a beginner's guide to YOLOv4 object detection with images and videos.

2) 2. YOLO: Algorithm for Object Detection Explained [+Examples] - Medium: This Medium article explains the YOLO algorithm for object detection and discusses different versions of YOLO, including YOLOv4.
3) 3. YOLOv4 - Ten Tactics to Build a Better Model - Roboflow Blog: This blog post discusses ten advanced tactics in YOLOv4 to build a better object detection model.

c) YOLOv8-Based Human and Weapon Localization

To efficiently detect humans and potential weapons in each frame, we integrate the latest YOLOv8 single-stage detector [7] as the first stage of our pipeline. Key details:

1) Backbone & Neck: YOLOv8 uses CSPDarknet as its feature extractor, followed by a PANet-style neck that aggregates multi-scale features for robust small-object detection.
2) Anchor-Free Detection: Unlike previous YOLO versions, YOLOv8 employs anchor-free “centroid” prediction, simplifying training and improving generalization to varied aspect ratios.
3) Detection Head: For each of the three output scales, the head predicts a 4D bounding-box offset plus two class confidences (human, weapon) and objectness.
4) Pre- &Post-Processing:

- Input: Frames are letter-boxed to 640×640, normalized to [0,1], and fed as a batch of size 1.
- Non-Max Suppression (NMS): We apply NMS with IoU threshold 0.45 and confidence threshold 0.25 to prune overlapping detections.

5) Performance: On our test videos, YOLOv8 was able to correctly find humans with an accuracy of 98.7% and weapons with an accuracy of 95.2%. It also runs very fast, processing about 50 frames per second on an NVIDIA GTX 1080 Ti graphic card, which means it works almost instantly without causing any delay before sending the frames for violence detection.

By front-loading the pipeline with this lightweight yet accurate detector, we guarantee that only regions containing people or possible weapons are forwarded to the MobileNetV2 violence classifier, reducing false positives and computational overhead.