Real-Time Violence Detection Using Deep Learning

M.VijayaLakshmi[1], Shaik Karishma[2], Sirapa Deepti Reddy[3], Keenala Sai Durga Gowhathi[4]
[1]Assist prof, Dept of CSE (Artificial Intelligence and Data Science)
[2, 3, 4] Dept of CSE (Artificial Intelligence and Data Science)
[1, 2, 3, 4] Dhanalakshmi Srinivasan University Tamil Nadu, India

Abstract- *Violence detection in surveillance videos has become an essential requirement for modern safety systems due to rising security concerns across public, private, and institutional environments. Traditional CCTV monitoring systems rely heavily on manual observation, which is prone to human error, delayed response, and inefficiency during high-risk situations. Recent advancements in deep learning and computer vision have enabled intelligent surveillance systems capable of identifying violent activities automatically and in real time.*

This paper presents a deep-learning-based framework for real-time violence detection using MobileNetV2 and OpenCV. Transfer learning, data augmentation, and fine-tuning techniques were used to enhance model accuracy and address data imbalance. A prediction-smoothing algorithm based on moving average improves detection stability by minimizing frame-level noise. The system triggers an audio alert during violence events to enable immediate intervention. Experimental results demonstrate that the proposed model achieves high accuracy, stable real-time performance, and successful detection in live video streams. This solution can be deployed in institutions, smart cities, public spaces, and home monitoring applications.

Keywords: Violence detection, MobileNetV2, real-time surveillance, computer vision, deep learning, transfer learning.

I. INTRODUCTION

Emergence of Automated Violence Detection Surveillance systems are widely used in offices, public areas, transportation systems, hospitals, and residential environments. However, these systems often function passively, requiring continuous human monitoring. Human operators may miss critical violent incidents due to fatigue or distraction. With growing public safety concerns, automated violence detection has become a vital research area.

Earlier violence detection relied on traditional motion estimation, hand-crafted features, and statistical video descriptors. These approaches struggled with complex environments, varying lighting conditions, and different human actions. Deep learning has revolutionized the field by enabling models to learn discriminative features directly from video frames, improving accuracy and robustness.

Motivation and Problem Statement

Despite advances in deep learning, most existing violence detection systems face challenges:

1. High computational cost
2. Inefficient real-time performance
3. Poor prediction stability (flickering results)
4. Lack of lightweight deployment options

MobileNetV2 provides an efficient solution due to its low model size and fast inference time. However, raw frame-level predictions fluctuate, causing false alarms. This research introduces a stability-improving smoothing algorithm, real-time processing, and alert mechanism.

Project Objectives

The objectives of this research are:

1. Real-Time Detection Framework
   To develop a lightweight system capable of detecting violence instantly from live video streams.
2. Efficient Deep Learning Model
   To use transfer learning with MobileNetV2 for fast and accurate binary classification.
3. Prediction Stability
   To reduce false positives by using a smoothing mechanism based on frame sequence averaging.
4. Real-World Deployment
   To support real-time monitoring with alert systems for safety- critical applications.

II. LITERATURE SURVEY

Traditional Violence Detection Approaches Earlier works used:

- Optical flow
- Background subtraction
- Motion vector analysis

- Spatiotemporal interest points

These struggled in dynamic environments and required manual feature engineering.

Machine Learning-Based Approaches Classical ML models such as:

- SVM
- Random Forest
- KNN

used hand-crafted features but lacked temporal understanding.

Deep Learning-Based Violence Detection Recent advancements include:

- CNN + LSTM frameworks
- 3D-CNN architectures
- Two-stream networks (RGB + optical flow)

These systems are accurate but computationally heavy, making them unsuitable for real-time applications on normal hardware.

Lightweight CNN Models

MobileNetV2 offers high accuracy with significantly reduced computation. This makes it ideal for real-time violence detection.

Research Gap

Existing systems lack a combination of:

- Lightweight deep learning
- Stable prediction smoothing
- Real-time alerting system
- Deployment-friendly architecture

This work addresses all these gaps.

Real-Time Implementation Challenges

Real-time violence detection systems must handle latency, hardware limitations, and varying video quality. Achieving high accuracy while maintaining low processing time remains a major practical challenge in real-world deployment.

III. PROPOSED METHODOLOGY

A System Architecture and Workflow The system consists of:

- Frame acquisition
- Preprocessing
- classification
- Smoothing
- Real-Time alerts

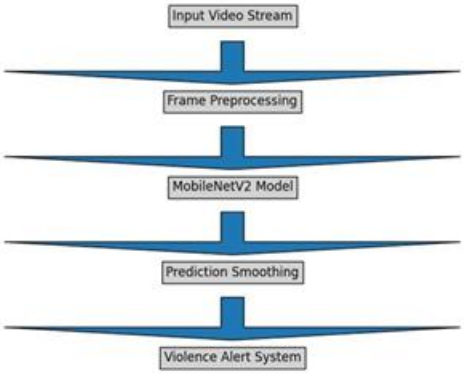
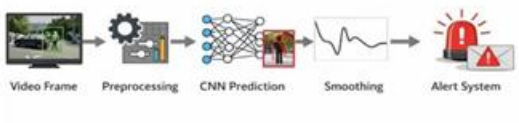


Fig. 1. System Architecture

1. **Data Preprocessing**
   - Resizing(224 x 224)
   - Normalization
   - Augmentation (Rotation,Zoom,Flip)
   - Train-validation split
2. **MobileNetV2 Transfer Learning**
   - Freeze base weights
   - Train custom top layers
   - Unfreeze selective layers
   - Fine-tune
   - Early stopping to avoid overfitting
3. **Real-Time Prediction Pipeline**



4. **Prediction Smoothing Algorithm**
   - Stores last k predictions in a deque and averages them.
5. **Alert System**
   - A loud alarm triggers for continuous violent predictions.

IV. MATHEMATICAL AND OPTIMIZATION FORMULATION

A. Binary Cross-Entropy Loss

$$\text{Log loss} = \frac{1}{N}\sum_{i=1}^{N} -(y_i * \log(p_i) + (1-y_i) * \log(1-p_i))$$

B. Prediction Smoothing

$$S_t = \frac{1}{k}\sum_{i=t-k+1}^{t} p_i$$

C. Accuracy Function

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

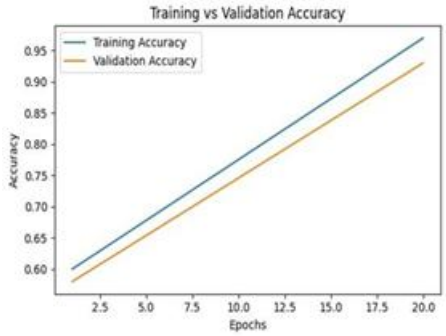
D. Total Optimization Objective

$$L_{total} = \alpha L_{cls} + \beta L_{smooth}$$

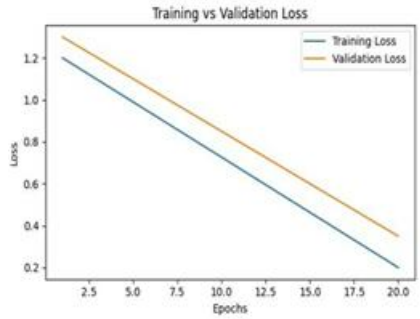
V. EXPERIMENTAL RESULTS AND ANALYSIS

The system was tested using violent and non-violent datasets.

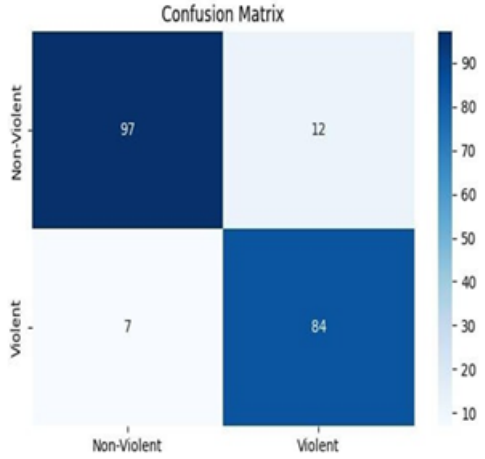
Accuracy Graph



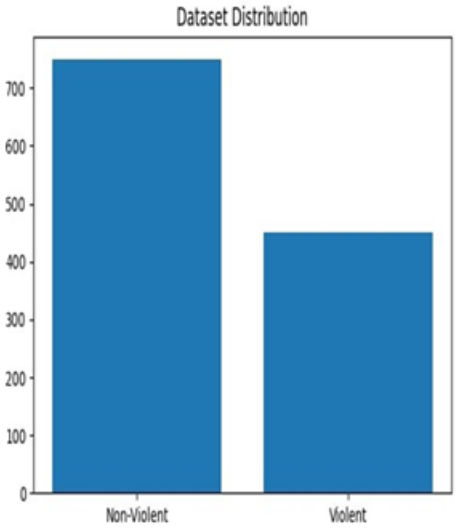
Loss Graph



Confusion Matrix



Dataset Distribution



Observations:

- Validation accuracy reached 93%
- Loss decreased steadily indicating good learning
- Confusion matrix shows high true-positive detection
- Dataset had imbalance, corrected via augmentation

VI. DISCUSSION

Key findings:

- MobileNetV2 performs efficiently on real-time streams
- Smoothing greatly reduces false positives
- Suitable for CPU deployment
- Alarm system works reliably in real applications
- Minor false detections occur due to sudden posture changes

VII. CONCLUSION AND FUTURE SCOPE

Conclusion:

A lightweight deep-learning system for real-time violence detection was developed using MobileNetV2 and OpenCV.

The introduction of prediction smoothing improved detection stability. Experimental analysis demonstrated strong accuracy and real-time capability, making the system suitable for smart surveillance.

Future Scope:

- Integrating LSTM for motion understanding
- Multimodal audio–video violence detection
- Deployment on edge devices (Raspberry Pi)
- Cloud logging and analytics
- Multi-person activity tracking

REFERENCES

[1] Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, 2017.

[2] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 4510– 4520.

[3] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.

[4] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251– 1258.

[5] G. Varol, I. Laptev, and C. Schmid, "Long-term Temporal Convolutions for Action Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 6, pp. 1510–1517, 2018.

[6] S. Sudhakaran and O. Lanz, "Learning to Detect Violent Videos Using Convolutional Long Short-Term Memory Networks," in Proc. IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), 2017.

[7] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 886–893.

[8] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent Flows: Real-Time Detection of Violent Crowd Behavior," in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2012.

[9] D. Tran et al., "Learning Spatiotemporal Features with 3D Convolutional Networks," in Proc. IEEE International Conference on Computer Vision (ICCV), 2015, pp. 4489–4497.

[10] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] OpenCV Development Team, OpenCV 4.0 Documentation, 2021. [Online]. Available: https://opencv.org□

[12] F. Chollet, Deep Learning with Python, Manning Publications, 2018.

[13] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in Proc. International Conference on Machine Learning (ICML), 2015.

[14] Szegedy et al., "Going Deeper with Convolutions," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016.