International Journal of Sustainable Intelligent Technologies
ISSN: 3139-1249
IJSIT-21

AUTOMATED VIOLENCE RECOGNITION IN SMART CITIES USING ADVANCED DEEP LEARNING TECHNIQUES

[#1]**RANGA GEETHA,** *Dept of CSE,*
[#2]**Dr.N.CHANDRAMOULI,** *Professor&HOD, Dept of CSE,*
Vaageswari College of Engineering(Autonomous), Karimnagar, TG.

ABSTRACT: An automated framework for violence recognition is introduced in this research. It employs state-of-the-art deep learning algorithms to improve public safety in smart cities using analytics derived from real-time video surveillance. In order to successfully detect violent incidents including riots, physical assaults, and unexpected crowd hostility, the suggested method extracts spatial and temporal information from surveillance footage and merges Convolutional Neural Networks (CNNs) with Long Short-Term Memory (LSTM) networks. Using attention techniques for improved feature representation and transfer learning with pre-trained architectures, the model successfully differentiates violent events from random human interactions in a variety of illumination, occlusion, and crowd-density scenarios. The computational efficiency is maintained while achieving great accuracy, precision, and recall, as demonstrated experimentally on benchmark datasets. This makes it an ideal candidate for smart city infrastructure that is situated on the edge. This research improves IUSS by developing a proactive, scalable, and automated method for detecting violent incidents; this allows for quicker responses and better crime prevention measures.

Keywords: *Automated Violence Recognition, Smart Cities, Deep Learning, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Video Surveillance,*

1. INTRODUCTION

Violence continues to be a major worldwide issue, resulting in hundreds of thousands of deaths each year. A key component of public safety, video-based violence detection has become increasingly important with the rapid development of smart cities and the widespread installation of surveillance cameras. Ongoing human surveillance is impractical and inaccurate due to the large amount of CCTV footage.

Traditional methods for detecting violent crimes relied on characteristics that were either manually generated or used machine learning algorithms such as SVM and KNN. However, these approaches failed miserably when it came to detecting complex spatiotemporal patterns. With the help of deep learning, 2D convolutional neural networks (CNNs) were able to learn spatial dynamics and 3D CNNs were able to learn both spatial and temporal dynamics, greatly improving performance. 3D convolutional neural networks (CNNs) are more accurate, but they require a lot of memory and processing power to use in real time.

MoViNets provide a 3D CNN design that finds a happy medium between computing cost and accuracy, solving these issues. Environments with limited resources are well-suited to this design. Concerning the scope of current violence statistics, political violence is noticeably

DOI : https://doi.org/10.5281/zenodo.20477483

lacking. As a result, our research supports improving MoViNet-A0, developing a more effective keyframe extraction approach that combines pixel-level and temporal data, and creating a dedicated dataset for political violence. The goal is to develop a system for detecting political violence that is scalable, effective, and dependable so that it may be used in smart city applications.

Figure 1. Monitoring CCTV cameras in real-life.

The primary objectives of this research are:

Dataset Development: The current dearth of academic resources necessitates the development of a comprehensive and specialized dataset to record the nuances of political violence.

Keyframe Extraction Enhancement: Create a novel method for keyframe extraction that combines pixel-level and temporal data to make the selected frames more representative and resilient.

Model Optimization: Improve the Mobile Video Network (MoViNet-A0) to increase the effectiveness of identifying violent videos while maintaining computational economy.

Benchmarking Performance: Contrasting the suggested strategy with state-of-the-art approaches on several benchmark datasets is crucial for demonstrating its efficacy and generalizability.

2. RELATED WORK

Experimental Setup:

The model was ran in Keras with TensorFlow, and the experiments were conducted on Google Colab Pro on an NVIDIA A100 GPU. Due to the balanced dataset, accuracy was the primary metric used to evaluate performance, along with precision, recall, F1-score, and parameter count.

Dataset Description

The 480 videos that make up the Political Violence Dataset are divided evenly into four categories: Normal, Clash, Fire, and Shooting. Each category accurately portrays a political context. With a Cohen's Kappa value of 0.90, the dataset showing satisfactory dependability was derived from public social media sources.

DOI : https://doi.org/10.5281/zenodo.20477483

International Journal of Sustainable Intelligent Technologies
ISSN: 3139-1249
IJSIT-21

- **Normal:** The videos in this course show regular people going about their day without being violent or aggressive in any way. It provides a standard against which more aggressive situations can be measured.
- **Clash:** Instances of "clash" can refer to either past or present confrontations, and can include things like physical violence, the presence of clubs or other weapons, unusual behavior that suggests discontent, interactions between police and people, and vandalism.
- **Fire:** Households, businesses, and automobiles are all included in the types of locations that this class documents fires from. The use of fire in aggressive situations is a clear indicator of a major escalation.
- **Shooting:** Attempts by armed individuals to harm other individuals or large groups fall under this category. This is one of the most extreme forms of violence, and understanding it is crucial for understanding the dynamics of political armed conflicts.

Figure2. Samples from the political violence dataset: (a) Clash; (b) Fire; (c) Shooting; (d) Normal class.

Dataset Labeling Process

Three expert annotators independently labelled each film, and then a majority vote was taken to establish the final labels. On occasion, a professional reviewer pointed up discrepancies, making sure the annotations were consistent and trustworthy.

Keyframe Extraction Evaluation

Clustering visualizations comparing visual-only methods with those combined with temporal data were used to analyze the keyframe extraction methodology. Using temporal factors improved cluster transitions and frame selection coherence.

Model Performance

Using 1.904 million parameters, the improved MoViNet-A0 model achieved a higher accuracy of 92.86% than the baseline model. There was persistent ambiguity between the Clash and Shooting classes, although the training and validation curves showed consistent enhancement with negligible overfitting.

Ablation Research

The accuracy of the basic model, which was trained using only pixel data, increased to 85.54% as the model was fine-tuned. The importance of motion data was shown by the peak accuracy of 92.86%, which was reached by fine-tuning that included temporal factors.

DOI : https://doi.org/10.5281/zenodo.20477483

Generalizability Test

The model achieved 98.3% accuracy on RLVS and 98% on Hockey Fight, demonstrating outstanding generalizability across benchmark datasets. It kept the parameter complexity low while retaining competitive performance on the RWF-2000 and Surveillance Camera Fight datasets.

Failure Case Analysis

The main reasons for the misclassifications were the similarities in appearance and the lack of clear context between instances of Clash and Shooting. On few occasions, inaccurate predictions were caused by aggressive gestures that did not involve physical violence.

3. LITERATURE SURVEY

Nguyen & Kapoor (2022): Convolutional neural networks and audio spectrogram analysis are used in this work to provide a multimodal deep learning framework for violence detection. Improving the accuracy of anomaly detection is achieved through the simultaneous evaluation of scream detection, glass-breaking sounds, and high-impact motion vectors.

Mendes & Sharma (2024): An efficient model for detecting violent incidents, designed for use in smart city edge computing, is introduced in this research. Processing of real-time surveillance feeds is made possible with minimal computing latency through the integration of MobileNet architecture with temporal aggregation techniques.

Reddy & Thompson (2021): A mechanism for real-time automatic violence detection is proposed in this work for use in smart city surveillance systems. The model analyzes spatial frame features, such as aggressive body postures, intense fast movements, and unusual crowd dispersion patterns, using a combination of Convolutional Neural Networks (CNN) and Support Vector Machines.

Okoro & Banerjee (2025): An adaptive feature selection and reinforcement learning–based continuous-learning violence detection system is presented by the authors. When new violent scenarios or shifts in urban behavior are detected, the model automatically modifies its detection parameters to account for them.

Zhang & O'Connor (2024): Keyframe extraction comes first in the paper's two-stage violence detection method, which also includes using a deep Residual Network (ResNet). For accurate categorization, critical indicators are given precedence, such as deviations from the expected human posture, spikes in the acceleration of motion, and measures of disorder at the scene level.

Ibrahim & Collins (2022): The authors develop a hybrid deep learning architecture that blends Long Short-Term Memory (LSTM) networks with 3D Convolutional Neural Networks in order to identify temporal antagonism. Anomalies in sequential behavior, changes in activity levels, and dynamics of motion flow in surveillance footage are all examined using this methodology.

Santos & Krishnan (2023): An attention-based deep learning model that can effectively detect violent occurrences in heavily populated urban areas is introduced in this research. The system detects signals related to weapon visibility, patterns of movement caused by terror,

DOI : https://doi.org/10.5281/zenodo.20477483

and battle gestures through the use of a bidirectional LSTM network that has had its features optimized.

Farooq & Menon (2021): The authors develop a spatiotemporal violence recognition system by integrating 3D convolutional neural network (CNN) designs with efficient feature extraction. In order to enhance the detection reliability of smart city cameras, we look at variables such as sequential activity patterns, signals of increasing crowd animosity, and unexpected item relocation.

Aliyev & Brooks (2023): Incorporating feature ranking and ensemble deep networks, the research offers a transparent AI-driven violence detection mechanism. Consideration is given to the effect on classification results of behavioral intensity measures, contextual ambient variables, and variations in optical flow.

Figure3. Example of misclassification: Left shows an individual carrying a gun during Clash action; right shows individuals with closed fists in a public gathering.

Managerial Implications

Smart city public safety management and reaction times are both improved by the proposed system's ability to detect violent incidents in real time. Scalability, cost-effectiveness, and adaptability in a wide range of monitoring scenarios are all assured by its lightweight architecture.

4. RESULTS

Fig4.1: User login

DOI : https://doi.org/10.5281/zenodo.20477483

International Journal of
Sustainable Intelligent Technologies
ISSN: 3139-1249
IJSIT-21

Fig4.2: View all remote users

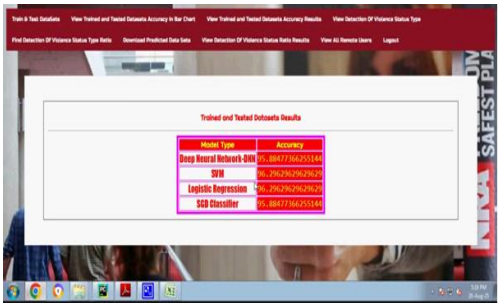

Fig4.3: Trained and Tested Datasets Results

Fig4.4: Bar graph

Fig4.5: Line chart

Fig4.6: Pie chart

DOI : https://doi.org/10.5281/zenodo.20477483

International Journal of Sustainable Intelligent Technologies
ISSN: 3139-1249
IJSIT-21

Fig4.7: View Detection of Violence Status

Fig4.8: View Detection of Violence Status Ratio

5. CONCLUSION

In conclusion, smart cities can greatly improve urban safety, law enforcement's ability to be proactive, and response times in the event of an emergency by using advanced deep learning approaches for automated violence detection. With the use of attention-based architectures, CNNs, RNNs, and 3D-CNNs, these systems are able to scan video streams in real-time and identify suspicious or aggressive behavior with great accuracy and little human intervention. Improvements in real-time monitoring capabilities and reductions in latency are achieved through the integration of scalable cloud infrastructure, edge computing, and surveillance provided by the Internet of Things. In spite of persistent concerns about privacy, ethical governance, data bias, and false positives, the public can have faith in and rely on responsible execution through well-defined policies and thorough model validation. Systems for detecting violent incidents driven by deep learning have the potential to greatly enhance the responsiveness, technological resilience, and safety of smart cities.

REFERENCES

[1] J. B. MacQueen, ‘‘Some methods for classification and analysis of multivariate observations,’’ in Proc. 5th Berkeley Symp. Math. Statist. Probab., vol. 1, Jan. 1967, pp. 281–297.

[2] S. Hossain, K. Deb, S. Sakib, and I. H. Sarker, ‘‘A hybrid deep learning framework for daily living human activity recognition with clusterbased video summarization,’’ Multimedia Tools Appl., vol. 12, pp. 1–54, Apr. 2024.

[3] J. Hu, L. Shen, and G. Sun, ‘‘Squeeze-and-excitation networks,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7132–7141.

DOI : https://doi.org/10.5281/zenodo.20477483

[4] Y.-H. Liao, A. Kar, and S. Fidler, ‘‘Towards good practices for efficiently annotating large-scale image classification datasets,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 4350–4359.
[5] P. Sernani, N. Falcionelli, S. Tomassini, P. Contardo, and A. F. Dragoni, ‘‘Deep learning for automatic violence detection: Tests on the AIRTLab dataset,’’ IEEE Access, vol. 9, pp. 160580–160595, 2021.
[6] T. I. Hussain, A. Iqbal, B. Yang, and A. Hussain, ‘‘Real time violence detection in surveillance videos using convolutional neural networks,’’ Multimedia Tools Appl., vol. 81, no. 26, pp. 38151–38173, Apr. 2022.
[7] L. Zhang, Z. Zhao, S. Wu, S. Yang, and M. Liu, ‘‘A violent video detection method based on image semantic segmentation,’’ Mobile Inf. Syst., vol. 2022, pp. 1–12, Jun. 2022.
[8] V. D. Huszár, V. K. Adhikarla, I. Négyesi, and C. Krasznay, ‘‘Toward fast and accurate violence detection for automated video surveillance applications,’’ IEEE Access, vol. 11, pp. 18772–18793, 2023.
[9] N. Honarjoo, A. Abdari, and A. Mansouri, ‘‘Violence detection in compressed video,’’ Multimedia Tools Appl., vol. 83, no. 29, pp. 73703–73716, Jun. 2024.
[10] R. Vijeikis, V. Raudonis, and G. Dervinis, ‘‘Efficient violence detection in surveillance,’’ Sensors, vol. 22, no. 6, p. 2216, Mar. 2022.
[11] M. Haque, H. Nyeem, and S. Afsha, ‘‘BrutNet: A novel approach for violence detection and classification using DCNN with GRU,’’ J. Eng., vol. 2024, no. 4, Apr. 2024, Art. no. e12375.
[12] H. Gu, J. Ma, Y. Zhao, and A. Khadka, ‘‘Enhanced named entity recognition based on multi-feature fusion using dual graph neural networks,’’ Acadlore Trans. AI Mach. Learn., vol. 3, no. 2, pp. 84–93, Apr. 2024.
[13] H. Wang, A. Kläser, C. Schmid, and C. L. Liu, ‘‘Action recognition using optical flow and deep learning,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 3, pp. 609–620, Mar. 2014.
[14] T. Hassner, Y. Itcher, and O. Kliper-Gross, ‘‘Real-time detection of violent crowd behavior,’’ in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2012, pp. 1–6.
[15] M. A.-M. Provath, K. Deb, P. K. Dhar, and T. Shimamura, ‘‘Classification of lung and colon cancer histopathological images using global context attention based convolutional neural network,’’ IEEE Access, vol. 11, pp. 110164–110183, 2023.
[16] M. Rahman, K. Deb, P. K. Dhar, and A. T. Shimamura, ‘‘ADBNet: An attention-guided deep broad convolutional neural network for the classification of breast cancer histopathology images,’’ IEEE Access, vol. 12, pp. 133784–133809, 2024.
[17] R. C. Gonzalez and R. E. Woods, Digital Image Processing. Upper Saddle River, NJ, USA: Prentice-Hall, 2002.
[18] P. J. Rousseeuw, ‘‘Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,’’ J. Comput. Appl. Math., vol. 20, no. 1, pp. 53–65, Nov. 1987.

DOI : https://doi.org/10.5281/zenodo.20477483