A Survey on Violence Detection in Surveillance Videos Using Artificial Intelligence

Atharva Wankhade, prof Snehil Jaiswal, Snehal Tingane

Abstract— **The function of surveillance systems in maintaining public order and public safety has grown in recent years. Video surveillance situations, such as those found in train stations, schools, and hospitals, require automatic detection of aggressive and suspicious behaviours to prevent any potential casualties that might result in social, economic, and ecological harm. The efficient use of automated violence detection by law enforcement is of critical importance. Detection of violence and weaponized violence in closed circuit television (CCTV) footage requires a comprehensive approach. In this paper, we presented review on various violence detection system in surveillance videos using different artificial intelligent techniques. As the intelligent violence detection system effort applies to industry and in terms of security, it is beneficial to society.**

Index Terms— **Violence Detection, Weaponized Violence Detection, Video Surveillance, Machine Learning, Deep Learning, Artificial Intelligence**

I. INTRODUCTION

Violence and gang-related activities can pose a serious threat to a city, particularly when authorities are unable to respond quickly enough to prevent further damage. In some cases, these incidents can result in loss of life and property, especially when weapons are involved [1]. Regrettably, incidents of road rage, gang related violence, and other spontaneous acts of violent crime frequently occur without prior warning or the ability for authorities to intervene proactively. These events pose a considerable challenge for law enforcement agencies and other relevant authorities. Unfortunately, the reporting of such incidents often occurs after the fact, leaving authorities with limited options for timely intervention and effective prevention [2].

Although surveillance systems have helped authorities identify instigators and culprits through recordings, it often takes too long to detect, search, and arrest someone after a crime is committed. To reduce turnaround time and increase efficiency, there is a growing need for automated detection and signalling systems. Since the breakthrough of deep learning [3] in the ImageNet 2012 competition, deep neural networks (DNNs) have become the go-to AI technology for automating such tasks. By leveraging DNNs and other AI techniques, smart cities worldwide can better detect and respond to violence, safeguarding lives and properties. The benefits of such technologies are clear; for instance, integrated surveillance systems equipped with advanced AI algorithms can analyse real-time video feeds from CCTV cameras to identify and alert authorities to potential violent incidents, enabling swift intervention. Furthermore, AI-powered predictive analytics can analyse various data sources, including social media feeds and sensor data, to identify patterns and trends associated with violence, enabling authorities to allocate resources strategically and prevent outbreaks of violence in specific areas. As these technologies evolve, they will play an increasingly important role in maintaining safety and security in our cities [4-6].

Current violence detection methods predominantly rely on spatiotemporal models to identify instances of violent activities within video footage. However, it is essential to recognize that violence encompasses a wide range of behaviours, spanning from physical altercations to gunfights. Treating all violent events equally may not effectively prioritize the severity or potential harm involved [7] [9-10]. To address this challenge, it becomes crucial to develop methods to detect weapons in surveillance footage.

II. LITERATURE REVIEW

B Omarov et.al. [1] aims to address the problems as state-of-the-art methods in video violence detection, datasets to develop and train real-time video violence detection frameworks, discuss and identify open issues in the given problem. In this study, they analyzed 80 research papers that have been selected from 154 research papers after identification, screening, and eligibility phases. As the research sources, we used five digital libraries and three high ranked computer vision conferences that were published between 2015 and 2021.

G. Sreenu et.al. [2] includes a deep-rooted survey which starts from object recognition, action recognition, crowd analysis and finally violence detection in a crowd environment. Majority of the papers reviewed in this survey are based on deep learning technique. Various deep learning methods are compared in terms of their algorithms and models. The main focus of this survey is application of deep learning techniques in detecting the exact count, involved persons and the happened activity in a large crowd at all climate conditions. Paper discusses the underlying deep learning implementation technology involved in various crowd video analysis methods. Real time processing, an important issue which is yet to be explored more in this field is also considered. Not many methods are there in handling all these issues simultaneously. The issues recognized in existing methods are identified and summarized.

Eknarin Ditsanthia et.al. [3] propose a novel representation learning approach to improve the detection rate of violent behaviours in videos. Their proposed approach consists of two parts. In the first part, they leverage features extracted from image-based deep convolution neural network to describe spatial information in a video frame. They also

Manuscript received March 22, 2024

Atharva Wankhade, G H Raisoni University Amravati
prof Snehil Jaiswal, G H Raisoni University Amravati
Snehal Tingane, G H Raisoni University Amravati

introduce a new kind of image features, named multiscale convolutional features, to handle variations in the video data. In the second part, a bidirectional long-short term memory (LSTM) is applied to learn a video-level classifier from both violent/non-violent video sequences. Experimental results on three challenging datasets demonstrate that accuracy improvements are achieved by the proposed detection method compared with the previous methods.

Toluwani Aremu et.al. [4] introduce the Smart-City CCTV Violence Detection (SCVD) dataset, specifically designed to facilitate the learning of weapon distribution in surveillance videos. To tackle the complexities of analyzing 3D surveillance video for violence recognition tasks, we propose a novel technique called, SSIVD-Net (Salient-Super-Image for Violence Detection). Our method reduces 3D video data complexity, dimensionality, and information loss while improving inference, performance, and explainability through the use of Salient-Super-Image representations.

Ali Mansour Al-Madani et.al. [5] presents a 3D convolutional neural network-based technique for detecting violence in videos. Accuracy is improved by employing machine and deep learning techniques in a suggested manner. Performance evaluations have shown that the suggested technique effectively identifies violence in video clips. Experimental data show that the proposed strategy outperforms existing methods for identifying crimes and violence in films. The pre-training models, Inception-V3, InceptionResNetV2, ViolenceNet, and ViolenceNet-OF, were trained on the four datasets.

S A Arun Akash et.al. [6] used deep learning as computer vision to predict and detect the action, properties from video. In real-time police reach violent destinations and start checking CCTV cameras, and investigate to proceed further. This study is deliberately designed to detect violent acts from CCTV cameras. The Inception – v3 and Yolo – v5 models detect the violent act, the number of persons involved, and also the weapons used in the situation.

Bermejo Nievas et.al. [7] introduce a new video database containing 1000 sequences divided in two groups: fights and non-fights. Experiments on this database and another one with fights from action movies show that fights can be detected with near 90% accuracy.

Aqib Mumtaz et.al. [8] proposed deep representation-based model using concept of transfer learning for violent scenes detection to identify aggressive human behaviours. The result reports that proposed approach is outperforming state-of-the-art accuracies by learning most discriminating features achieving 99.28% and 99.97% accuracies on Hockey and Movies datasets respectively, by learning finest features for the task of violent action recognition in videos.

J.V. Vidhya et.al. [9] converted the given input video into frames and the preprocessing is done at the frame level. For Feature extraction the 2D convolutional neural network (Conv2D) is used and it adapts the layers of VGG-19 net architecture with global average pooling and learns the spatial information in the given video. Those extracted features are then combined using Long Short-Term Memory (LSTM) and it learns about temporal information from the video. The model is validated using the Hockey data set and a loss of 0.02 and accuracy of 98 is achieved.

Muzamil Ahmed et.al. [10] proposed framework, the keyframe extraction technique eliminates duplicate consecutive frames. This keyframing phase reduces the training data size and hence decreases the computational cost by avoiding duplicate frames. For feature selection and classification tasks, the applied sequential CNN uses one kernel size, whereas the inception v4 CNN uses multiple kernels for different layers of the architecture. For empirical analysis, four widely used standard datasets are used with diverse activities. The results confirm that the proposed approach attains 98% accuracy, reduces the computational cost, and outperforms the existing techniques of violence detection and recognition.

Romas Vijeikis et.al. [11] present a novel architecture for violence detection from video surveillance cameras. The proposed model is a spatial feature extracting a U-Net-like network that uses MobileNet V2 as an encoder followed by LSTM for temporal feature extraction and classification. The proposed model is computationally light and still achieves good results—experiments showed that an average accuracy is 0.82 ± 2% and average precision is 0.81 ± 3% using a complex real-world security camera footage dataset based on RWF-2000.

A. Z. Kouzani et.al. [12] examines the latest technological innovations for tackling domestic violence. It describes a range of technology platforms and tools, and discusses their capabilities and shortcomings. Firstly, the review methodology is given including the aims and objectives of the review, the search strategy, and the selection of sources. Next, the technological innovations in the context of domestic violence are defined.

J. Su et.al. [13] show competitive violence detection results using a general action recognition CNN without modifying the original architecture. Experimental results on three publicly available benchmark datasets show that the proposed method outperforms other sophisticated techniques designed specifically to detect violence in videos.

Elly Matul Imah et.al. [14] presents violence detection using the visual geometry group network-16 (VGGNet-16)-based deep transfer learning feature extraction, combined with ensemble decision fusion learning. Ensemble decision fusion learning is a kind of ensemble learning method. It combines classifiers from multiple models and datasets. The majority of voting connects the classifier’s output is used to decision fusion in this study.

III. CONCLUSION

This paper presented a comprehensive review of the latest technological innovations for addressing domestic violence. The review explored various technologies such as AI models, including machine learning and particularly deepneural networks architectures, ambient sensors, smartphones and applications, wearable devices and sensors, online privacy and security methods, digital platform features, anti-stalking tools, and virtual reality platforms. The study also analysed these technologies and described their capabilities, limitations, and applications in tackling different aspects of domestic violence. Among the technologies, AI models were highlighted as both widely used and highly effective in addressing domestic violence.

REFERENCES

[1] Omarov B, Narynov S, Zhumanov Z, Gumar A, Khassanova M. State-of-the-art violence detection techniques in video surveillance security systems: a

systematic review. PeerJ Comput Sci. 2022 Apr 6;8:e920. doi: 10.7717/peerj-cs.920. PMID: 35494848; PMCID: PMC9044356.

[2] Sreenu, G., Saleem Durai, M.A. Intelligent video surveillance: a review through deep learning techniques for crowd analysis. J Big Data 6, 48 (2019). https://doi.org/10.1186/s40537-019-0212-5

[3] E. Ditsanthia, L. Pipanmaekaporn and S. Kamonsantiroj, "Video Representation Learning for CCTV-Based Violence Detection," 2018 3rd Technology Innovation Management and Engineering Science International Conference (TIMES-iCON), Bangkok, Thailand, 2018, pp. 1-5, doi: 10.1109/TIMES-iCON.2018.8621751.

[4] Aremu, Toluwani & Li, Zhiyuan & Alameeri, Reem & Khan, Mustaqeem & El Saddik, Abdulmotaleb. (2023). SSIVD-Net: A Novel Salient Super Image Classification & Detection Technique for WeaponizedViolence.10.48550/arXiv.2207.12850.

[5] Ali Mansour Al-Madani, et.al. “Real-Time Detection of Crime and Violence in Video Surveillance using Deep Learning”, 2022 Proceedings of the First International Conference on Advances in Computer Vision and Artificial Intelligence Technologies (ACVAIT 2022), pp. 431-441, https://doi.org/10.2991/978-94-6463-196-8_33.

[6] Akash, S & Moorthy, R & Esha, K & Narayanaraju, Nathiya. (2022). Human Violence Detection Using Deep Learning Techniques. Journal of Physics: Conference Series. 2318. 012003. 10.1088/1742-6596/2318/1/012003.

[7] Bermejo Nievas, E., Deniz Suarez, O., Bueno García, G., Sukthankar, R. (2011). Violence Detection in Video Using Computer Vision Techniques. In: Real, P., Diaz-Pernil, D., Molina-Abril, H., Berciano, A., Kropatsch, W. (eds) Computer Analysis of Images and Patterns. CAIP 2011. Lecture Notes in Computer Science, vol 6855. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-23678-5_39

[8] Mumtaz, A. B. Sargano and Z. Habib, "Violence Detection in Surveillance Videos with Deep Network Using Transfer Learning," 2018 2nd European Conference on Electrical Engineering and Computer Science (EECS), Bern, Switzerland, 2018, pp. 558-563, doi: 10.1109/EECS.2018.00109.

[9] Vidhya, J. V. and R. Annie Uthra. “Violence detection in videos using Conv2D VGG-19 architecture and LSTM network.” (2021).

[10] Ramzan, Muhammad & Khan, Hikmat & Iqbal, Saqib & Khan, Muhammad & Choi, Jungin & Nam, Yunyoung & Kadry, Seifedine. (2021). Real-Time Violent Action Recognition Using Key Frames Extraction and Deep Learning. Computers, Materials and Continua. 69. 2217-2230. 10.32604/cmc.2021.018103.

[11] Vijeikis, Romas, Vidas Raudonis, and Gintaras Dervinis. 2022. "Efficient Violence Detection in Surveillance" Sensors 22, no. 6: 2216. https://doi.org/10.3390/s22062216.

[12] Z. Kouzani, "Technological Innovations for Tackling Domestic Violence," in IEEE Access, vol. 11, pp. 91293-91311, 2023, doi: 10.1109/ACCESS.2023.3306022.

[13] J. Su, P. Her, E. Clemens, E. Yaz, S. Schneider and H. Medeiros, "Violence Detection using 3D Convolutional Neural Networks," 2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Madrid, Spain, 2022, pp. 1-8, doi: 10.1109/AVSS56176.2022.9959393.

[14] Elly Matul Imah et.al. “Child Violence Detection in Surveillance Video Using Deep Transfer Learning and Ensemble Decision Fusion Learning”, 2022, International Journal of Intelligent Engineering and Systems, Vol.15, No.3. DOI: 10.22266/ijies2022.0630.38