2nd International Conference on Scientific and Academic Research
March 14-16, 2023 : Konya, Turkey
All Sciences Proceedings
http://as-proceeding.com/

Violence Activity Detection Classification - A Review

Muhammad Awais*, Sara Durrani [2]

[1]Software Engineering Department, Capital University of Science and Technology Islamabad, Pakistan
[2]Software Engineering Department, Capital University of Science and Technology Islamabad, Pakistan

*(mawaiskhan1808@gmail.com) Email of the corresponding author

Abstract – With the emerging trends of different automated surveillance systems for the security of people, activities like violence activity detection have become an active area of research. We observe several criminal and abnormal violent activities in our daily lives that need can be detected on spot to avoid a bigger violent event. To prevent from violence and different kind of harmful activity its need to work on accuracy of that automated surveillance. This work aims to provide a systematic literature review on state-of-the-art violent activity detection methods, datasets needed to develop those frameworks for training and testing and, identify gaps through discussion that can be filled with further proposed solutions. In this study, we have compared recent papers from 2017 to 2023 in this domain for their data classes, open-source availability and statistics. Further, the methodologies are divided into domains i.e. Dataset that used by previous researchers and get results to detect violence activity in public and private place with different accuracy, Real-time violent activity detection in video using traditional approaches, machine learning approaches and deep learning approaches. In the end, we have identified the issues and gaps in the existing literature that can create a potential difference for future researchers in this domain.

Keywords – Violence activity, KNN, SVM, Machine Learning, Deep learning, Decision Tree, CNN

I. INTRODUCTION

Physical force is used in violent behavior when the goal is to harm, maim, or even kill another person. Abusive acts of violence, such as murders, accidents, thefts, and drug dealing, frequently lead to lifetime mental traumas in addition to physical harm (and in some severe cases, death) [1]. Public officials are gravely concerned about the rise in armed conflict, armed protests, and criminal activity. The use of explosive weapons and lethal personal weapons, such as pistols, shotguns, automatic machine guns, revolvers, shotguns, and rocket launchers, is a current issue for human rights around the world [2]. The need for security and property protection in public areas and on private property is growing today. Police and other security agencies want a general overview of the object they are securing as well as information specific to that object. Thus, numerous cameras and other surveillance equipment must be placed for safety. It is necessary to detect suspicious activity around-the-clock, seven days a week. The expert eye may fail to notice risky or suspicious behaviors in any given situation. Suspicious, odd, or aberrant behavior refers to unusual behavior. These are special activities that involve using physical force to harm an object, where someone is hurt or murdered, or where theft takes place [3]. Computer vision-based violence detection methods examine the footage captured by security cameras. These cameras and other surveillance tools have been deployed in a variety of locations for public safety during the past few years, including educational institutions, hospitals, banks, markets, streets, etc. to keep an eye on people's movements. Monitoring includes analyzing people's behaviors to determine whether they are suspicious or not. It is quite difficult to notice suspicious activity around-the-clock or to locate it in vast

databases of recorded videos. Several techniques have been developed to identify human actions in the real world for this goal. These techniques aid in identifying suspicious activity in surveillance films [4].

Convolution Neural Networks (CNNs) and Deep Learning (DL) algorithms are evolving quickly, enabling the development of precise models that address real-world issues in the fields of medicine, agriculture, traffic management, threat management, activity classification, object classification, and autonomous vehicles. Approaches for classifying human activity and detecting objects have a considerable impact on how video surveillance systems characterize human behavior [2].

The automated detection of human behavior in video surveillance systems makes it simple to locate questionable object activity. Places like airports, train stations, banks, offices, and exam rooms. With the help of artificial intelligence, machine learning, and deep learning, it is possible to automatically identify human activity in public areas. Computers that use artificial intelligence can think like humans. Learning from training data and making predictions about future data are crucial aspects of machine learning. Deep learning is employed because huge databases and GPU (Graphics Processor Unit) processors are available today.

Public safety and security will be provided by computer vision combined with video surveillance. Environment modelling, motion detection, moving object classification, tracking, behavior understanding and interpretation, and data fusion from many cameras are some of the techniques used in computer vision [5].

Intelligent video surveillance has gradually replaced conventional video surveillance as computer vision and intelligence have advanced. A significant area of research in the realm of computer vision is behavior recognition. The precise semantics of behaviors can be automatically understood by analyzing the body language and movement of the people in the situation [6].

Our research contribution is classification of different machine learning and deep learning algorithm that used worldwide to detect the violence activity and accuracy level of that algorithm and dataset with statics that used to train and test model.

II. Literature Review

The recognition of complicated sequential visual patterns makes violence detection from video data a difficult challenge. Traditional low-level characteristics, which are used by the majority of approaches, are ineffective at identifying such complicated patterns and are challenging to use in real-time monitoring. We introduce a deep-learning-based 3D CNN model to learn complicated sequential patterns to accurately forecast violence in light of the limitations of the existing methodologies [7]. For feature extraction, there are primarily two primary methods used: computing optical flow data from the videos and computing deep convolutional neural network-based representations. Owing to convolutional neural networks' (CNN's) shown effectiveness in a variety of computer vision

In recent studies, CNN-based applications are strongly favored. Long Short-term Memory (LSTM) are utilized to model temporal information since they are good at remembering associations between successive frames. In conclusion, because of its high performance, CNN + LSTM network is frequently employed in action recognition [8].

By using the feature pyramid network (FPN) to identify humans in aerial imagery, using Scatter Net hybrid deep learning to evaluate each identified human's stance (SHDL). Finally, the angles between the limbs of the support vector machine-projected stance were used to identify the people engaging in the aggressive behavior (SVM) [9].

Inertial sensors like pressure, stretch, accelerometers, cameras, etc. have been employed in a variety of ways to identify human activity using machine learning techniques including support vector machine (SVM), random forest (RF), and artificial neural network (ANN).

Naive Bayes (NB), Decision Tree, Hidden Markov Model (HMM), etc. Wearable accelerometers were

crucial in identifying typical actions including standing, walking, and sitting [10].

There is considerable interest in the automatic detection of crowd or interpersonal violence in videos. An end-to-end deep neural network model for identifying violence in videos is proposed in this paper. The proposed model uses a sequence of fully connected layers for classification and a pre-trained VGG-16 on ImageNet as a source of temporal and spatial feature extraction, respectively. The acquired accuracy is almost cutting edge. As part of our contribution, we also present a brand-new benchmark called Real-Life Violence Situations, which includes 2000 brief videos split into 1000 violent and 1000 nonviolent ones [11].

Common methods for categorizing various activities are given and addressed. Several classifier techniques for activity categorization, including k-NN, decision trees, support vector machines (SVM), and neural networks, have been proposed by numerous researchers. In order to get better outcomes, decision trees, K-NN classifiers, and SVM classifiers have been utilized [12].

III. Materials and Method

A. Dataset Description

This study [13] has tested two publicly available brute force data sets, namely the Crowd Violence and Hockey data sets, in order to confirm the accuracy of brute force detection. In the hockey dataset, 500 violent videos are present. The crowd violence dataset's films were gathered from actual crowd incidents on YouTube. The data collection includes 123 non-violent videos and 140 videos with a resolution of 320 x 240 pixels. In order to confirm the facial recognition system's accuracy [13].

The other dataset that get from [14] consists of 350 MP4 video files (H.264 codec) with an average length of 5.63 seconds; the shortest and longest videos range in length from 2 to 14 seconds. All of the clips have a 1920 x 1080-pixel resolution and a 30 frames per second frame rate. Directories are used to organize the dataset.

Given the enormous number of persons involved in the violent incidents, this dataset focuses on crowd violence. The majority of the films in this dataset were gathered from violent incidents that happened during football games. This dataset includes 246 videos. Each video's first 20 frames are used as network inputs [15].

The video “Violent-Flows" was released and features aggressive mob behavior that is used in [16]. The videos used to create the clips were taken from YouTube and featured various crowd fighting scenarios along with violent acts. There are 246 videos in this dataset overall, each with a frame resolution of 320 by 240 pixels. Each video lasts for 1 to 6 seconds.

Ten diverse group activities, including running, chasing, following someone, and fighting, are included in this dataset. The dataset consists of four lengthy videos, each measuring 640×480 pixels, with two distinct camera angles and locations. There are roughly 200,000 frames in all of the videos. There are many activities in every video segment [16].

The resolutions of the video collection that we used in this study, which was compiled from YouTube, range from 1024 pixels to 72 pixels. While higher quality slows down the training process, we initially took 30 frames per second from the films and down sampled them to 28 * 28 pixels. Afterwards, these frames are taken from both violent and nonviolent videos and inputted into the network [17].

Table 1. Dataset Summary

Years	**Categories**	**Resolution**	**Statics**	**length**
2021	Violent, Non violent	320x 240	763 videos	-
2017	Violent	1920x1080	350 MP4	2 to 14 seconds/30 frames per second
2021	Violent	320x240	246	1 to 6 seconds/20
2021	Fighting	640x480	-	-
2019	Violent, Non Violent	1024x72	-	/30

B. Classification of Violence Activity detection

Violence Detection using Machine Learning techniques:

The activities have been divided into categories using a supervised K-NN classifier. It operates based on the shortest distance between the prediction point and the data point. A data point that is close to A conventional Euclidian distance is used to define an instance. Since K is the only parameter that needs to be adjusted for the K-NN method, adjusting K between 1 and 20 results in an optimal value of K = 1 for the best accuracy. Both the training set and the test set were produced using a five-fold cross-validation technique in this instance. To determine the classification accuracy in the test case, the estimated classes are put side by side with the actual classes. The performance of the classifiers is calculated using the accuracy term. The overall accuracy for our data set is 85.9%.

The decision tree is a hierarchical model used to answer the problem, with each leaf node denoting a class label and the inside nodes of the tree representing attributes.

Building a model to predict and categories whether a fresh set of data comes from an attack series or the typical sequence is our aim. A 93.8% total accuracy rate.

Shin et al. proposed the supervised algorithm known as SVM (2014). This method is used to locate the hyper-plane that has the biggest possible margin between the data points. A linear SVM classifier can be used to categories a small number of two features. The extended SVM is effective for non-linear classification. It is accomplished by using the Kernel technique to the classification model (Mantyjarvi et al., 2001). It uses a kernel function to transform the incoming data into the desired form. The evaluation of the classification accuracy is 98.8%. [12].

Violence Detection using Deep Learning techniques:

The conventional fight detection techniques build intricate handcrafted features from the input using domain knowledge. Deep models, on the other hand, can take immediate action and automatically extract the features. For the purpose of detecting violence in films without using prior knowledge, Ding et al. introduced the new 3D convent's technique. The set of video frames are convolutional using a 3D CNN, which allows the input's motion data to be retrieved. The back-propagation method is used to compute gradients after the model has been trained using supervised learning. The Hockey dataset is used in experiments, and the results demonstrate that the suggested method outperforms it without relying on manually created features in terms of accuracy.

To identify violence in videos, a deep neural network-based technique is developed. In order to extract the features from a video's frame level, CNN is utilized. Then, these features are added together using an LSTM version that makes use of convolutional gates. ConvLSTM and CNN may take localized spatio-temporal information from the videos and use them to do local motion analysis. It is also suggested to feed the model that decodes the changes made to the videos with the nearby frame differences. Three well-known datasets—hockey, films, and violent-flows—are used in the experiments. The suggested model outperforms state-of-the-art techniques like ViF+OViF, ViF, three streams + LSTM, and others in terms of accuracy, according to the results.

For security reasons, an improved monitoring system is necessary for the detection of violent activity in order to prevent social, economic, and ecological harm. For this reason, the triple-staged end-to-end deep learning violence detection architecture is suggested. First, in surveillance video streams, people are identified using a lightweight CNN model to avoid and minimize the extensive processing of useless frames. To extract the spatiotemporal properties of these sequences and feed them to the Softmax classifier, an order of 16 frames containing recognized people is delivered to 3D CNN. Finally, a neural networks optimization tools and an open visual inference created by Intel are used to optimize the 3D CNN model. An intermediate depiction of the trained model is created, and it is changed for execution at the end

platform for the detection of violence. An alarm is sent to the nearby security agency or police station to prompt action once violence is detected. The Violent Crowd, Hockey, and Violence in Movies datasets are used in experiments. The experiment's findings show that the suggested method outperforms cutting-edge techniques like ViF, AdaBoost, SVM, Hough Forest and 2D CNN, sHOT, and others in terms of accuracy, precision, recall, and AUC [4].

Table 2. Comparision

Method	Accuracy
KNN	85.9%
Decision Tree	93.8%
SVM	98.8%

IV. CONCLUSION

There is need of surveillance cameras in every private public sector that detect human behaviour and activity to keep protection from any kind of violence so many researchers purposed many machine learning and deep learning techniques to detect human behaviour from videos. The basic purpose of review is to explore all techniques of violence activity detection in surveillance system and all information of machine learning and deep learning algorithm with data set that are used worldwide to get accuracy of system. Our contribution is to review the techniques of violence activity detection through system.

ACKNOWLEDGMENT

Thankfully, we are aware of our parents' affection for us. Last but not least, we would like to thank our family and friends whose prayers made it possible for us to finish this project.

REFERENCES

[1] Thaipisutikul, T., Tuarob, S., Pongpaichet, S., Amornvatcharapong, A., & Shih, T. K. (2021, January). Automated classification of criminal and violent activities in Thailand from online news articles. In *2021 13th International Conference on Knowledge and Smart Technology (KST)* (pp. 170-175). IEEE.

[2] Bhatt, A., & Ganatra, A. (2023). Weapon operating pose detection and suspicious human activity classification using skeleton graphs. *Mathematical Biosciences and Engineering*, *20*(2), 2669-2690.

[3] Vrskova, R., Hudec, R., Sykora, P., Kamencay, P., & Benco, M. (2020, October). Violent Behavioral Activity Classification Using Artificial Neural Network. In *2020 New Trends in Signal Processing (NTSP)* (pp. 1-5). IEEE.

[4] Ramzan, M., Abid, A., Khan, H. U., Awan, S. M., Ismail, A., Ahmed, M., ... & Mahmood, A. (2019). A review on state-of-the-art violence detection techniques. *IEEE Access*, *7*, 107560-107575.

[5] Chole, S., Tiwari, R. N., Siddique, S., Jain, P., & Mane, S. DETECTING SUSPICIOUS ACTIVITIES IN SURVEILLANCE VIDEOS USING DEEP LEARNING METHODS.

[6] Yao, H., & Hu, X. (2023). A survey of video violence detection. *Cyber-Physical Systems*, *9*(1), 1-24.

[7] Ullah, F. U. M., Ullah, A., Muhammad, K., Haq, I. U., & Baik, S. W. (2019). Violence detection using spatiotemporal features with 3D convolutional neural network. *Sensors*, *19*(11), 2472.

[8] Aktı, Ş., Tataroğlu, G. A., & Ekenel, H. K. (2019, November). Vision-based fight detection from surveillance cameras. In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1-6). IEEE.

[9] Srivastava, A., Badal, T., Garg, A., Vidyarthi, A., & Singh, R. (2021). Recognizing human violent action using drone surveillance within real-time proximity. *Journal of Real-Time Image Processing*, *18*, 1851-1863.

[10] Randhawa, P., Shanthagiri, V., & Kumar, A. (2020). Violent activity recognition by E-textile sensors based on machine learning methods. *Journal of Intelligent & Fuzzy Systems*, *39*(6), 8115-8123.

[11] Soliman, M. M., Kamal, M. H., Nashed, M. A. E. M., Mostafa, Y. M., Chawky, B. S., & Khattab, D. (2019, December). Violence recognition from videos using deep learning techniques. In *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)* (pp. 80-85). IEEE.

[12] Randhawa, P., Shanthagiri, V., Kumar, A., & Yadav, V. (2020). Human activity detection using machine learning methods from wearable sensors. *Sensor Review*, *40*(5), 591-603.

[13] Wang, P., Wang, P., & Fan, E. (2021). Violence detection and face recognition based on deep learning. *Pattern Recognition Letters*, *142*, 20-24.

[14] Bianculli, M., Falcionelli, N., Sernani, P., Tomassini, S., Contardo, P., Lombardi, M., & Dragoni, A. F. (2020). A dataset for automatic violence detection in videos. *Data in brief*, *33*, 106587.

[15] Sudhakaran, S., & Lanz, O. (2017, August). Learning to detect violent videos using convolutional long short-term memory. In *2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS)* (pp. 1-6). IEEE.

[16] Asad, M., Yang, J., He, J., Shamsolmoali, P., & He, X. (2021). Multi-frame feature-fusion-based model for violence detection. *The Visual Computer*, *37*, 1415-1431.

[17] Sumon, S. A., Shahria, M. T., Goni, M. R., Hasan, N., Almarufuzzaman, A. M., & Rahman, R. M. (2019).

Violent crowd flow detection using deep learning. In *Intelligent Information and Database Systems: 11th Asian Conference, ACIIDS 2019, Yogyakarta, Indonesia, April 8–11, 2019, Proceedings, Part I 11* (pp. 613-625). Springer International Publishing.