INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

HOG Ensembled Boosting Machine Learning Approach for Violent Video Classification

Snehil G Jaiswal[1]*, Sharad W Mohod[2], Dinesh Sharma[3]

1 Research Scholar, Sant Gadge Baba Amravati University Amravati and Registrar, G H Raisoni University, Amravati, Maharashtra, India
2 Professor and Head of Department, Department of Electronics and Telecommunication Engineering, Prof. Ram Meghe Institute of Technology & Research, Amravati, Maharashtra, India
3 Assistant Professor, Department of Electronics and Telecommunication Engineering, Chandigarh College of Engineering and Technology, Chandigarh, India

OPEN ACCESS

Received: 16-07-2023
Accepted: 03-08-2023
Published: 12-09-2023

Citation: Jaiswal SG, Mohod SW, Sharma D (2023) HOG Ensembled Boosting Machine Learning Approach for Violent Video Classification. Indian Journal of Science and Technology 16(34): 2709-2718. https://doi.org/10.17485/IJST/v16i34.1777

* **Corresponding author**.

mr.snehil.jaiswal@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Background: With the proliferation of machine learning and its applications in a variety of spheres that are important to humans in their day-to-day lives, there is a pressing need for automatic detection models that can identify abnormal behaviors or acts of violence. **Methods:** This study examines a machine learning model that uses ensemble boosting and histograms of oriented gradients (HOG) to detect violent content from a feature vector with a single parameter. **Findings:** The tests performed on two benchmark datasets, such as the Hockey Dataset and the Peliculas dataset, reveal a high level of performance accuracy for the classification of violent videos. The experiment findings show that the suggested violence detection model performs well in terms of average metrics, with accuracy, precision, and recall being 90.50%, 91.80%, and 89.70%, respectively. **Novelty and applications:** The proposed method is capable of striking a balance between high performance and a limited number of parameters, and as a result, it can be implemented with a minimal investment of computational resources.

Keywords: Violence detection; Computer Vision; Action Recognition; Machine Learning; Histogram of Oriented Gradients (HOG); Ensemble Boosting

1 Introduction

Recent advances in video and image processing have been unprecedented because it is important to find complex content for a variety of applications and purposes, such as searching, summarizing, and recognizing actions[(1)]. Object and motion detection technology has come a long way in terms of development. These strategies can be coupled to construct a system that can successfully detect violent acts that may occur in everyday situations. Surveillance cameras and other surveillance equipment have become increasingly affordable in recent decades, making them ideal for keeping places safe. Movements and actions of individuals are monitored manually at public places such as marketplaces, streets, and banks. Violent acts have become a big threat to world

security, yet it is difficult and unrealistic to manually evaluate videos and uncover every suspicious scene in actual time [2].

Computer vision and machine learning are among the most effective approaches to monitor and detect violence in surveillance videos, as they are capable of storing, processing, analyzing and deriving quick results from large-scale databases. The competitive field of research is considered as human activity recognition which focuses on modelling detection systems and human activity detection [3]. The identification and classification of violence in video is a subset of action detection, and violent scenes / content in video can be identified with the help of feature extraction algorithms for both the visual and the auditory domains (if available).

Aggressive behavior in humans is defined as the practice of violence. Through the study of these potentially harmful visual patterns, it is possible to devise a variety of descriptors to represent various features that can be used to identify the patterns. These parameters mainly correspond to many features of the video, which includes the time, flow, acceleration, and appearance of the image, amongst others [4].

The detection and categorization of violence is a challenging task because of the variety of human movements that occur within the classroom, the shifting of the background, as well as changes in scale and perspective [5]. An additional barrier that the system designed to identify violent acts must overcome is the need for timely detection and monitoring of violent acts. As a result of this, the framework that was developed to identify violent content in videos is intended to be a method that is quick, dependable, and robust, and consequently to operate continuously in real time [6]. Because of this, Computer vision and machine learning techniques are needed to recognize and categorize violent video for intelligent surveillance.

Primarily the methods of violence detection and classification can be categories on the basis of features extracted and classifier used. Different local and global features are extracted and methods like support vector machine (SVM), machine learning and deep learning are used as classifier in the methodologies proposed in recent year [7].

A novel approach of fight scene detection using motion blobs features is proposed in [8]. Firstly, the absolute difference between two successive frames of video sequence is computed, which is then converted into binarized difference. Using the binarized difference between two frames, motion blobs are marked on scenes involving fight (violence) and non-fight (non-violence). Feature vector is computed using the statistical parameters extracted from motion blob such as distance between blobs, area, centroid and perimeter.

The machine learning algorithms such as KNN, Adaboost classifies the video as fight or non-fight using feature vector. The experiments are performed on dataset such as Hockey Fight, Movies and UFC-1. When compared to contemporary fight detection technologies, the categorization accuracy given is not at par. Figure 1 depicts the Fast Fight Detection system's generalized architecture.

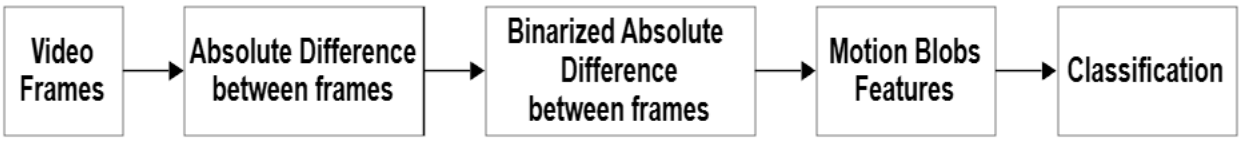

Fig 1. Framework in Fast Fight Detection Method [6]

To target problem of violent activity detection, an approach using estimation of motion vector is proposed in [9]. In order to determine the region's motion vector descriptor, which is subsequently used by the support vector machine to determine whether the activity is violent or not, the motion vectors that were retrieved from each frame and between each frame were examined. Since, the majority of the dataset available for activity detection involves only simple videos of individual activity, the researchers built a new dataset named VVAR10 on the basis of samples taken from UFC-50, YouTube and other previously available dataset. VVAR10 contains videos including actions such as Fighting, Punching, Hammering, Pursuing labeled as positive video clips (violent class) and videos with actions such as jumping and walking as negative clips (non-violent class). The proposed method when evaluated on VVAR10 database shows a fair accuracy with good computational speed.

Common methods for detecting fights require expert knowledge to construct intricate, manual features. However, deep models are quick to respond and can automatically pull relevant details. Using a novel 3D convNets technique, Ding et al. [10] were able to detect violent content in videos with no prior knowledge. Using the input to learn about the motion of the objects in the frame, a 3D convolutional neural network (CNN) is used to estimate the convolution on the video frames. After a model has been trained with supervised learning, gradients can be calculated with the back-propagation method. The studies are carried out using the Hockey dataset, and the outcomes demonstrate that the suggested strategy outperforms handcrafted features in terms of accuracy.

Table 1. Analysis of Methods Reported in Literature

Method	Feature extraction and classification method	Dataset used	Accuracy Achieved
Analysis of motion measure vector for detecting fight scene in video [8]	statistical parameters extracted from motion blob such as distance between blobs, area, centroid and perimeter with KNN and Adaboost	Hockey Fight, Movies and UFC-1	Approximately 90%
SVM with the radial basis integrated with statistical parameters of region motion vectors descriptor [9]	Region Motion Vector descriptor, with support vector machine	VVAR10 dataset (Researchers Developed)	95 % on VVAR10 and less on other datasets
Violence detection using 3D CNN [10]	Within the convolutional layers, 2D convolution is used to extract neighborhood-specific features.	Hockey Dataset	85% to 90%
Violence detection using CNN and deep audio features [11]	Mel Filter-Bank (MFB) features with CNN and SVM	Dataset made using MediaEval 2015 Affective Impact of Movies task video clips	Approx. 90%

Depending on the audio data in the videos, a method for identifying violent scenes is proposed [11]. In addition to its classification abilities, CNN is also able to extract deep acoustic features. As an initial input feature for the CNN, the delta and delta-delta features of the 40-dimensional Mel Filter Bank (MFB) is employed. Finally, the video is cut up into manageable chunks. MFB features are broken up into three feature channels, making it easier to explore the details of the local features. The next step is to use CNN to symbolize attributes. To make SVM classifiers, CNN-based features are used. Then, every segment of footage is put through a process to identify any instances of violence. The segment-level detections are pooled using either a maximum or minimum pooling to arrive at the final result. Experimental findings show that the proposed approach outperforms the three initial approaches (audio only, visuals only, and audio learning fusion plus visual) on the mediaEval dataset.

The goal of the present research is to streamline the identification process of violent videos by providing the system with pre-processed frames that has just the minimal elements of features necessary to highlight violent acts committed by human beings.

2 Methodology

The methods employed in this study is thoroughly explained in this section. Training and testing are the two phases of the proposed methodology. The system learns to detect the category of violent content present in a video during the training phase by training classifiers with a single low-level feature extracted from the training dataset. During the testing phase, the system is evaluated by calculating the system's accuracy in detecting violence in a given video. Each of these phases is explained in detail along with the basic steps.

Our proposed method for classifying violent videos makes use of low-level features and a simple classifier, and it is inspired by the insights of a number of different technologies that are summarized in section 1. Figure 3 provides a visual representation of the proposed methodology's overarching structure. The functioning of the present method is discussed in more detail below in relation to Figure 2.

2.1 Basic Operations

Due to noisy cameras, low-quality recording equipment, and a lack of storage space and transmission bandwidth, preprocessing operations are required to enhance the quality of video data acquired by devices like closed-circuit television. These operations enhance the quality of the video and image, as well as upgrade the digitized video and image, enabling increasingly precise data to be extracted and analyzed [12,13].

With pixel values ranging from 0 to 255 (8-bit unsigned integers), the built-in rgb2gray function of Matlab transforms RGB pictures to grayscale images. Figure 4 (a) shows an example video frame from the dataset, and Figure 4 (b) shows the same frame after it has been converted to grayscale. It is necessary to perform preprocessing in order to cut down on the amount of time needed for model training and, as a consequence, to speed up model inference. Video sequences can contain violent actions that involve motions, and to find eventful frames, an estimate of the peak signal-to-noise ratio (PSNR) between two

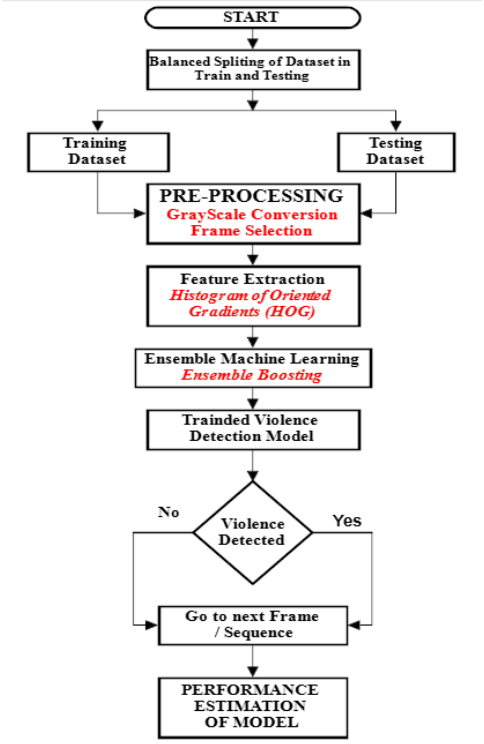

Fig 2. Proposed algorithm for classification of violent video using HOG and Ensemble Boosting

consecutive frames is used[14]. Motions are required to contribute violent or any other action, and there are no human actions that do not involve motions.

2.2 Histogram of Oriented Gradients (HOG)

HOG, frequently referred to as the Histogram of Oriented Gradients, is an example of a feature descriptor that is frequently utilized in the process of extracting features from image data. The HOG algorithm divides an image into a number of smaller sections, then analyses the gradients and orientation of each individual section. In the final step, it creates the histogram of these sections by making use of the gradients and orientation values[15].

The HOG feature descriptor is responsible for counting the number of instances of gradient orientation that can be found in particular regions of an image.

Consider a portion of the image depicted in the preceding Figure 3 (a) and assume the pixel (u, v) to be the target pixel. For the desired pixel, formula (1) and (2) may be used to determine the gradient vector as shown below.

$$\nabla f = \begin{bmatrix} f(u+1,v) - f(u-1,v) \\ f(u,v+1) - f(u,v-1) \end{bmatrix} \tag{1}$$

$$\nabla f = \begin{bmatrix} g_u \\ g_v \end{bmatrix} \tag{2}$$

Where, g_u and g_v is the gradient in x and y direction respectively.

Equation (2) may be used to calculate the gradient magnitude and gradient orientation as shown below.

$$g = \sqrt{(g_u)^2 + (g_v)^2} \tag{3}$$

$$\theta = \tan^{-1}\left(\frac{g_u}{g_v}\right) \tag{4}$$

The foundation of feature extraction for Histogram of Oriented Gradients is detailed in following Algorithm as shown in Figure 4 .

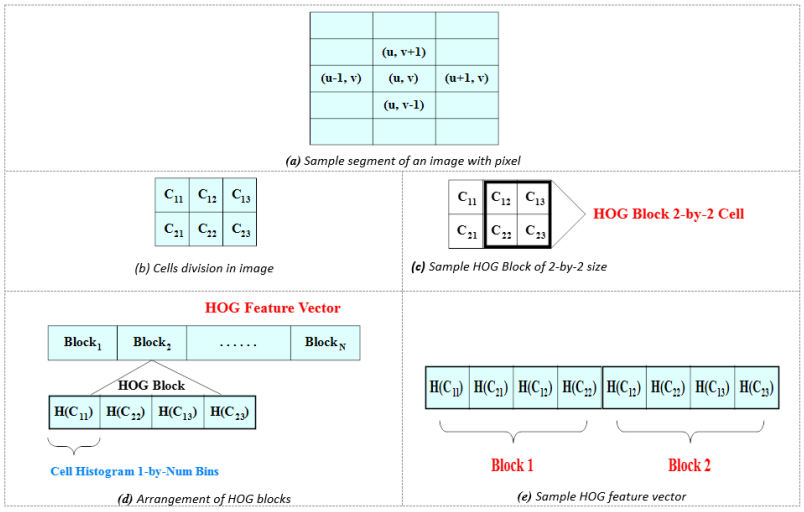

Fig 3. HOG feature vector formation steps

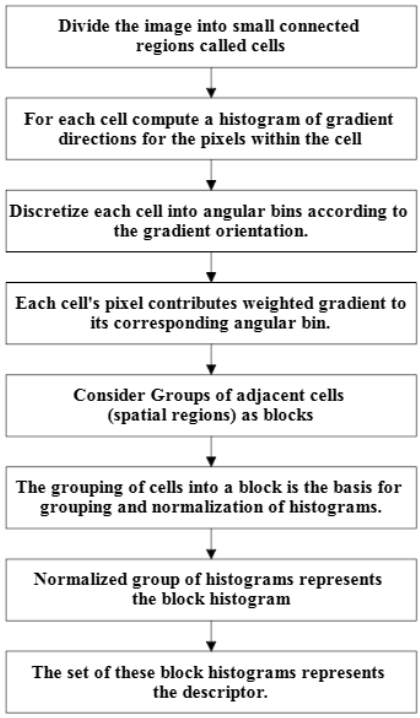

Fig 4. Algorithm from HOG computation of image

HOG Feature Vectors arrangement demonstrates algorithm implementation scheme in Figure 3. The six-cell Figure 3 (b) shown below illustrates this point.

Each HOG block would be 2 cells by 2 cells if the Block Size was set to [2],[2]. All of the cell sizes are specified in pixels as shown in Figure 3 (c). Then HOG blocks are used to organize the HOG feature vector. Figure 3 (d) illustrates the 1-by-NumBins cell histogram, H(Cyx). The HOG feature vector is depicted in the Figure 3 (e) above, demonstrating a 1-by-1 cell overlap between groups. The algorithm extracts HOG features and train a classifier to estimate the label of the test image. Predicting the test image's class based on the HOG characteristics, is referred to as a prediction.

2.3. Classifier

Several models (weak learners) learn how to resolve the same issue using ensemble learning, which combines the findings to provide excellent results[16]. The main assumption is that this can result in models that are more reliable and accurate.

Misclassified observations receive greater weights with each iteration of boosting algorithms like AdaBoostM1 and LogitBoost. It's possible for these to reach extremely high weight values. When this occurs, the boosting algorithm may ignore the vast majority of the training data in favor of the few misclassified observations. As a result, general accuracy in classification tends to drop[17]. In such situation, good solution is using robust boosting (RobustBoost). This algorithm does not assign nearly all of the data weight to observations that have been misclassified very severely. It has the potential to produce a higher average accuracy of classification. The RobustBoost algorithm can only be used when coupled with Matlab's Optimization Toolbox. Unlike AdaBoostM1 and LogitBoost, RobustBoost does not minimize a specific loss function. Instead, it optimizes the proportion of observations where the margin of classification is greater than a certain threshold. The classifier is constructed using RobustBoost algorithm in Matlab.

3 Results and Discussion

The experimental analysis of an algorithm developed using an ensemble machine learning technique and a feature vector built using a single low-level feature for the goal of classifying and detecting violent video is reported in this section. We evaluate performance of proposed algorithm and compares it to existing methods for violent video detection.

3.1 Dataset

The proposed system is experimentally evaluated using two widely-used reference datasets, the Hockey Fight Dataset and the Peliculas Dataset. Both of these datasets are distinct from one another in a variety of aspects, including the degree of variation in background information, occlusion, movements, and lighting conditions. Figure 5 displays sample video frames from Hockey and Peliculas datasets for violent and non-violent classes.

3.1.1Hockey Fight Dataset

This dataset is comprised of both fight and nonfight scenes, both of which were extracted from video footage of hockey matches. The fight scenes are kept in category of violent clips, while the nonfight in non-violent. The dataset contains a total of 1000 video samples, 500 of which are from violent sequences and the remaining 500 from non-violent sequences. The length of videos typically falls somewhere between 1.5 and 2 seconds, and the frame rate is typically set at 25 frames per second. The background in each of the videos looks to be the same since they are all from the same collection of ice hockey events.

3.1.2 Peliculas Dataset

Scenes of violence and nonviolence from popular Hollywood films, football games, and other events are included in this dataset. The dataset contains 200 video samples total, 100 of which are from violent sequences and the remaining 100 from nonviolent ones. Due to the fact that each of the clips is from a different movie, the variation in the background, foreground, movements of the people in the video are available. Depending on the scene, the video clip can be anywhere from 25 to 30 frames per second in length.

3.2 Parameters for Performance Assessment

Performance evaluation is a critical task in machine learning applications. Confusion metrics are used to evaluate the effectiveness of a developed model. The values for TP (true positive), FP (false positive), FN (false negative) and TN (true negative) are considered for the computation of Accuracy, Precision and Recall Rate using following standard formulas[14] as

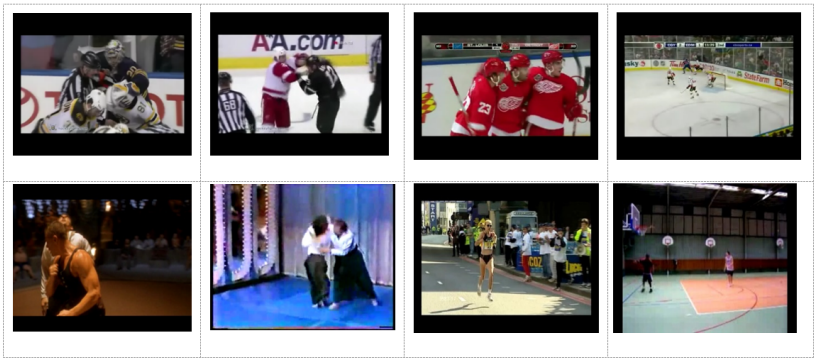

Fig 5. Hockey datasets (top row) and Peliculas datasets (bottom row) contain sample frames from video sequences of the violent and non-violent classes

given below in (5), (6) and (7).

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

$$\text{Pr ecision} = \frac{TP}{TP+FP} \tag{6}$$

$$\text{Re}c\text{ all} = \frac{TP}{TP+FN} \tag{7}$$

Where

- TP: The proportion of violent videos that are actually violent.
- TN: The proportion of truly non-violent videos that have been designated as such.
- FP: The proportion of violent videos that were actually classified nonviolent.
- FN: The proportion of violent videos that have been labelled as non-violent.

A Windows 10 machine with a maximum 2.00 GHz Intel(R) Core-i3-5005U CPU, 4 GB of RAM, and no graphics processing unit (GPU) was used for the experiments. The algorithm is implemented on Matlab with the prominent use of computer vision and statistical & machine learning toolboxes.

Table 2. Experimental results for different frame selections on standard dataset

Number of Frames under Selection	Hockey Dataset				Peliculas dataset			
	Time required for feature extraction (in seconds)	Accuracy (in %)	Precision (in %)	Recall (in %)	Time required for feature extraction (in seconds)	Accuracy (in %)	Precision (in %)	Recall (in %)
10	439.65	85	83.87	86.66	235.90	75	74.19	76.66
20	542.56	89.66	92.66	87.42	318.65	88.33	92.59	83.33
30	718.20	90	90.66	89.47	410.45	91.66	93.10	90

Table 2 above summarizes the experimental results with different frame size given as an input for two-class classification using as Ensemble RobustBoost aggregation method when a decision tree has a maximum of 50 decision splits. The study analyzes computation time for extracting Histogram of Oriented Gradients and constructing feature vectors.

The reported evaluation parameters such as Accuracy, Precision and Recall Rate are calculated using the confusion matrix drawn for each and every case under consideration. Average computation time is reported after performing multiple iterations of the experiment. Figure 6 (a) shows the Peliculas dataset's confusion matrix and ROC curve, while Figure 6 (b) displays the Hockey Fight dataset's results.

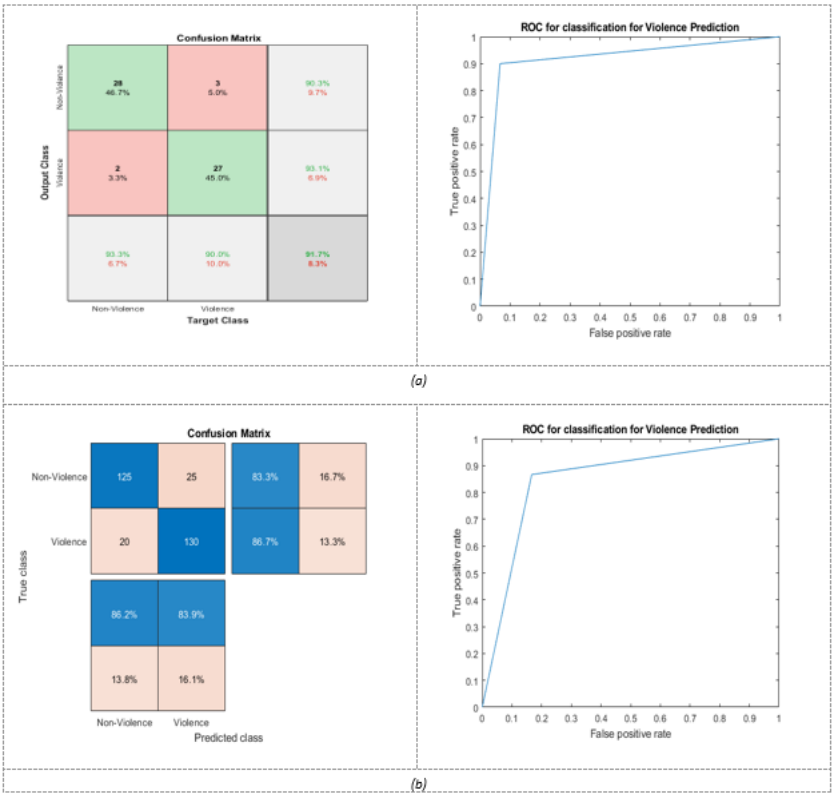

Fig 6. Confusion matrix and ROC for 30 frames in case of (a) Peliculas dataset and (b) Hockey Fight dataset

The comparison of algorithm performance with literature techniques with the highest accuracy configuration is being also done. The suggested approach for detecting violence on the Hockey dataset and the Peliculas dataset was compared with well-known methods that have been documented in the literature in Table 3. The experimental findings are nearly cutting-edge.

Table 3. Comparision With Other Methods

Method and Classifier	**Accuracy (%)**
HOG + Random Forest [18]	86
HOG + histogram intersection kernel [19]	91.5
HOF + histogram intersection kernel [19]	90.9
MoSIFT + histogram intersection kernel [19]	88.6
Multimodal Contrastive Learning [20]	84.03
optical flow, RGB and audio features [21]	84.54
Dynamic Image + Inception-Resnet-V2 [22]	93.3
YOLO [23]	81.2
Haar Cascade Algorithm [24]	84.6
Fuzzy Histogram of Optical Flow Orientations (FHOH) Multi-Resolution Local Binary Patterns (MLBP) + ensemble boosting [14]	88.66
Proposed Method	91.7

4 Conclusion

Using a machine learning model with a single parameter-based feature vector, this work introduces a novel method for the identification and categorization of violent videos. Despite being computationally inexpensive, the results indicated an average precision of 91.80% and accuracy of 90.50%. However, it underperformed in improving recall rate in comparison to precision rate. Violence and non-violence datasets reported in several work often have large temporal variations, but some non-violence actions also have large variation in temporal dimension. We will look at ways to include interference information into non-violent datasets in order to improve the accuracy of violence detection and anti-interference recognition. Future research will compare the suggested strategy with existing deep learning techniques, with a focus on extracting spatiotemporal information from video and determining the viability of the present approach for more general action detection tasks in other computer vision domains.

Acknowledgement

The authors thank the researchers who contributed their efforts into the field of action recognition and violent action detection using video processing that gave motivation to take this work. Authors also thank Sant Gadge Baba Amravati University for providing the required technical support for performing the experiments.

References

1) Kaur G, Singh S. Violence Detection in Videos Using Deep Learning: A Survey. In: Advances in Information Communication Technology and Computing;vol. 392. Springer Nature Singapore. 2022;p. 165–173. Available from: https://doi.org/10.1007/978-981-19-0619-0_15.
2) Thakkar K, Kadiya K, Suthar M, Chauhan MJ. Anomaly Detection In Surveillance Video. 2021. Available from: https://www.irjet.net/archives/V8/i5/IRJET-V8I5486.pdf.
3) Miah P, Haque AA, Imran AA, Hassan MR, Rahman R. Violent activity detection through surveillance camera using deep learning. 2023. Available from: http://hdl.handle.net/10361/18331.
4) Min FU, Ullah MS, Obaidat A, Ullah K, Hijji M, Baik SW. A Comprehensive Review on Vision-Based Violence Detection in Surveillance Videos. *ACM Computing Surveys*. 2023;55. Available from: https://doi.org/10.1145/3561971.
5) Kulbacki M, Segen J, Chaczko Z, Rozenblit JW, Kulbacki M, Klempous R, et al. Intelligent Video Analytics for Human Action Recognition: The State of Knowledge. *Sensors*. 2023;23(9):4258–4258. Available from: https://doi.org/10.3390/s23094258.
6) Ramzan M, Abid A, Khan HU, Awan SM, Ismail A, Ahmed M, et al. A Review on State-of-the-Art Violence Detection Techniques. *IEEE Access*. 2019;7:107560–107575. Available from: https://doi.org/10.1109/ACCESS.2019.2932114.
7) Sreenu G, Durai MAS. Intelligent video surveillance: a review through deep learning techniques for crowd analysis. *Journal of Big Data*. 2019;6(1):48. Available from: https://doi.org/10.1186/s40537-019-0212-5.
8) Gracia IS, Suarez OD, Garcia GB, Kim TK. Fast Fight Detection. *PLOS ONE*. 2015;10(4):e0120448. Available from: https://doi.org/10.1371/journal.pone.0120448.
9) Xie J, Yan W, Mu C, Liu T, Li P, Yan S. Recognizing violent activity without decoding video streams. *Optik*. 2016;127(2):795–801. Available from: https://doi.org/10.1016/j.ijleo.2015.10.165.
10) Ding C, Fan S, Zhu M, Feng W, Jia B. Violence Detection in Video by Using 3D Convolutional Neural Networks. In: Advances in Visual Computing. Springer International Publishing. 2014;p. 551–558. Available from: https://doi.org/10.1007/978-3-319-14364-4_53.
11) Mu G, Cao H, Jin Q. Violent Scene Detection Using Convolutional Neural Networks and Deep Audio Features. In: Cheng, H, editors. Communications in Computer and Information Science;vol. 663. Springer Singapore. 2016;p. 451–463. Available from: https://doi.org/10.1007/978-981-10-3005-5_37.
12) Jaiswal SG, Mohod SW. Implementation of Violence Detection System using Soft Computing Approach. 2021. Available from: https://doi.org/10.1007/978-981-15-8335-3_56.
13) Jaiswal SG, Mohod SW. Recapitulating the Violence Detection Systems. 2019. Available from: https://doi.org/10.1007/978-981-13-8715-9_25.
14) Jaiswal SG, Mohod SW. Classification Of Violent Videos Using Ensemble Boosting Machine Learning Approach With Low Level Features. *Indian Journal of Computer Science and Engineering*. 2021;12(6):1789–1802. Available from: https://doi.org/10.21817/indjcse/2021/v12i6/211206165.
15) Aslan MF, Durdu A, Sabanci K, Mutluer MA. CNN and HOG based comparison study for complete occlusion handling in human tracking. *Measurement*. 2020;158:107704. Available from: https://doi.org/10.1016/j.measurement.2020.107704.
16) Shaout A, Crispin B. Streaming Video Classification Using Machine Learning. *The International Arab Journal of Information Technology*. 2020;17(4):677–682. Available from: https://doi.org/10.34028/iajit/17/4a/13.
17) De Paiva BBM, Pereira PD, De Andrade CMV, Gomes VMR, Souza-Silva MVR, Martins KPMP, et al. Potential and limitations of machine meta-learning (ensemble) methods for predicting COVID-19 mortality in a large inhospital Brazilian dataset. *Scientific Reports*. 2023;13(1):3463. Available from: https://doi.org/10.1038/s41598-023-28579-z.
18) Das S, Sarker A, Mahmud T. Violence Detection from Videos using HOG Features. In: 2019 4th International Conference on Electrical Information and Communication Technology (EICT). IEEE. 2019;p. 1–5. Available from: https://doi.org/10.1109/EICT48899.2019.9068754.
19) Nievas EB, Suarez OD, García GB, Sukthankar R. Violence Detection in Video Using Computer Vision Techniques. In: Real, P, Diaz-Pernil, D, Molina-Abril, H, et al., editors. Computer Analysis of Images and Patterns;vol. 6855. Springer Berlin Heidelberg. 2011;p. 332–339. Available from: https://doi.org/10.1007/978-3-642-23678-5_39.
20) Yang L, Wu Z, Hong J, Long J. MCL: A Contrastive Learning Method for Multimodal Data Fusion in Violence Detection. *IEEE Signal Processing Letters*. 2023;30:408–412. Available from: https://doi.org/10.1109/LSP.2022.3227818.
21) Xiao Y, Wang L, Wang T, Lai H. Scoreformer: Score Fusion-Based Transformers for Weakly-Supervised Violence Detection. In: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. 2023;p. 1–5. Available from: https://doi.org/10.1109/ICASSP49357.

2023.10097219.
22) Jain A, Vishwakarma DK. Deep NeuralNet For Violence Detection Using Motion Features From Dynamic Images. In: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT). IEEE. 2020;p. 826–831. Available from: https://doi.org/10.1109/ICSSIT48917.2020.9214153.
23) Lopez DJ, Lien CC. Real-Time Human Violent Activity Recognition Using Complex Action Decomposition. In: 2020 International Computer Symposium (ICS). 2020;p. 360–364. Available from: https://doi.org/10.1109/ICS51289.2020.00078.
24) Teja MGSKS, Reddy MR, Aishwarya R. Man-on-Man Brutality Identification on Video data using Haar Cascade Algorithm. In: 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE. 2020;p. 274–278. Available from: https://doi.org/10.1109/ICICCS48265.2020.9120872.