International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I8P231
Volume 70 Issue 8, 292-309, August 2022

Original Article

A Novel Approach for Real Time Multi-Scene Violent Activities Recognition with Modified ResNet50 and LSTM

Devang Jani[1], Anand Mankodia[2]

[1]*Ganpat University, Ganpat Vidyanagar, Mehsana. 384001, Gujarat, India.*

[2]*EC, Ganpat University - U. V. Patel College of Engineering, Ganpat Vidyanagar, Mehsana. 384001. Gujarat, India.*

[1]devangjani645@gmail.com

Received:10 June 2022 Revised:11 August 2022 Accepted:17 August 2022 Published: 27 August 2022

***Abstract** - Day by day, the demand for autonomous video surveillance systems has been escalating due to inefficient manual inspection power of identifying anomalies in recorded videos by human beings. Currently, most video surveillance systems use manual video inspection after detecting suspicious activities or trying to use manual inspection of videos once the complaint regarding anomalous, violent or suspicious activities is filed at a particular area or location. Implementation of the real-time scanning of the video stream from the multiple cameras at the central level and a single camera at the edge level is a very big challenge due to the requirement of GPU, computational hardware as well a large amount of computation power with different provocations for the mutual type of human abnormal activities behaviors. The proposed methods represented in this paper provide a novel idea about real-time recognition of nine different mutual violent actions and Normal nonviolent actions using modified deep learning models, namely ResNet50 in association with LSTM. The proposed method provides Nine different diversified violent activities, specifically Attacking, Fighting, hitting with an object, Kicking, Punching, Pushing, Shooting with a Gun, Slapping, stabbing with a knife and one Nonviolent activity that is Normal class. A total of ten violent & nonviolent classes with an accuracy of 87.60% were developed and tested using TensorFlow, Keras and Supercomputing facilities. The Proposed Method is unaccustomed to multi-class violent activity recognition in a real-time environment. Real-time violent activities recognition is a summons as the violent recognition algorithms available to date can provide the decision regarding the events that occurs in the video, whether violent or nonviolent. It can work in short-length recorded videos in a non-real-time environment, a post-effect type of processing that cannot prevent future violent activities as the proposed methods can. The design of the Multi-class Violent recognition model is an arduous and much more time-consuming task due to the performance of each class affecting the overall accuracy and efficiency of the model. It is also annoying and tiresome because it requires continuous time of some days without interruption for the model's training.*

***Keywords** - Resnet50, LSTM, Violent and Nonviolent Classes, Surveillance System, Real-Time.*

1. Introduction

The Internet is becoming a fundamental requirement for people from teenage to old due to the increased habits of social media and other entertainment platforms. The necessity of CCTV surveillance systems is also created for monitoring the premises like the shopping mall, Industries, Educational Institutes, Traffic areas etc., where suspicious activities have been reported in past decades and also highly advocated as the activities related to crime are growing very fast from past few decades [1,3]. The Surveillance system to catch suspicious activities, exclusively violent activities available to date, is either fully manual or semi-manual. The manual testing of the recorded videos is performed by a human being using manual visualisation of the video, which is a poor and ineffective method for identifying suspicious activities, especially when the quality of the recorded video is poor. The length of the video is very high due to the weakness of human beings for continuous observation of the recorded videos with the same efficiency and accuracy [4]. Humans cannot perform as much as machines with high accuracy and for a longer time. However, the development of these types of automatic suspicious activity recognition is a novel task whether it may develop for real-time or non-real-time. Still, it will always provide more accurate recognition of the desired activities with more accuracy, less time and less error than human beings. Fig.1 shows the processing diagram including ten different classes at the final output with labels and classified images for expression of classes and works in detail [2]. The dataset used for ten different classes was created on its own and downloaded from YouTube and other social media. The violent datasets available on online platforms have a variety of poor, average and good-quality videos. It requires many days to identify and download desired contents from contrasting mediums as per as the distinctive violent classes are concerned. Approximately 15GB datasets were collected for violent classification using different datasets mediums like YouTube, social media, Github etc. We required an equal number of image training for the model to train the proposed model. In actual cases, we had a deficiency of images for training for slapping and pushing classes. We have created the remaining datasets for the same on our own. These downloaded and created datasets were converted into images before it is given for training, testing and validation purposes.

Around 35 thousand images were converted from Video to Image at the first level. The heart of the overall research problem is the modified ResNet50 and LSTM network. Still, the most important task is preprocessing to get the maximum desired output for exact classification and accuracy. It is a most tedious, time-consuming, and boring task due to the micro-level functioning required for each image. If the preprocessing task is not done precisely, it may redirect the overall classification with mislabeling and misbalancing at the final output classification. Preprocessing plays a major role when dealing with the research for multi-class violent classification. The datasets that we have planned to use for training the proposed model have images with different resolutions, variations in size and varieties in quality. It also requires the labeling of each image for a particular class after scaling the images concerning size, resolution and overall quality.

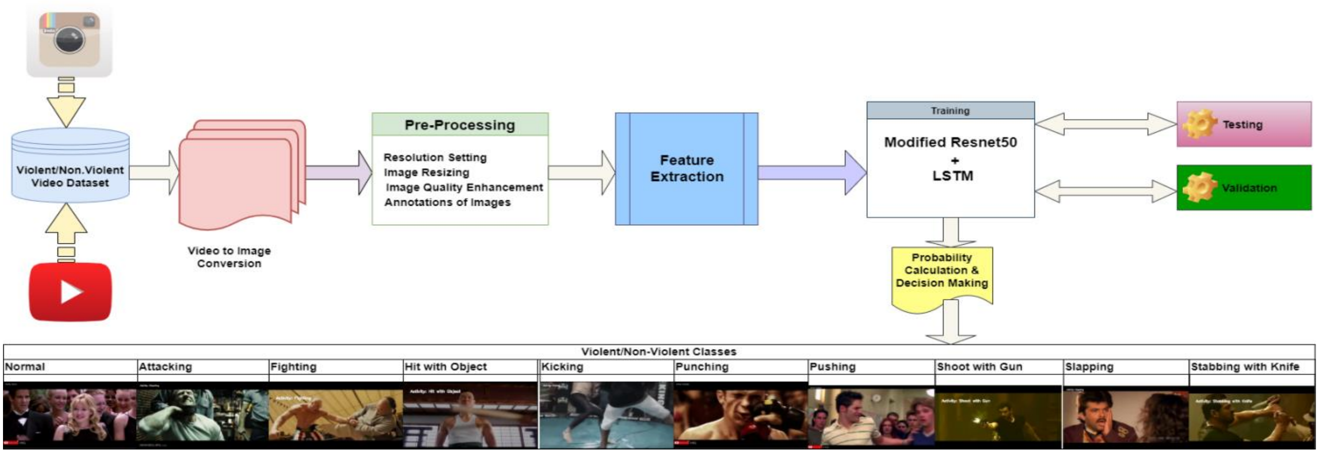

Fig. 1 Processing Diagram in detail with ten different Classes

Labelling the images for different classes is called annotations of the image. The annotation of individual images is a very stuffy, crucial and sluggish task. The deep learning models can learn by feature extraction, which the model can do itself. The learning rate is susceptible to the quality and quantity of the images provided for the training. The training, testing and validation process were performed many times with refinement in datasets and modifications in deep learning models to get the desired accuracy, especially for the multi-class problem. Fig.1 also consists of demonstrations of each class with an image label as outputs after real experiments on testing videos available on YouTube.

Section 1 is about the introduction of the research problem. Section 2 summarises recent literature related to our work. Section 3 discusses the Experiments and Results of ten different classes for nine violent classes and one Normal Class. Section 4 discusses the conclusion and future direction with limitations and challenges for current work.

2. Literature Review

The scenario of recognising Violent activities using deep learning models is created due to many limitations of handcrafted featured traditional methods for real-time applications. The deep learning models have self-learning capabilities by making proportional weighting factors designation and different learning parameters[5,6]. In recent years many papers have been published in reputed journals with non-real-time application and for two class outputs by denotation of labeled as violent or nonviolent only. Table-1 shows the summary of recent papers with comparative analysis of object/event detection methods, feature extraction methods, accuracy, number of classes provided as an output of proposed work, analysis of algorithm whether it can work in real-time environments or not and the dataset which was used to train the model. The major challenges are using multiple datasets with different domain-specific environments with multiple distinguished violent classes for training the model in real-time applications.

Two class violence or non-violence detection was developed especially for fight detection as violence by using optical flows generated from consecutive images for the training of CNN, LSTM and VGG16 with UCF Crimes Dataset, Hockey Fight Dataset, Películas Dataset, Surveillance Camera Fight Dataset separately with 87% accuracy for non-real-time applications[7]. Two class violence and non-violence model was developed for non-real-time applications with an accuracy of 97% using a crowd violence dataset by training CNN [8]. Design of SVM AdaBoost by traditional handcrafted feature extraction method was developed for Four classes, namely Hitting, kicking, fighting and Punching, with 79% accuracy in non-real time recognition application using Crowd Violence Database, Hockey fight dataset, violent flows database[9]. Two Class violence and non-violence classification model by training of ResNet and LSTM was developed with 89% accuracy using KTH, Hockey fight dataset and Violent-Flows separately for non-real-time violence detection applications[10]. Using the Violent Interaction dataset, the concept for fight detection was introduced for violence detection in pre-recorded football match videos by training of LSTM deep learning model with an accuracy of 91%[11]. Two class model with detection of violence and non-violence was developed by training CNN with BEHAVE dataset and videos downloaded from YouTube with CCTV violence footage [12]. The feature extraction method used as long-range Spatio-temporal features for the training of SepConvLSTM with RWF-2000 dataset for non-real-time application was developed with 89% accuracy [13]. The model was developed for detecting stabbing as violence by training CNN using the Real-Life Violence Situation Dataset for non-real-time applications with 90% accuracy [14]. A Two-Class violence and non-violence detection algorithm was developed, which can be used at the edge of CCTV surveillance systems using the UCF-101 dataset and UCF-Crime dataset separately for the training of the LSTM deep learning model with an accuracy of 75%[15].

Most of the work in past years was focused on the denotation of output as to whether the input video or images contain violence. It just provides an idea from the tested video about the presence or absence of violence in a non-real-time environment. It can perform on recorded video only. Real-time analysis with good accuracy is a major challenge, especially when dealing with multi-class problems. The model's overall efficiency and accuracy depend on each class's performance. As the number of classes increases, the overall accuracy may affect because the overall accuracy depends on each class's individual accuracy. Conclusively designing the model for multi-class violence in a real time environment is a big challenge due to many issues. Our research problem deals with nine different violent classes and one Normal nonviolent class by facing a real-time environment with an accuracy of 87.60%.

Table 1. Summary of Recent Papers with Model Used, Accuracy, No. of Violent Classes Addressed, Model Implementation for Real-Time/Non-Real Time and Data Sets.

Method	Object & Event Detection Method	Feature Extraction Method	Accuracy	No. of Classes (Labels) Addressed	Real-Time/Non-Real Time	Dataset
Vision-based Fight Detection from Surveillance Cameras (2020)[7]	optical flows generated from consecutive images	CNN+ Bidirectional LSTM, VGG16	87%	2 (Fight & Non-Violence Class)	Non-Real Time	UCF Crimes Dataset, Hockey Fight Dataset, Peliculas Dataset, Surveillance Camera Fight Dataset
Violence detection using Spatio-temporal features with 3D CNN (2019)[8]	Pre-train mobile-Net CNN model	3D CNN	97%	2 (Violence or Non-Violence)	Non-Real Time	Crowd Violence Database
A Comparative Analysis of Different Violence Detection Algorithms from Videos(2020)[9]	LTP, ViF, OViF, ViF +OViF	SVM+ AdaBoost	79%	4 (Hitting some object, Kicking, Fighting, Punching)	Non-Real Time	Crowd Violence Database Hockey fight dataset, violent flows database
Video Surveillance for Violence Detection Using Deep Learning(2020)[10]	gradient descent optimisation algorithm (RMSprop)	Residual networks (ResNets) ConvLSTM	89%	2 (Violence or Non-Violence)	Non-Real Time	KTH, Hockey fight dataset, and Violent-Flows were trained on NVIDIA GTX1080Ti GPU
Real-time Violence Detection Framework for Football Stadium comprising of Big Data Analysis and Deep Learning through Bidirectional LSTM(2019)[11]	HOG Feature Extractor	Bidirectional Long Short-Term Memory (BDLSTM)	91%	2 (Violence or Non-Violence)	Non-Real Time	Violent Interaction dataset

A Multi-Temporal Framework For Human Violent Event Analysis Of Video Surveillance Using CNN[12]	Pre-train mobile-Net CNN model	CNN	90%	2 (Violence or Non-Violence)	Non-Real Time	BEHAVE YouTube CCTV Footage
Efficient Two-Stream Network for Violence Detection Using Separable Convolutional LSTM(2021) [13]	long-range Spatio-temporal features	SepConv LSTM	89%	2 (Violence or Non-Violence)	Non-Real Time	RWF-2000
Detecting stabbing by a deep learning method from surveillance videos(2019)[14]	Pertained CNN	CNN	90%	2 (Violence or Stabbing)	Non-Real Time	Real-Life Violence Situation Dataset
Using Images for Real-Time Violence Detection in the Edge(2020)[15]	optical flows generated from consecutive images	LSTM	75%	2 (Violence or Non-Violence)	Real-Time	UCF-101, UCF-Crime dataset

3. Experimental Results & Discussions

We have explored different strategies to discover the features' saliency from different deep learning models to detect violence in videos. ResNet50+LSTM was optimised and fine-tuned at the FC layer and implemented to check the parameters like accuracy, Precision, Recall, F1-Score, the time required to generate class labels for different classes, and quality of recognition of violent classes etc. This Deep Learning algorithm is used to extract features from the frames of the videos. In our experiments, the extracted features have been fed into a fully connected network which detects violence at the frame level. Furthermore, we have applied attention to the features extracted from the frames through a spatial transformer network which also enables transformations like rotation, translation and scale called data augmentation. In the end, the features extracted from the optimised and fine-tuned ResNet50 model proved more salient in detecting violence. These ResNet50 features, in association with LSTM, provide an accuracy of 87.60%, which is better than the other models we have experimented with. A dataset has been created with a consistence of violent and nonviolent videos from different well-known datasets cited in reputed research papers, namely the Hockey Fight dataset, UCF crime dataset, crowd violence, violence flow dataset, NTU Human Action Recognition datasets, violence scene dataset, VSD 2015 dataset, violence videos downloaded from movies and other violent videos from YouTube and Real-life violence videos from social media. We also create a few datasets for slapping and Pushing classes due to the lack of data available for training.

3.1. Experimental setup for Experiments

- The Training of ResNet50+LSTM was carried out on Param Shavak SuperComputing Facility available at GEC Rajkot, Gujarat, India, under Gujcost Project.
- The SuperComputer features are 2 - 5 Tera-Flops peak computing power with 8 TB of storage, 64 GB RAM, 2 multicore CPUs each with a minimum of 12 cores, and 2 numbers of accelerator cards - NVIDIA K40 accelerator card and NVIDIA P5000 for deep learning were used to generate Results.

At the first level, 35 thousand Images were converted from the videos explained in the introduction. At the second level, after preprocessing, a total of 12000 images by separation of 1200 images of each class for labeling and annotations were assembled. These 12000 images were divided into 70% for training, 15% for testing and the remaining 15% for testing.

Tabel 2. Summary of Implemented ResNet50+LSTM Model for Ten Classes and Accuracy

Deep Learning Model	No. of Class es	No. of Images used for training	Class Labels	Precision	Recall	F1-Score	Accuracy
ResNet50+LSTM	10	12000	Attacking	0.87	0.82	0.84	87.60%
			Fighting	0.94	0.88	0.90	
			Hit with Object	0.87	0.97	0.92	
			Kicking	0.91	0.95	0.93	
			Normal	0.86	0.84	0.84	
			Punching	0.83	0.91	0.87	
			Pushing	0.93	0.93	0.93	
			Shoot with Gun	0.82	0.82	0.82	
			Slapping	0.88	0.84	0.86	
			Stabbing with Knife	0.82	0.89	0.85	

Fig. 2 Loss/Accuracy during Training and Validation of ResNet50+LSTM Model

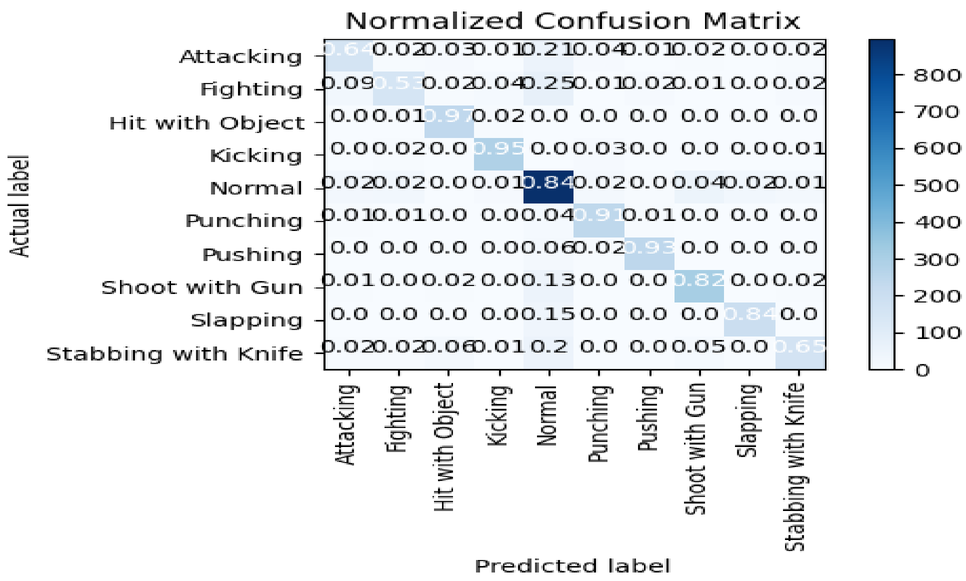

Fig. 3 Confusion Matrix of ResNet50+LSTM for Ten Classes

Table 2 summarises important parameters, particularly Precision, Recall, F1-Score of each class and overall accuracy of the modified Resnet50+LSTM deep learning model trained for Multi-class violent recognition. Precision is the measurement of the true positives to the total number of true positives and false positives. A recall is a measurement of the sensitivity of the models, while F1-Score is the harmonic mean of Precision and Recall.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$Recall = \frac{No. of\ Correctly\ Predicted\ Positive\ Instances}{No. of\ Total\ Positive\ Instances\ in\ the\ Datasets}$$

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

The Accuracy of the Model is identifying the number of correct predictions. It depends on the performance of each class.

$$Accuracy = \frac{(True\ Positives + True\ Negatives)}{(True\ Positives + True\ Negatives + False\ Positives + False\ Negatives)}$$

Fig.-2 describes the graph of Loss/Accuracy for the number of Epochs generated in a group of Images or training and validation of the model. It describes the visual performance of training accuracy, validation accuracy, training loss and validation loss for 300 Epochs. Fig.-3 describes the confusion matrix for ten different classes for different activities, particularly Attacking, Fighting, Hitting with Objects, Kicking, Normal Class, Punching, Pushing, Shoot with a Gun, Slapping and Stabbing with the knife. The confusion matrix provides an idea about the evaluation of the functions of the classification model. It denotes how many percentages of images are wrongly classified compared to what was and is wrongly classified in which class.

Fig.4 (a) to Fig.4 (j) shows the sample results as screenshots of recognised violent activities and normal activity, specifically Attacking, Fighting, Hit with Objects, Kicking, Normal Class, Punching, Pushing, Shoot with a Gun, Slapping and Stabbing with knife respectively during testing.

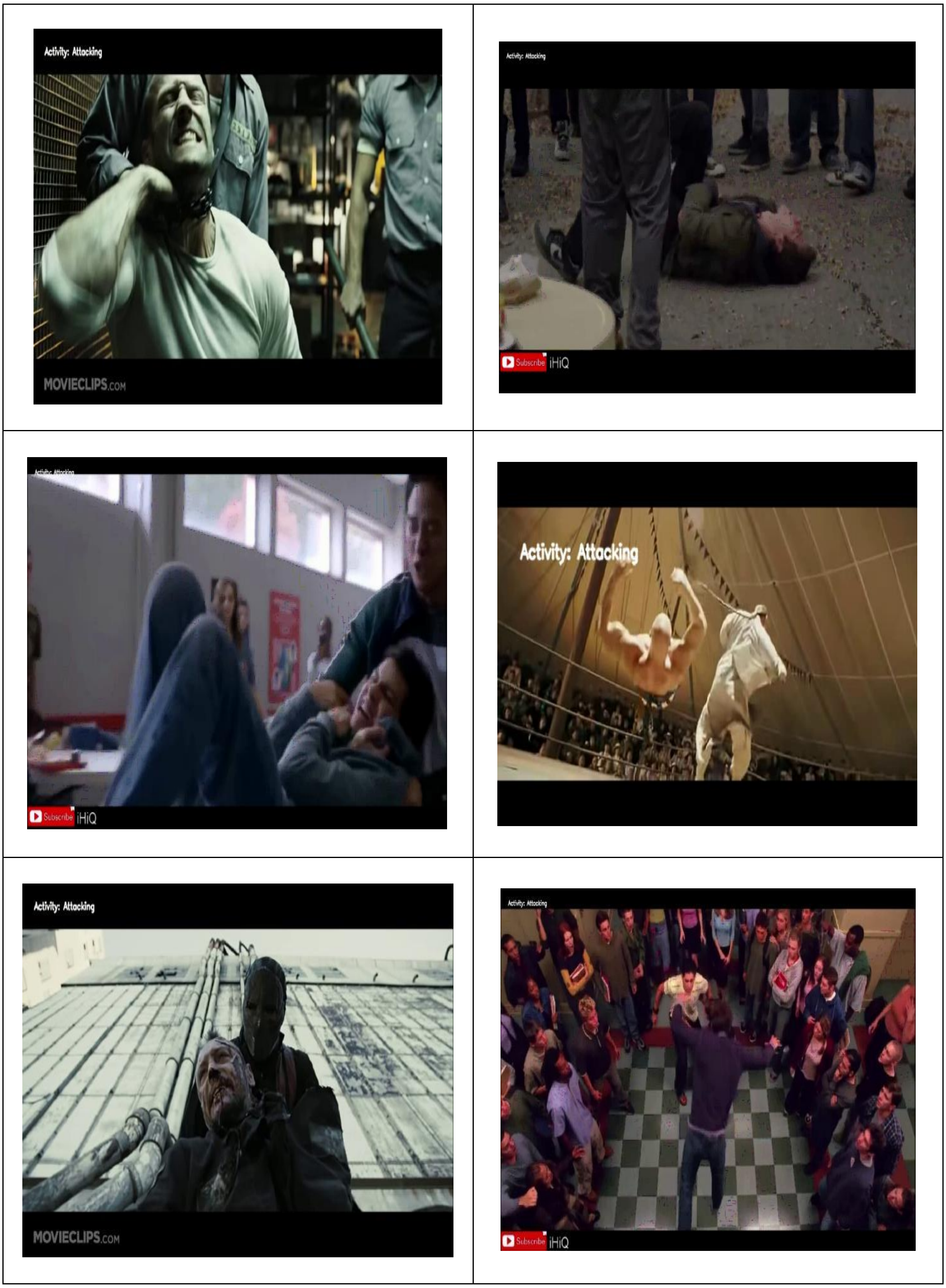

Fig. 4(a) Recognition of "Attacking" activity by the Proposed Model

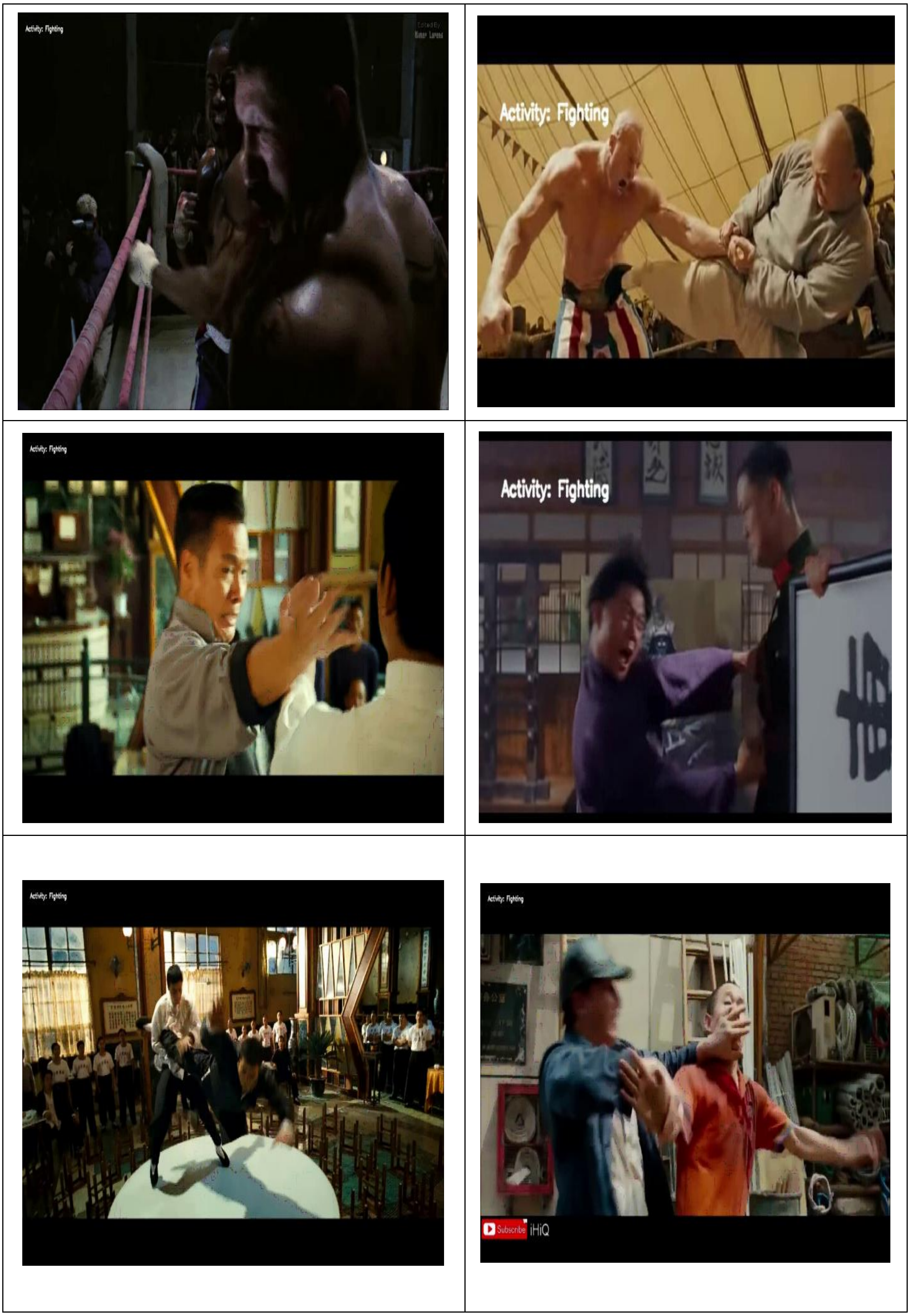

Fig. 4(b) Recognition of "Fighting" activity by the Proposed Model

Fig. 4(c) Recognition of "Hit with Object" activity by the Proposed Model

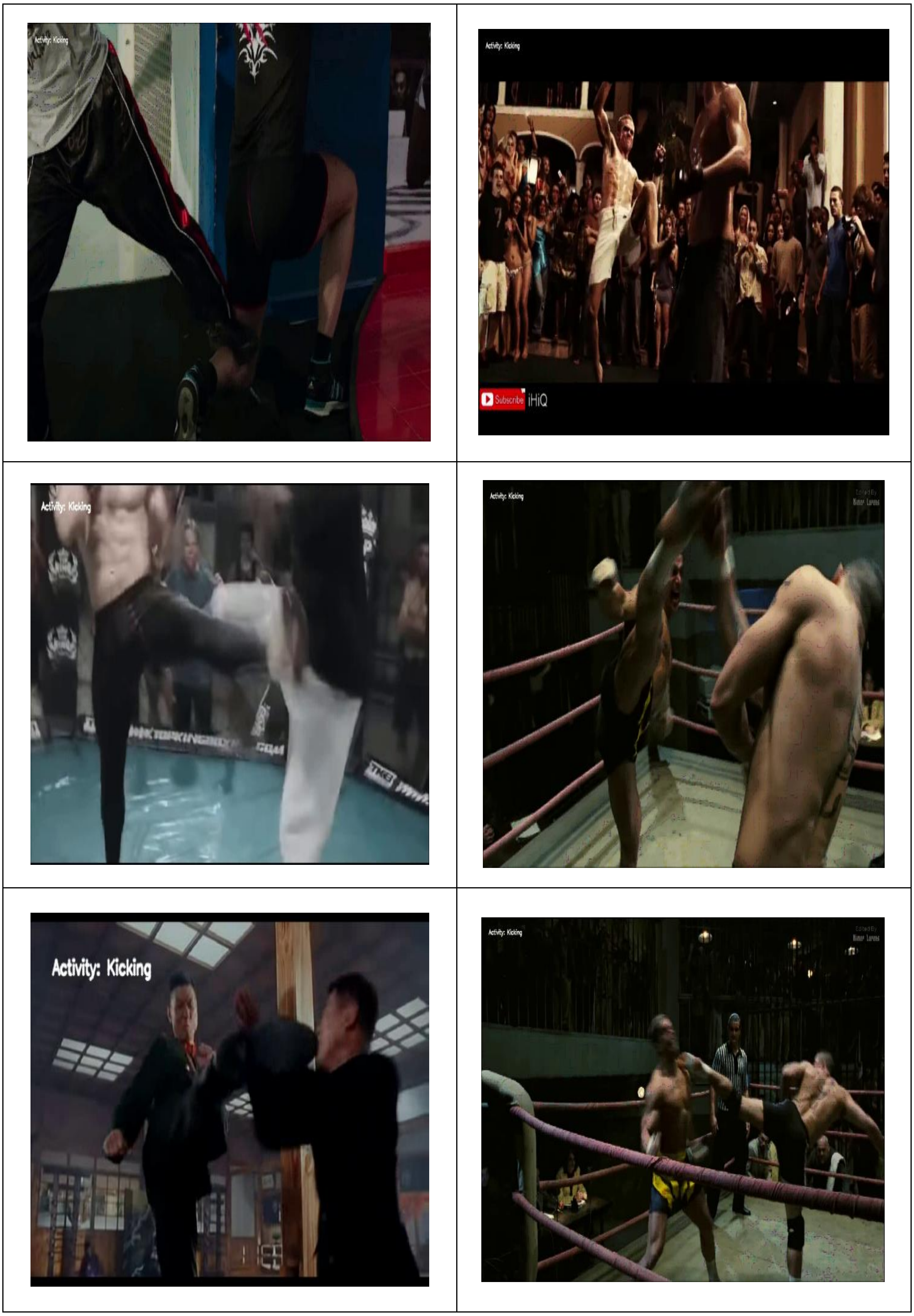

Fig. 4(d) Recognition of "Kicking" activity by the Proposed Model

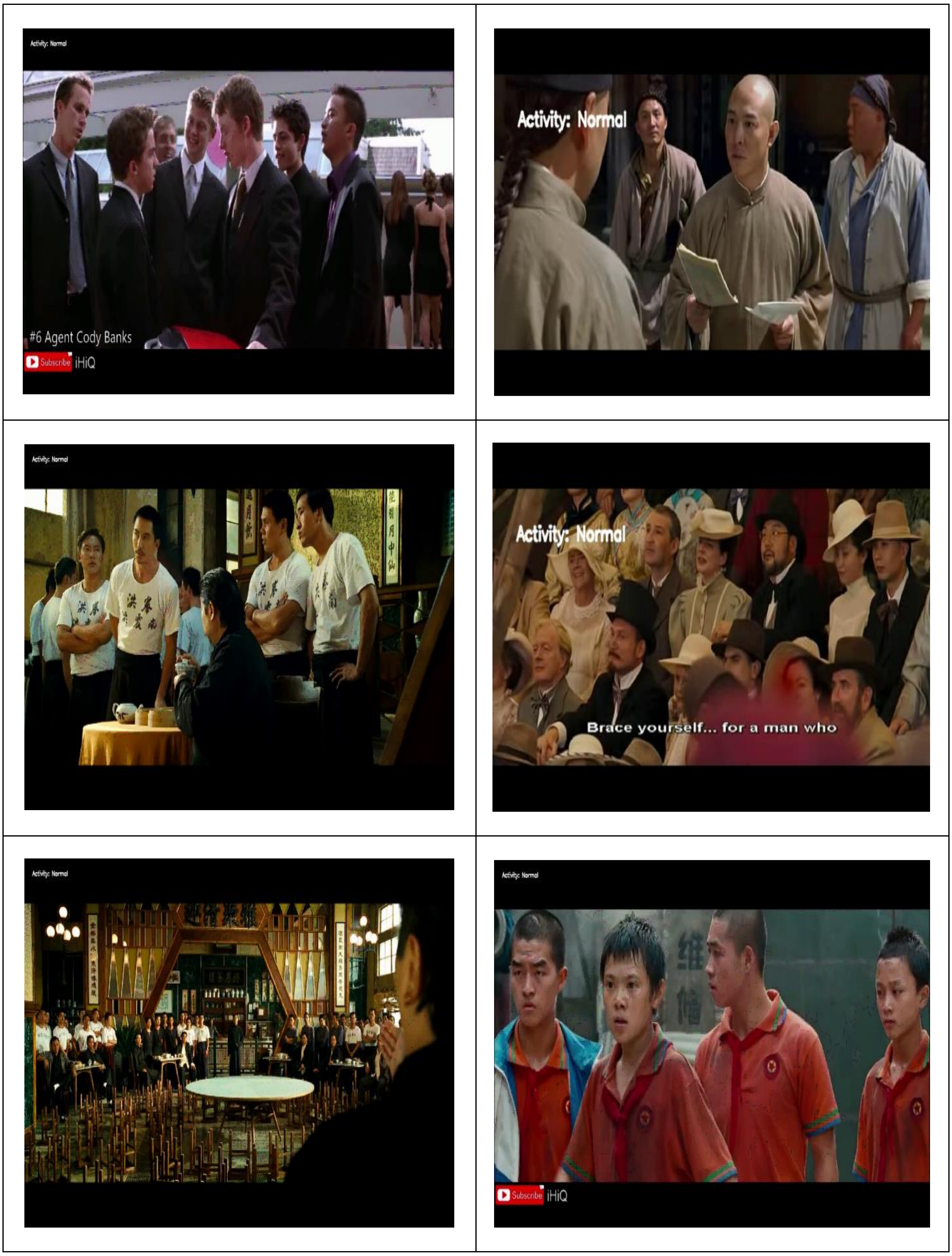

Fig. 4(e) Recognition of "Normal" activity by the Proposed Model

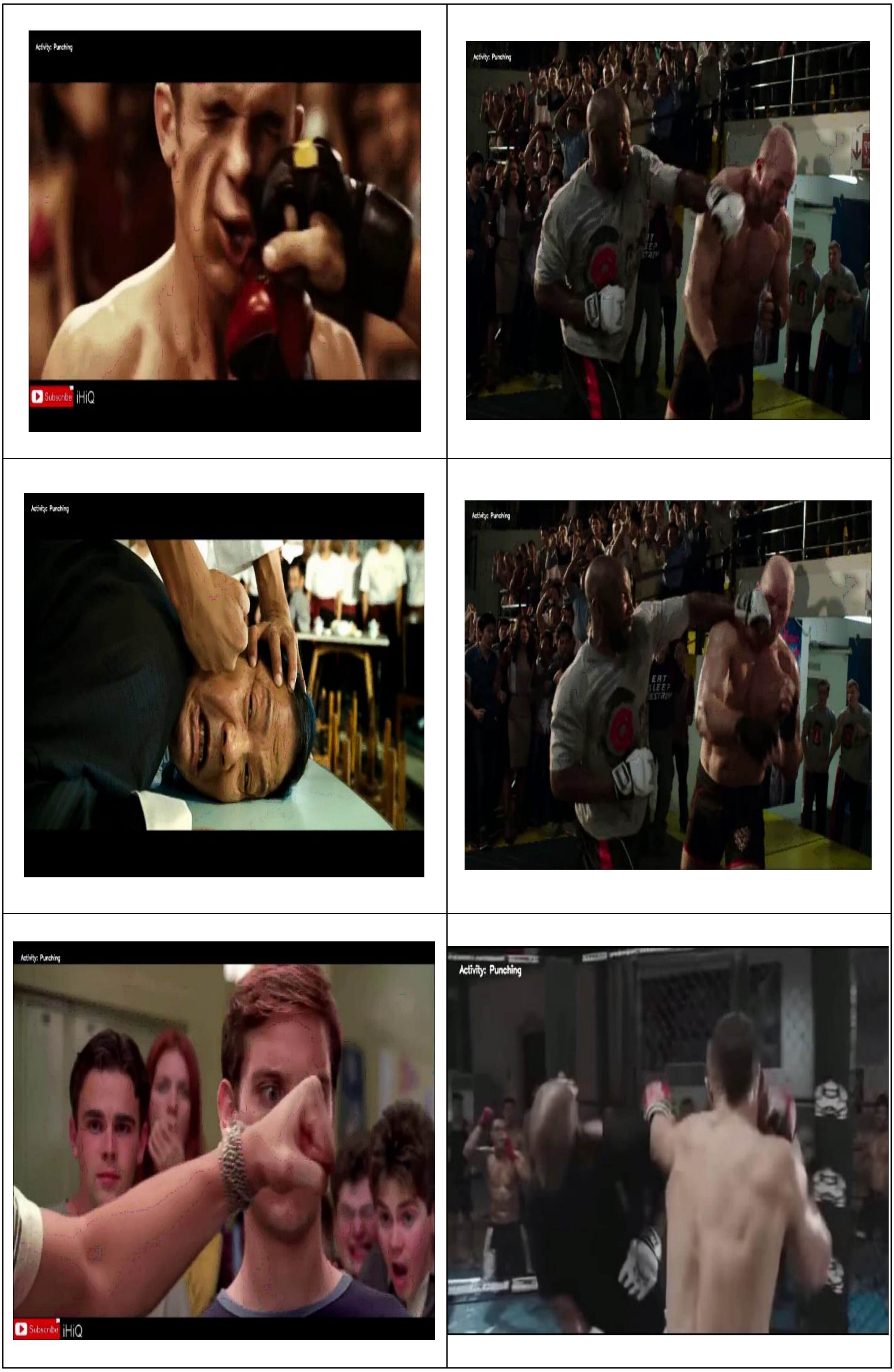

Fig. 4(f): Recognition of "Punching" activity by the Proposed Model

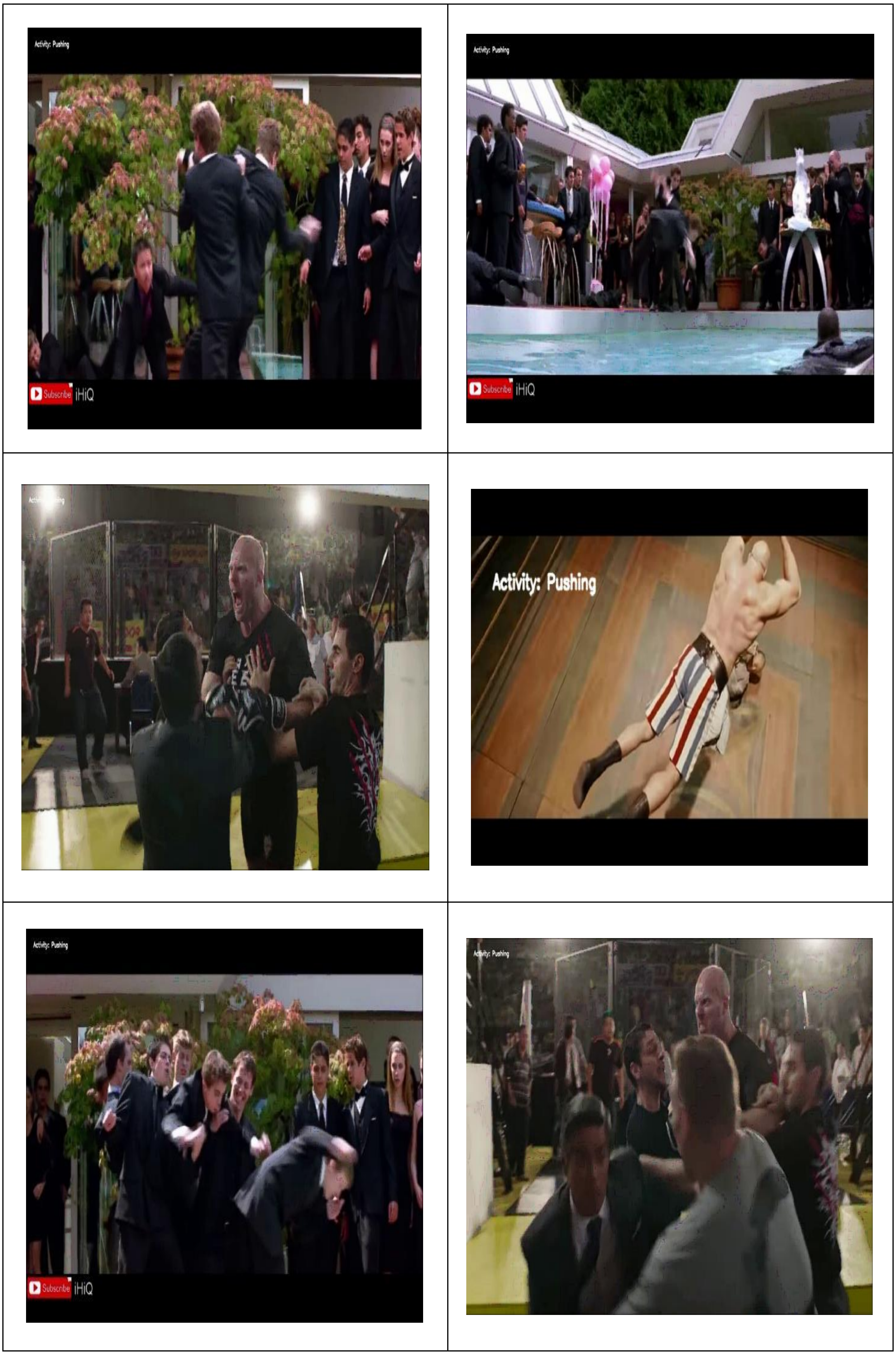

Fig. 4(g) Recognition of "Pushing" activity by the Proposed Model

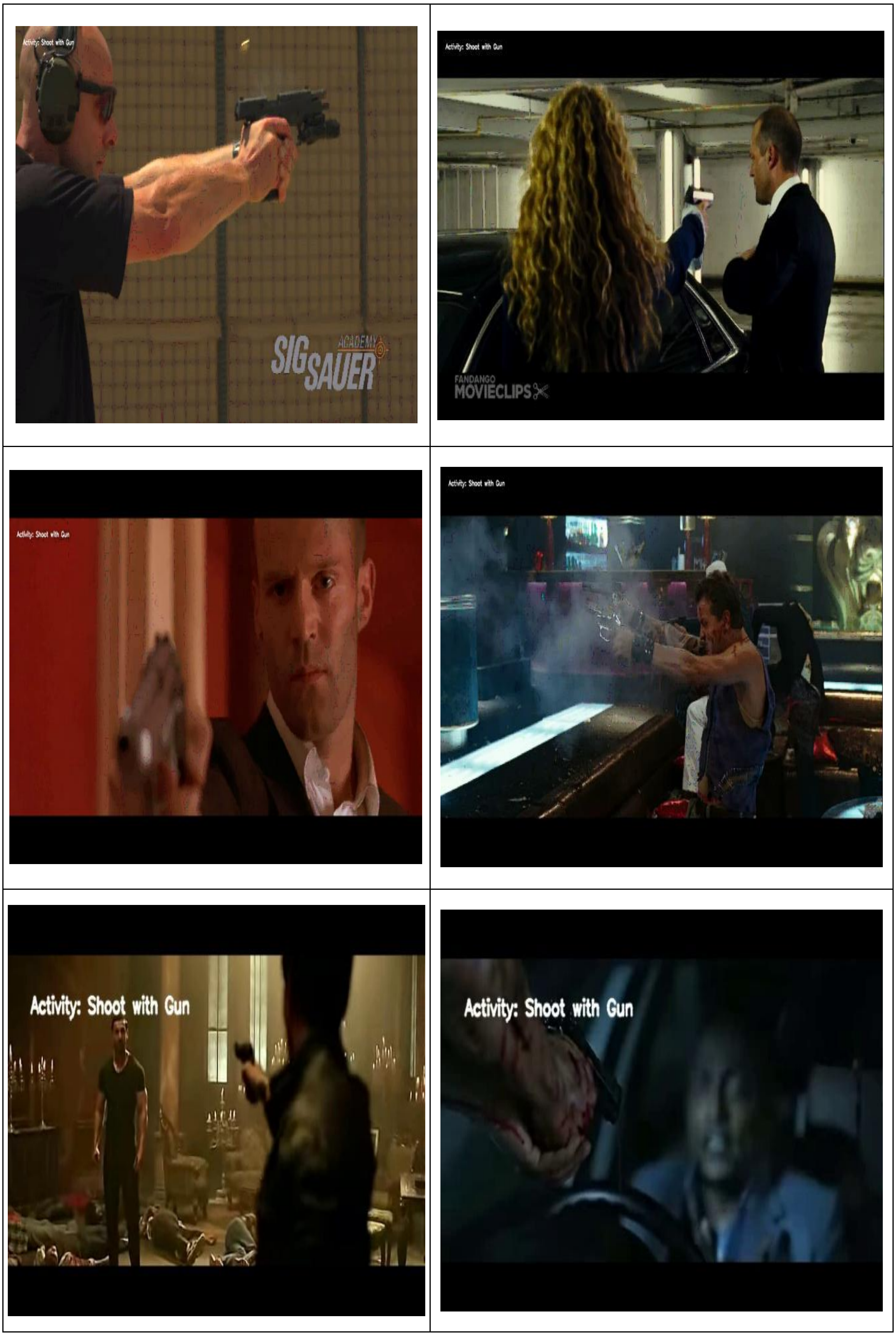

Fig. 4(h) Recognition of "Shoot with Gun" activity by the Proposed Model

Fig. 4(i) Recognition of "Slapping" activity by the Proposed Model

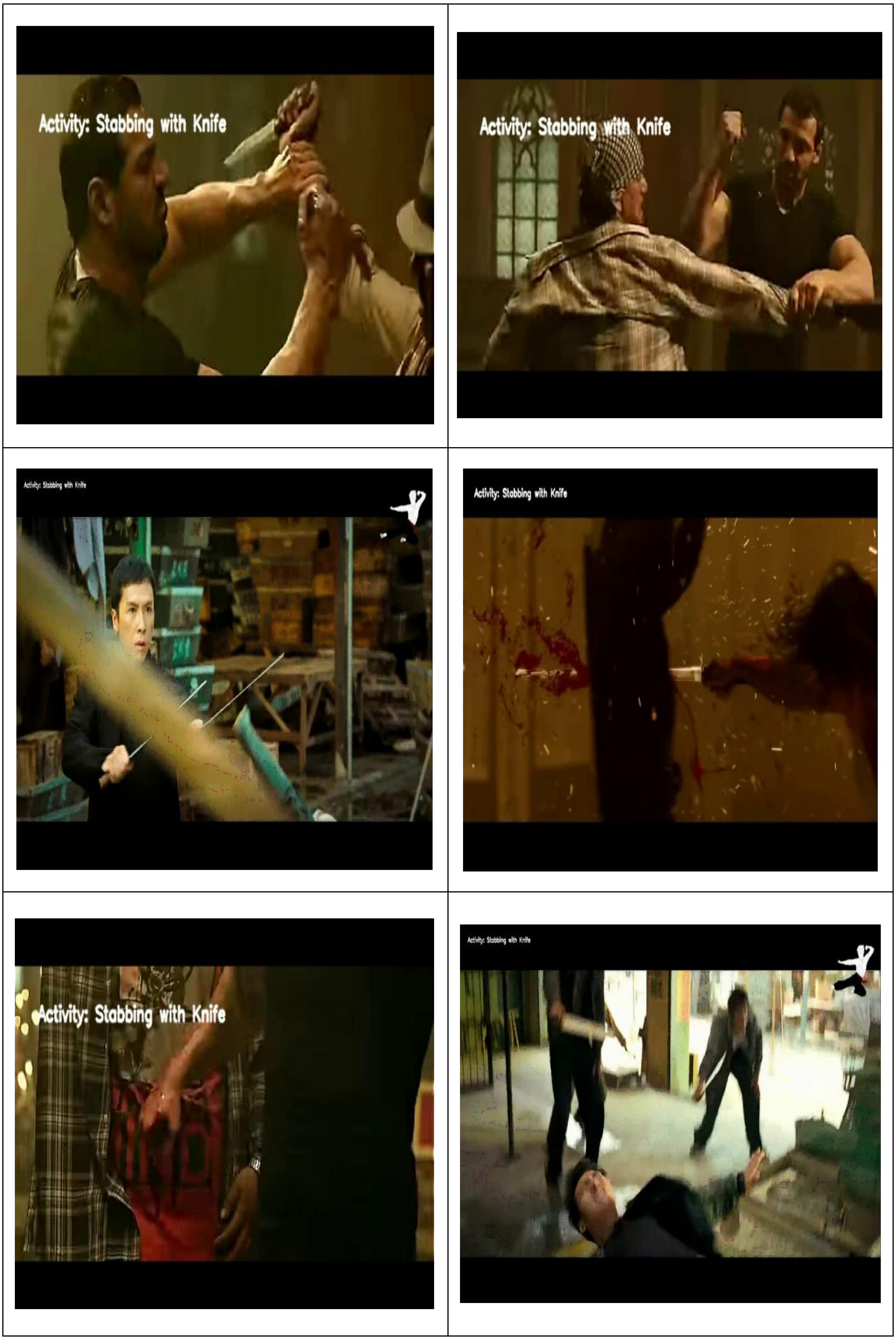

Fig. 4(j) Recognition of "Stabbing with Knife" activity by the Proposed Model

4. Conclusion and Future Scope

The existing work acknowledges the availability of violence frames in pre-recorded video with average accuracies for non-real-time environments. The work proposed in the different papers for real-time applications can address only two classes with average accuracy. The research work discussed in the paper is innovative and alive with challenges because it can address ten different classes with considerable accuracy of 87.60%, which can be applied to real-time environments. The development of Modified ResNet50 with the LSTM model for multi-class addressing was tedious and complex due to the lack of good quality datasets in multiple violent domains with ten different classes. The violent content in the videos is harmful to children and teenagers. Suppose the videos are not viewed under the supervision of the adult people by the children or teenagers who contain violent information. In that case, it may misguide them and encourage them to copy or recreate the same event in the surrounding environment. As a result, easy accessibility of harmful, violent content may spread the violence in children and teenagers. After recognising violence in future, we can localise violent recognised content. Later Recognition and Localisation of the violent activities, we can blur or block those violent contents to save children and teenage people from watching harmful violent contents.

References

[1] D. Jani and A. Mankodia, "A Review on State of the Art Abnormal Activity Recognition Approaches," *International Journal of Emerging Trends in Engineering Research,* vol. 9, no. 3, pp. 182–188, 2021. Doi: 10.30534/ijeter/2021/05932021.

[2] D. Jani and A. Mankodia, "ITEE Journal Deep Learning Based Multi Class Human Violent Action Recognition," 2021. [Online]. Available: http://www.iteejournal.org/v10no5oct21_pdf2.pdf

[3] D. Jani and A. Mankodia, "Comprehensive Analysis of Object Detection and Tracking Methodologies from Surveillance Videos," 2021. doi: 10.1109/ICCMC51019.2021.9418373.

[4] Hardik Bhatt, Anand Mankodia," A Comprehensive review on Content Based Image Retrieval System: Features & Challenges," In: K. Kotecha et al. Eds., *Data Science and Intelligent Applications, Lecture Notes on Data Engineering and Communications Technologies,* vol. 52, pp. 63-74, 2020. https://doi.org/10.1007/978-981- 15-4474-3_7

[5] A. P. Mankodia, S. K. Shah, "Design of User Interface for Shot Boundary Detection using Classical Methods", proceeding of *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering,* vol. 1, no. 7, pp. 287-293, 2013.

[6] A. P. Mankodia, S. K. Shah, "Analysis of Histogram Based Shot Segmentation Techniques for Video Summarization", proceeding of *International Journal of Electronics Communications and Electrical Engineering,* vol. 4, no. 4, pp. 2-11, 2014.

[7] Ș. Aktı, G. A. Tataroğlu, and H. K. Ekenel, "Vision-based Fight Detection from Surveillance Cameras," 2020. Doi: 10.1109/IPTA.2019.8936070.

[8] F. U. M. Ullah, A. Ullah, K. Muhammad, I. U. Haq, and S. W. Baik, "Violence Detection using Spatiotemporal Features with 3D Convolutional Neural Network," *Sensors* Switzerland, vol. 19, no. 11, 2019. Doi: 10.3390/s19112472.

[9] P. Vashistha, J. P. Singh, and M. A. Khan, "A Comparative Analysis of Different Violence Detection Algorithms from Videos," in *Lecture Notes in Networks and Systems,* Springer. vol. 94, pp. 577–589, 2020. Doi: 10.1007/978-981-15-0694-9_54.

[10] M. Sharma and R. Baghel, "Video Surveillance for Violence Detection Using Deep Learning," in *Lecture Notes on Data Engineering and Communications Technologies,* Springer Science and Business Media Deutschland GmbH, vol. 37, pp. 411–420, 2020. Doi: 10.1007/978-981-15-0978-0_40.

[11] S. R. Dinesh Jackson et al., "Real Time Violence Detection Framework for Football Stadium Comprising of Big Data Analysis and Deep Learning Through Bidirectional LSTM," *Computer Networks*, vol. 151, pp. 191–200, 2019. Doi: 10.1016/j.comnet.2019.01.028.

[12] B. Pradeepa, A. Viji, and J. J. Anthesious, "A Multi-Temporal Framework for Human Violent Event Analysis in Video Surveillance using CNN."

[13] Z. Islam, M. Rukonuzzaman, R. Ahmed, Md. H. Kabir, and M. Farazi, "Efficient Two-Stream Network for Violence Detection Using Separable Convolutional LSTM," 2021. Doi: 10.1109/IJCNN52387.2021.9534280.

[14] "*IEEE Instrumentation and Measurement Society and Institute of Electrical and Electronics Engineers*," IST 2019 : 2019 IEEE International Conference on Imaging Systems & Techniques, Abu Dhabi, UAE, 2019 .

[15] A. Pandey, T. Samarth, and S. Raghuram, "Using Images for Real-Time Violence Detection in the Edge," in *Lecture Notes in Electrical Engineering,* vol. 656, pp. 677–685, 2020. Doi: 10.1007/978-981-15-3992-3_58.

[16] Deniz, Oscar & Serrano Gracia, Ismael & Bueno, Gloria & Kim, Tae-Tyun, "*Fast Violence Detection in Video*," VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications, vol. 2, 2014.

[17] Ribeiro, Pedro &Audigier, Romaric & Pham, Quoc-Cuong, "RIMOC, A Feature to Discriminate Unstructured Motions: Application to Violence Detection for Video-Surveillance," *Computer Vision and Image Understanding*, vol. 144, 2015. 10.1016/j.cviu.2015.11.001.

[18] V. Machaca Arceda, K. Fernández Fabián and J. C. Gutíerrez, "Real Time Violence Detection in Video," *International Conference on Pattern Recognition Systems (ICPRS-16),* pp. 1-7, 2016. Doi: 10.1049/ic.2016.0030.

[19] Qian, Huimin& Mao, Yaobin& Xiang, Wenbo& Wang, Zhiquan, "Recognition of Human Activities using SVM Multi-Class Classifier," *Pattern Recognition Letters*, vol. 31, pp. 100-111, 2010. 10.1016/j.patrec.2009.09.019.

[20] Al-Naji A, Chahl J, "Detection of Cardiopulmonary Activity and Related Abnormal Events Using Microsoft Kinect Sensor," *Sensors* Basel, vol. 18, no. 3, pp. 920, 2018. Doi: 10.3390/s18030920.

[21] Wang, Siqi& Zeng, Yijie& Liu, Qiang& Zhu, Chengzhang& Zhu, En& Yin, Jianping, "Detecting Abnormality without Knowing Normality: A Two-stage Approach for Unsupervised Video Abnormal Event Detection," pp. 636-644, 2018. 10.1145/3240508.3240615.

[22] H. Alvari, S. Sarkar and P. Shakarian, "Detection of Violent Extremists in Social Media," *2nd International Conference on Data Intelligence and Security (ICDIS),* pp. 43-47, 2019. Doi: 10.1109/ICDIS.2019.00014.

[23] Mu, Guankun& Cao, Haibing &Jin, Qin, "Violent Scene Detection Using Convolutional Neural Networks and Deep Audio Features," vol. 663, pp. 451-463, 2016. 10.1007/978-981-10-3005-5_37.

[24] Gargi Sharma, and Gourav Shrivastava. "Crop Disease Prediction using Deep Learning Techniques - A Review" *SSRG International Journal of Computer Science and Engineering,* vol. 9, no. 4, Apr. 2022, pp. 23-28. *Crossref,* https://doi.org/10.14445/23488387/IJCSE-V9I4P104

[25] M.A. Muthiah, E. Logashamugam, N.M. Nandhitha "Performance evaluation of Googlenet, Squeezenet, and *Resnet50* in the classification of Herbal images" *International Journal of Engineering Trends and Technology,* vol. 69, no. 3, pp. 229-232, 2021.

[26] Zulaiha Parveen A, Mr.T.Senthil Kumar "The Deep Learning Methodology For Improved Breast Cancer Diagnosis In MRI" *International Journal of Computer and Organization Trends* vol. 11 no. 3, pp. 11-14, 2021.

[27] Zhijie Yang, Hongbin Huang "Garbage Classification Based On Deep Residual Weakly Supervised Learning Model" *International Journal of Recent Engineering Science* vol. 7, no. 3, pp. 47-51, 2020.