TOWARD SUBJECTIVE VIOLENCE DETECTION IN VIDEOS

Bruno Peixoto, Bahram Lavi, João Paulo Pereira Martin,
Sandra Avila, Zanoni Dias, and Anderson Rocha

Institute of Computing, University of Campinas (Unicamp), Campinas, São Paulo, Brazil

ABSTRACT

Violence detection in videos aims to identify whether a violent action occurred within a video stream. Effective tools for intelligent video analysis are highly demanded, specially to determine violence in video streams. Such solution could have applications in detecting inappropriate behaviors in video feeds, aiding law-enforcement in forensic cases, protecting children from accessing inappropriate online content and helping parents making informed decisions about what their kids should watch. Prior art on violence detection, particularly recently proposed deep learning based ones, seeks to identify violence in videos as a whole, without considering breaking down the subject into some of its underlying concepts. In this paper, we explore a different methodology of violence detection, which relies upon two deep neural network (DNNs) frameworks to learn spatial-temporal information on video clips under different scenarios — subjective- and conceptual-based. We leverage deep feature representations for each specific concept, and aggregate them by training a shallow neural network as a binary-classification problem to describe violence as a whole. Finally, we show that using more specific concepts is an intuitive and effective solution, besides being complementary to form a more robust definition of violence.

Index Terms— computer vision, violence classification, deep-learning, semantic concept detection, forensic computing

1. INTRODUCTION

Semantic violence detection is an important capability for the issue of video analysis in filtering sensitive media contents. It can also provide a useful tool to protect users from receiving undesired media from any kind of sources and, in conjunction with intelligence video surveillance systems, to detect inappropriate behaviors and aid law-enforcement in forensic examination cases. Moreover, it can prevent content from being uploaded to social media, forums or educational platforms; or on the other hand, prevent it from being shown in specific places such as schools and workplaces.

Undeniably, hundreds of hours of video are uploaded every minute through the Internet, while handling and analyzing them are, accordingly, heavily time consuming. Violence is considered as one of the sensitive media that is very subjective to define and, as such, leads to different interpretations. Early exposure of violence on scenes of media content might be not suited for everyone, decisively, for underage persons. Typically, "automatic" solutions in prior art rely on extracting discriminant spatial-temporal features in videos in an attempt to identify violence, which turns this task into one very active research area. Proof of interest is also apparent on the competition "MediaEval Affect Task", which aims to identify violence in movies [1].

In the task of violence detection, some works consider only a specific subject at a time, such as *fights* [2, 3], while in [4], a combination of both *explosions* and *blood* were considered. Most recent solutions involved the use of deep-learning techniques to extract features and combine them with spatial-temporal descriptors [5, 6, 7, 8, 9]. In this work, following the benchmark work in the MediaEval 2013 VSD dataset [10], the adopted definition of violence was that a scene is violent if "one would not let an eight-year old child see" [10].

Here, we aim to address the violence detection task by breaking down the subjective concept of violence into more specific concepts such as: *Blood*, *Cold Arms*, *Explosions*, *Fights*, *Fire*, *Firearms*, *Gunshots*. The idea of breaking down violence into different subjects is not novel. Cheng et al. [11] identified audio signatures with slightly varied events that could identify different kinds of violence, such as explosions, gunshots, and car crashes. Inspired by this process, we utilize two deep-learning techniques to explore the presence of violence in videos by considering the performance separately on each of the used networks in terms of subjective violence. We then take into account violence as a single higher level concept in order to analyze the performance behavior of both networks. Finally, we perform a fusion of the concepts of violence to identify only the subject of violence, and compare performance on different scenarios.

The contribution of this paper can be summarized as follows. First, we consider three different scenarios for violence detection in videos: (i) conceptual-based violence detection to identify a desired violence concept in videos; (ii) a setup in which only the violence as a unique concept is used regardless of the first scenario; and (iii) regarding to the first scenario, we consider the fusion of concepts as a whole to identify the more high-level concept of violence. Second, we utilize two deep neural networks (DNNs), which are robust on learning high level spatial-temporal information from raw image data. The two networks are the 3D-based convolutional neural network (commonly called C3D) and the joint of CNN with long short-term memory (CNN-LSTM). Both networks have been originally proposed for human action recognition [12]. Finally, we design a shallow neural network to combine the feature maps obtained from the above networks to address the third scenario of this study. This work extends upon our preliminary work [13] in the methodology on subjective violence classification. In this study, we aim to address an end-to-end classification approach, by considering two DL frameworks, separately, under different scenarios as discussed above.

This work was funded by the Coordination for the Improvement of Higher Level Education Personnel (CAPES) under the grant Capes DeepEyes, and the So Paulo Research Foundation (FAPESP) under the grant DéjàVu 2017/12646-3 and grants 2013/08293-7, 2015/11937-9, 2017/16246-0, and 2017/16871-1. Additional funding was awarded by the National Counsel of Technological and Scientific Development (CNPq) under grants 400487/2016-0 and 425340/2016-3 and the French Committee for the Evaluation of Academic and Scientific Cooperation with Brazil, CAPES-COFECUB, grant 831/15.

The remainder of this paper is organized as follows. We first summarize recent related work in the area of violence detection in Sec. 2. In Sec. 3, we discuss the utilized two deep learning techniques in violence classification problem. In Sec. 4, we evaluate their effectiveness on *MediaEval-2013-VSD* data set. And finally, we conclude the paper with directions for future work in Sec.5.

2. BACKGROUND

In this section, we first describe recent violence detection approaches aimed at analyzing violent segments in videos. We then summarize deep learning techniques proposed for violence classification in videos.

Violence detection methods. The violence detection problem had its first works derived from solutions of action recognition. To deal with this task, some works proposed the Bag of Visual Word (BoVW) approach [3, 14]. In [3], low-level features are generated with an image descriptor similar to Space-Time Interest Points (STIP) [15] and classifies the feature vectors via Support Vector Machine (SVM). In [14], local spatial-temporal features are used for classification. Clarin et al. [16] proposed to use local interest-point based approaches to detect fights as subjective violence. A novel descriptor was proposed in [17] for real-time crowd violence detection.

DNN techniques on violence detection. To the best of our knowledge, only a few published papers have utilized DL frameworks for the actual violence detection problem. In [5, 6, 7, 8, 9, 18], DL techniques were utilized as a high-level feature extractor method aiming to identify the correct class label by using a supervised learning algorithm (e.g., SVM). Ding et al. [19] relied upon a 3D convolutional based network, which interprets violence as *fights*, and trained the network on a Hockey games data set. In [20], a three-stream DNN framework was proposed for detecting violence under the subject of person-to-person violence setup. They adopted 2D CNN-based network for each stream, namely acceleration, spatial, and temporal streams, to learn the spatial information, while a Long Short-Term Memory (LSTM) Neural Network was used on top of three streams to learn temporal information.

Remarkably, most of the above mentioned works only rely on evaluating their methodology on a specific concept of violence (e.g., fights) without considering the myriad of possible different concepts for violence. However, such a complex concept could be very challenging for a DNN to grasp with. Instead, here we aim to separately leverage two DNN frameworks to learn smaller and more objective concepts, and finally compare the behavior of both networks with each other. We additionally design a feature fusion network in order to learn deep feature obtained by the utilized DNNs.

3. METHODOLOGY

We briefly explain both DNNs we rely upon in this work in order to learn an end-to-end classification solution for violence under various subjective concepts. Each network can respectively extract different types of violence information from raw videos, and eventually identify the desired class label. Following this section, we first discuss two DNNs for learning spatial-temporal information from raw images of video, which are utilized under different setups. We then explain our feature fusion network, which aims to learn the violence definition from obtained high-level features of the DNNs.

3.1. 3D Convolutional network

Unlike typical CNN-based frameworks, 3D convolutional networks are basically trained on frame sequences of video clips. This type of network is able to learn correlations directly in the 3D space, where in the convolutional processing of CNN it has the ability of computing features from both spatial and temporal domains. The network consists of eight convolution layers, which can be categorized into five groups; the first two groups are comprised of a single convolutional layer , while the last three groups are followed by two sequential convolution layers The network is followed by 3 fully-connected layers and the number of classes (2 in our case).

We followed the configuration of hyperparameters of each layer as in [21]. We use $3 \times 3 \times 3$ filters for all convolutional layers. Each convolutional layer is followed by a rectified linear unit (ReLU) and a max-pooling layer. Max-pooling filters are of size $2 \times 2 \times 2$ except in the first layer, where it is $2 \times 2 \times 1$. The size of convolution output is kept constant by padding 1 pixel in all 3D domains. Filter stride for all is set 1 for convolution and 2 for pooling operations. Fully connected layers are followed by ReLU layers. Softmax layer at the end of the network outputs class scores.

3.2. CNN-LSTM network

Considering the great success of the CNN-LSTM architecture on jointly learning spatial-temporal information on hand gesture recognition problem [22], we adopt this network for the task of violence detection; whereas the network is trained on sequence-level classification problem. The CNN-LSTM network consists of the following distinct layers: at the beginning of the network, two convolutional layers are followed by max-pooling layers. The max-pooling layer halves the width and height of feature maps passed through the convolutional layers. Then a flatten layer is used to reduce each vector to one dimension. The output of flatten layers is then fed into an LSTM layer by decreasing output sizes. Finally, a softmax function yields the final probability of the network to determine whether or not the violence concept occurred in the video sequence analyzed.

3.3. Feature-fusion network

Clearly, it is possible to design a network to learn spatial-temporal information only on the subject of violence rather than going through each concept separately. However, the number of data samples for the task of violence detection is relatively large, due to large amount of video frames. At this part of our study, we present a strategy to design a network, which can independently learn the final decision from the output weights obtained from the binary classification networks. This solution can be used to reduce both computational and memory foot print.

We design a joint network, which aims to leverage concepts of violence and, ultimately, identify the subject of violence. It is noted that the fusion network is not utilized in an early stage; this mainly because our fusion network is trained independently on the feature representations generated by the two DNNs. Therefore, we are interested on the features from a trained model, which gained better performance as its originally learned for raw image data. For fusion, we designed a shallow network with three hidden layers and a softmax layer at the top of the network to predict whether violence is present within the sequence of frames. Fig. 1 shows the whole pipeline of our fusion network.

After a grid search of number of neurons for each hidden layer, the best trade-off with respect to the performance of the network for each hidden layer is selected 512, 128, 32 neurons, respectively

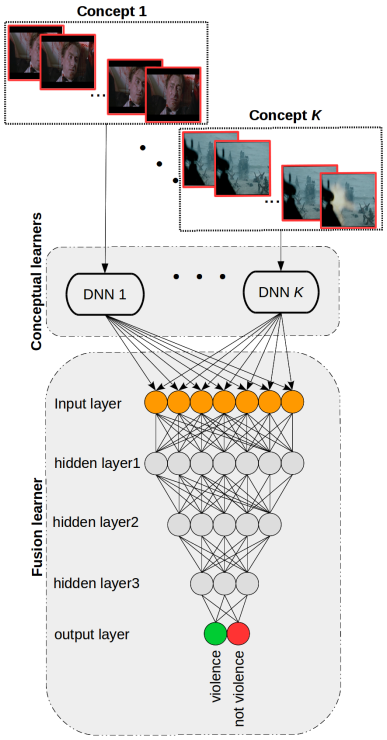

Fig. 1. Pipeline of the proposed feature fusion network. Also, two instances from the used data set are presented, in which the raw image videos are only used for both DNNs in early-stage, while the generated feature representations is used for fusion task.

from first to last hidden layer. The input of the network is the feature weights obtained from the last fully-connected layer from which the final features are obtained for the model trained on each concept of violence. The size of the input network is dependent on the number of neurons of the desired layer in the main network.

4. EXPERIMENTS

We evaluate the performance of the utilized networks (C3D and CNN-LSTM) on benchmark MediaEval-2013-VSD data set. We first explain the setting of the used data set on violence detection. We then describe the implementation details of our frameworks, as well as our feature fusion network. Finally, we present and discuss the achieved performances over different experimental strategy.

Dataset. MediaEval-2013-VSD data set [10] is a benchmark data set on violent scenes which contains of 25 Hollywood movies of diverse genres. The data set provides shot segmentation from the movies, where each segment has been manually annotated in order to distinguish whether or not physical violent occurred within the scenes for each movie. The definition of violence used by the competition is that a scene is violent if "one would not let an eight-year old child see". The data set released with already separated partitions in training and testing sets. The training set includes 18 movies while the test set comprised with 7 movies Remarkably, among all the scenes, only 20% of them have been categorized as violent. Although the data-set provides annotations for individual concepts (e.g., blood, fights, etc.), these annotations are only available for the training set.

Implementation details. We implemented the architecture for the aforementioned networks using Keras DL library on Python. We then carried out our experiments by using Tensorflow toolbox as the DL platform on NVIDIA GeForce GTX 1080 Ti GPU. Each network was trained for a binary classification of each individual concept. For the sake of fair comparison, we separated five of the available training set movies, two for validation during training, and three for testing, and kept the same division throughout all experiments. We also scaled all video frames of the data set to the size of $128 \times 128 \times 3$ pixels for our networks inputs.

For the C3D model, we followed [21] to train the network by the stochastic gradient descent (SGD) algorithm, training on video clips of 32 frames. Due to the large amount of negative samples, we balanced the training set for each network selecting all the positive samples relevant to each concept and choosing an equal amount of clips for the negative class The network ran for 100 epochs, each one comprising of 1200 steps of training in 10 randomly selected video clips per batch. To train our CNN-LSTM model, we used the Root Mean Square Propagation (RMSProp) algorithm to train is a similar fashion, feeding the network with 10 batches of 32 sequential frames for 100 epochs, each one running for 1200 steps of training.

In the case of feature fusion network, we carried out the experiments separately for each main network (C3D and CNN-LSTM). For C3D network, we used the output weights of the last fully-connected layer, with feature size of 4096, for each input of 32 frames. While the output of LSTM from CNN-LSTM network is used (with feature size of 512) as the input of our fusion network for each individual frame. Therefore, we set the different input size for fusion network with respect to each experiment. We carry out the experiments on the same data samples as originally used for training the main networks. For each experiment on feature fusion network, we train the network for 150 epochs.

We notice that the behavior of fusion network was better, in terms of performance, when the network was trained with same optimizer used by the main network. We therefore use SGD, and RMSprop for the features map obtained by C3D, and CNN-LSTM networks, respectively. We trained the network with the batch size of 32, and learning rate of 0.0001.

Results. Here we present the results in terms of the performance of each network. We use metrics to quantify our experiments on frame-level classification problem. The metrics are averages of true positive (TP), true negative (TN), and normalized accuracy (Norm.

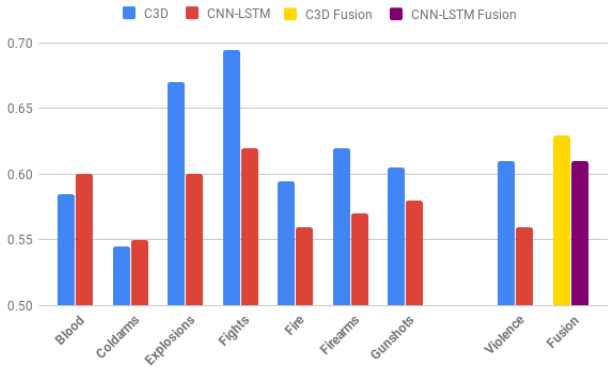

Fig. 2. Summary of the Normalized Accuracy of testing set evaluated with C3D, CNN-LSTM, and fusion models. The figure is best viewed in color.

	C3D						CNN-LSTM					
	Validation			Testing			Validation			Testing		
	TP	TN	Norm. Acc.	TP	TN	Norm. Acc.	TP	TN	Norm. Acc.	TP	TN	Norm. Acc.
Blood	0.47	0.64	0.56	0.46	0.71	0.59	0.52	0.62	0.57	0.79	0.41	0.60
Coldarms	0.54	0.48	0.51	0.56	0.53	0.55	0.72	0.36	0.54	0.96	0.14	0.55
Explosions	0.71	0.44	0.58	0.62	0.72	0.67	0.87	0.36	0.612	0.81	0.38	0.6
Fights	0.53	0.66	0.60	0.69	0.70	0.70	0.82	0.25	0.54	0.84	0.4	0.62
Fire	0.52	0.68	0.60	0.54	0.65	0.60	0.69	0.42	0.56	0.7	0.4	0.56
Firearms	0.45	0.66	0.56	0.59	0.65	0.62	0.60	0.59	0.60	0.77	0.36	0.57
Gunshots	0.62	0.72	0.67	0.57	0.64	0.61	0.61	0.52	0.57	0.63	0.52	0.58
Avg.	***0.55***	*0.61*	*0.58*	*0.57*	*0.67*	*0.62*	*0.69*	*0.46*	*0.57*	***0.79***	*0.37*	*0.58*
Violence	0.48	0.68	0.58	0.55	0.57	0.56	0.61	0.51	0.56	0.64	0.49	0.56
Feature fusion	0.52	**0.71**	**0.61**	**0.58**	**0.69**	**0.63**	**0.74**	**0.53**	**0.63**	0.74	**0.49**	**0.61**

Table 1. Accuracy of validation and testing for C3D and CNN-LSTM network. Also, the performance of each network under different scenarios. The evaluation metrics are: Average of true positive (TP), true negative (TN), and normalized accuracy. Best results in bold.

Acc.) over different experiments. Table 1 reports the average accuracy and cost function for both validation and testing on C3D network. It also reports the average value across whole concepts.

Additionally, Fig. 2 shows a summarized view on the normalized accuracy only in the case of evaluating the performance of the networks on testing step. At the following we discuss the achieved performances by comparing the behavior of networks with each other.

Discussion. Training a DNN to detect violence is a hard task, mainly due to the highly subjective nature of the theme. We chose to train independent networks for more specific concepts in order to find a more robust method to solve this problem — a divide-to-conquer approach. During the experiments, the machine memory constraints allowed small batches to be processed in parallel, reducing the training time considerably.

We were able to achieve some interesting results when aggregating concepts with another network for final decision making. Our experiments with C3D showed that the network was able to identify better concepts such as explosions and fights, that have a high correlation with movement. While the more still concepts such as blood and cold arms did not perform so well for this network. This can be explained by the nature of the concept and how the network processes its input. While we were feeding the network with video clips of 32 frames (the movies themselves were shot in 25 frames per second), these related concepts are more susceptible to only appear in a small portion of the clip, and this can negatively influence training.

Since we are motivated to detect violence in general, we compare two methods: (i) Training the models for the violence concept, which had a theoretical average testing accuracy of 62% with C3D and 58% with CNN-LSTM; and (ii) training another network to combine all specific concepts in order to detect violence from the features extracted for each individual network. With this method, we achieved 63% of classification accuracy for the testing set with C3D network and 61% accuracy with the CNN-LSTM network, both results outperforming each of our networks individually when trained to detect the high-level concept of violence directly (the one without the combination of individual concepts). This shows the separation of concepts for detecting violence leads to better results than just trying to detect violence directly.

Comparing both models, we had similar results for the fusion network in the testing set. Figure 2 depicts a comparison of the testing set for each network. C3D tends to achieve better results for almost all concepts and for the fusion itself. The difference between the movement-based concepts and the still ones is more apparent in the C3D network as well, while the CNN-LSTM model remains more stable with its accuracy across all concepts.

When we compare the accuracies between the testing and validation set, we find it interesting that the C3D model had a smaller difference from the validation to the testing set. We interpret this as a higher robustness of this model, in comparison with the significant higher accuracy for the validation set in the CNN-LSTM model.

The fusion network also shows its relevance when we compare its results with the average accuracy of all the individual concepts. For example, in the testing set for CNN-LSTM, the average accuracy of the networks for individual concepts was 58%, while the accuracy of the respective fusion network was 61%.

A network trained directly for the violence detection problem – i.e., without relying on individual concepts and their fusion – yields 56% accuracy for both C3D and CNN-LSTM compared to the 63% and 61% of our proposed method in each network, respectively.

5. CONCLUSIONS AND FUTURE WORK

Detecting violence is very important for many applications, but its high subjectivity makes it hard for a generalized model to have high accuracy. Our method tries to incorporate more specialized concepts on what can be classified as a violent scene in the foreground of violence detection. This opens up the field to not only identify violence, but what kind of violence is involved in the scene. Also, it simplifies the concepts the neural network has to deal with, making the experiments more easily comparable.

Using two networks intended to detect features related to the passage of time and movement yielded us better results in concepts that were closely related to movement, such as *fights* and *explosions*, while more static concepts such as *blood* and *cold arms* followed close behind in terms of accuracy. In the future, combining features learned in these networks with another network, more suited for the detection of objects in still images, can lead us one step further in solving this difficult problem. Our experiments already show that specializing a network in different concepts is appropriate to detect violence than training a network to detect a higher-level concept (Our results of 63% vs. 56%). We can also train different concepts in different network models if needed be.

It is also interesting to note that C3D had a better true-negative rate, while CNN-LSTM had a better true-positive rate, signaling that maybe a combination of both networks can achieve a higher accuracy. Future work will also be devoted to develop this combination in order to boost accuracy for each individual concept as well as the fusion of them.

6. REFERENCES

[1] Claire-Hélène Demarty, Cédric Penet, Guillaume Gravier, and Mohammad Soleymani, "The mediaeval 2012 affect task: violent scenes detection," in *Working Notes Proceedings of the MediaEval 2012 Workshop*, 2012.

[2] Fillipe DM De Souza, Guillermo C Chavez, Eduardo A do Valle Jr, and Arnaldo de A Araújo, "Violence detection in video using spatio-temporal features," in *Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2010, pp. 224–230.

[3] Enrique Bermejo Nievas, Oscar Deniz Suarez, Gloria Bueno García, and Rahul Sukthankar, "Violence detection in video using computer vision techniques," in *International conference on Computer analysis of images and patterns*. Springer, 2011, pp. 332–339.

[4] Jian Lin and Weiqiang Wang, "Weakly-supervised violence detection in movies with audio and video based co-training," in *Pacific-Rim Conference on Multimedia*. Springer, 2009, pp. 930–935.

[5] Qi Dai, Rui-Wei Zhao, Zuxuan Wu, Xi Wang, Zichen Gu, Wenhai Wu, and Yu-Gang Jiang, "Fudan-huawei at mediaeval 2015: Detecting violent scenes and affective impact in movies with deep learning.," in *MediaEval*, 2015.

[6] Vu Lam, Sang Phan Le, Duy-Dinh Le, Shin'ichi Satoh, and Duc Anh Duong, "Nii-uit at mediaeval 2015 affective impact of movies task.," in *MediaEval*, 2015.

[7] Ionut Mironica, Bogdan Ionescu, Mats Sjöberg, Markus Schedl, and Marcin Skowron, "Rfa at mediaeval 2015 affective impact of movies task: A multimodal approach.," in *MediaEval*, 2015.

[8] P Marin Vlastelica, Sergey Hayrapetyan, Makarand Tapaswi, and Rainer Stiefelhagen, "Kit at mediaeval 2015-evaluating visual cues for affective impact of movies task.," in *MediaEval*, 2015.

[9] Yun Yi, Hanli Wang, Bowen Zhang, and Jian Yu, "Mic-tju in mediaeval 2015 affective impact of movies task.," in *MediaEval*, 2015.

[10] Claire-Helene Demarty, Bogdan Ionescu, Yu-Gang Jiang, Vu Lam Quang, Markus Schedl, and Cedric Penet, "Benchmarking violent scenes detection in movies," in *Content-Based Multimedia Indexing (CBMI), International Workshop*. IEEE, 2014, pp. 1–6.

[11] Wen-Huang Cheng, Wei-Ta Chu, and Ja-Ling Wu, "Semantic context detection based on hierarchical audio models," in *in ACM SIGMM International workshop on Multimedia information retrieval*. ACM, 2003, pp. 109–115.

[12] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu, "3d convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2013.

[13] Bruno Malveira Peixoto, Sandra Avila, Zanoni Dias, and Anderson Rocha, "Breaking down violence: A deep-learning strategy to model and classify violence in videos," in *Proceedings of the International Conference on Availability, Reliability and Security*. ACM, 2018, p. 50.

[14] Fillipe Souza, Eduardo Valle, Guillermo Chávez, and Arnaldo de A Araújo, "Color-aware local spatiotemporal features for action recognition," in *Iberoamerican Congress on Pattern Recognition*. Springer, 2011, pp. 248–255.

[15] Ivan Laptev, "On space-time interest points," *International journal of computer vision*, vol. 64, no. 2-3, pp. 107–123, 2005.

[16] C Clarin, J Dionisio, M Echavez, and P Naval, "Dove: Detection of movie violence using motion intensity analysis on skin and blood," *Phillipine Comput. Sci. Congr.*, vol. 6, pp. 150–156, 2005.

[17] Tal Hassner, Yossi Itcher, and Orit Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2012, pp. 1–6.

[18] Qing Xia, Ping Zhang, JingJing Wang, Ming Tian, and Chun Fei, "Real time violence detection based on deep spatio-temporal features," in *Chinese Conference on Biometric Recognition*. Springer, 2018, pp. 157–165.

[19] Chunhui Ding, Shouke Fan, Ming Zhu, Weiguo Feng, and Baozhi Jia, "Violence detection in video by using 3d convolutional neural networks," in *International Symposium on Visual Computing*. Springer, 2014, pp. 551–558.

[20] Zhihong Dong, Jie Qin, and Yunhong Wang, "Multi-stream deep networks for person to person violence detection in videos," in *Chinese Conference on Pattern Recognition*. Springer, 2016, pp. 517–531.

[21] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *IEEE International Conference on Computer Vision*, 2015, pp. 4489–4497.

[22] Eleni Tsironi, Pablo Barros, Cornelius Weber, and Stefan Wermter, "An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition," *Neurocomputing*, vol. 268, pp. 76–86, 2017.