Analyze the Presence of Violence and a Particular Event of Violence by Weapon Detection using Deep Learning

Neeraj Patil[1], Deepak Pitamabare[2], Amruta Kumbhar[3], Prachi Sabale[4], Prof. N.S.Devekar[5]

[1,2,3,4]Students, Dept. of Computer Engineering, AISSMSCOE, Pune, India
[5]Professor, Dept. of Computer Engineering, AISSMSCOE, Pune, India

Abstract - Nowadays human security against violence is one of the major concern. Violence detection techniques analyze the surveillance camera videos. Over the last few years, these cameras and other surveillance equipment are installed at sensitive areas like ATMs, Government offices, Schools, Hospitals, etc. For the security against violence, there is a need of generating timely and automated alerts to concern officials to take further action.

The aim is to develop a system for monitoring and analyzing the video streams from surveillance cameras and making the decision about violence in real-time. Frame level features in a video are extracted by employing a CNN after which they are compounded by utilizing a variant of LSTM which in turn makes use of convolutional gates .CNN and LSTM are together used for the analysis of local motion in a video.

The Deep learning technique for violence detection is used to classify the violent recognition on the base of data set and extracted features using more convolutional layers. After assessing contemporary circumstances in the world, it is imperative that there is existential and paramount need of exploiting automated visual surveillance for detecting weapons, which will enhance effectiveness of security operations.

We aim for weapon detection with the help of Single Shot detector (SSD) which uses multi-scale convolutional feature and takes only one shot to detect multiple objects .Single shot detectors are potentially faster and precisely accurate and more applicable for object detection in videos.

Therefore with the help of video surveillance we can analyze the presence of violence and a particular event of violence by weapon detection using Deep learning techniques.

Key Words: Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), DataSet, Deep Learning, TensorFlow, SSD, Transfer Learning.

1. INTRODUCTION

Security is a major in every domain, due to rise in violent activities and weapons. Earlier surveillance systems were more dependent on human operator. Now, because of surveillance cameras installed at various public places like offices, hospitals, schools, highways, etc. it can be helpful for capturing useful actions and movements for event prediction and online monitoring. Having an automated system with the ability to recognize the occurrence of violence in videos with realtime response will enable authority holders to increase safety and take appropriate decisions. Violence is an abnormal behaviour and those actions can be identified through smart surveillance system using which we can prevent further fatal accidents.The fundamental goal is to collect categories and recognize the most prominent and effective methods or techniques that are used in violence and anomalous activity detection using deep learning approach. The aim is to develop an intelligent surveillance system which detects violence or weapons in given video frame using a deep supervised learning approach.

The model incorporates a pre-trained convolution Neural Network (CNN) connected to Convolutional LSTM layer. The model takes the raw video as an input which is subsequently converted into frames and output is a binary classification of violence or non-violence label. Therefore with the help of video surveillance we can analyze the presence of violence and a particular event of violence by weapon detection using Deep learning techniques.

2. Goals and Objectives

In recent years, violence has become a major issue across the globe. So, there is a need to detect violence automatically without human intervention and generate alerts timely to the control crew.

Developing a technique for the automatic analysis of surveillance videos in order to identify the presence of violence and weapon identification.

To develop a system having high accuracy, less false alerts and low computational cost in monitoring and analyzing the video streams from surveillance cameras and making the decision about violence in real-time.

3. Literature Survey

Various strategies have been proposed by researchers dealing with the problem of detection of violence from video surveillance. All the existing techniques can be divided into classes depending on the basic idea -

1. Inter-frame changes: Frames containing violence undergo massive variations because of fast motion due to fights.

2. Local motion in videos: The motion change patterns taking place in the video is analyzed.

3. Several other methods follow the techniques used in action recognition, i.e. to identify spatio-temporal interest points and extract features from these points.

“Violence Detection Using Spatiotemporal features” Fath U Min Ullah, Amin Ullah, Khan Muhammad, Ijaz Ul Haq and Sung Wook Baik.The three-tiered structure is forced into this program. Detection of a person using CNN performed in the first phase, in the second phase, Frame sequence provided by 3D CNN training model again in the third phase transferred to SoftMax separator. With comparative analysis again final prediction made, slide window works better as compared to SVM. The OPENVINO toolkit was used and model modeling and growth system performance.

Professor Ali Khaleghi and Prof. Mohammad Shahram Moin ku in their paper " Improved Anomaly Detection in Surveillance Videos Based on A Deep Learning Method” developed a program that finds normal and unusual video. The first data preparation step, input video separated by frame and the next pre-processing step removes the background. The removal process is done manually or automation that creates a behavioral structure of the data that modeling and feature detection is available. Items are later obtained using CNN and the final decision was made in two based classifier.

The hand-crafted feature based techniques used methods like bag of words, histogram, improved Fisher encoding, etc. for aggregating the features across the frames. Recently various models using long short term memory(LSTM) RNNs are developed for addressing problems involving sequences like MT, speech recognition, caption generation and video action recognition . The LSTM was introduced in 1997 to combat the effect of vanishing gradient problem which was plaguing the deep learning community. The LSTM incorporates a memory unit which contains in-formation about the inputs the LSTM unit has seen and is regulated employing a number of fully-connected gates

Dong et al proposed the thought of using LSTM for feature aggregation for violence detection.The method is based on LSTM encoding and late fusion which consists of extracting features employing a convolutional neural network from raw pixels, optical flow images and acceleration flow maps.

Recently, Xingjian et al. replaced the fully-connected gate layers of the LSTM with convolutional layers and used this improved model for predicting precipitation now casting from radar images with improved performance. This newer model of the LSTM is called as convolutional LSTM (convLSTM). Later, it’s been used for predicting optical flow images from videos and for anomaly detection in videos. By replacing the fully-connected layers within the LSTM with convolutional layers, the convLSTM model is capable of encoding spatio-temporal information in its memory cell.

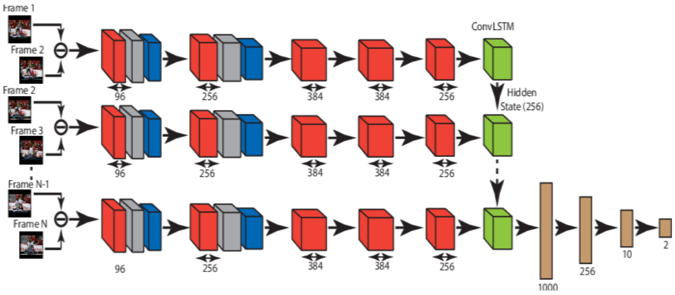
Prof. Prakhar Singh and Prof. Vinod Pankajakshan Introduced a method using standard features removed from the inclusion video in their paper "An In-Depth Learning Approach Based on Unwanted Visibility in Watching Videos". The Convolutional Neural Network (CNN) stack is used to extract a feature from the video input sequence frames. The Convolutional Long Short-Term Memory (convLSTM) stack is then used to predict future sequences and later the CNN transmission stack is used to predict future video sequences. Combined error is compared to the limit and the category is determined.

Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg presented way for detecting objects in images by employing a single deep neural network. Their approach, named SSD(Single Shot Detector), discretizes the output space of bounding boxes into collection of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of every object category in each default box and produces adjustments to the box to raised match shape. Additionally, the network combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes.The SSD model is straightforward relative to methods that needs object proposals because it completely eliminates proposal generation and subsequent pixel or feature resampling stage and encapsulates all computation in a very single network. This makes SSD easy to train,to integrate into systems that needs a detection component.

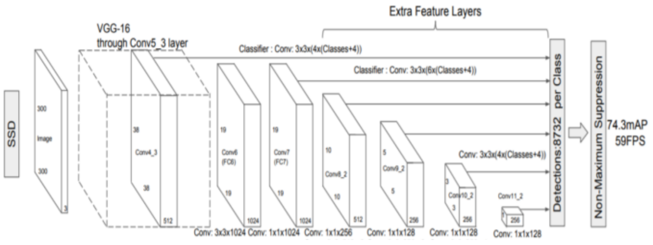
4. PROPOSED SYSTEM

The aim is to develop an intelligent surveillance system which detects violence or weapons in given video frame using a deep supervised learning approach. The fundamental goal of the proposed study is to develop an end-to-end trainable deep neural network model for classifying videos in to violent and non-violent ones.

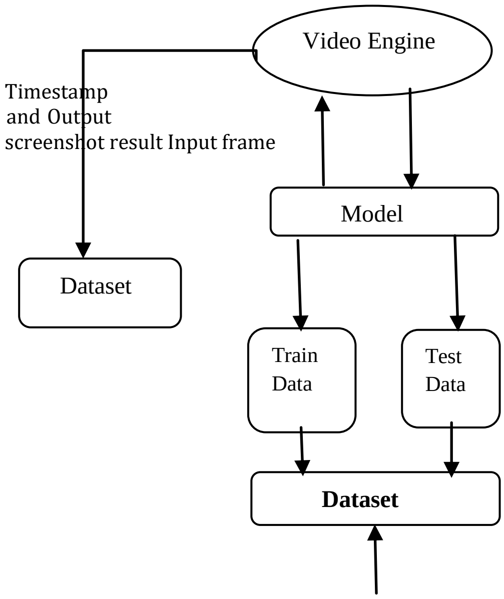
The network consists of a series of convolutional layers followed by max pooling operations for extracting discriminant features and convolutional long short memory (convLSTM) for encoding the frame level changes, that characterizes violent scenes existing in the video.

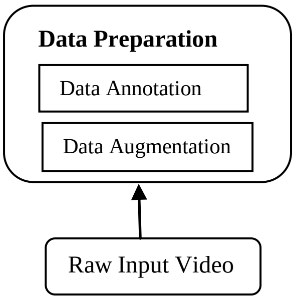


We aim for weapon detection with the help of Single Shot detector (SSD) which uses multi-scale convolutional feature and takes only one shot to detect multiple objects.It is significantly faster in speed and high-accuracy object detection algorithm.Instead of using sliding window, SSD divides the image using a grid and have each grid cell be responsible for detecting objects in that region of the image.



5. SYSTEM ARCHITECTURE





System architecture consists of 5 modules namely Data preparation module which includes a data annotation and data augmentation , dataset, Deep Leaning model, video engine and database. This system is implemented in python and TensorFlow as a backend. User gives video file as an input and system gives output as video classification as violent or nonviolent. System supports .mp4 and video formats.

Modules: This system is divided into five parts according to functions performed by individuals.

Data Preparation: Data preparation is the act of manipulating raw data into a form that can readily and accurately be analysed, This module will be dealing with raw video data. It consists of two submodules Data Augmentation and Data Annotation.

Data Augmentation: It is a method of augmenting the available data. Data augmentation is a strategy that enables practitioners to significantly increase the diversity of data available for training models, without collecting new data. Data augmentation techniques such as cropping, padding, and horizontal flipping are commonly used to train large neural networks Main purpose of augmentation is to increase the size of available dataset.

Data Annotation: Data annotation basically is the process of adding metadata to a dataset .Since the system is based on supervised learning, annotation is an important module which labels the data. Data annotators helps us to categorize things. They can work with things like videos, advertisements, photographs and other types of material. They assess the content and then attach tags to the content.

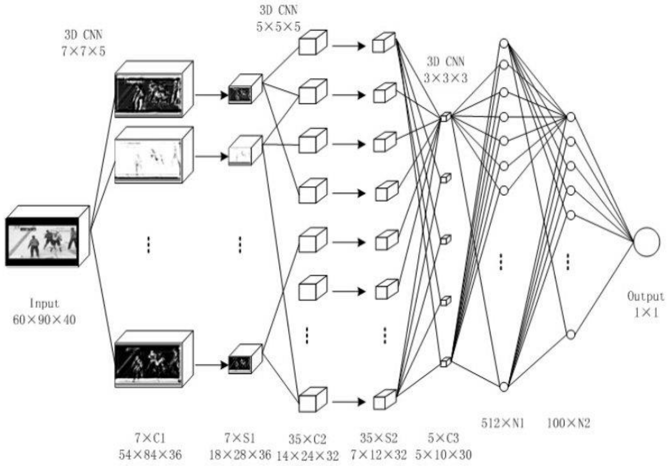
Dataset: Dataset consist of data prepared by data preparation module. Dataset is further split into training and testing.

Deep Learning Model: Deep learning models are built using neural networks. A neural network takes in inputs, which are then processed in hidden layers using weights that are adjusted during training. Then the model spits out a prediction. The weights are adjusted to find patterns in order to make better predictions. This is the deep learning model trained using input dataset. This model will be invoked by video engine and model will classify input as violent or non-violent.

Video Engine: This video engine is an interface between user and deep learning model. The video engine will take the input from user and will pass it through the DL model.

Database: This database contains timestamp and screenshot of already trained activities identified by system.

Basic Overview :



6. FUTURE SCOPE

The planned system solely detects suspicious human behaviour and particularly presence of guns.

As present and future work, we are evaluating reducing the number of false positives, by preprocessing the videos, i.e. increasing their contrast and luminosity, and also by enriching the training set with weapons in motion. Detection of fire and different weapons will be enforced in future.

Real-time autonomous drone surveillance system can be developed to identify violent individuals in public areas.

7. CONCLUSIONS

Thus, we have proposed an efficient framework which can detect violence in sensitive areas and detect weapons by analyzing the video streams collected from the surveillance cameras. The proposed model consists of a convolutional neural network (CNN) for frame level feature extraction followed by feature aggregation in the temporal domain using convolutional long short term memory(convLSTM).

By combining CNN with LSTM, the accuracy increases to a certain margin as compared to pure transfer learning models. For weapon detection, our system uses Single Shot Detector (SSD) which is precisely more accurate than the traditonal method.

Therefore with the help of video surveillance we can analyze the presence of violence and a particular event of violence by weapon detection using Deep learning techniques.

8. REFERENCES

[1] Swathikiran Sudhakaran, Oswald Lanz,” Learning to Detect Violent Videos using Convolutional Long ShortTerm Memory,” 978-1-5386-2939-0/1720 IEEE 2017

[2] Prakhar Singh, Vinod Pankajakshan, “A Deep Learning Based Technique for Anomaly Detection in Surveillance Videos, “Twenty Fourth National Conference on Communications 2018.

[3] S.M. Rojin Ammar Md. Tanvir Rounak Anjum Md. Touhidul Islam,“Using Deep Learning Algorithms to Detect Violent Activities”.

[4] Lyu, Y., Yang, Y., “Violence detection algorithm based on local spatio-temporal features and optical flow,”. 2015 International Conference on Industrial InformaticsComputing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), pp. 307– 311. IEEE, December 2015.

[5] Wei Liu et al., “SSD: Single Shot MultiBox Detector”, European Conference on Computer Vision, Volume 169, pp 20-31 Sep. 2017.

[6] Jain, H., Vikram, A., Mohana, Kashyap, A., & Jain, A. (2020). Weapon Detection using Artificial Intelligence and Deep Learning for Security Applications. 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC). doi:10.1109/icesc48915.2020.9155832.

[7] https://ieeexplore.ieee.org/document/9014714
https://arxiv.org/pdf/1702.05147.pdf

[8] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong and W.-c. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting 2015.

[9] Ali Khaleghiand Mohammad Shahram Moin, “Improved Anomaly Detection in Surveillance Videos Based on a Deep Learning Method, “978-1-5386-5706-5/18 IEEE 2018.

[10] www.tensorflow.org/tutorials/images/transfer_learning.

[11] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Dar-rell. Long-term recurrent convolutional networks for visual recognition and description 2015.

[12] Lyu, Y., Yang, Y., "Violence detection algorithm based on local spatio-temporal features and optical flow,". 2015 International Conference on Industrial InformaticsComputing Technology, Intelligent Technology, Industrial Information Integration (ICIICII), pp. 307– 311. IEEE, December 2015

[13] Ruben J Franklin et.al., "Anomaly Detection in Videos for Video Surveillance Applications Using Neural Networks," International Conference on Inventive Systems and Control,2020.