International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 12 | Dec 2019 www.irjet.net p-ISSN: 2395-0072

Review of Violence Detection System using Deep Learning

Vishwajit Dandage[1], Hiemanshu Gautam[2], Akshay Ghavale[3], Radhika Mahore[4],
Prof. P.A. Sonewar[5]

[1,2,3,4]*Student, Dept. of Computer Engineering, SKNCOE, Pune, India*
[5]*Professor, Dept. of Computer Engineering, SKNCOE, Pune, India*

---***---

Abstract - *Nowadays, tremendous growth is observed in research of surveillance system. The surveillance cameras installed at various public places like offices, hospitals, schools, highways, etc. can be used to capture useful actions and movements for event prediction, online monitoring, goal driven analysis and also in intrusion detection. The goal of this article is to develop an independent system for automatic analysis to detect presence of violence and non-violence using Deep Learning techniques. We have put forward a deep neural network for significant detection accuracy. A Convolutional Neural Network (CNN) is used to extract frame level features from a video which are then aggregated using a variant of Long Short-Term Memory (LSTM) that uses convolutional gates. CNN and LSTM are together used for the analysis of local motion in a video. The overall performance of proposed model is calculated in terms of accuracy on three datasets.*

***Key Words*: CNN, LSTM, RNN, Inception, GoogLeNet, VGG, AlexNet, Data-Set, Deep Learning, TensorFlow**

1. INTRODUCTION

Over the previous few decades, outstanding infrastructure growths are noticed in security-related problems throughout the world. So, with multiplied demand for security, video-based surveillance has become a crucial space for the analysis. An intelligent video surveillance system primarily censored the performance, happenings, or ever-changing information typically in terms of human beings, vehicles or the other objects from a distance by means of some equipment (usually digital camera). The scopes like prevention, detection and intervention that have led to the development of real and consistent video surveillance systems are capable of intelligent video processing competencies.

Earlier surveillance systems were more dependent on human operator. Now a days because of better efficiencies and reliability, automated systems are preferred and in accordance of security it is very helpful to detect violence. Violence is an abnormal behavior and those actions can be

Researchers have taken CNN and 3D CNN into consideration. They proposed CNN model for detection of person in video and by using CNN processing time is reduced. After that images with persons are fed to 3D CNN model which was trained on spatiotemporal features and final predictions were made. Furthermore, they are planning to propose edge

identified through smart surveillance system using we can prevent further fatal accidents. On large scale this system can be implemented in several locations like streets, parks, medical centers and alert authorities about the violent activities.

2. LITERATURE SURVEY

Prof. Ali Khaleghi and Prof. Mohammad Shahram Moin in their paper “Improved Anomaly Detection in Surveillance Videos Based on A Deep Learning Method” has developed a system that detects normal and abnormal video. In the first data preparation step, the input video is divided into frames and the next pre-processing step removes the background. The feature extraction step is performed manually or automatic that forms the behavioral structure of the data that is modelled and the feature representation is obtained. Later the objects are detected using CNN and the final decision is made by a two-class based classifier.

Mr. Antreas Antoniou has considered two approaches for surveillance system, detection approach and classification approach. Detection approach consist of background subtraction and optical flow, while classification approach is based on Neural Networks. He has proposed two custom architectures for intelligent surveillance system. Custom architecture 1 is inspired from VGGNet and AlexNet. In that fully connected layers are reduced to 1024 * 1024 *7 from 4096 * 4096 * 1000.

Custom Net 2 is also known as ParaNet. It is inspired from GoogLeNet and VGGNet. It has separate pipeline for each class.

Swathikiran Sudhakaran and Oswald Lanz presents trainable deep neural network model. The proposed model consists of a convolutional neural network for frame level feature extraction and convolutional long short-term memory for feature aggregation in the temporal domain. The model uses three different datasets which results in improved performance.

intelligence for violence recognition work in IOT using smart devices for quick responses.

Prof. Prakhar Singh and Prof. Vinod Pankajakshan has presented an approach to detect anomalies using general features extracted from the input video in their paper

"A Deep Learning Based Technique for Anomaly Detection in Surveillance Videos". A Convolutional Neural Network (CNN) stack is utilized to extract feature from the video frames of input sequence. Then a Convolutional Long Short-Term Memory (convLSTM) stack is used to predict future motion sequence and later the stack of transpose CNN is used to predict the future video sequence. The computed error is then compared with the threshold and the class is determined.

Table 1: Previous Survey

Sr No.	TITLE	METHODOLOGY	OBSERVATIONS
1	"Improved Anomaly Detection in Surveillance Videos Based on a Deep Learning Method ", by Ali Khaleghi and Mohammad Shahram Moin [1]	The proposed methodology of this paper consists of three main components to detect normal and abnormal incidents. The first component is pre-processing which is used to estimate and remove the background that eliminates computing cost and time. The second component is feature extraction and learning component. This component obtains three features as appearance, density and motion features. In the last detection component, the trained data is given to classifiers to determine the class. It includes simple deep classifier, simple linear classifier and Auto-Encoders to determine the final decision.	In this paper a new methodology of deep learning is used to detect anomaly in video surveillance. To evaluate the system, the famous UCSD dataset is used for anomaly detection. Moreover, this model implements training phase independently so that it can also be used as a pre-trained network in other implementations.
2	"A General-Purpose Intelligent Surveillance System for Mobile Devices using Deep Learning", Antreas Antoniou Plamen Angelov [2]	CustomNet 1: Activation using PRelu activation function. Only 7 types of classes are considered ParaNet: Parallelism of GoogLeNet, Smaller receptive fields of AlexNet. Idea of Multi Task Learning	CustomNet 1 had a very fast training of only 3 hours compared to GoogLeNet's 10 hours. In addition, the GoogLeNet trained used PReLU instead of ReLU. Real time detection can be implemented using combination of supervised and unsupervised learning using powerful GPUs such as Nvidia Tegra X1
3	"Learning to Detect Violent Videos using Convolutional Long Short-Term Memory" Swathikiran Sudhakaran and Oswald Lanz University of Trento, Trento, Italy Fondazione Bruno Kessler, Trento, Italy [3]	convolutional neural network: frame level feature extraction. convolutional long short-term memory: feature aggregation in the temporal domain.	This paper shows that a network trained to model changes in frames performs better than a network trained using frames as inputs and comparative study between the traditional fully-connected LSTM and convLSTM.
4	"Violence Detection Using Spatiotemporal features with 3D Convolutional Neural Network." Fath U Min Ullah, Amin Ullah, Khan Muhammad, Ijaz Ul Haq and Sung Wook Baik [4]	Three-staged framework enforced in this system. Person detection using CNN is done in 1st stage, in second stage, frame sequence provided to 3D CNN model for training and in third stage it is forwarded to SoftMax classifier.	By comparative analysis and final prediction made, sliding window performs better as compared to SVM. OPENVINO toolkit is used to optimize model and to increase performance of system
5	"A Deep Learning Based Technique for Anomaly Detection in Surveillance	1.CNN based feature extraction- 2 convolutional layers are used for extracting high level features like	In this paper OpenCV, Keras and Theano in Python is used. The performance of the system was

	Videos", by Prakhar Singh and Vinod Pankajakshan [5]	objects, people and the input dimensionality is downsized to reduce subsequent layer complexity. 2.Temporal feature extraction-ConvLSTM and 3D-CNN is utilized. 3.Prediction of video frames-Transposed CNN (deconvolution) is used to reconstruct an image from the features that were extracted in the previous steps. This step has 2 layers to compute output sequence, one does up sampling by factor 3*3 and the other does up sampling by factor 2*2.	evaluated based on 2 datasets UCSD Ped1 and UCSD Ped2. This approach does on depend on any handcrafted features.

4. PROPOSED SYSTEM

The aim of our proposed system is to develop an intelligent surveillance system which detects violence in given video frame. The intelligent surveillance system first learns features and then trains on those learned features. It detects violence in given video and if violence is detected in frames, it will send alert to respective authorities and store detected frame in local database. The input for the system will be a video frame and output will be in binary form that is either violent frame or non-violent frame. The system is created in such a way that it helps the user by determining whether the crime occurs or not in a short video sequence. This system can help Government Agencies to response faster. In our system, we are using TensorFlow GPU libraries to use GPU along with CPU to build the system that is capable of exploiting parallelization for fast processing. Detecting the presence of violence using a high population, high dimensional dataset is challenging due to limited dataset. Faster RCNN algorithm is found to be able to identify person in given video and then LSTM classification identifies the violence detection in given video frame and send alert if violence is detected.

5. SYSTEM ARCHITECTURE

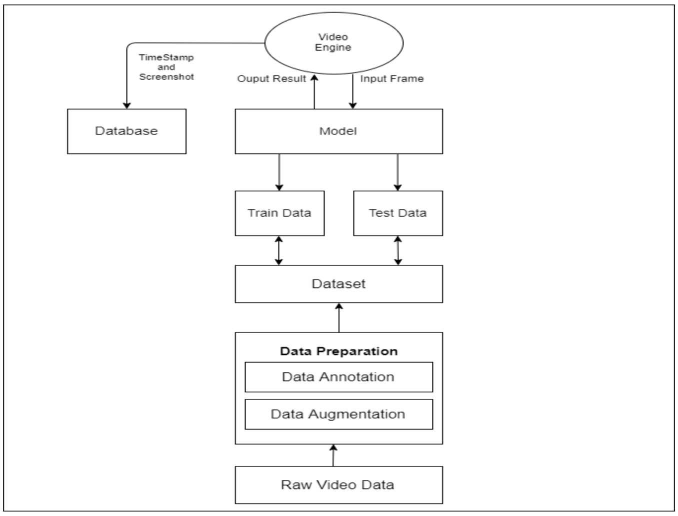

Fig 1: System Architecture

System architecture consist of 5 modules namely Data preparation module, 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 is violent or non-violent. System supports .mp4 and .avi video formats.

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

Data Preparation: This module deals 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. Main purpose of augmentation is to increase the size of available dataset.

Data Annotation: System is based of supervised learning so annotation is an important module which labels the data.

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

Deep Learning Model: 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 module 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. It has feature of alerting govt authorities if any suspicious activity is detected.

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

6. FUTURE SCOPE

The planned system solely detects suspicious human behavior and presence of guns. Detection of fire and different weapons will be enforced in future. Cloud readying can be done.

7. CONCLUSIONS

This paper highlights that a convolutional neural network which leverages transfer learning with long short-term memory networks outperforms all the other variance of convolutional neural networks.

By combining CNN with LSTM, the accuracy increases to a certain margin as compared to pure transfer learning models. The system provides a simple graphical user interface to interact with deep learning model.

REFERENCES

[1] 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.

[2] Antreas Antoniou Plamen Angelov, “A General-Purpose Intelligent Surveillance System for Mobile Devices using Deep Learning, “International Joint Conference on Neural Networks (IJCNN) 2016.

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

[4] Fath U Min Ullah, Amin Ullah, Khan Muhammad, Ijaz Ul Haq and Sung Wook Baik, “Violence Detection Using Spatiotemporal Features with 3D Convolutional Neural Network, “Sensors, 19, 2472; doi:10.3390/s19112472 2019.

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

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

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

[8] C. Piciarelli, C. Micheloni, and G. L. Foresti, “Trajectory-based anomalous event detection,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 18, no. 11, pp. 1544–1554, Nov 2008.

[9] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, “Anomaly detection in crowded scenes,” in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 2010, pp. 1975–1981.

[10] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, Neural computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] P. Bilinski and F.Bremond. Human violence recognition and detection in surveillance videos. In AVSS, 2016.

[12] D. Chen, H. Wactlar, M.-y. Chen, C. Gao, A. Bharucha, and A. Hauptmann. Recognition of aggressive human behavior using binary local motion descriptors. In International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), 2008.

[13] http://joshua-p-r-pan.blogspot.com/2018/05/violence-detection-by-cnn-lstm.html

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

[15] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6479846/