Grenze International Journal of Engineering and Technology, Jan Issue

A Systematic Review of Abnormal Behavior Detection: Convolution Neural Networks and Long Short Term Memory Models

Kande Archana[1] and Dr. Kamakshi Prasad[2]
[1]Research Scholar @ JNTUH College of Engineering Hyderabad
Email: kande.archana@gmail.com
[2]Professor & BOS Chairman @ JNTUH College of Engineering Hyderabad
EmailId: kamakshiprasad@jntuh.ac.in

***Abstract*—Abnormal behavior detection is interesting topic computer vision and this is required in surveillance system. Various researches were carried out in abnormal behavior detection using machine learning methods. In this paper, the recent papers in abnormal behavior detection is reviewer with advantages and limitations. Many researches in abnormal behavior detection is based on the Convolutional Neural Network (CNN) for feature extraction and classification. Some of the researchers were involves in applying Long Short Term Memory (LSTM) model for the abnormal behavior detection. Autoencoder models were applied for spatiotemporal analysis of abnormal behavior detection. The CNN models have limitations of overfitting problem and imbalance data problem. The LSTM models have vanishing gradient problem that affects the performance of the model. The hybrid model of CNN and LSTM have high computational complexity and requires high resources for computation. The review shows that VGG-19 model have higher performance in abnormal behavior classification compared to other CNN models. An efficient model is required for abnormal behavior detection to overcome overfitting and vanishing gradient problem.**

***Index Terms*— Abnormal behavior detection, Autoencoder, Convolutional Neural Network, Hybrid model, and Long Short Term Memory.**

I. INTRODUCTION

In traditional surveillance systems, human operators are assigned to monitor cameras in public areas to identify abnormal events from monitor screens. Increases in number of surveillance equipment, the demand for accurate and automated abnormal behavior detection is increases since manual monitoring is very inefficient [1]. Abnormal behavior detection in video sequence is difficult task due to the features analysis between normal and abnormal, and also dependence of context scenario [2]. Various fields like traffic monitoring, medical monitoring, and security guards were uses video surveillance. Anomaly detection plays important role among these research fields to discovering various irregularities [3]. So, intelligent surveillance system is highly required to detect the abnormality and alert the human only at the time of abnormality in recorded video. Anomaly detection and event recognition are two paths of intelligent video analysis in research area [4]. Video anomaly detection has capable to process in crowded scenes and uncrowded scenes to detect abnormal.

Grenze ID: 01.GIJET.8.1.520

Abnormal event identification in crowded scenes has additional challenge due to occlusions [5].
Feature extraction, abnormal detection and abnormal localization are three common steps in video-based abnormal behavior detection. Recently, deep learning shows good performance in many areas of computer vision domain [6]. Most of existing abnormal detection algorithm have high computational complexity and has lower performance in real time system [7]. An efficient computer vision method is required to identify anomalous activity to classify normal and abnormal events effectively without human interference. Such a model is significantly helpful in monitoring and also reduce the human effort to manually maintain observation on a 24-hour basis [8]. CNN based models have been applied for the detection of abnormality. This has the limitations of overfitting, high reconstruction error and imbalance data problem [9, 10]. In this paper, the CNN and LSTM based models were reviewed with advantages and limitations in abnormal behaviour detection.This paper is organized as taxonomy of abnormal behaviour detection is given in section 2, comparative analysis of various models are given in section 3, problem statement is given in section 4 and conclusion is given in section 5.

II. TAXONOMY OF ABNORMAL BEHAVIOUR DETECTION

Automatic abnormal behaviour detection is important process for surveillance system and helps in effective monitoring for public safety. Most of abnormal behaviour detection method is based on CNN based models due to its efficiency in feature analysis. Some of the models are based on LSTM due to handling of given features. The dictionary learning is applied with CNN models to improve the learning performance. The taxonomy of the abnormal behaviour detection models were given in Figure 1.

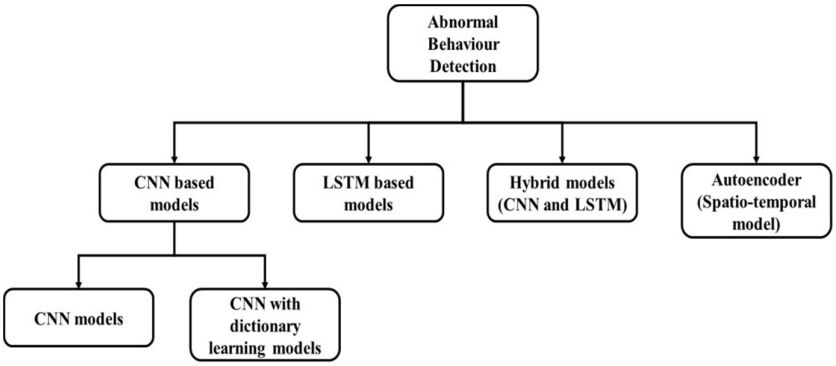

Figure 1. The taxonomy of abnormal behaviour detection models

The CNN model [11], LSTM model [12], hybrid model [13], The CNN with dictionary learning [14], and Autoencoder model [15] were explained with advantages and limitations in this section.
Dhamari, *et al*. [11] applied CNN model of Visual Geometry Group – 19 (VGG-19) to extract descriptive features. Binary SVM model is constructed based on the extracted features for the classification. Three parameters such as equal error rate, area under the curve and accuracy were used to test the performance of the model. University of Minnesota (UMN) and University of California, San Diego Pedistrain1 (UCSD-PED1) datasets were used to test the performance of the developed VGG-19 model. The VGG-19 hand crafted features were compared with other CNN models for abnormal behaviour detection. The VGG-19 model has higher performance compared to existing CNN models of VGG-16, AlexNet, ResNet50, and GoogleNet. The VGG-19 model suffers from the limitations of overfitting problem and imbalance data problem.
Ullah, *et al*. [12] extracted spatial CNN features from the dataset and applied to the attention based LSTM model for abnormal behaviour detection. The residual block of LSTM model with CNN features provides effective learning methods. The Avenue, UMN and UCF-Crime datasets were used to test the performance of the developed method. The developed method has higher performance in abnormal behaviour detection than existing methods. The developed method suffers from the limitation of overfitting and vanishing gradient problem.
Wan, *et al*. [13] applied hybrid method of LSTM and CNN model for public safety of intelligent transportation system. The superframe segmentation is applied to reduce the redundant features in the dataset. The developed method has higher performance in the detection process compared to existing methods. The SumMe and Hollywood2 datasets were used to test the performance of the developed method in detection. The feature reduction process helps to reduce the overfitting problem in hybrid method. The computational complexity of the hybrid method is high and requires more resources for classification.

Chu, *et al*. [14] applied 3D CNN model with dictionary learning to analysis the spatio-temporal information in abnormal behaviour detection. The unsupervised feature learning in this model is guided by the handcrafted features of sparse coding results. The statistical information is used to define multi-level similarity relationship between inputs. The multi-level similarity structure of quadruplet concept is used to construct a generalized triplet loss for training process. The robust and rich feature representation is obtained based on joint optimization between unsupervised feature learning and sparse coding. The Avenue, subway, and UCSD datasets were used to test the performance of 3D CNN with dictionary learning method. The developed method has higher performance in abnormal behaviour detection compared to existing methods. The developed method suffers from the limitations of overfitting problem due to redundant features in classification.
Chandrakala, *et al*. [15] applied residual spatiotemporal autoencoder with skip connection for abnormal behaviour detection. The skip connection in memory guided network helps to avoid the loss of meaningful information such as foreground patterns with memorizing significant normality patterns. The CUHK-Avenue and LV datasets were used to test the performance of the developed method. The developed method has higher performance in abnormal behaviour detection compared to existing methods. The developed method suffers from the limitation of vanishing gradient problem that affects the performance of the model.

III. Datasets used in Abnormal Behaviour Detection

This section provides the details of datasets used in abnormal behaviour detection researches.**USCD:** USCD [1, 2, 3, 4, 6, 7, 9, 10] is common dataset followed to test the performance of abnormal behavior detection. The USCD dataset consists of two subsets USCD ped 1 and USCD ped 2. The pixel-level ground truth is presented in the dataset to evaluate the performance of the models. The USCD ped 1 dataset consists of 34 short clip for training and 36 short clip for testing. Each clip has 200 frames and frame size is 158 × 238 pixel. In USCD ped 2, 16 short clips are present for training and 12 short clips are present for testing. Each clip has 150 to 180 frames in 240 × 360 resolution.**CUHK Avenue:** CUHK Avenue [1, 6, 10] is second commonly used dataset to test the performance of abnormal behavior detection. The 16 training clips and 21 testing clips are present in this dataset in 360 × 640 resolution. Apart from these two datasets, PETS 2009 [1], UMN [7], subway [3], and coal mining [2] datasets were used to test the performance of the model.

IV. Comparative Analysis

In pattern recognition and computer vision, abnormal event detection is important topic. The abnormal event detection is required in surveillance to increases the security and public safety. This section provides the details of recent methods in abnormal event detection with its limitations. The CNN and LSTM based models were analyzed with advantages and limitations in Table I.

TABLE I. The Comparative Analysis of CNN and LSTM Models in Abnormal Behaviour Classification

Author(s)	Methods	Advantages	Limitations	Results
Li, *etal*. [1] (2020)	Low rank and compact coefficient dictionary learning method is proposed for abnormal event detection. Joint optimization method is applied to obtain compact coefficient low-rank dictionary. A reconstruction cost is applied to detect abnormal frames in the model. The PETS 2009, UCSD and CUHK Avenue datasets was used to evaluate the performance of this model.	The developed method requires the less number of training data compared to existing deep learning methods. The efficiency of the abnormal event detection is high.	Computational time of the developed method is high. Detection is required to be tested on various scenes.	The developed method has 88.68 % AUC and existing Conv-WTA+SVM method has 82.1 % AUC.
Ye,*etal*. [2](2020)	The subgraph calibration method based on hybrid modulation is applied for the abnormal event classification. The CNN and LSTM model is applied to extract spatio temporal features of video frames. The SVM is applied for expectation	The CNN and LSTM model provides effective feature selection performance. The classification of feature expectations subgraph improves	The CNN and LSTM model creates overfitting problem in the detection process. The developed method has lower performance complex video surveillance due to low efficiency of graph kernel model.	The developed method has accuracy of 99.55 % and existing 2D CNN-LSTM-SVM has 99.01 % accuracy.

	subgraph for abnormal event detection. The UCSDped1 and a coal mining video datasets were used to test the developed method performance.	the performance of the classification.		
Zhou, *et al*. [3] (2019)	Hybrid autoencoder of LSTM encoder-decoder is applied for spatio-temporal context for abnormal event detection. The ConvAE is applied in the first stage and LSTM with Backpropagation through time (BPTT) is applied in second stage for anomaly detection. The UCSD Ped, Avenue and Subway datasets were used to test the performance of the developed method.	The developed method has higher performance in abnormal event detection compared to existing methods.	The overfitting problem in the developed method affects the performance of the classification.	The hybrid model has 98.4 % AUC and the Pure model has 98.1 % AUC.
Ramchandran and Sangaiah, *et al*. [4] (2019)	The deep learning framework of convolutional autoencoder-ConvLSTM model to improve the performance of abnormal event detection. The Avenue, UCSD ped1 and UCSD ped2 datasets were used to test the performance of the classification.	The developed method has higher performance in abnormal event detection compared to existing methods.	The developed method has lower efficiency in large dataset.	The developed method has 90.7 % accuracy in avenue dataset, 98.4 % accuracy in UCSD ped1 dataset and 98.5 % accuracy in UCSD ped2 dataset.
Feng, *et al*. [5] (2018)	Applied attention guided LSTM model for the spatial and temporal location of fall events. YOLO v3 model is applied to detect the pedestrians in the video. The Fall event detection dataset was created to test the performance of developed method.	The CNN-LSTM model provides the effective performance in the classification of the fall event.	The overfitting problem is created in the model due to many features are extracted from the dataset.	The developed method has average precision of 94.8 % and the existing method has 93.5 % average precision.
Xia and Li, [6] (2020)	The deep learning framework is applied in this model to learn the historical features and current features of normal video. In testing phase, the predicted features differ from the actual features are considered as normal. The VGG16 network of CNN and third pooling layer output is used to effectively capture appearance features. The UCSD Ped2 and CUHK Avenue datasets were used to evaluate the performance of developed method.	The developed model has good performance in the classification.	The VGG16 model requires high computational power for the classification.	The developed method has AUC of 87.3 % and existing U-Net has 95.4 % AUC.
Fan, *et al.* [7] (2020)	The spatio temporal CNN model with simple structure and low computational complexity is applied for abnormal event classification. Supervised training model is applied to improve the performance of classification. The UCSD and UMN datasets were used to test the performance of developed method.	The developed method has capable of compute in CPU without require GPU. The cost of the model is low and provides the effective performance in abnormal classification.	The developed method has the limitation of overfitting problem that affects the performance of the detection.	The proposed method has 97 % AUC and existing method has 78.1 % AUC.
Ullah,*etal*. [8]	The pre-trained CNN model is applied	The developed	The developed method has	The developed

(2021)	with Bi-LSTM model for feature extraction and classification. The UCF-Crime and UCFCrime2Local datasets were used to test the performance of the model.	method has considerable performance in abnormal event detection. The CNN has effective feature extraction and LSTM model has higher performance in classification.	lower performance for low disparity.	method has 89.05 % AUC and existing method has 80.96 % AUC.
Tang, *et al*. [9] (2019)	An end-to-end framework is developed to conduct future frame prediction and reconstruction sequentially. Two U-Net model were used in the generator model for the detection. The USCD, CUHK Avenue and ShanghaiTech datasets were used to test the performance of the model.	The developed method has higher performance compared to existing methods in abnormal event detection.	The overfitting problem in U-Net model affects the performance of the detection.	The proposed method has 93.6 % AUC and existing method has 93 % AUC.
Deepak, *et al*. [10] (2020)	Applied residual spation temporal autoencoder for anomaly detection. The reconstruction loss is measured to detect the irregularities. The CUHK-Avenue, UCSD-Ped2, and Live Videos datasets were used to test the performance of developed method.	The spatio temporal autoencoder has higher performance compared to existing method.	The developed method has lower performance in large dataset.	The developed method has 83 % AUC and existing method 78 % AUC.

V. Problem Statement

Abnormal behavior detection is interesting topic in computer vision and required in surveillance system. From the review of recent methods in abnormal behavior detection, some limitations have been discussed in this section.

1. Most of the researches in abnormal behavior detection involves in applying the CNN models due to its high efficiency in feature analysis. The CNN models are suffering from the limitations of overfitting and data imbalance problem.
2. Some of the researches involves in applying the LSTM model to overcome the limitations of CNN models due to its efficient feature handling. The LSTM models have limitations of vanishing gradient problem and low efficiency in feature selection.
3. The CNN with dictionary learning have been applied by some researchers to improve the efficiency of the learning process. The dictionary learning method further increases the overfitting in the model.
4. Some of the researchers involves in applying hybrid models of CNN and LSTM to improve the feature learning and overcome overfitting problem. The hybrid models have high computational complexity and requires more resources for computation.

VI. Conclusion

Deep learning methods plays an important role in the abnormal behavior detection for public safety. CNN based models were highly applied in abnormal behavior detection due to its efficiency in feature analysis. In this paper, CNN and LSTM based models in abnormal behavior detection were reviewed with advantages and its limitations. The VGG-19 model have higher efficiency in abnormal behavior detection compared to other CNN models such as VGG-16, AlexNet and GoogleNet. The CNN based models have the limitations of overfitting and imbalance data problem in classification. The LSTM based models have vanishing gradient problem that affects the performance of the model. The hybrid model of CNN and LSTM have high computational complexity and requires high resource for computation. An efficient model is required to improve the performance of abnormal behavior detection and overcome the overfitting problem.

REFERENCES

[1] Li, A., Miao, Z., Cen, Y., Zhang, X.P., Zhang, L. and Chen, S., 2020. Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning. Pattern Recognition, 108, p.107355.

[2] Ye, O., Deng, J., Yu, Z., Liu, T. and Dong, L., 2020. Abnormal event detection via feature expectation subgraph calibrating classification in video surveillance scenes. IEEE Access, 8, pp.97564-97575.

[3] Zhou, F., Wang, L., Li, Z., Zuo, W. and Tan, H., 2020. Unsupervised learning approach for abnormal event detection in surveillance video by hybrid autoencoder. Neural Processing Letters, 52(2), pp.961-975.

[4] Ramchandran, A. and Sangaiah, A.K., 2020. Unsupervised deep learning system for local anomaly event detection in crowded scenes. Multimedia Tools and Applications, 79(47), pp.35275-35295.

[5] Feng, Q., Gao, C., Wang, L., Zhao, Y., Song, T. and Li, Q., 2020. Spatio-temporal fall event detection in complex scenes using attention guided LSTM. Pattern Recognition Letters, 130, pp.242-249.

[6] Xia, L. and Li, Z., 2021. A new method of abnormal behavior detection using LSTM network with temporal attention mechanism. The Journal of Supercomputing, 77(4), pp.3223-3241.

[7] Fan, Z., Yin, J., Song, Y. and Liu, Z., 2020. Real-time and accurate abnormal behavior detection in videos. Machine Vision and Applications, 31(7), pp.1-13.

[8] Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M. and Baik, S.W., 2021. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. Multimedia Tools and Applications, 80(11), pp.16979-16995.

[9] Tang, Y., Zhao, L., Zhang, S., Gong, C., Li, G. and Yang, J., 2020. Integrating prediction and reconstruction for anomaly detection. Pattern Recognition Letters, 129, pp.123-130.

[10] Deepak, K., Chandrakala, S. and Mohan, C.K., 2021. Residual spatiotemporal autoencoder for unsupervised video anomaly detection. Signal, Image and Video Processing, 15(1), pp.215-222.

[11] Al-Dhamari, A., Sudirman, R. and Mahmood, N.H., 2020. Transfer deep learning along with binary support vector machine for abnormal behavior detection. IEEE Access, 8, pp.61085-61095.

[12] Ullah, W., Ullah, A., Hussain, T., Khan, Z.A. and Baik, S.W., 2021. An Efficient Anomaly Recognition Framework Using an Attention Residual LSTM in Surveillance Videos. Sensors, 21(8), p.2811.

[13] Wan, S., Xu, X., Wang, T. and Gu, Z., 2020. An intelligent video analysis method for abnormal event detection in intelligent transportation systems. IEEE Transactions on Intelligent Transportation Systems.

[14] Chu, W., Xue, H., Yao, C. and Cai, D., 2018. Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos. IEEE Transactions on Multimedia, 21(1), pp.246-255.

[15] Chandrakala, S., Srinivas, V. and Deepak, K., 2021. Residual Spatiotemporal Autoencoder with Skip Connected and Memory Guided Network for Detecting Video Anomalies. Neural Processing Letters, pp.1-16.