Survey on Anomaly Detection for Video Surveillance

Prof. Rokade M.D[1], Bora Tejashri[2]

[1]*monikarokade4@gmail.com*
[2]*boratej283@gmail.com*
[3] *Dept. of Computer Engineering, SPCOE,SPPU, Maharashtra, India*

Abstract– *An suspicious activity is any observed action that could specify a person may be involved in a crime or about to commit a criminality. Anomaly detection is nothing but detecting suspicious activity. Surveillance cameras are one of the best solution used for the security in various places (e.g public places, outside the home etc.) But in video surveillance, detecting anomalous events such as traffic accidents, crimes or illegal activities is the most critical task. Present system needs man power for monitoring the system as detecting and identifying criminal and abnormal activity is so difficult. So in this paper we are surveying on anomaly detection for video surveillance using concepts of deep learning, RNN etc.*

***Key Words*: Video Surveillance, Anomaly detection, CNN, Recurrent neural network, Deep learning.**

1. INTRODUCTION

Anomaly detection has been broadly studied in computer vision because of its probable applications in video surveillance, action recognition and scene understanding, etc. An anomaly detection system would really help in reduction human labor and time. However, anomaly detection is still an tremendously challenging job because of the unbounded assets of anomaly. In real applications, compared with normal events, anomaly is exceptional and it is extremely costly to collect abnormal events; On the other hand, it is infeasible to gather all possible abnormal actions.

Consequently for a typical anomaly detection dataset, only normal situations are given in a training set. To recognize whether an abnormal event happens, a common methodology is to exploit regular patterns in terms of appearance and motion on the training set. Any patterns that disagree with these regular ones would be categorized as irregular ones.

Dictionary learning based approaches have demonstrated their achievement for anomaly detection. In these methods, learning a dictionary to encrypt all normal events on the training set and an abnormal event would end result in a large reconstruction error. Though, the optimization of sparse coefficients is very time consuming, which turn into the bottleneck of dictionary learning based for anomaly detection approaches. Additional, features direct the enactment of anomaly detection, while dictionary learning based approaches are generally based on hand-crafted features, which may not be optimal for video representation. In recent times, in light of the countless success of deep learning in many computer vision tasks, it has been lead to the anomaly detection.

Specifically, an Auto-Encoder is learnt on the normal training data underneath an assumption that consistent data can be recreated by themselves while irregular ones can't. However, such answer to the solution is based on a 3D Convolutional Neural Network (CNN), while earlier work has shown that extracting form and motion information independently with a two-stream network is a better solution for feature extraction in videos. Further, such a solution either one takes a video cube as its input, or regular/irregular frames in this cube may affect the grouping of each other. To avoid this, video cubes have to be experimented by centering the cube over all frames, which is computationally costly.

In this paper, we are also studying a sparse coding based approach for anomaly detection. More specially, a dictionary is learnt to encode regular patterns in terms of appearance, and features resultant to normal measures be sparsely reconstructed by this dictionary with a minor reconstruction error. Further, to progress the smoothness of prediction over adjacent frames, a temporally-coherent term is imposed. Then we reach at a Temporally-coherent Sparse Coding (TSC) formulation.

It is fascinating that our TSC formulation can be understood as one special stacked Recurrent Neural Network: the optimization of sparse coefficients to an Iterative Soft-thresholding Algorithm(ISTA) algorithm relates with to a stacked network, and the temporally-coherent term makes the reconstruction coefficients of current frame depend on that of earlier frame. In order to straightly optimize the reconstruction coefficients rather than elaborately selecting the hyper-parameters in TSC, we offer to optimize all parameters in sRNN instantaneously, which avoids the nontrivial hyper-parameter selection in TSC. In addition, sRNN is a feed-forward network that would significantly speed up the anomaly prediction in testing phase.

2. Related Work

2.1 Problem Statement

There has been lot of problems regarding security.
The great rise of security threats is a big challenge our country. To know how to overcome with the security problem we are studying the paper on anomaly detection.

3. Process for Anomaly detection

There are different generally used stages in a process to detect anomalous events. They are -

3.1 Feature extraction

In this type of extraction, we can extract hand-crafted or learnt features on a training set. Initial work operates the low-level trajectory features to represent the unvarying patterns[24]. However, these techniques are not robust in complex or congested scenes. In order to resolve this problem, spatial temporal features, like histogram of oriented gradients, the histogram of oriented flows, are usually used. Based on those spatial-temporal features, model the normal patterns with a Markov random field (MRF) fit the regular histograms of optical flow in local regions with an exponential spreading. To represent the limited optical flow patterns, utilize a combination of probabilistic PCA simulations.

3.2 Model selection and anomaly prediction

Dictionary learning based methodologies are widely referred in anomaly detection. A essential assumption of these approaches is that any feature can be linearly denoted as a linear permutation of basis of a dictionary which encodes unvarying patterns on the training set. Usage of the reconstruction error to fix whether a frame is abnormal or not. The reconstruction error, such as minimum square error, doesn't take sparsity term into thought, and in fact, it does benefit the anomaly detection accuracy. To avoid this, propose two solutions, i.e. maximum coordinate (MC) and non-zero concentration (NC), to detect anomaly. However, sparse reconstruction based methods are usually time-consuming in the optimization of sparse coefficients. To solve this problem, advise to discard the sparse constraint and learn multiple dictionaries to code the patches at multiple scales, which inevitably brings additional costs in the training phase.

3.3 Deep learning based anomaly detection

Deep learning approaches have verified its successes for image classification, object recognition, as well as anomaly detection. We can propose a 3D convolutional Auto-Encoder (Conv-AE) to model the regular structures, however, 3D convolution cannot characterize the spatial and temporal data very well, as shown in the activity recognition [13]. In light of the capability of convolutional neural networks (CNN) to study spatial features and the strong capability of recurrent neural network (RNN) and long short term memory (LSTM) to model temporal forms, make efforts to leverage a convolutional LSTMS Auto-Encoder (ConvLSTM-AE) to characterize both appearance and motion information. Although RNNs or LSTMs are strong and effective for processing sequential data, they are in fact "black box" whose internal structures are hard to be interpreted. Recently, show that a special type of RNN in reality enforces a sparse constraint on the features. Encouraged by the work of sparse coding based anomaly detection and interpretable RNN, we propose a TSC and its sRNN complement for anomaly detection.

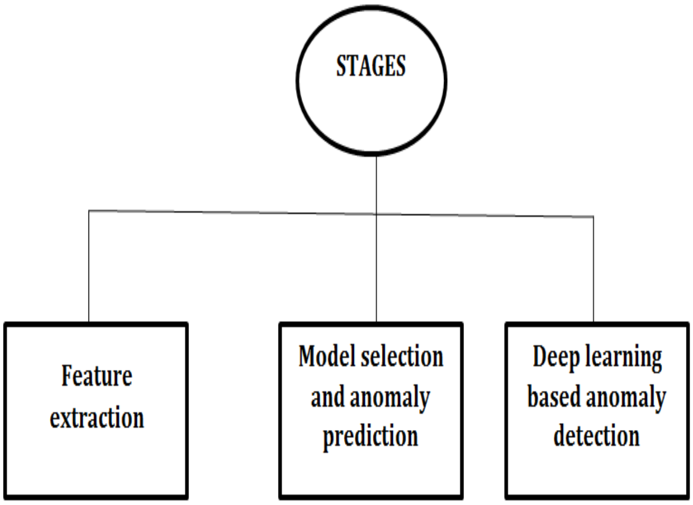

Figure 3.1: Different stages

4. Algorithms

In this survey paper, we are going to look different algorithm which may be used for future research.

4.1 Recurrent neural network

A RNN is a class of artificial neural networks where connections among nodes form a directed graph alongside a temporal series. This allows it to demonstration temporal dynamic behavior. Derivative from neural networks, RNNs can use their inner state (memory) to procedure variable length orders of inputs. This sorts them applicable to tasks like undivided, etc.

The term "recurrent neural network" is used generally to refer to two broad classes of networks with a alike common structure, where one is finite impulse and the other one is infinite impulse. Equally classes of networks show temporal dynamic behavior.

A finite instinct recurrent network is a directed acyclic graph that can be unrolled and exchanged with a strictly feedforward neural network, while an infinite impulse recurrent network is a directed cyclic graph that can't be unfolded.

Both finite impulse and infinite impulse recurrent networks can have supplementary stored states, and the storage can underneath direct control by the neural network. The storage can also be swapped by another network or graph, if that integrates time delays or has feedback loops. Such controlled states are mentioned to as gated state or gated memory, and are part of long short-term memory networks (LSTMs) and gated recurrent units.

4.2 Convolution neural network

In deep learning, CNN, or ConvNet is a class of deep neural networks, furthermost generally applied to examining visual imagery. They are also recognized as shift invariant or space invariant artificial neural networks (SIANN), based on the shared-weight architecture of the convolution kernels that scan the unseen layers and translation invariance features. They are applicable in image and video recognition, recommender systems, image classification, Image segmentation, medical image analysis, natural language processing, brain-computer interfaces, and financial time series.

CNNs are regularized types of multilayer perceptrons. Multilayer perceptrons typically mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next. The "fully-connectedness" of these networks makes them prone to overfitting information. Typical means of regularization include varying the weights as the loss function gets reduced while randomly trimming connectivity. CNNs take a unalike approach towards regularization: they take benefit of the hierarchical pattern in data and gather patterns of increasing complication using smaller and simpler patterns imprinted in the filters. Therefore, on the measure of connectedness and complexity, CNNs are on the minor extreme.

5. OBJECTIVE

The main objectives of this survey paper are -

1. To study about anomaly detection.
2. To know different methodologies through which we can detect the abnormal activity.

6. CONCLUSIONS

In this paper, we did the survey for anomaly detection using different concepts and methods. We studied about TCS framework for anomaly detection which reserves the similarities between the frames within the normal/abnormal actions, RNN, CNN in short. We have seen different stages used for anomaly detection.

ACKNOWLEDGEMENT

Authors want to acknowledge Principal, Head of Computer department and the guide for all the support and help rendered. To express profound feeling of gratitude to the regarded guardians for giving the motivation required to the finishing of paper.

REFERENCES

[1] Monika D.Rokade, Dr. Yogesh Kumar Sharma, "Identification of Malicious Activity for Network Packet using Deep Learning ", in International Journal of Advanced Science and Technology, 2020.

[2] Joey Tianyi Zhou, Jiawei Du, Hongyuan Zhu, Xi Peng, Yong Liu, Rick Siow Mong Goh, "AnomalyNet: An Anomaly Detection Network for Video Surveillance, 2019.

[3] W. Luo, W. Liu, and S. Gao, "A revisit of sparse coding based anomaly detection in stacked rnn framework," in The IEEE International Conference on Computer Vision (ICCV), Oct 2017

[4] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 733–742.

[5] M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette, "Real-time anomaly detection and localization in crowded scenes," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2015.

[6] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in matlab ," in Proceedings of the IEEE international conference on computer vision, 2013, pp. 2720–2727.

[7] B. Zhao, L. Fei-Fei, and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011, pp. 3313–3320.

[8] S. Wu, B. E. Moore, and M. Shah, "Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 2054–2060.

[9] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 1975–1981.

[10] L. Kratz and K. Nishino, "Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 1446–1453.

[11] F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, "Anomalous video event detection using spatiotemporal

context," Computer Vision and Image Understanding, vol. 115, no. 3, pp. 323–333, 2011.

[12] J. Kim and K. Grauman, "Observe locally, infer globally: a spacetime mrf for detecting abnormal activities with incremental updates," in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 2921–2928.

[13] J. R. Medel and A. Savakis, "Anomaly detection in video using predictive convolutional long short-term memory networks," arXiv preprint arXiv:1612.00390, 2016.

[14] Y. S. Chong and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," in International Symposium on Neural Networks. Springer, 2017, pp. 189–196.

[15] J. T. Zhou, M. Fang, H. Zhang, C. Gong, X. Peng, Z. Cao, and R. S. M. Goh, "Learning with annotation of various degrees," IEEE Trans. Neural Netw. Learning Syst.

[16] A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Fast L1-Minimization algorithms and an application in robust face recognition: A review," EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2010-13, February 5 2010.

[17] M. D. Zeiler, "Adadelta: an adaptive learning rate method," arXiv preprintarXiv:1212.5701, 2012.

[18] T. Xiao, C. Zhang, H. Zha, and F. Wei, "Anomaly detection via local coordinate factorization and spatio-temporal pyramid," in Asian Conference on Computer Vision. Springer, 2014, pp. 66–82.

[19] Z. Zhu, J. Wang, and N. Yu, "Anomaly detection via 3d-hof and fast double sparse representation," in Image Processing (ICIP), 2016 IEEE International Conference on. IEEE, 2016, pp. 286–290.

[20] H. Mousavi, M. Nabi, H. K. Galoogahi, A. Perina, and V. Murino, "Abnormality detection with improved histogram of oriented tracklets," in International Conference on Image Analysis and Processing. Springer, 2015, pp. 722–732.