Anomaly Detection in Human Behavior using Video Surveillance

Neha Sharma, Pradeep Kumar D, Rohit Kumar, Shiv Dutt Tripathi

Abstract**:** ***Conventional static surveillance has proved to be quite ineffective as the huge number of cameras to keep an eye on most often outstrips the monitor's ability to do so. Furthermore, the amount of focus needed to constantly monitor the surveillance video cameras is often overbearing. The review paper focuses on solving the problem of anomaly detection in video sequence through semi-supervised techniques. Each video is defined as sequence of frames. The model is trained with goal to minimize the reconstruction error which later on is used to detect anomaly in the test sample videos. The model was trained and tested on most commonly used benchmarking dataset- Avenue dataset. Experiment results confirm that the model detects anomaly in a video with a reasonably good accuracy in presence of some noise in dataset.***

Keywords**:** ***video surveillance, anomaly detection, semi-supervised learning, unusual activity, video processing, abnormal behavior.***

I. INTRODUCTION

With the increasing number of anti-social activities that have been taking place, security has been given utmost importance off late and it is paramount that every citizen plays his share in warranting the safeguarding of our society.

Many organizations have mounted CCTV cameras for the constant monitoring and invigilation of people in public areas and their interactions. For a developed country such as ours, with a population of 1.33 billion, every person is captured by a camera ~ 70 times a day. A lot of video is spawned and stored for a certain time duration. A 704x576 resolution image which is recorded at 25 frames per second will produce roughly 20GB data per day. Since constant monitoring of data by humans to judge if the events are abnormal is a near impossible task and requires a colossal workforce and constant attention and awareness, it calls for a need to automate the same. Also, there is a necessity to show which frame and which parts of the video contain the unusual activity which can aid in faster judgment of that anomalous activity being abnormal.

Further, the interpretation of an anomaly rests on the context in which it is used. A video event is termed as an anomaly when something unusual happens in the video frame which does not confer to the usual norms. With the expeditious surge in video data, there is an alarming urgency not just for identification of objects but also for detecting the unusual objects withal.

Types of anomalies :

- **Point Anomaly**: A lone instance of data is said to be anomalous if it is too far off from the remainder instances. *Business use case :* Detecting credit card fraud based on the amount spent from that card.
- **Contextual Anomaly**: The abnormality in this case is context definitive. This type of anomaly is frequent in time-series data. *Business use case:* Spending $200 on grocery and food every day during the holiday season is typical, but may be irregular otherwise.
- **Collective Anomalies:** A set of data samples that simultaneously helps in discovering anomalies.

Manual detection of anomalous and abnormal events in long series of video data, such as surveillance tapes, requires a great deal of manpower that might not be available at all times to all organisations.

Video data, by itself, is demanding to model and represent owing to its high dimensionality, noise and broad variety of interactions. Anomalies are also exceedingly circumstantial and the definition can be ambiguous. Now, there are copious successful cases where anomaly detection has worked well[1,2,7]. However, these methods work by exploiting labelled data which is infeasible and costly. One must record and classify past events and then train the model. This demands for an approach that is increasingly feasible to implement and doesn't burden the programmer

II. LITERATURE SURVEY

This section discusses the diverse research papers that are of consequence to this work and present the underlying features and inferences in them.

Khawaja M. Asim, Asifullah Khan, and Iqbal Murtza in [7] present a supervised approach for dealing with the problem of detecting anomalies in videos. Taking into account the pixel based approach for identifying anomalies, the authors make use of k-means clustering algorithm, accompanied by a-posteriori probability based model and zone crossway technique for discovering abnormalities.

The algorithm consists of the following steps :

i. Selection of points of interest

ii. Interest points description

iii. Feature vector clustering

iv. Construction of an ensemble of key-points

Revised Manuscript Received on December 30, 2019.

* Correspondence Author

Neha Sharma, Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore, India.

Mr. Pradeep Kumar D., Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore, India.

Shiv Dutt Tripathi, Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore, India.

Rohit Kumar, Computer Science and Engineering, Ramaiah Institute of Technology, Bangalore, India.

v. Model training and subsequent testing for accuracy
vi. Region junction

The technique regards ordinary events as the ones having much greater probabilities of occurrence. Thickly sampled points are delivered to a probabilistic model through the k-means clustering to attain the probabilities of episodes.

The k-means bundling technique is used to quantize the vectors, and is one of the most prevalent methods for cluster analysis. The algorithms dispenses n sample points in the feature space, randomly selects k points as cluster means, and then allocates every observation to the closest mean. The means of each cluster are updated iteratively.

Let there be n given observations, [x1, x2,..,xn], wherein each xi depicts an observation which is a d-dimensional vector. The clustering algorithm partitions the n observation into a user-defined “k” number of clusters where k<=n. Mathematical representation of the algorithm is shown in equation 1.

$$\arg_s \min \sum_{i=1}^{k} \sum_{x_j \in s_i} \left\| x_j - \mu_i \right\|^2 \qquad (1)$$

A threshold value is applied for differentiating the anomalous events from ordinary ones. The ultimate results of abnormal event detection which are attained from multiple scales are put together using region assimilation. An amalgamation of the outcomes of multi scale unusual event discovery using zone assimilation helps reduce false positive vigorously. The method is tested on the standard UCSD dataset, detecting anomalies with great success.

Chong and Tay in [1] talk about the most common feature for anomaly detection- video feature representation. Ample research has been done in finding the skillful anomaly detection but finding anomaly in videos is still an open challenge.This is because of large variations of its environment, human continual movement, high space-time complexity and the complex dimensionality of video data. The author also mentions that it is awfully difficult for any anomaly detectors to go with the supervised approach because one will have to train the model for every possible situations and the model will have incredible intricacies. Therefore,the author suggests that the semi-supervised approach could help in learning of the video data.

The author also discusses some semi supervised algorithms and optical flow-based descriptor in contrast to trajectory extraction which requires identifying and tracking of objects, while optical flow methods do not depend upon any such preconditions.

In [2], Sabokrou and Fathy discuss another novel approach for anomaly detection. The authors’ proposed method can detect real-time anomalies in crowded scenes. Their work treats a video as a collection of cubic blotches which are outlined using certain descriptors. Gaussian classifiers, which are extremely simple are used to demarcate the abnormal and normal events.

The algorithm commences with the creation of a sequence of frames from captured video and the classification of frames as local and global patches using Gaussian distributions after which the final decision is made regarding the nature of those frames.

In [3], B. Ravi Kiran and Ranjit Parakkal talk about a semi-supervised approach to detecting aberrations in videos.

Semi-supervised learning is a sphere of machine learning which capitalizes on unlabeled data alongside a small measure of labeled data. It is a blend of supervised and unsupervised learning approaches. Research has shown that a large amount of unlabeled data, when used with a small fraction of labeled data for training, can produce great improvements in the learning accuracies of machine learning models.

The procurement of labeled data for learning problems usually requires a skilled person or an experiment to determine the class labels associated with the instances. The cost linked with the labeling process may thus render a thoroughly labeled training set unattainable. In contrast, the acquisition of unlabeled data is comparably inexpensive. In such situations, a semi-supervised learning can prove to be of great practical significance.

The authors first review the deep convolutional architectures for representation of features along with the generative and predictive models for the task of detecting unusual patterns in the video footages. For representing an unusual activity, the authors make use of an “anomaly-mask” that highlights a suspicious activity in the given frame.

For reconstruction modeling, the paper discusses several dimensionality reduction techniques including :

Principal Component Analysis(PCA) : PCA finds the most important dimensions from a vast set of dimensions that are most useful for our model. It determines the dimensions where there is maximum variance indicating the most important feature for our model, which allows us to drop the redundant ones.

Taking X as the input matrix having a non-zero mean, we find orthogonal projections that disassociate features in the training data using equation 2:

$$\min_{W^T W = I} \| X - (XW)W^T \|_F^2 = \| X - \hat{X} \|_F^2 \qquad -(2)$$

Here, WTW=I represents an orthogonal reconstruction of input matrix X, and the projection XW represents a vector in lower dimensional space. This reduction in dimensionality captures the anomalous behavior in the samples, since they are not that well reassembled. The Mahalanobis distance between the reconstruction and the original input gives the anomaly score.

Autoencoders : An auto-encoder is an artificial neural network which is used to study effective data coding in unsupervised fashion. They achieve to learn a portrayal of a hotchpotch of data for dimensionality contraction by training the model to discount useless or unnecessary data, often termed as noise. Now, along with the reduction lateral, a reconstruction lateral is also learnt, wherein the auto-encoder generates a representation closing resembling the original input from the reduced input which is devoid of irrelevant features or dimensions.

These reconstruction based predictive and generative models erect representations to reduce the errors in reconstruction in learning models.

III. PROBLEM FORMULATION

The following section discusses the problem of detecting unusual activities in frames effectively. The semi-supervised approach is the one that best suits our case. Even though supervised learning methods are the standard, and provide considerably good results, they are just not feasible for large datasets. A camera generates hundreds of gigabytes of video per day, and this video needs to be processed prior to the application of machine learning algorithms. The training dataset requires labelling, and this renders the task extremely time consuming and almost impractical. This leads us towards the unsupervised learning approach, but this method does not provide great results. So we finally stumble upon a hybrid method that takes the best of both worlds and provides the accuracy of supervised models, and the ease of practicality of the unsupervised ones.

A. An Illustrative example

The datasets considered in this work are listed as follows :

- The UCSD dataset [4] consists of videos of people walking on pavements and footpaths where the occurrence of motor vehicles even a bicycle correspond to anomalous events taking place.
- Strolling of humans in unusual locations also amounts to an anomalous activity taking place.
- In Avenue Dataset[5], anomalies correlate to odd activities - a person propelling an object in the air, like papers or a bag.
- In Subway dataset, people moving in the incorrect direction are taken as anomalies.
- In recent times, LV dataset has been used in a simulated environment for the laborious task of detecting suspicious acts in real time streaming videos.

Fig. 1 depicts the working of the desired anomaly detection model on two datasets : UCSD and Avenue. In UCSD, since the presence of vehicles such a cycles or cars, or even people who are not on foot is considered as an anomaly, we see that the man cycling on the pavement in row one and the skateboarder in the row two are identified as doing something abnormal.

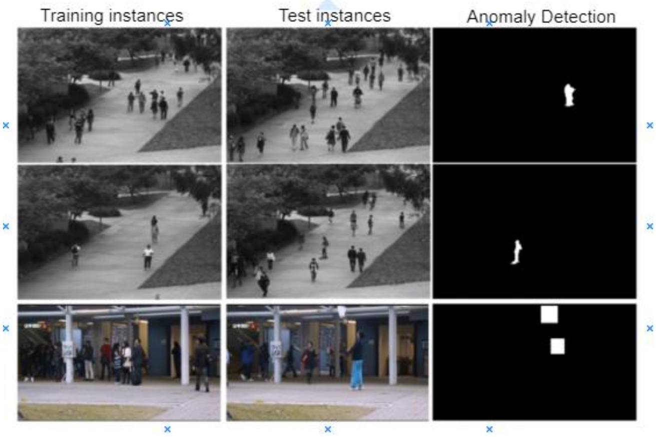

Fig 1. UCSD dataset (top two rows) , portrays the appearance of a cyclist or a skate-boarder on the pavement as an unusual activity. In the Avenue dataset(bottom row), the throwing of papers into the air by a person accounts for an anomaly.

The third row depicts the Avenue dataset wherein the propelling of an object in the air is identified as an anomalous event taking place. Since the person in the frame is throwing a bunch of papers, this act is identified by the model as an anomaly.

IV. METHODOLOGY

The method used finds its basis on the difference between the older bunch of frames and the most recent ones to detect anomaly in the given video. The model is first trained on the normal videos (without any abnormal activities) with the goal in mind to diminish the reconstruction error betwixt the input - output video sequences constructed by the trained model with the help of an autoencoder. Once the model is compliant, reconstruction error for anomalous events far exceeds that of the normal ones. By setting up a threshold value on the produced error, the model is able to detect abnormality in the scene.

Steps involved:

A. Data Preprocessing:

In this, the first step is to convert the input video into frames and then resize it to 227x227. Next, each frame is normalized by scaling the pixel value between 0 and 1, post which the images are transformed to grey scale for reduction in dimensionality. We then clip out negative values if any and finally store them in numpy array for further processes.

B. Feature Studying:

Feature Learning relates to intuitively discovering the depictions that are required for detecting features and classifications from unprocessed data. convolutional spatio-temporal auto-encoder[8] is used for learning frequently occurring patterns in the training videos. The architecture comprises two parts – the dimensional decoder and the encoder as depicted in Fig 2.

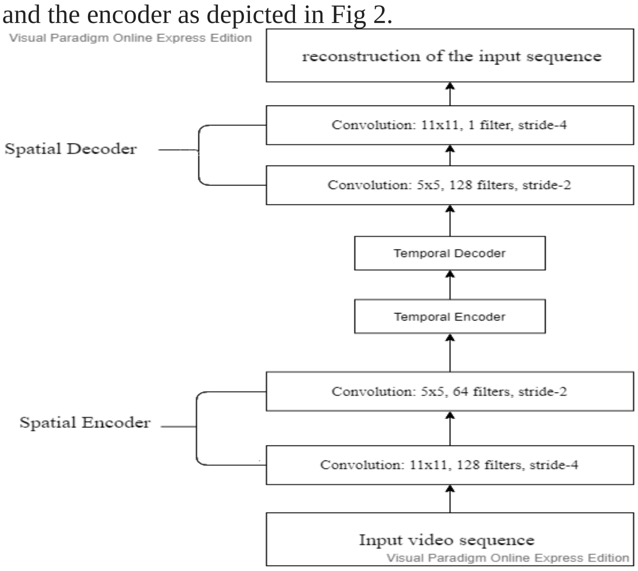
Visual Paradigm Online Express Edition
reconstruction of the input sequence
Convolution: 11x11, 1 filter, stride-4
Spatial Decoder
Convolution: 5x5, 128 filters, stride-2
Temporal Decoder
Temporal Encoder
Convolution: 5x5, 64 filters, stride-2
Spatial Encoder
Convolution: 11x11, 128 filters, stride-4
Input video sequence
Visual Paradigm Online Express Edition

Fig 1.The output is a reconstruction of the input frame of length T. The dimension is reduced as we go from input and we get back the output with reconstruction value that we tend to decrease in case of training the model using backpropagation algorithm

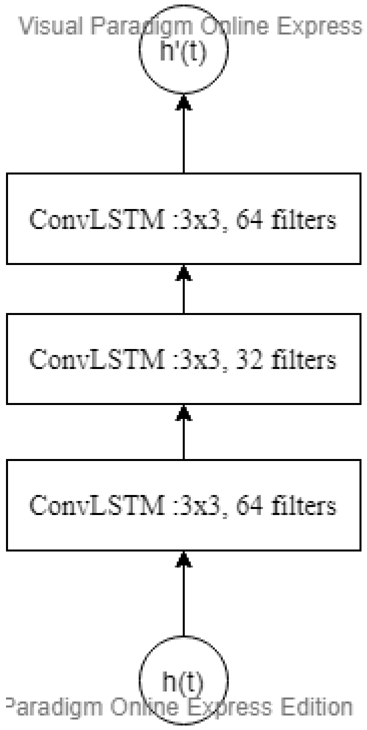
Visual Paradigm Online Express
h'(t)
ConvLSTM :3x3, 64 filters
ConvLSTM :3x3, 32 filters
ConvLSTM :3x3, 64 filters
h(t)
Paradigm Online Express Edition

Fig 2:The zoomed-in architecture of temporal encoder-decoder which comprises of 3 ConvLSTM layers at time t.

C. Evaluation Metric:

To evaluate the performance of the model we feed in the test data comprising of unseen video footage and analyze if the model detects any anomaly, making sure that false alarm rate is low. Through measuring the distance (Euclidean) amidst the input and reconstructed frames, the model flags the video as Anomalous or not depending upon the threshold value that is set during the testing period.

V. EXPERIMENT

A. Dataset:

The training phase makes use of a most commonly used dataset for benchmarking : the Avenue dataset [7] subsisting 16 and 21 training and testing clips respectively. The extent of each clip is bounded by a minute or two.

Anomalous events in the video clips are the ones where we can find people running haphazardly or carrying out any activity that does not conform to the general rules of walking, where normalcy is described by simply strolling down the street or climbing a flight of stairs. Actions such as throwing objects in the air or a run-in are defined to be suspicious.

Predicaments to the dataset include an unstable camera or the presence of sparse outliers.

B. Model Parameters:

The model was trained by curtailing the restoration error of input volumes. We make use of Adam optimizer to set the learning rate impromptu depending on the model. Mini-batches of size 64 and the model was trained for 30 epochs. The spatial or dimensional encoder – decoder made use of an activation function based on hyperbolic tangents.

C. Environment and API's used:

Anaconda was used as the developing environment. Major APIs used in the experiment - numpy, keras, tensorflow, scipy.

VI. RESULT

The data was trained on a GTX 1050 GPU for 30 epochs. Each epoch took an ETA of 30 min approximately. Sequential model from Keras API was used to train and test on the Avenue dataset. The model detected anomaly from an anomalous video successfully with an accuracy of 0.77.

VII. CONCLUSION

This research paper applies the deep learning approach of machine learning to tackle the problem of detecting and analysing abnormalities in video data. Convolutional LSTM and dimensional factor picker were utilised for solving the problem. The Convolutional LSTM is best suited for the above mentioned issue because its structure is inherently convolutional and thus helps in dealing with irregularities in the data. For the experiment, Keras' predefined layers were used. A thorough model was built by extracting dimensional features of the data and training it on several instances. Although the model detects anomalies from the benchmark dataset, real world scenarios could be more complex and thus there are chances of false alarm. Future work could be done on reducing the false alarm in case of complex environment.

REFERENCES

1. Chong, Yong Shean, and Yong Haur Tay. "Modeling representation of videos for anomaly detection using deep learning: A review." arXiv preprint arXiv:1505.00523 (2015).W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.
2. Sabokrou, Mohammad, Mahmood Fathy, Mojtaba Hoseini, and Reinhard Klette. "Real-time anomaly detection and localization in crowded scenes." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 56-62. 2015.B. Smith, "An approach to graphs of linear forms (Unpublished work style)," unpublished.
3. Kiran, B., Dilip Thomas, and Ranjith Parakkal. "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos." *Journal of Imaging*4, no. 2 (2018): 36.
4. Mahadevan, Vijay, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. "Anomaly detection in crowded scenes." In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1975-1981. IEEE, 2010. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style)," *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan, 1982, p. 301].
5. Lu, Cewu, Jianping Shi, and Jiaya Jia. "Abnormal event detection at 150 fps in matlab." In *Proceedings of the IEEE international conference on computer vision*, pp. 2720-2727. 2013.
6. Asim, Khawaja M., Iqbal Murtza, Asifullah Khan, and Naeem Akhtar. "Efficient and supervised anomalous event detection in videos for surveillance purposes." In *2014 12th International Conference on Frontiers of Information Technology*, pp. 298-302. IEEE, 2014.
7. http://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html
8. Taylor, Graham W., Rob Fergus, Yann LeCun, and Christoph Bregler."Convolutional learning of spatio-temporal features." In European conference on computer vision, pp. 140-153. Springer, Berlin, Heidelberg, 2010.

AUTHORS PROFILE

Neha Sharma is an undergraduate student pursuing her bachelor's degree in Computer Science Engineering from MS Ramaiah Institute of Technology, Bangalore. Her areas of interest include machine learning, deep learning and data analytics.

Mr. Pradeep Kumar D. is working as an Assistant Professor in Computer Science Department of MS Ramaiah Institute of Technology. He received the degree of B.E. in the year 2009 from Visvesvaraya Technological University in Computer Science. He also received his M.Tech degree from the same university in the year 2011. His areas of interest include data mining, data sciences, big data, machine learning and internet of things.

Shiv Dutt Tripathi is an undergraduate student pursuing his bachelor's degree in Computer Science Engineering from Ramaiah Institute of Technology, Bangalore. His areas of interest include TensorFlow, Web Development and Competitive Coding.

Rohit Kumar is an undergraduate student pursuing his bachelor's degree in Computer Science Engineering from Ramaiah Institute of Technology, Bangalore. His areas of interest include Machine Learning, Data Science and Backend Development.

Retrieval Number: B3133129219/2019©BEIESP
DOI: 10.35940/ijeat.B3133.129219
Journal Website: www.ijeat.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication