Anomaly Detection

Anjali Mahindrakar[1], Asavari Shejwal[1], Aditya Chavan[1], Vrushabh Panchal[1], Keerti Kharatmol[2]

[1]BE Student, Department of Computer Engineering, K.C. College of Engineering and Management Studies and Research, Thane Maharashtra, India

[2]Assistant Professor, Department of Computer Engineering, K.C. College of Engineering of Engineering and Management Studies and Research, Thane Maharashtra, India

ABSTRACT – We have represented an effective method for detecting anomalies in videos. Anomaly detection in crowded localities are very difficult due to complex mechanics of a human crowd. We have incorporated a methodology for detecting and locating anomalous activities in video frames of crowded scenes using spatial temporal encoder architecture. This allows us to extract features from both spatial and temporal dimensions by performing spatial–temporal convolutions, thereby, both the appearance and motion information encoded in continuous frames are extracted. Therefore, we've used a spatial temporal encoder CNN model for anomaly detection in video surveillance. This model consists of two main modules: spatial feature representation and temporal evolution of the spatial features. Experimental results on Avenue and UCSD datasets confirm the detection accuracy of our method in terms of frames per seconds.

KEYWORDS - Spatio-temporal auto-encoder, Convolutional neural network, Reconstruction error, Thresholding, Auto-encoder and Decoder, Event Count.

I. INTRODUCTION

Due to increasing usage of surveillance systems in railway stations, airports, roads or malls, anomaly detection has become widespread. There is an increasing need not only for recognition of objects and their behaviours, but also in particular for detecting abnormal behaviour in the large body of ordinary data. The large volumes of video data are challenging to analyse and model due to its high dimensionality, noise, and wide variety of events. Anomalies are often defined in relation to a specific context or environment. What might be considered abnormal in one context may be completely normal in another. For example, running in a restaurant would be considered an anomaly because it goes against the norms and expectations of that particular setting. However, running at a park would be considered normal because it is a common activity in that context. It's important to consider the context when analysing data and identifying anomalies. Further, the definition of anomaly can be ambiguous and often not precisely defined. Therefore, it is difficult for machine learning methods to identify video patterns that produce anomalies in real-world applications. A behaviour can be accurately described by identifying and characterising the types and locations of feature points present, without relying on global appearance, posture, nearby motion or occlusion. This approach requires the detection and description of a rich set of features, which can be used to distinguish between different behaviours. Overall, the approach emphasises the importance of feature detection and description in accurately characterising and comparing different behaviours.

II. PROBLEM STATEMENT

Anomaly detection refers to the task of identifying unusual or rare events, observations or patterns that deviate from the expected or normal behaviour of a system or process. The goal of anomaly detection is to distinguish between normal and abnormal behavior, which can be useful for detecting fraud, errors, faults, anomalies, or attacks in various domains such as finance, healthcare, manufacturing, cybersecurity, and more. The problem is to develop an effective and scalable anomaly detection system that can detect anomalies in real-time or near real-time, with high accuracy and low false-positive rates, while being able to handle large-scale and high-dimensional data sets. We aim to accomplish this solution by our proposed Anomaly Detection machine learning model.

III. LITERATURE SURVEY

1. **Doll´ar, P., Rabaud, V., Cottrell, G., Belongie, S.: Behaviour recognition via sparse spatio-temporal features** states a new method for recognizing human behaviours from video footage. The method uses a mathematical technique called sparse coding to identify important visual features of the behaviour and then analyses these features over time to recognize the behaviour. The researchers tested the method on several datasets of human behaviours, including walking, jogging, and waving, and found that it achieved high accuracy in recognizing these behaviours. This method could have applications in fields such as security and surveillance, where automated behaviour recognition systems could help identify potential threats.

2. **Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.:Learning temporal regularity in video sequences:** It presents a method of unsupervised learning by learning the temporal patterns present in videos without the need for explicit supervision or labels. The proposed approach uses a deep neural network architecture to predict the next frame in a sequence based on the previous frames. The model can be used for tasks such as video prediction, action recognition, and anomaly detection. The approach is evaluated on several benchmark datasets, demonstrating its effectiveness in learning and exploiting temporal regularities in video sequences.

3. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes: The paper "Anomaly detection in crowded scenes" presented at the IEEE Conference on Computer Vision and Pattern Recognition focuses on detecting abnormal behaviour in crowded environments using video surveillance. The authors propose a novel approach that utilises deep neural networks to learn the normal behaviour patterns of the crowd and identify any deviations from it as anomalies. The proposed method achieves better results compared to existing approaches and has the potential to enhance public safety in crowded places.

IV. PROPOSED SYSTEM

General features: the steps involved in abnormal event detection in videos using spatiotemporal auto-encoder:

I. Collect and preprocess video data: The first step involves collecting the video data and preprocessing it to obtain meaningful features. This may involve techniques such as resizing, cropping, and normalisation.

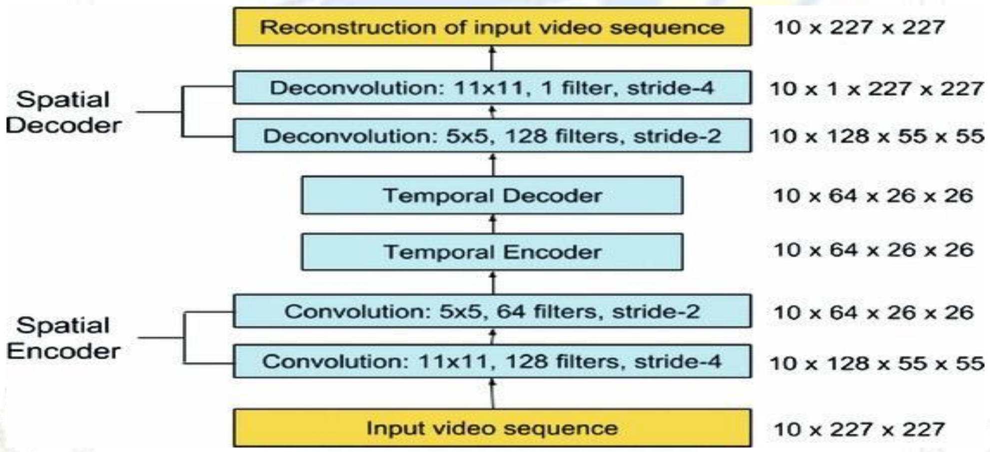
II. Train the spatiotemporal auto-encoder: A spatiotemporal auto-encoder is a deep learning architecture that learns to encode spatial and temporal information from the video data. This involves training the model on a large dataset of normal video sequences.

III. Extract features: Once the auto-encoder has been trained, it can be used to extract features from video sequences. These features capture the normal patterns and variations in the video data.

IV. Define a threshold for abnormality: The next step involves defining a threshold for abnormality. This threshold is used to distinguish between normal and abnormal video sequences. Any sequence that falls outside this threshold is considered abnormal.

V. Test the model: The final step involves testing the model on new video sequences to detect abnormal events. This involves feeding the video data through the auto-encoder to extract features, and then comparing these features to the threshold for abnormality. If the features fall outside this threshold, the video sequence is flagged as abnormal.

Overall, the flowchart for abnormal event detection in videos using spatiotemporal auto-encoder involves collecting and preprocessing video data, training the auto-encoder, extracting features, defining a threshold for abnormality, and testing the model on new video sequences.



The method proposed in this article is based on the idea that the most recent frames of video will be different from older frames when an abnormal event occurs. The authors develop an end-to-end model consisting of a spatial feature extractor and a temporal encoder-decoder to learn the temporal patterns of the input volume of frames. The model is trained with normal scenes to minimise reconstruction error. When tested with abnormal scenes, the system can detect them by thresholding on the reconstruction error produced by each testing input volume.

Our approach consists of three main stages:

1. Preprocessing: The task of this stage is to convert raw data to the aligned and acceptable input for the model. Each frame is extracted from the raw videos and resized to 227 *227.

2. Feature Learning: We propose a convolutional spatiotemporal autoencoder to learn the regular patterns in the training videos. Our proposed architecture consists of two parts— spatial autoencoder for learning spatial structures of each video frame, and temporal encoder-decoder for learning temporal patterns of the encoded spatial structures.
2.1 Autoencoder: Autoencoders, as the name suggests, consist of two stages: encoding and decoding.
2.2 Spatial Convolution: a convolutional network is to extract features from the input image.

3. Regularity Score : we can evaluate our models performance by feeding in testing data and check whether it is capable of detecting abnormal events while keeping false alarm rate low.

V. CONCLUSION

As we've studied, Anomaly Detection is the task of identifying unusual patterns or events in a given dataset. In the context of spatial temporal data, such as sensor readings over time or location-based data with timestamps, anomaly detection can be framed as a spatial temporal sequence outlier detection problem. This means that we are looking for sequences of data points that are significantly different from the normal patterns observed in the dataset. In this model, we've used semi-supervised learning - which takes long video segments as input containing only normal events in a fixed view. As a result, the outcome obtained in our deep learning model states whether an anomaly is detected or not in a particular video sequence in terms of frames per second. Hence, through this paper, we would like to help our society by implementing a video surveillance system based on our proposed deep learning Anomaly Detection model which could be helpful for security purposes in various sectors such as healthcare, cybersecurity, for military and defence and for private as well as public localities.

VI. REFERENCES

1. Doll´ar, P., Rabaud, V., Cottrell, G., Belongie, S.: Behaviour recognition via sparse spatio-temporal features
2. Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.:Learning temporal regularity in video sequences
3. Mahadevan, V., Li, W., Bhalodia, V., Vasconcelos, N.: Anomaly detection in crowded scenes
4. Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L.: Large-scale video classification with convolutional neural networks.
5. Lu, C., Shi, J., Jia, J.: Abnormal event detection at 150 fps in matlab.
6. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model
7. Patraucean, V., Handa, A., Cipolla, R.: Spatio-temporal video autoencoder with differentiable memory.