International Information and Engineering Technology Association
Advancing the World of Information and Engineering

International Journal of Safety and Security Engineering
Vol. 13, No. 2, April, 2023, pp. 205-211
Journal homepage: http://iieta.org/journals/ijsse

A Review of Deep Learning Algorithms for Anomaly Detection in Videos

Yuvarani Sadatcharam[1], Dinakaran Muruganadam[2*]

[1] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore 632014, India
[2] School of Computer Science and Engineering, Vellore Institute of Techology, Chennai 600127, India

Corresponding Author Email: dinakaran.m@vit.ac.in

https://doi.org/10.18280/ijsse.130203

Received: 12 February 2023
Accepted: 9 April 2023

Keywords:
anomaly detection, video surveillance, detection techniques

ABSTRACT

Anomaly detection in video assists in the resolution of a wide range of problems. A robust anomaly detection model is necessary due to the growing use of surveillance cameras in both indoor and outdoor settings. As a result, numerous strategies have been proposed in this field. Anomaly detection has already been the subject of several surveys. This survey focuses on deep learning approaches based on video anomaly detection. we categorize the various Deep Learning approaches according to their objectives like score based, future frame-based, Clasiification and reconstruction error based approaches. Additionally, it discusses evaluation criteria and commonly used datasets. We also suggest some possible directions future directions for research.

1. INTRODUCTION

Abnormal detection in video is the task of finding anomaly events in the video. Anomalies do not have a confirmed pattern. It deviated from the normal pattern [1]. Anomalies are also called an exception, outliers, abnormalities, and irregularities. For example cars in a pedestrian way, people loitering for a long time, and people running in a different direction. Video anomalies are scene-dependent which means anomalies in one scene may not be anomalies in the next scene. For example, in the pedestrian way car Cyclist is considered an anomaly but in the traffic scene, cars are considered a normal event. Normal videos are used for training purposes. Which scenes deviating from the normal event are an anomaly. Surveillance cameras help to monitor human activity and prevent crime scenes. A human cannot able to watch 24/7 and alter if something is wrong. The unexpected event is not happening regularly [2]. It's a very rare event. If monitoring human activity most of the time video contains normal activity. Difficult to identify an abnormal event. An automated anomaly detection system ensures public safety. Researchers have been focusing to create an algorithm for an unusual event in video.

Video anomaly detection has single-scene videos and multi-scene videos. Single-scene video contains location-dependent [1]. Most of the anomaly datasets are single scenes. For example, a person walking on grass normal event in a pedestrian event, another scene is the restricted area. So dependent on the location the anomaly event will differ in a single scene. Anomaly detection in multi-scene contains normal videos from different scenes. Here the variety of scenes and activity captured in the scene not formulating a single model. Multi-scene video anomaly detection does not apply to single-scene videos where normal and abnormal events are the same for the entire scene. But in a single scene anomaly will vary dependent on location. Available dataset for single scene videos CHUK Avenue, UCSD Ped1, Ped2, UMN (Lawn, Indoor, Plazza), Subway entrance and exit, Street Scene. Multi Scene video datasets are Shanghai Tech and UCF-Crime. In CHUK Avenue and Street Scene, video frames are overlapping with neighborhood frames. But it can be handled by a single scene.

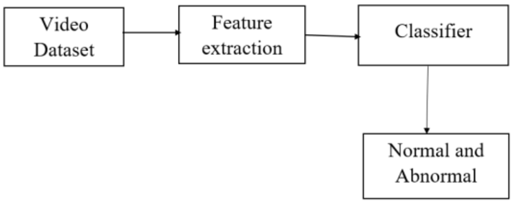

Figure 1. The flow of anomaly detection

Generally, the anomaly is classified into point anomaly, collective anomaly, and contextual anomaly. In-point anomalies data point far away from usual data. Example skater on a pedestrian road. Contextual-based anomaly is dependent on context. The Group of data together leads to an anomaly called a collective anomaly. Example panic event. Anomalies are categorized into short-term motion-only anomalies, appearance-only anomalies, long-term trajectory anomalies, and group anomalies. The unusual object that appears in a scene is called an appearance-only anomaly. Short-term motion only anomalies unexpected objects moving in the scene. Where appearance and short-term anomalies are called local anomalies. Long-term trajectory anomalies unexpected object trajectory in the scene. Unexpected object interaction in the scene is called group anomaly. Figure 1 represents the flow of anomaly detection. In computer vision problems, the detection of anomalies is a challenging task [3]. Research facing challenges in video surveillance.

Lack of anomaly sample in the video: anomaly detection task is different from general classification. Anomaly events are less in training data. It's very difficult to use a supervised algorithm because of data imbalance. Therefore, it is not possible to predict all anomaly events in one model.

Computation and storage: most of the abnormal detection algorithm involves high computational resource. This problem makes the infeasible use of the real world. But real-time anomaly detection is required.

Uncertainty: generally anomalies are unexpected events, where not conform with expected behavior. Here the boundary between usual and unusual events is very thin. In real-world problems, classifying the usual and unusual events are not clearly defined for example some scene considers usual in one video, but in another scene that abnormal.

Noisy data: Video cameras are fixed everywhere to improve security and they are fixed everywhere like parks, shopping malls even personal houses. Collecting video surveillance data is easy, but annotating manually is a time-consuming process. Poor quality of data undoubtedly leads to less accuracy.

The rest of this article is structured as follows, In section 2 describes the publically available dataset for video anomaly detection. Section 3 categorizes various deep learning approaches according to their objectives. Section 4, discusses widely used evolution metrics. Section 5 Discuss the current state of research in this field and make some suggestions for future research directions.

2. DATASETS

The commonly used dataset for video anomaly detection [4, 5] discussed in this section. Existing benchmark dataset shown in Table 1.

UCSD:

The UCSD dataset is widely used for anomaly detection in video. UCSD consist of two subset ped1 and ped2. This dataset consists of footage from the pedestrian road taken from a stationary camera. The dataset includes normal and abnormal events. Normal events are the person walking. Abnormal events are car, cyclist, wheelchair, etc. ped1 consists of 34 training clips and 36 testing clips with 234 ×159 resolution. Ped2 consists of 16 training clips and 12 testing clips with 360 × 240 resolution. Ped1 and ped2 contain frame-level and pixel-level ground truth.

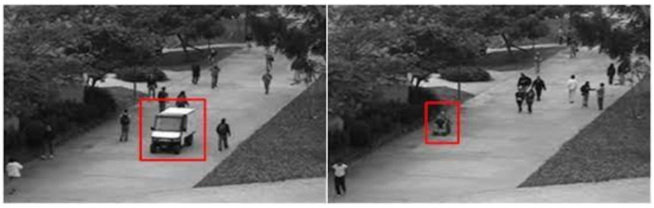

Figure 2. Example of UCSD dataset anomaly

In Figure 2, left side shows the car as an anomaly on the right side person with the wheelchair is an anomaly.

In UMN dataset consist of three different scenes. They are the lawn scene, indoor scene, and plaza scene. Using a stationary camera video has been captured with 30 fps at a resolution of 320× 240. This dataset consists of panic events. Normal vents are people walking in a different direction. Abnormal events are running, and panic events. Sample frame of UMN dataset represent in Figure 3.

CHUK Avenue:

Avenue dataset captured from CHUK Campus Avenue. It consists of 16 training clips and 21 testing clips with 640× 360 resolution. A training video has only normal events, testing has both normal and abnormal events. The dataset has a different camera position and angle from other datasets, camera slightly shacked. Normal events are walking. 47 abnormal events are in the avenue, it includes running, bag dropping, and playing with bags and paper. Figure 4 represent the anomaly in CHUK Avenue dataset.

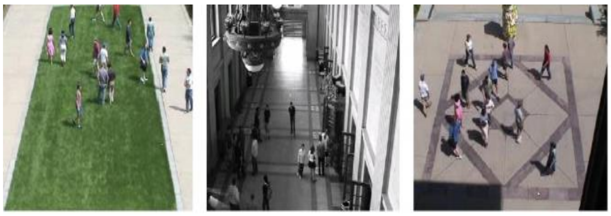

Figure 3. Example of UMN dataset

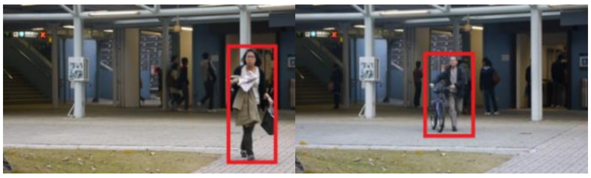

Figure 4. Example of CHUK Avenue dataset

Subway entrance and exit:

The subway dataset consists of two subsets like entrance and exit gates with 512×384 resolution. Here no specific training and testing video. The video has both usual and unusual events. Abnormal events are wrong direction crossing, without payment entering into gate. Example of subaway entacnce dataset represent in Figure 5.

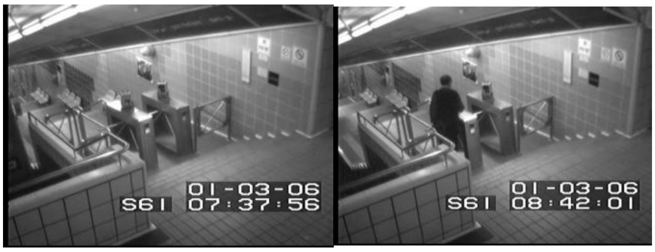

Figure 5. Example of subway dataset

Table 1. Benchmark datasets

Dataset	Total No. of Video	Frames	Resolution	Abnormal activity
UCSD ped1	70	14000	238×158	Car, biker wheelchair, etc
UCSD ped2	28	4560	360×240	Car, biker wheelchair, etc
UMN (Lawn, Indoor, Plaza)	5	1450	320×240	Panic event
CHUK Avenue	37	30652	640×360	Running, throwing an object, running loitering
Subway entrance	1	72401	512 ×384	Loitering, avoiding payment, wrong direction
Subway exit	1	136524	512 ×384	Loitering, avoiding payment, wrong direction
Shanghai Tech	437	317398	856×480	Sudden motion, fighting, running, etc

Dataset	Total No. of Video	Frames	Resolution	Abnormal activity
UCF crime	1900	~13.8M	320×240	Explosion, road accident, assault abuse, etc
Street Scene	81	203257	1280×720	Jaywalking, person exits car

Shanghai Tech dataset:

The Shanghai Tech campus dataset contains 330 training clips and 107 testing clips with 856 × 480. The dataset contains 13 different scenes. Comparing the other dataset Shanghai Tech challenging dataset. The dataset has 10 anomalous events per second. This video has been taken from different camera points and light conditions. The usual event is walking and unusual events like biking and skating. Sample frame of Shanghai Tech dataset represent in Figure 6.

Figure 6. Example of Shanghai Tech dataset

Street Scene:

The most recently added dataset for anomaly detection. The activity of bike lane, pedestrian walking, and two Lane Street was captured in the scene. It consists of 46 training clips and 35 testing clips. It's a challenging dataset because of its variety of activities like bike riding, pedestrian walking, and moving background. Unusual activities are illegal U-turns. Sample frames of Street Scene dataset show in Figure 7.

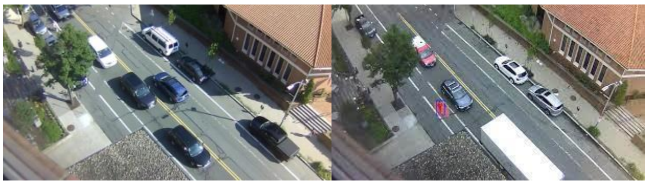

Figure 7. Example of Street Scene

UCF Crime:

UCF crime contains real-world unusual activities like an accident, fighting, stealing, explosion, abuse, etc. total duration of the datasetv120 hours. Datasets are divided into training and testing clips. The training dataset has 810 anomalies events and 800 normal events. In the testing dataset, 150 are usual events and 180 are abnormal events with 240 × 320 resolution. Figure 8 shows anomaly in UCF crime dataset.

Figure 8. Example of UCF crime dataset

3. DEEP LEARNING IN ANOMALY DETECTION FOR VIDEOS

Deep learning algorithms focus to create a new architecture for a specific problem [2]. Most of the algorithms and similar to each other. Because of this categorizing the algorithm based on a final objective like reconstruction error, future frame prediction, score based, and classification. A quick Summary of all these techniques are provided in Table 2.

3.1 Reconstruction error

Reconstruction-based anomaly detection techniques have already been used by the study [6]. The reconstruction error of the normal sample is low or closer to the training data. But for abnormal data reconstruction error is expected high. The special type of neural network which able to reconstruct the original input to compact representation is called an autoencoder. Most of the paper's goal is reconstruction error-based anomaly detection than using autoencoders [7]. Anomaly detection in a crowded scene based on low rank and compact coefficient feature. Feature space extracted by the histogram of optical flow projection. In the training stage coefficient of low rank was obtained by joint optimization of nuclear norm l2,l l2,l norm applied to the testing sample to get reconstruction vectors. Reconstruction error of abnormal sample deviated from a normal sample. Reconstruction cost is introduced in this paper.

3D fully convolution auto encoder used for the detection of spatiotemporal and temporal irregularity in an end-to-end manner [8]. Raw pixels as input to the deep residual conditional generative adversarial network. DR-cGAN learns objects of interest [9]. For training, DRc GAN takes input as the frame of a normal event to give corresponding dense optical flow information. For testing compute the reconstruction error for the local pixel between the synthesized and real optical flow. To remove the false positive rate online hard negative mining is used. Semantic region merging makes the abnormal object a full output frame. Two-stream deep spatial-temporal auto encoder explored for anomaly detection [10]. Spatial stream DSTAE and temporal stream DSTAE extract the appearance and motion pattern respectively. Based on joint reconstruction error fusing the spatial and temporal information and detecting anomalies.

The authors [11] present a hybrid autoencoder. LSTM autoencoder fails to deal with global context anomaly because of fixed dimension representation. A hybrid auto-encoder extracting spatial and temporal information also improves the capability of the decoder using a shortcut connection. Abnormal detection based on reconstruction error. The authors [12] explored work on residual Spatio-temporal autoencoder for abnormal pattern detection. Input video segment passed to residual autoencoder. It consists of Conv LSTM and 3D convolution and deconvolution layer and learning the pattern of normal activities, generating reconstructed video segment. In testing, input passes into the residual spatiotemporal autoencoder and finds reconstruction loss. Normal frames have low reconstruction cost.

3.2 Classification

Classification-based anomaly detection methods are solving the problem of data imbalance. Motion and appearance features are extracted and fed into the classifier for finding an abnormal event. Aggregation of ensemble model used for crowded anomaly detection [13]. The semantic feature is learned using an ensemble of fine-tuned CNN. Different levels of features fed for SVM classifier. Then posterior probability is used to predict anomaly events.

Extracted motion information by background subtraction and finding attention regions [14]. The region is fed into a 3D convolution neural network to classify normal and abnormal events. Pre-trained convolution neural network used for extraction of spatiotemporal features from the sequence of frames [15]. The feature itself captures the anomalies event information also. The feature fed into Bi-directional Long Short-Term Memory to classify abnormal and normal events. Anomaly detection in human behavior is an important computer application problem [16]. Explored work based on the spatiotemporal correlation of gradient-based features. Here discriminative classifier is used for classifying the violent behavior in crowded scenes. It reduces the dimension of activity representation and computation time.

Video converted into the frame and fine tune them as staked grayscale 3 channel image (SG3I) and RGB images [17]. The volume of interest with motion and the volume of interest of relative have were extracted then data augmentation and resizing were done. After that images pass through adoptive pre-trained 2D CNN used for extracting spatial and temporal information. This 2D CNN implements lighter than 2D CNN which achieves high accuracy with low computational cost. The study [18] presented Motion information Images computed by optical flow. MII is based on magnitude, and angle difference computed between optical flow vectors. MII fed into CNN for learning normal and abnormal events.

The authors [19] present work for spatiotemporal cascade autoencoder for anomaly detection. Gradient and optical flow cuboid generated from the raw input video. Then the cuboids pass into STAAE to extract the motion and appearance anomaly score in two streams based. Fuse the appearance and motion abnormality score. Based on the abnormal score remove the normal cuboids. The abnormal cuboids input to the two streams ST CAE and calculate the abnormal score of appearance and motion based on reconstruction loss. Finally, the patches are classified into normal and abnormal events.The authors [20] presented IBaggedFCNet for video anomaly detection. Inception V3 is used for feature extraction. 3 layers fully connected neural network with bagging ensemble generated prediction score for classification. Here the different combinations of the approach are presented like C3D + 3 layer fully connected layer with bagging and inception v3 without bagging.

The authors [21] present a paper on pedestrian crowd detection and segmentation using multisource feature descriptors. Input images divided into grid blocks then compute the appearance feature using multisource feature descriptor namely local binary pattern, Fourier analysis, and gray level co-occurrence matrix. After that concatenating the all-feature vectors. A long feature vector passes into the SVM classifier. In the same process in the testing phase, a long feature vector is generated and mapped into a learned classifier to generate a confidence score. Finally, the Gaussian kernel applies for smoothing the output.

3.3 Future frame detection

The anomaly event not have conformed patterns. The study [22] suggested future frame prediction approaches. A specific type of neural network is a generative adversarial network. Bidirectional retrospective generation adversarial network used for anomaly detection [23]. Combination of bidirectional prediction and retrospective prediction mine help bidirectional temporal information between input and predicted frame. Gradient loss and intensity loss are calculated between the input frame and the predicted frame. Losses are used for spatial constraint. 3D convolution neural network combined with sequence discriminator to capture the long-term temporal relationship between predicted and input frames. Motion and appearance constraints lead to future frame prediction for a normal event. This network can differentiate normal and abnormal events.

TransAnomaly is a combination U-Net and video vision transformer (ViViT) [24]. Here ViViT captures temporal information and is capable of predicting the video. To avoid the influence of irrelevant factors during anomaly detection PSNR was calculated based on a sliding window. The study [25] explored work on anomaly detection using multi-scale features. The consecutive frame of an input passes into the encoding phase. encoding phase has a context module and ConvGRU. The context model used semantic information about the image. Here extracting features more on spatial information. Using optical flow is a consuming process to obtain temporal information instead of the ConvGRU module used. Finally combining the spatial and temporal information calculates the abnormal score and also the spatiotemporal window score between predicted frames and ground truth help to distinguish normal and abnormal events.

Table 2. Summary of method and contribution

Year	Author	Type	Main Contribution
2020	Li et al. [7]	Reconstruction error	Low rank and Compact coefficient dictionary learning
2020	Yan et al. [8]	Reconstruction error	3D FCAE
2022	Ganokratanaa et al. [9]	Reconstruction error	Deep Residual conditional Generative Adversarial Network
2021	Li et al. [10]	Reconstruction error	Two stream deep spatio temporal auto encoder
2021	Deepak et al. [12]	Reconstruction error	Residual spatio temporal auto encoder
2020	Singh et al. [13]	Classification	Fine tuned ConvNet
2020	Nasaruddin et al. [14]	Classification	3DCNN
2021	Ullah et al. [15]	Classification	CNN + bidirectional LSTM
2020	Direkoglu [18]	Classification	Motion Information image and Convlutional Neural Network
2020	Li et al. [19]	Classification	Spatio temporal casecade auto encoder
2020	Zahid et al. [20]	Classification	IBaggedFCNet
2021	Yang et al. [23]	Future frame prediction	Bidirectional Retrospective Generation Adversarial Network
2021	Yuan et al. [24]	Future frame prediction	U-Net + Video Vision Transformer

Year	Author	Type	Main Contribution
2021	Cai et al. [25]	Score + future frame prediction	Multiscale feature + ConvGRU
2019	Li et al. [26]	Future frame prediction	Spatiotemporal Unity Networking
2022	Kim et al. [30]	Anomaly score	Cross U net
2020	Pang et al. [33]	Anomaly score	End to end model

The authors [26] present work on Spatio-temporal based anomaly detection. Its combination U-Net with ConvLSTM.here introduced a new regular score function that calculates prediction error for both current and prediction frames. Based on prediction error events are distinguished between normal and abnormal events. The authors [27] presented video predicted framework for anomaly detection. An auto encoder as a generator and combining dense residual network and self-attention. High quality of future frame predicted using constraints on motion in the video.

3.5 Score based

Some researchers instead of going for frame prediction and reconstruction error, abnormal score-based anomaly detection [28]. The score decides the video or segment's normal or abnormal event. Video anomaly detection has a problem of instability on a different dataset [29]. To overcome this problem explores the work based on bidirectional prediction. Predict the target frames using forward and backward prediction on sub-networks. The loss function is different between the real target frame and the predicted frame. Here also Anomaly score was estimated based on the sliding window scheme. The study [30] explored the work on anomaly detection using the Cross U Net framework. This framework uses two sub-networks, the tired layer output is combined with the corresponding layer, then the output of sub network passes to the input for the next layer. Frame anomaly score estimated by cascade sliding window method.

3D convolutional autoencoder extracts the spatial and temporal features for the normal event of training videos [31]. Feature extraction is based on autoencoders which learn effectively in an unsupervised manner. Group the 3D spatiotemporal feature into a normality cluster. To remove the sparse cluster, one Class SVM classifier was used to classify between usual and unusual events based on the normality score. The study [32] explored work on building a defense mechanism to detect an abnormal event in an adversarial attack. A deep auto-encoder is used to extract spatial and temporal features from raw data. The reconstruction of video volume performed on learned features. Then structural distortion-based abnormal score is generated. Abnormal data lies in high distortion space compared to the normal event.

The study [33] explored the work of highly dependent on manually labeled data for training in anomaly detection. Initially unlabeled frames input and generated pseudo normal and abnormal frames. Then pass into pre-trained ResNet-50 to extract the features finally fed into a fully connected layer in an end-to-end manner and compute the anomaly score of all frames. The process is repeated and updates the anomaly score. The authors [34] presented a paper for anomaly detection based on temporally coherent sparse coding. Here temporal coherent preserves the similarity between frames. In this work, feature extraction is done by ResNet with multi patches at multi scales. Temporal and spatial features of 21 patches fed into anomaly detection module where temporal, spatial Special Recurrent Neural Network(SRNN) generate the normal score.

4. EVOLUTION METRICS

In evaluation metrics section discussing widely used evolution metrics of the paper that has been presented in this paper. Generally, two criteria metrics are followed in most of the papers. It was introduced in the study [35]. The first criterion is a frame-level criterion. These metrics determine by using temporal labels. The next criteria are a pixel-level criterion. Some of the paperwork with pixel and frame level criteria. Both criteria use the area under the curve (AUC) of the receive operating characteristic curve (ROC) to compute the final performance of the model. The true positive rate and the false positive rate is mentioned in Eqns. (1) and (2) [36]. Eq. (1) gives the proportion of correct predictions in the positive class. Eq. (2) gives the proportion of incorrect prediction in the positive class.

$$\text{True Positive Rate} = \frac{\text{No.of true positives frames}}{\text{True Positive+False Negative}} \quad (1)$$

$$\text{False Positive Rate} = \frac{\text{No.of false positives frames}}{\text{False Positive+True Negative}} \quad (2)$$

In ROC plot y-axis denote true positive rate and x-axis denote false positive rate. The values of each point are taken differently from the classification threshold. A higher AUC of ROC value indicates that the model is performing well for the problem. The main advantages of metrics include scale-invariant and threshold invariants. It is not considering absolute value for prediction and finding how the prediction is performing and ranking. The Strength also acts as the weakness scale invariant not performing well if well-accurate probabilities occurred. It does not work with optimizing metrics. The equal error rate (EER) is computed over the ROC. EER tells the misclassification frame in percentages if the false positive rate is equal to the miss rate. In frame level criterion use False positive rate= 1- True positive rate. If it’s the pixel level criterion will be 1-EER. In the case of frame level criterion, the algorithm considers it as correct when the anomalous pixel not overlapping with the spatial region. Pixel level criterion also does not consider the overlapping with ground truth. So the study [37] came up with a new metric for evaluation. They proposed track-based detection and region-based detection criterion for object tracking and detection. The track-based detection computes the false positive rate per frame as opposed to track based detection rate (TBDR) mentioned in Eqns. (3) and (4).

$$\text{Track} - \text{based detection rate} = \frac{\text{No.of anomalous tracks detected}}{\text{Total no of anomalous tracks}} \quad (3)$$

$$\text{False Positive Rate} = \frac{\text{No.of false positive regions}}{\text{Total No.of frames}} \quad (4)$$

The false positive region per frame is used to measure the region-based detection criterion as opposed to the region-based detection rate (RBDR) Ramachandra and Jones (2020). RBDR define in Eq. (5).

$$\text{Region} - \text{based detection rate} = \frac{\text{No.of anomalous regions detected}}{\text{Total no of anomalous regions}} \quad (5)$$

Note the anomalous tracks are correctly detected if ground truth has an intersection over union (IoU) above the threshold α [2]. Likewise, the region was predicted as anomalous in a false positive frame. If ground truth has IoU above the threshold of β.

5. DISCUSSION

This paper discussed the different methodologies to find an anomaly. Several methods are simple and complex architecture. Eventually, anomaly detection hard task to identify. Different methods grouped like reconstruction error, future frame prediction, score-based, and classification [2]. The variety of approaches presented by the researcher explored different techniques for an anomaly. Referring to all the paper common thing is appearance and motion information. In the research community, extracting spatial and temporal features plays a vital role in anomaly detection. Most of the paper uses the deep learning model which automatically learns the feature. Research still focuses on a robust feature extraction model.

Research concentrating on end-to-end model creation. Instead of using the separate component for feature extraction and classification use end-to-end fashion to detect anomalies. The main advantage of the end-to-end model easily applicable to a real-life problem. The actual pipeline model is very difficult to use in the real-life problem. An end-to-end model needs a large amount of dataset to implement but older datasets pedestrian (UCSD), and panic event (UMN) have less amount of data. So difficult to implement in older data. This problem was solved by the study [5, 37]. An important issue with video datasets, it is a hard task to annotate and collect the data, and that is the reason researchers are not concentrating on creating large data. This problem insisted to go for an unsupervised or weakly supervised approach.

In the evolution, the authors [37] presented an evolution metric for frame level and pixel level criterion but it's not taken as the performance of the model due to the reason mentioned in the paper. Here need for a more robust evolution metric that would more effective to use in the future. Spatial aspect evolution metrics are needed in the future to understand which frames cause the anomalies.

The anomaly detection research field developed a lot in terms of better results and different methodologies. Incorporating the Spatio-temporal information and archiving excellent results. But real-life applications need the extract robust spatiotemporal information. Here seen definition of anomaly varies concerning the author, anomaly also varies with the respective context. If creating a larger dataset with real-life scenes or events. Further adding different fields to the problem is helpful to identify the anomaly.

6. CONCLUSION

This paper discussed recent technique for anomaly detection in video. This categorizes the techniques based on the final step to identifying anomalies like future frame detection, scoring model, classification, and reconstruction error. Here additionally provided the benchmark dataset which was recently used for anomaly detection. If increase the size of the data its can apply to real-life scenarios. The main issue with a large dataset is annotation. Because of the problem, research uses the unsupervised and weakly supervised model, and it helps to identify the anomaly with a small amount of learning.

The future scope of research might include robust feature extraction modeling on spatial and temporal, studying the recent publishing paper for large datasets and creating an end-to-end model.

REFERENCES

[1] Ramachandra, B., Jones, M.J., Vatsavai, R.R. (2020). A survey of single-scene video anomaly detection. IEEE transactions on pattern analysis and Machine intelligence, 44(5): 2293-2312. https://doi.org/10.1109/TPAMI.2020.3040591

[2] Suarez, J.J.P., Naval Jr, P.C. (2020). A survey on deep learning techniques for video anomaly detection. arXiv preprint arXiv:2009.14146. https://doi.org/10.48550/arXiv.2009.14146

[3] Ren, J., Xia, F., Liu, Y., Lee, I. (2021). Deep video anomaly detection: Opportunities and challenges. In 2021 international conference on data mining workshops (ICDMW), Auckland, New Zealand, pp. 959-966. https://doi.org/10.1109/ICDMW53433.2021.00125

[4] Patil, N., Biswas, P.K. (2016). A survey of video datasets for anomaly detection in automated surveillance. In 2016 Sixth International Symposium on Embedded Computing and System Design (ISED), Patna, India, pp. 43-48. https://doi.org/10.1109/ISED.2016.7977052

[5] Sultani, W., Chen, C., Shah, M. (2018). Real-world anomaly detection in surveillance videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6479-6488.

[6] Popoola, O.P., Wang, K. (2012). Video-based abnormal human behavior recognition—A review. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42(6): 865-878. https://doi.org/10.1109/TSMCC.2011.2178594

[7] Li, A., Miao, Z., Cen, Y., Zhang, X. P., Zhang, L., Chen, S. (2020). Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning. Pattern Recognition, 108: 107355. https://doi.org/10.1016/j.patcog.2020.107355

[8] Yan, M., Meng, J., Zhou, C., Tu, Z., Tan, Y.P., Yuan, J. (2020). Detecting spatiotemporal irregularities in videos via a 3D convolutional autoencoder. Journal of Visual Communication and Image Representation, 67: 102747. https://doi.org/10.1016/j.jvcir.2019.102747

[9] Ganokratanaa, T., Aramvith, S., Sebe, N. (2022). Video anomaly detection using deep residual-spatiotemporal translation network. Pattern Recognition Letters, 155: 143-150. https://doi.org/10.1016/j.patrec.2021.11.001

[10] Li, T., Chen, X., Zhu, F., Zhang, Z., Yan, H. (2021). Two-stream deep spatial-temporal auto-encoder for surveillance video abnormal event detection. Neurocomputing, 439: 256-270. https://doi.org/10.1016/j.neucom.2021.01.097

[11] Zhou, F., Wang, L., Li, Z., Zuo, W., Tan, H. (2020). Unsupervised learning approach for abnormal event detection in surveillance video by hybrid autoencoder.

Neural Processing Letters, 52: 961-975. https://doi.org/10.1007/s11063-019-10113-w

[12] Deepak, K., Chandrakala, S., Mohan, C.K. (2021). Residual spatiotemporal autoencoder for unsupervised video anomaly detection. Signal, Image and Video Processing, 15(1): 215-222. https://doi.org/10.1007/s11760-020-01740-1

[13] Singh, K., Rajora, S., Vishwakarma, D.K., Tripathi, G., Kumar, S., Walia, G.S. (2020). Crowd anomaly detection using aggregation of ensembles of fine-tuned convnets. Neurocomputing, 371: 188-198. https://doi.org/10.1016/j.neucom.2019.08.059

[14] Nasaruddin, N., Muchtar, K., Afdhal, A., Dwiyantoro, A.P.J. (2020). Deep anomaly detection through visual attention in surveillance videos. Journal of Big Data, 7(1): 1-17. https://doi.org/10.1186/s40537-020-00365-y

[15] Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M., Baik, S.W. (2021). CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. Multimedia Tools and Applications, 80: 16979-16995. https://doi.org/10.1007/s11042-020-09406-3

[16] Deepak, K., Vignesh, L.K.P., Chandrakala, S.J.I.E. (2020). Autocorrelation of gradients based violence detection in surveillance videos. ICT Express, 6(3): 155-159. https://doi.org/10.1016/j.icte.2020.04.014

[17] Mehmood, A. (2021). Efficient anomaly detection in crowd videos using pre-trained 2d convolutional neural networks. IEEE Access, 9: 138283-138295. https://doi.org/10.1109/ACCESS.2021.3118009

[18] Direkoglu, C. (2020). Abnormal crowd behavior detection using motion information images and convolutional neural networks. IEEE Access, 8: 80408-80416. https://doi.org/10.1109/ACCESS.2020.2990355

[19] Li, N., Chang, F., Liu, C. (2020). Spatial-temporal cascade autoencoder for video anomaly detection in crowded scenes. IEEE Transactions on Multimedia, 23: 203-215. https://doi.org/10.1109/TMM.2020.2984093

[20] Zahid, Y., Tahir, M.A., Durrani, N.M., Bouridane, A. (2020). IBaggedFCNet: An ensemble framework for anomaly detection in surveillance videos. IEEE Access, 8: 220620-220630. https://doi.org/10.1109/ACCESS.2020.3042222

[21] Basalamah, S., Khan, S.D. (2020). Pedestrian crowd detection and segmentation using multi-source feature descriptors. International Journal of Advanced Computer Science and Applications, 11(1): 707-713. https://doi.org/10.14569/IJACSA.2020.0110187

[22] Liu, W., Luo, W., Lian, D., Gao, S. (2018). Future frame prediction for anomaly detection–a new baseline. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6536-6545. https://doi.org/10.1109/CVPR.2018.00684

[23] Yang, Z., Liu, J., Wu, P. (2021). Bidirectional retrospective generation adversarial network for anomaly detection in videos. IEEE Access, 9: 107842-107857. https://doi.org/10.1109/ACCESS.2021.3100678

[24] Yuan, H., Cai, Z., Zhou, H., Wang, Y., Chen, X. (2021). TransAnomaly: Video anomaly detection using video vision transformer. IEEE Access, 9: 123977-123986. https://doi.org/10.1109/ACCESS.2021.3109102

[25] Cai, Y., Liu, J., Guo, Y., Hu, S., Lang, S. (2021). Video anomaly detection with multi-scale feature and temporal information fusion. Neurocomputing, 423: 264-273. https://doi.org/10.1016/j.neucom.2020.10.044

[26] Li, Y., Cai, Y., Liu, J., Lang, S., Zhang, X. (2019). Spatio-temporal unity networking for video anomaly detection. IEEE Access, 7: 172425-172432. https://doi.org/10.1109/ACCESS.2019.2954540

[27] Zhang, W., Wang, G., Huang, M., Wang, H., Wen, S. (2021). Generative adversarial networks for abnormal event detection in videos based on self-attention mechanism. IEEE Access, 9: 124847-124860. https://doi.org/10.1109/ACCESS.2021.3110798

[28] Landi, F., Snoek, C.G., Cucchiara, R. (2019). Anomaly locality in video surveillance. arXiv preprint arXiv:1901.10364. https://doi.org/10.48550/arXiv.1901.10364

[29] Chen, D., Wang, P., Yue, L., Zhang, Y., Jia, T. (2020). Anomaly detection in surveillance video based on bidirectional prediction. Image and Vision Computing, 98: 103915. https://doi.org/10.1016/j.imavis.2020.103915

[30] Kim, Y., Yu, J., Lee, E., Kim, Y. (2022). Video anomaly detection using Cross U-Net and cascade sliding window. J. King Saud Univ. Comput. Inf. Sci., 34, 3273-3284.

[31] Wu, R., Li, S., Chen, C., Hao, A. (2021). Improving video anomaly detection performance by mining useful data from unseen video frames. Neurocomputing, 462: 523-533. https://doi.org/10.1016/j.neucom.2021.05.112

[32] Sharma, M.K., Sheet, D., Biswas, P.K. (2020). Spatiotemporal deep networks for detecting abnormality in videos. Multimedia Tools and Applications, 79: 11237-11268. https://doi.org/10.1007/s11042-020-08786-w

[33] Pang, G., Yan, C., Shen, C., Hengel, A.V.D., Bai, X. (2020). Self-trained deep ordinal regression for end-to-end video anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12173-12182. https://doi.org/10.1109/CVPR42600.2020.01219

[34] Luo, W., Liu, W., Lian, D., Tang, J., Duan, L., Peng, X., Gao, S. (2019). Video anomaly detection with sparse coding inspired deep neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(3): 1070-1084. https://doi.org/10.1109/TPAMI.2019.2944377

[35] Li, W., Mahadevan, V., Vasconcelos, N. (2014). Anomaly detection and localization in crowded scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(1): 18-32. https://doi.org/10.1109/tpami.2013.111

[36] Bradley, A.P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognition, 30(7): 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

[37] Ramachandra, B., Jones, M. (2020). Street scene: A new dataset and evaluation protocol for video anomaly detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2569-2578. https://doi.org/10.1109/wacv45572.2020.9093457