PLOMS Journal of
Artificial Intelligence
(PLOMS AI)

PLOMS AI

Review Paper

A Study of Crowd Abnormal Events Understanding in Surveillance Videos

Mousumi Yeasmin Benzir
Department of Computer Engineering, Izmir Institute of Technology, Turkey
Correspondence: benzir247@gmail.com
Received: 15-07-2022; Accepted: 16-08-2022; Published: 09-09-2022

Abstract: Crowd abnormal events detection in surveillance videos is a common topic in computer vision. For better security and safety, automatic video surveillance systems can detect and record abnormal activities at public and private places. However, traditional methods based on optical flow or segmentation cannot show good detection performance. On the other hand, deep learning based solutions for crowd unusual events detection showed better performance than those of conventional machine learning. This paper includes the latest deep learning models for crowd abnormal events detection in surveillance videos and their overall performance study.

Keywords: abnormal events; deep learning; optical flow; surveillance videos

I. Introduction

The detection of both abnormal (e.g., [1]–[11]) and normal (e.g., [12]–[14]) video events is one of the main targets of a surveillance camera system. Surveillance systems can detect and track objects using either laser scanned data points [15]–[20] or videos [21]–[24]. Automatic video surveillance systems are highly expected, as we do not need to manually monitor the abnormal crowd events. Nowadays, approaches of deep learning achieved far significant advances than those of traditional for detecting crowd abnormal activities using videos from surveillance systems. Deep learning approaches work on multiple layers of artificial neural network to empower machines for making decisions. Although detection of abnormal activities of crowd in real-world surveillance videos is very important, it is a challenging task as the prior knowledge about the anomalies is normally extremely limited. Besides, there is no common explanation for abnormal events and it is commonly depended on the scene under consideration. To take these challenges, a great number of deep learning based approaches were proposed in the literature during last decade. Accordingly, many surveys have already been conducted on the basis of those methods. For examples, Afiq et al. [25] performed a review on classifying abnormal behavior in crowd scene; Khan et al. [26] demonstrated the seminal research works on crowd management; Suarez et al. [27] presented a survey of deep learning solutions for anomaly detection in surveillance videos; and Braham et al. [28] did a comparative study for crowd event analysis.

However, there is a lack of study with the most recent approaches in those surveys. This study aims to give an extra insight of the most recent deep learning based crowd anomaly detection methods.

The rest of this study follows as: Section II briefs several crowd datasets; Section III bespeaks on various crowd anomaly detection methods; Section IV hints key research challenge; and Section V concludes the paper.

II. Most Common Crowd Datasets

There exist various crowd datasets to detect abnormal activities from videos, among them most famous datasets are UCSD (University of California San Diego) [29], UMN [30], Subway [31], ImageNet [32], CUHK (Chinese University of Hong Kong) Campus Avenue [33], ShanghaiTech Campus [34], and UCF-Crime [35].

- The UCSD dataset was recorded from a stationary camera. This dataset is divided into 2 subsets called Pedestrian 1 (Ped1) and Pedestrian 2 (Ped2). In Ped1, there exists an acute angle between the camera view and sidewalk, and the camera height is lower than that in Ped2 [36]. Abnormal activities are bikers, skaters, carts, wheelchairs, and people walking off the pedestrian ways.

- The UMN dataset is one of the crowd abnormal activity testing datasets from the University of Minnesota. It is a synthetic dataset [37]. The aim of this dataset is to correctly detect the change in the movement of the crowd. In each video, motion pattern is completely unstructured [38]. An anomaly is indicated if everyone starts running instantaneously.

- Subway dataset was obtained by two cameras in an underground train station. This dataset has two long videos for subway-entrance and subway-exit scenes. Both videos are annotated at frame-level and have similar types of anomalies, which are wrong direction walking, loitering, and avoiding payment [33].

- ImageNet dataset consists of over 15 million labeled highresolution images belonging to approximately 22000 categories with variable-resolution. The images were collected from the online and labeled by human labelers using Amazon's Mechanical Turk crowd-sourcing tool.

- CUHK-Avenue dataset was recorded at CUHK Campus Avenue. The 16 training videos capture normal cases, whereas 21 testing videos include both normal events and abnormal cases marked in rectangles. The abnormal events are running, walking in opposite directions, throwing objects, and loitering [39].

- ShanghaiTech Campus dataset was collected from ShanghaiTech University campus considering 13 different scenes with various lighting conditions and camera angles. This dataset has 130 abnormal events in 13 scenes [40]. It is one of the massive and most challenging datasets available for anomaly detection in videos [41].

- UCF-Crime dataset consists of 1900 long and untrimmed real-world surveillance videos. Unusual activities are abuse, arrest, arson, assault, road accident, burglary, explosion, fighting, robbery, shooting, stealing, shoplifting, and vandalism [42].

III. Methods of Abnormal Video Event Detection

Abnormality detection is commonly termed as an outlier detection problem. The methods of crowd abnormal events can be divided into two primary groups namely traditional and deep learning as shown in Fig. 1.

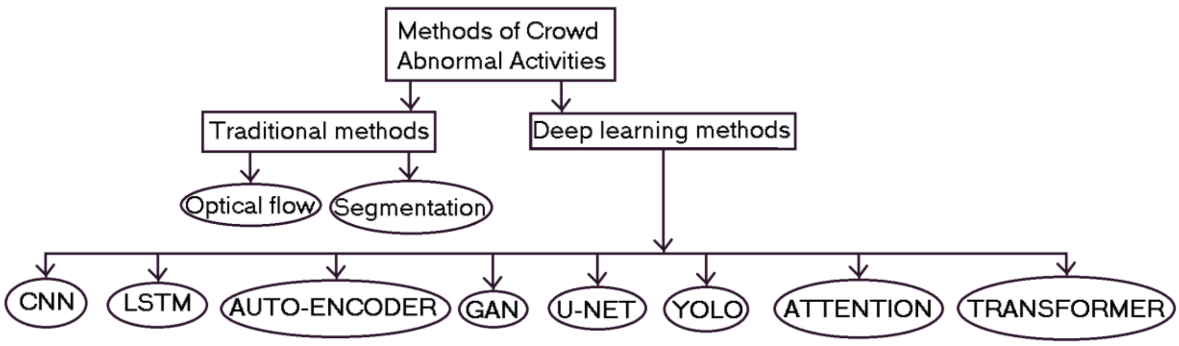

Figure 1. Classification of abnormal video event detection methods.

A. *Traditional Methods*

Traditional methods usually use optical flow and/or segmentation based techniques.

1. *Optical flow based methods*
A lot of crowd abnormal activity detection methods are based on optical flow technique. For example, Ihaddadene et al. [1] presented a tool that automatically detected abnormal situations in crowded scenes in real time. Their approach analyzed the general motion aspect, instead of tracking subjects one by one, by detecting abnormal optical flow patterns of tracked KLT points. Mehran et al. [6] introduced a method for detecting and localizing abnormal behaviors in crowd videos using Social Force model. They used a grid of particles, which was placed over the image and it was advected with the space-time average of optical flow. Sharif et al. [43] suggested an approach to detect an abnormal situation in a crowd scene. Their approach estimated sudden changes and abnormal motion variations in a set of interest points. The number of tracked points of interest was reduced by using a mask that corresponds to the hot areas of the built motion heat map. Optical flow technique tracked the points of interest. There were sufficient variations in the optical flow patterns in a crowd scene in case of abnormal situations.

2. Segmentation based methods
Optical flow can be unreliable and global comparisons of optical flow can lead to erroneous results. When optical flow representations are not powerful enough to detect anomalous occurrences, segmentation based methods can be used. For example, Mahadevan et al. [44] considered three properties for the design of a localized video representation suitable for anomaly detection in such scenes: (1) joint modeling of appearance and dynamics of the scene, and the abilities to detect, (2) temporal, and (3) spatial abnormalities. Their model for normal crowd behavior was based on mixtures of dynamic textures and outliers under their model were labeled as anomalies.

B. *Deep Learning Based Methods*

There exist various kinds of deep learning based methods used in crowd abnormality detection.

1. CNN-based Methods
CNN was coined by Yann LeCun in the 1980s. Nowadays, a CNN is a very popular model in computer vision. It is chiefly consisted of convolution layers, activation function, pooling layers, and fully connected layers. There are two well-known options in CNN during training images. First option is to train the domain specific problem statement from the scratch. The second option is to use pre-trained model, which is usually called the transfer learning [45]. Hyperparameters of CNN are variables including the number of hidden layers, the learning rate, the batch size or the number of epochs. To select a suitable CNN model is important in the trained model [46]. Adam optimizer [47] is frequently used for CNN. Fine-tuning takes a pre-trained model for a fixed task and then tweaking it to make it performing another similar job. For example, Singh et al. [48] utilized an ensemble of different fine-tuned CNNs based on the hypothesis that dissimilar CNN models learn many levels of semantic. Zahid et al. [49] utilised videos into 60 frame-clips to localize abnormality considering Inception-v3 [50] along with a pretrained feature extractor of 3DCNN [51]. Hu et al. [52] applied a pre-trained 3D VGGNet16 [53] model to detect and localize abnormality from crowd scenes. Hao et al. [36] used 3D ResNet [54] to detect crowd video abnormal activities.

2. LSTM-based Methods
An LSTM keeps unique units called memory blocks in the recurrent hidden layer. Each memory block in the original architecture contained an input gate and an output gate. The input gate manages the flow of input activations into the memory cell. The output gate supervises the output flow of cell activations into the rest of the network. However, the forget gate was attached to the memory block [55]. LSTM can be more suitable for temporal information modeling. For example, Xia et al. [56] used LSTM [57] to decode the historical feature sequences with temporal attention for predicting the

features. Moustafa et al. [58] utilized a LSTM based approach for pathway and crowd anomaly detection, where crowd scene was divided into a number of static overlapped spatial regions.

3. Auto-Encoder-based Methods

Auto-Encoder is used to learn efficient codings of unlabeled data in deep models for transfer learning. It consists of two main parts called an encoder and a decoder. The encoder depicts the input into the code, whereas the decoder uses the code to a demonstrate of the input. For unsupervised anomaly detection cases, the auto-Encoder was trained on normal activities by reducing their reconstruction error [59], and then, the thresholded reconstruction error was applied for recognizing anomalies. The reconstruction error can be low for the normal activities, but the reconstruction error becomes high for the abnormal activities [60], [61]. Auto-Encoder can be used in 2D or 3D applications [60], [62], [63].

4. GAN-based Methods

The GAN [64] contains two adversaries named Generator and Discriminator. The generator considers noise as input and generates samples. The discriminator gets samples from the generator as well as training data. It should differentiate two data source. In the training phase, the generator learns to produce a sample that is close to its ground truth. The discriminator learns how to distinguish the generated data from its ground truth. Usually, GAN models are popular for image generation and video prediction, more specifically in anomaly detection [65]. Wang et al. [66] used the generation error of a generative neural network to detect anomalies. Chen et al. [67] utilized an end-to-end pipeline named noisemodulated GAN for video anomaly detection. Tang et al. [68] used the PatchGAN discriminator [69] to predict the broad locations of abnormal events. Zhong et al. [70] used a kind of P-GAN [71] for anomaly detection in videos.

5. U-Net-based Methods

A U-Net is a U-shaped model transformed from a fully convolutional network [72]. Ronneberger et al. [73] introduced the first classical U-Net for biomedical image segmentation. U-Net has a great role in frame prediction. The consecutive frames of one clip of surveillance video normally have the same background and the similar foreground [74]. Park et al. [75] used a U-Net [73] to skip connections between the encoder and the decoder boost generation ability by preventing gradient vanishing and achieving information symmetry. Chen et al. [74] applied a U-Net [73] based bidirectional prediction model for anomaly detection.

6. YOLO-based Methods

The YOLO (You Only Look Once) [76] is a pre-trained object detection tool [77]. It can process many frames per second on a GPU. It can provide the same or even better accuracy as compared to ResNet [78]. YOLO has several versions. YOLOv3 [79] detector was applied to extract patches from current-frame. Shine et al. [80] selected anomaly candidates by analyzing 14 background frames per video using YOLOv3 detector [79]. Doshi et al. [81] got bounding box (location) and the class probabilities (appearance) for each object detected in a given frame using YOLOv4 [76].

7. Attention-based Methods

Attention mechanism can rapidly extract key features from small amounts of data [82]. Attention-based model helps to perform the neural network dynamically shift so that the overall decision making can be more reliable [83]. Recently, attention-based methods are applied in many computer vision based applications for image segmentation [84] and classification [85]. Zhou et al. [83] proposed an attention map by putting together mask map and background for anomaly detection in video surveillance.

8. Transformer-based Methods

Vaswani et al. [86] used transformer-based method to solve sequence-to-sequence tasks. Feng et al. [87] demonstrated a convolutional transformer for predicting future frame based on past frames in video anomaly detection. Yuan et al. [88] used the video vision transformer [89] for video prediction.

Table 1. Summary of deep learning based crowd abnormality detection methods.

Reference	Method	Dataset	Mean ACC	Mean AUC
Singh [48]	CNN-based	UCSD [29], CUHK Avenue [33]	92.7%	0.923
Zahid et al. [49	CNN-based	UCSD [29], CUHK Avenue [33], ShanghaiTech Campus [34]	—	0.765
Hu et al. [52]	CNN-based	UCSD [29], UMN [30]	—	0.965
Hao et al. [36]	CNN-based	UCSD [29], CUHK Avenue [33], ShanghaiTech Campus [34]	—	0.850
Xia et al. [56]	LSTM-based	UCSD Ped2 [29], CUHK-Avenue [33]	—	0.911
Moustafa et al. [58]	LSTM-based	UMN [30]	—	0.965
Shi et al. [59]	Auto-Encoder-based	6000 trajectories	90%	—
Asad et al. [63]	Auto-Encoder-based	UCSD [29], CUHK Avenue [33], etc.	—	0.888
Yang et al. [62]	Auto-Encoder-based	UCSD [29], CUHK-Avenue [33], etc.	—	0.912
Wang et al. [66]	GAN-based	UCSD [29], CUHK-Avenue [33]	—	0.919
Chen et al. [67]	GAN-based	UCSD [29], CUHK-Avenue [33]	—	0.891
Tang et al. [68]	GAN-based	UCSD [29], CUHK Avenue [33], ShanghaiTech Campus [34]	—	0.835
Zhong et al. [70	GAN-based	UCSD [29], CUHK Avenue [33], ShanghaiTech Campus [34]	—	0.849
Park et al. [75]	U-Net-based	UCSD Ped2 [29], CUHK Avenue [33], ShanghaiTech Campus [34]	—	0.840
Chen et al. [74]	U-Net-based	UCSD [29], CUHK Avenue [33]	—	0.904
Doshi et al. [81]	YOLO-based	UCSD Ped2 [29], CUHK Avenue [33], ShanghaiTech Campus [34], etc.	—	0.780
Zhou et al. [83]	Attention-based	UCSD [29], CUHK-Avenue [33]	—	0.887
Feng et al. [87]	Transformer-based	UCSD Ped2 [29], CUHK Avenue [33], ShanghaiTech Campus [34]	—	0.869
Yuan et al. [88]	Transformer-based UCSD [29]	CUHK-Avenue [33]	—	0.900

Table 1 makes a short description of the deep learning based crowd abnormality detection methods, where ACC and AUC represent accuracy and area under the receiver operating characteristic curve, respectively. CNN-based models are relatively easy to realize and quick to implement, while Transformer-based models are new for crowd abnormality detection. Based on the datasets and training conditions, the performance scores of ACC and AUC might be varied. As a result, specially the mean AUC scores of the models in Table 1 are accepted for many applications of computer vision and pattern recognition.

IV. Existing Open Challenges

Although deep learning based solutions for crowd abnormal activity detection showed significantly better than traditional solutions, various challenges exist as a huddle in this research area. Some common challenges are discussed below.

- *Definition of crowd abnormal event*: The definition of abnormal event is totally subjective. Based on the time and place, the same event can be normal or abnormal. This is one of the severe challenges for crowd abnormal activity detection.
- *Less number of datasets*: Deep learning methods need a lot training data, but the existing datasets are not enough to do accurate training or testing.
- *Lack of computing power*: Crowd abnormal activity detection methods need to process huge amount of video data, but the accessible GPU processing is generally less.
- *Low quality of videos*: Because of the long distance of cameras, the produced videos of political rally, religious events, and airport arrival are small and hence the quality of video is sometimes very poor.
- *Short video segments*: There is a common assumption that each test video segment consists of an abnormal activity. For this assumption, the length of the video segments should be as long as possible. But the video segments of many existing benchmark datasets are a few minutes long only.

V. Conclusions

This paper discussed the most advanced deep learning models for crowd abnormal events detection in surveillance videos. The performance of models was studied. No single model achieved absolute performance due to many dimensional challenges. Common challenges were highlighted.

Acknowledgment: This work was a part of the author's BSc course entitled CME499 Industrial Practice II.

VI. References

[1]. N. Ihaddadene, M. H. Sharif, and C. Djeraba, "Crowd behaviour monitoring", in Proceedings of the 16th International Conference on Multimedia, Vancouver, British Columbia, Canada, October 26-31, 2008, pp. 1013–1014..

[2]. M. H. Sharif, N. Ihaddadene, and C. Djeraba, "Covariance matrices for crowd behaviour monitoring on the escalator exits", in Advances in Visual Computing, 4th International Symposium (ISVC), Las Vegas, NV, USA, vol. 5359, 2008, pp. 470–481.

[3]. M. H. Sharif and C. Djeraba, "A simple method for eccentric event espial using Mahalanobis metric", in Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 14th Iberoamerican Conference on Pattern Recognition (CIARP), Guadalajara, Jalisco, Mexico, vol. 5856, 2009, pp. 417–424.

[4]. S. Mahmoudi, M. Sharif, N. Ihaddadene, and C. Djeraba, "Abnormal event detection in real time video", in International Workshop on Multimodal Interactions Analysis of Users in a Controlled Environment (ICMI), 2008, pp. 1–4.

[5]. M. H. Sharif, N. Ihaddadene, and C. Djeraba, "Covariance matrices for crowd behaviour monitoring on the escalator exits", in Advances in Visual Computing, 4th International Symposium (ISVC), Las Vegas, NV, USA, vol. 5359, 2008, pp. 470–481.

[6]. R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model", in CVPR, Miami, Florida, USA, 2009, pp. 935–942.

[7]. M. H. Sharif and C. Djeraba, "Exceptional motion frames detection by means of spatiotemporal region of interest features", in Proceedings of the International Conference on Image Processing (ICIP), Cairo, Egypt, 2009, pp. 981–984.

[8]. M. H. Sharif, S. Uyaver, and C. Djeraba, "Crowd behavior surveillance using Bhattacharyya distance metric", in Second International Symposium on Computational Modeling of Objects Represented in Images (CompIMAGE), Buffalo, NY, USA, May 5-7, 2010, pp. 311–323.

[9]. M. H. Sharif, N. Ihaddadene, and C. Djeraba, "Finding and indexing of eccentric events in video emanates", J. Multim., vol. 5, no. 1, pp. 22–35, 2010.

[10]. M. H. Sharif and C. Djeraba, "An entropy approach for abnormal activities detection in video streams", Pattern Recognition, vol. 45, no. 7, pp. 2543-2561, 2012.

[11]. M. H. Sharif, "An eigenvalue approach to detect flows and events in crowd videos", Journal of Circuits, Systems, and Computers, vol. 26, no. 7, pp. 1750110:1-50, 2017.

[12]. M. H. Sharif and C. Djeraba, "PedVed: Pseudo Euclidian distances for video events detection", in Advances in Visual Computing, 5th International Symposium (ISVC), ser. LNCS, vol. 5876, 2009, pp. 674–685.

[13]. M. H. U. Sharif, A. K. Saha, K. S. Arefin, and M. H. Sharif, "Event detection from video streams", International Journal of Computer and Information Technology, vol. 1, no. 1, pp. 108–114, 2011.

[14]. M. H. U. Sharif, S. Uyaver, and M. H. Sharif, "Ordinary video events detection", in Computational Modelling of Objects Represented in Images - Fundamentals, Methods and Applications III, Third International Symposium (CompIMAGE), Rome, Italy. CRC Press, 2012, pp. 19–24.

[15]. F. Galip, M. Caputcu, R. H. Inan, M. H. Sharif, A. Karabayir, S. Kaplan, M. Ozuysal, B. Sengoz, A. Guler, and S. Uyaver, "A novel approach to obtain trajectories of targets from laser scanned datasets", in 18th International Conference on Computer and Information Technology (ICCIT), 2015, pp. 231–236.

[16]. F. Galip, M. H. Sharif, M. Caputcu, and S. Uyaver, "Recognition of objects from laser scanned data points using SVM" in the First International Conference on Multimedia and Image Processing (ICMIP), 2016, pp. 28-35.

[17]. M. Sharif, H. Shehu, F. Galip, I. Ince, and H. Kusetogullari, "Object tracking from laser scanned dataset", International Journal of Computer Science Engineering Techniques, vol. 3, no. 6, pp. 19–27, 2019.

[18]. M. H. Sharif, "Particle filter for trajectories of movers from laser scanned dataset", in Third Mediterranean Conference on Pattern Recognition and Artificial Intelligence (MedPRAI), ser. Communications in Computer and Information Science, vol. 1144. Springer, 2019, pp. 133–148.

[19]. Z. Su, S. Li, H. Liu, and Z. He, "Tree skeleton extraction from laser scanned points", in IGARSS - IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 6091–6094.

[20]. M. H. Sharif, "Laser-based algorithms meeting privacy in surveillance: A survey", IEEE Access, vol. 9, pp. 92 394–92 419, 2021.

[21]. M. H. Sharif, F. Galip, A. Guler, and S. Uyaver, "A simple approach to count and track underwater fishes from videos", in 2015 18th International Conference on Computer and Information Technology (ICCIT), 2015, pp. 347–352.

[22]. M. H. Sharif, "A numerical approach for tracking unknown number of individual targets in videos", Digital Signal Processing, vol. 57, pp. 106–127, 2016.

[23]. H. Ahn and H. J. Cho, "Research of multi-object detection and tracking using machine learning based on knowledge for video surveillance system", Pers. Ubiq. Comput., vol. 26, no. 2, pp. 385–394, 2022.

[24]. M. H. Sharif, Sundry Applications and Computations of sin() & cos(). Richardson, TX 75081, USA.: PLOMS LLC., 2021, ISBN 978-1-63802-003-5.

[25]. A. A. Afiq et al., "A review on classifying abnormal behavior in crowd scene", J. Vis. Commun. Image Represent., vol. 58, pp. 285–303, 2019.

[26]. K. Khan,W. Albattah, R. U. Khan, A. M. Qamar, and D. Nayab, "Advances and trends in real time visual crowd analysis", Sensors, vol. 20, no. 18, p. 5073, 2020.

[27]. J. J. P. Suarez and P. C. N. Jr., "A survey on deep learning techniques for video anomaly detection", CoRR, vol. abs/2009.14146, 2020.

[28]. M. B. Braham, J. Weber, G. Forestier, L. Idoumghar, and P. A. Muller, "Recent trends in crowd analysis: A review", Machine Learning with Applications, vol. 4, pp. 1–30, 2021.
[29]. A. B. Chan, Z. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking", in CVPR, Anchorage, Alaska, USA, 2008.
[30]. U. of Minnesota, “Detection of unusual crowd activities in both indoor and outdoor scenes,” 2021, http://mha.cs.umn.edu/proj_events.shtml#crowd.
[31]. A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors", IEEE Trans. PAMI, vol. 30, no. 3, pp. 555–560, 2008.
[32]. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. F. Fei, "Imagenet: A large-scale hierarchical image database", in CVPR, Miami, Florida, USA, 2009, pp. 248–255.
[33]. C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB", in ICCV, Sydney, Australia, 2013, pp. 2720–2727.
[34]. W. Luo, W. Liu, and S. Gao, "A revisit of sparse coding based anomaly detection in stacked RNN framework", in International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 341–349.
[35]. W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos", in Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 6479–6488.
[36]. Y. Hao, J. Li, N. Wang, X. Wang, and X. Gao, "Spatiotemporal consistency-enhanced network for video anomaly detection" Pattern Recognit., vol. 121, p. 108232, 2022.
[37]. K. Lloyd, P. L. Rosin, A. D. Marshall, and S. C. Moore, "Detecting violent and abnormal crowd activity using temporal analysis of grey level co-occurrence matrix (glcm)-based texture measures", Mach. Vis. Appl., vol. 28, no. 3-4, pp. 361–371, 2017.
[38]. F. L. Sanchez, I. Hupont, S. Tabik, and F. Herrera, "Revisiting crowd behaviour analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects", Inf. Fusion, vol. 64, pp. 318–335, 2020.
[39]. F. Zhou, L. Wang, Z. Li, W. Zuo, and H. Tan, "Unsupervised learning approach for abnormal event detection in surveillance video by hybrid autoencoder", Neural Process. Lett., vol. 52, no. 2, pp. 961–975, 2020.
[40]. U. of Minnesota, "ShanghaiTech Campus dataset (Anomaly Detection)", 2021, https://svip-lab.github.io/dataset/campus_dataset.html.
[41]. K. Doshi and Y. Yilmaz, "Rethinking video anomaly detection - a continual learning approach", in IEEE Winter Conference on Applications of Computer Vision (WACV), USA, 2022, pp. 3961–3970.
[42]. U. of Central Florida, "Video Anomaly Dection Dataset", 2021, https://svip-lab.github.io/dataset/campus_dataset.html.
[43]. M. H. Sharif, N. Ihaddadene, and C. Djeraba, "Crowd behaviour monitoring on the escalator exits", in 2008 11th International Conference on Computer and Information Technology, 2008, pp. 194–200.
[44]. V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes", in CVPR, San Francisco, CA, USA, 2010, pp. 1975–1981.
[45]. G. Tripathi, K. Singh, and D. K. Vishwakarma, "Crowd emotion analysis using 2d convnets", in 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 2020, pp. 969–974.
[46]. R. F. Mansour, J. Escorcia-Gutierrez, M. Gamarra, J. A. Villanueva, and N. Leal, "Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model", Image Vis. Comput., vol. 112, p. 104229, 2021.
[47]. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization", in International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015.
[48]. K. Singh, S. Rajora, D. K. Vishwakarma, G. Tripathi, S. Kumar, and G. S. Walia, "Crowd anomaly detection using aggregation of ensembles of finetuned convnets", Neurocomputing, vol. 371, pp. 188–198, 2020.
[49]. Y. Zahid, M. A. Tahir, N. M. Durrani, and A. Bouridane, "Ibaggedfcnet: An ensemble framework for anomaly detection in surveillance videos", IEEE Access, vol. 8, pp. 220 620–220 630, 2020.
[50]. C. Szegedy,W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", in CVPR, Boston, MA, USA, 2015, pp. 1–9.

[51]. D. Tran, L. D. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks", in ICCV, Santiago, Chile, 2015, pp. 4489–4497.
[52]. Z. Hu, L. Zhang, S. Li, and D. Sun, "Parallel spatial-temporal convolutional neural networks for anomaly detection and location in crowded scenes", J. Vis. Commun. Image Represent., vol. 67, p. 102765, 2020.
[53]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", in 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015.
[54]. K. Hara, H. Kataoka, and Y. Satoh, "Learning spatio-temporal features with 3d residual networks for action recognition", in ICCV Workshops 2017, Venice, Italy, 2017, pp. 3154–3160.
[55]. F. A. Gers, J. Schmidhuber, and F. A. Cummins, "Learning to forget: Continual prediction with LSTM", Neural Comput., vol. 12, no. 10, pp. 2451–2471, 2000.
[56]. L. Xia and Z. Li, "A new method of abnormal behavior detection using LSTM network with temporal attention mechanism", J. Supercomput., vol. 77, no. 4, pp. 3223–3241, 2021.
[57]. S. Hochreiter and J. Schmidhuber, "Long short-term memory", Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.
[58]. A. N. Moustafa and W. Gomaa, "Gate and common pathway detection in crowd scenes and anomaly detection using motion units and LSTM predictive models," Multim. Tools Appl., vol. 79, no. 29-30, pp. 20689--20728, 2020.
[59]. H. Shi, X. Xu, Y. Fan, C. Zhang, and Y. Peng, "An auto encoder network based method for abnormal behavior detection", in The 4th International Conference on Software Engineering and Information Management, Yokohama Japan, Y. Li and H. Nishi, Eds., 2021, pp. 243–251.
[60]. M. Hasan, J. Choi, J. Neumann, A. K. R. Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences", in CVPR, Las Vegas, NV, USA, 2016, pp. 733–742.
[61]. D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. van den Hengel, "Memorizing normality to detect anomaly: Memoryaugmented deep autoencoder for unsupervised anomaly detection", in ICCV, Seoul, South Korea, 2019, pp. 1705–1714.
[62]. F. Yang, Z. Yu, L. Chen, J. Gu, Q. Li, and B. Guo, "Human-machine cooperative video anomaly detection", Proc. ACM Hum. Comput. Interact., vol. 4, no. CSCW3, pp. 1–18, 2020.
[63]. M. Asad, J. Yang, E. Tu, L. Chen, and X. He, "Anomaly3d: Video anomaly detection based on 3d-normality clusters", J. Vis. Commun. Image Represent., vol. 75, p. 103047, 2021.
[64]. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets", in Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems, Quebec, Canada, 2014, pp. 2672–2680.
[65]. X. Xia, X. Pan, N. Li, X. He, L. Ma, X. Zhang, and N. Ding, "Gan-based anomaly detection: A review", Neurocomputing, 2022.
[66]. Z. Wang, Z. Yang, and Y. Zhang, "A promotion method for generation error-based video anomaly detection", Pattern Recognit. Lett., vol. 140, pp. 88–94, 2020.
[67]. D. Chen, L. Yue, X. Chang, M. Xu, and T. Jia, "NM-GAN: noisemodulated generative adversarial network for video anomaly detection", Pattern Recognit., vol. 116, p. 107969, 2021.
[68]. Y. Tang, L. Zhao, S. Zhang, C. Gong, G. Li, and J. Yang, "Integrating prediction and reconstruction for anomaly detection", Pattern Recognit. Lett., vol. 129, pp. 123–130, 2020.
[69]. P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks", in CVPR, Honolulu, HI, USA, 2017, pp. 5967–5976.
[70]. Y. Zhong, X. Chen, J. Jiang, and F. Ren, "A cascade reconstruction model with generalization ability evaluation for anomaly detection in videos", Pattern Recognit., vol. 122, p. 108336, 2022.
[71]. W. Liu,W. Luo, D. Lian, and S. Gao, "Future frame prediction for anomaly detection - A new baseline", in CVPR, Salt Lake City, UT, USA, 2018, pp. 6536–6545.
[72]. J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", in CVPR 2015, Boston, MA, USA, 2015, pp. 3431–3440.
[73]. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", in International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), Munich, Germany, vol. 9351, 2015, pp. 234–241.

[74]. D. Chen, P. Wang, L. Yue, Y. Zhang, and T. Jia, "Anomaly detection in surveillance video based on bidirectional prediction", Image Vis. Comput., vol. 98, p. 103915, 2020.
[75]. C. Park, M. Cho, M. Lee, and S. Le, "Fastano: Fast anomaly detection via spatio-temporal patch transformation", in IEEE Winter Conference on Applications of Computer Vision (WACV),Waikoloa, HI, USA, 2022, pp. 2249–2259.
[76]. J. Redmon, S. K. Divvala, R. B. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection", in CVPR, Las Vegas, NV, USA, 2016, pp. 779–788.
[77]. K. Doshi and Y. Yilmaz, "Continual learning for anomaly detection in surveillance videos", in Conference on Computer Vision and Pattern Recognition, CVPRWorkshops, Seattle,WA, USA, 2020, pp. 1025–1034.
[78]. K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks", in European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, vol. 9908, 2016, pp. 630–645.
[79]. J. Redmon and A. Farhadi, "Yolov3: An incremental improvement", CoRR, vol. abs/1804.02767, 2018.
[80]. L. Shine, V. M. A, and C. V. Jiji, "Fractional data distillation model for anomaly detection in traffic videos", in Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, 2020, pp. 2581–2589.
[81]. K. Doshi and Y. Yilmaz, "A modular and unified framework for detecting and localizing video anomalies", in IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2022, pp. 3982–3991.
[82]. W. Zhang, G. Wang, M. Huang, H. Wang, and S. Wen, "Generative adversarial networks for abnormal event detection in videos based on selfattention mechanism", IEEE Access, vol. 9, pp. 124 847–124 860, 2021.
[83]. J. T. Zhou, L. Zhang, Z. Fang, J. Du, X. Peng, and Y. Xiao, "Attentiondriven loss for anomaly detection in video surveillance", IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 12, pp. 4639–4647, 2020.
[84]. L. Chen, Y. Yang, J. Wang, W. Xu, and A. L. Yuille, "Attention to scale: Scale-aware semantic image segmentation", in Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 3640–3649.
[85]. F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification", in Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6450–6458.
[86]. A. Vaswani et al., "Attention is all you need", in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems, USA, 2017, pp. 5998–6008.
[87]. X. Feng et al., "Convolutional transformer based dual discriminator generative adversarial networks for video anomaly detection", in ACM Multimedia Conference, Virtual Event, 2021, pp. 5546–5554.
[88]. H. Yuan, Z. Cai, H. Zhou, Y. Wang, and X. Chen, "Transanomaly: Video anomaly detection using video vision transformer", IEEE Access, vol. 9, pp. 123 977–123 986, 2021.
[89]. A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale", CoRR, vol. abs/2010.11929, 2020.