REVIEW

Anomaly behavior detection analysis in video surveillance: a critical review

Sanjay Roka,[a] **Manoj Diwakar**,[a,*] **Prabhishek Singh**,[b] **and Pragya Singh**[c]

[a]Graphic Era Deemed to be University, Computer Science and Engineering Department, Dehradun, Uttarakhand, India

[b]Bennett University, School of Computer Science Engineering and Technology, Greater Noida, Uttar Pradesh, India

[c]IIIT Allahabad, Department of Management Studies, Allahabad, Uttar Pradesh, India

Abstract. Anomaly detection is one of the most researched topics in computer vision and machine learning. Manual detection of an oddity in a video costs significant time and money, so there is a need for an autonomous detection system that can analyze the process and detect the anomaly in the majority of captured video datasets. Through an in-depth study on the recently published works on anomaly detection, a review is prepared to highlight the various tasks performed in abnormal behavior detection. Descriptions along with the pros and cons of various machine-learning and non-machine-learning techniques are discussed in depth. Similarly, more concentration is given to the generation adversarial network (GAN), and a comprehensive description of its design for achieving a better abnormality detection rate is provided. Moreover, a comparison of various state-of-the-art approaches on the basis of their methodologies, advantages, and disadvantages is given. We further quantitatively analyze some of the recent robust approaches at the frame level on the UCSD Ped1 dataset, with the GAN-based model achieving an astonishing performance. We provide various suggestions on how to further increase the performance of GAN for abnormal behavior detection in surveillance videos. [DOI: 10.1117/1.JEI.32.4.042106]

Keywords: anomaly detection; multivariate Gaussian fully convolution adversarial autoencoder; generative adversarial network.

Paper 221328SS received Nov. 17, 2022; accepted for publication Feb. 2, 2023; published online Mar. 7, 2023.

1 Introduction

Due to the heavy demands of security, the installation of surveillance cameras in both public and private places, such as shopping malls, airports, streets, railway stations, etc., has increased tremendously in recent decades. The main aim of these cameras is to capture and thus prevent ongoing abnormal behavior, i.e., assaults, road accidents, robberies, pedestrian fighting, traffic congestion, etc. Employing manual labor to monitor surveillance cameras has several limitations. (1) A physical presence in front of monitoring camera is time consuming and a monotonous task. (2) Monitoring 24/7 means staff must be present 24 h/day, because the occurrences of abnormal activities are unexpected. (3) The manual analysis and processing of large numbers of videos recorded from the cameras is also time consuming. Consequently, to solve the aforementioned problems, there is a need for a fully automated system that can learn, recognize, and alert the user to any suspicious actions in the recorded video. A system that automatically analyzes, processes, and detects abnormal events in videos captured from a surveillance camera is called an abnormal behavior detection system. Anomaly detection is one of the most popular, challenging, and widely researched topics in the areas of computer vision and machine learning. It is the process of recognizing any rare events, activities, or items of concern from the majority of the data processed because of their varying features. Changes in data pattern or unobserved activities are also considered to be anomalies.[1]

*Address all correspondence to Manoj Diwakar, manoj.diwakar@gmail.com

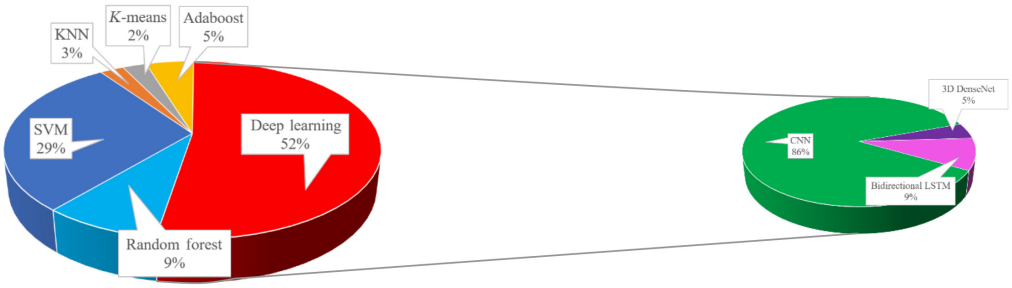

Fig. 1 Distribution of anomaly detection techniques.

There are three main forms of anomaly detection techniques: unsupervised,[2–4] supervised[5,6] and semi-supervised or weakly supervised.[7,8] In this paper, an effort has been made to elaborate on some of the recently used approaches for the detection of abnormal behavior in videos. Despite the various techniques, some challenges still arise due to various factors. First, the actual definition of an anomaly in the real world does not exist; the definition of anomalous behavior varies according to the application area. Second, due to the lack of a publicly available dataset, it is not possible to capture all anomalies in the videos. Most of the anomalous behavior detection systems are designed based on the training sample through which the system is made to learn the pattern of normal behavior, and later they detect the anomaly in the test sample using the learned pattern. Presently, very few benchmark datasets are available for anomaly detection; these include CUHK Avenue,[9] ShanghaiTech,[10] and UCSD.[11] Some of the popular algorithms that can be employed for anomaly detection are shown in Fig. 1. The concept for the figure was taken from the work of Omarov et al.[12] Nawaratne et al.[13] used the above dataset and designed a robust system using incremental spatio/temporal learner (ISTL) to detect the abnormal behaviors in real time. It is an unsupervised learning approach that uses active learning with fuzzy aggregation to regularly update and recognize new anomalies. Using both spatial and temporal features for anomaly detection is a good idea as it helps to boost the anomaly detection rate.

Recently, Wang et al.[3] presented the double-flow convolution, long short-term memory variational autoencoder (DF-ConvLSTM-VAE) model to improve the performance of the network in detecting the abnormal activities. Abnormal activities arise within the appearance and motion, so using a separate network to learn the patterns of appearance and motion is more effective. The model of Fan et al.[14] was slightly modified by inserting the ConvLSTM layer below the convolution layer in the encoder part of the VAE. Convolution and ConvLSTM layers were used to extract appearance and motion-based features, respectively. Instead of separate network, a single network was used to learn the patterns from these features, and the structure of the decoder was similar to that of the work in Fan et al.[14] During the test phase, the reconstruction error probability was used to detect the abnormality. Using a single network to capture the appearance/motion features reduces the computational complexity, and the model can train faster. In addition, in an unsupervised approach,[3,14] manually labeling the data is no longer required. In contrast to previous unsupervised approaches, Gong et al.[15] detected the abnormal behavior in a supervised manner using a local distinguishability aggrandizing network (LDA-Net) that contains two modules. The human detection module contains the you-only-look-once (YOLO) algorithm to capture the segmented patches of a specific human and forward them to the anomaly detection module to learn the motion features of each person. In the anomaly detection module, a primary binary classification sub-branch and an auxiliary distinguishability aggrandizing sub-branch are used to jointly detect and recognize the anomalies. A novel inhibition loss function is applied in the auxiliary sub-branch to lower the false classification rate in imbalanced datasets. This type of technique boosts the anomaly detection rate because only the objects detected by YOLO are processed, while the undetected objects are discarded. Consequently, the computation complexity and false detection rates are drastically reduced, and real-time performance is achieved.

Similarly, Cho et al.[2] used an implicit two-path AE (ITAE) that contains the two encoders to extract implicitly appearance and motion features and a single decoder to merge these features to

learn the pattern of normal video samples. Zhang et al.[7] used the weakly supervised framework based on a transformer for video anomaly detection. A vision transformer is a new concept for image analysis that has shown the remarkable performance in anomaly detection. Considering the position of the camera and the large amount of background information, the authors in Ref. 16 proposed a method based on the spatial-temporal cuboid of interest with varying sizes of cell structures. The features of different scale objects were extracted using the optical flow and varying sizes of cell grids. Bigger cells were near the camera (the bottom area in the image), and vice versa; consequently, the amount of background data and computational cost were reduced. A parallel 3D convolution neural network (CNN) was used to learn the spatio/temporal features of the cuboid. The model performance was good in detecting both global and local anomalies. The objects far from and near the camera are the same but differ in size; moreover, the object near the camera appears to move faster than the far objects. Therefore, varying the cell size structure is a good idea to capture the objects efficiently. Yang et al.[17] detected abnormal events using a bidirectional retrospective generation adversarial network (BR-GAN) that supports training in an end-to-end manner. It encompasses a generator, a frame discriminator, and a sequence discriminator. The sequence discriminator is designed using the 3D CNN to capture the long-term motion information between frame sequences. GANs are the powerful model that can be trained in an unsupervised way, and they can generate the very sharp images through adversarial training.

1.1 *Anomalies Detection Techniques*

The techniques used for the anomaly detection can be categorized as shown in Fig. 2.

1.1.1 *Statistical-based technique*

Statistical-based techniques implement the calculation of the statistical components, such as mean and standard deviation, data distribution, and finding the probabilities, to generate the behavior profile. Both parametric and nonparametric techniques are implemented for designing the statistical-based model.

Parametric techniques. In parametric techniques, a fixed number of parameters is used to design the probability model. Here, we have to make certain assumption regarding the distribution of data with which we are working. Parameters that can be used for the normal

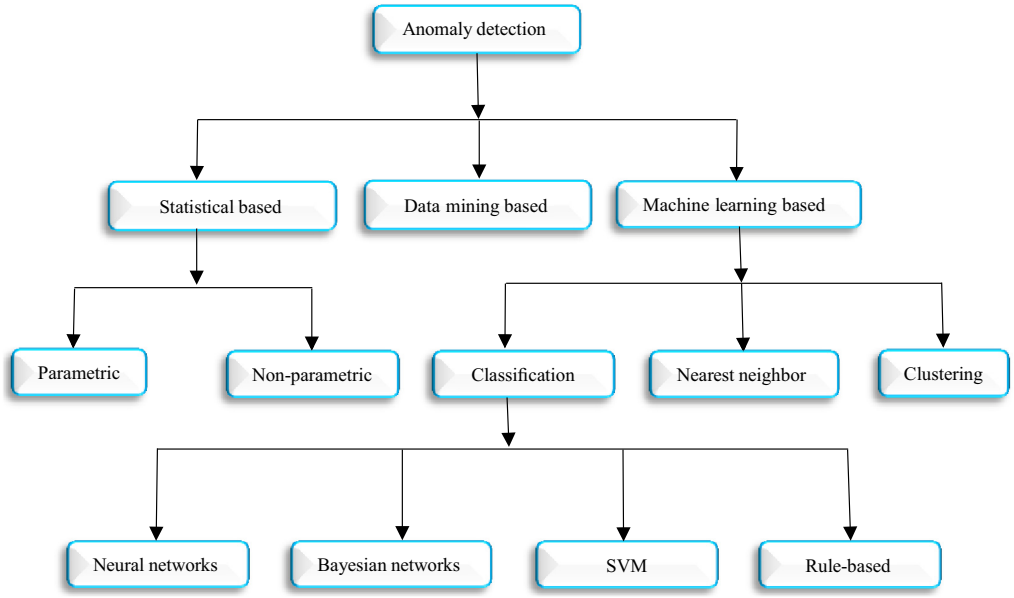

Fig. 2 Overview of anomaly detection techniques.

distribution are the mean and standard deviation. Examples of this type of technique are logistic regression, naïve Bayes model, and so forth. Based on the distribution type, this technique can be subdivided into Gaussian- and regression-based models. In Gaussian-based models, the parameter is calculated through maximum likelihood estimation,[18] and its distribution belongs to a Gaussian distribution. The testing methods, such as paired/unpaired *t*-test, Pearson correlation, etc., can be implemented to check whether the data instance is anomalous or not.

Nonparametric Technique. In nonparametric techniques, there is no need for a fixed number of parameters or an assumption in the distribution of the data. Hence, it is also known as a distribution-free technique. This type of technique is used widely due to its low complexity and distribution-free skill. Tests that can be used in this method are the Kruskal-Wallis test and Mann-Whitney U test. Examples of this technique are the *K*-nearest neighbor (KNN) and the decision tree model. To calculate the probability distribution, the concept of a histogram can be used. When this concept is implemented, it calculates the normal data and generates the histogram bin. If any plotted data instance falls within this bin, then it is an anomaly,[19] and an anomalous score is generated.[20]

1.1.2 *Data mining-based technique*

The process of discovering the meaningful patterns and rules from a large dataset is known as data mining. It is used to design machine learning-based models and artificial intelligence applications such as a search engine. It also has application in the areas of anomaly detection, sentiment analysis, spam filtering, qualitative data mining, etc. Application of data mining in anomaly detection includes the detection of oddity or uncommon events and includes the steps of learning, clustering, classification, and regression. In the past, many researchers have used this technique for oddity detection.[21]

1.1.3 *Machine learning-based technique*

Machine learning is a subset of artificial intelligence that predicts some defined output when the input is provided to it. It has wide application in the areas of anomaly detection, handwriting recognition, networks, natural language processing, robot locomotion, search engines, and so forth. It has the ability to detect the anomalies and can be partitioned into three groups: classification, nearest-neighbor, and clustering. Some recent works in the area of machine learning can be reviewed in Refs. 12 and 22–29.

Classification. Data points are divided into separate classes based on their properties.

1. Neural networks

A neural network is a collection of multiple processes linked with each other and working with the data asynchronously. It is also a collection of algorithms that mimic all of the operations occurring in the human brain that helps to determine the association between the tremendous amount of data. Based on the labels in the dataset, it can be implemented for unsupervised or supervised learning. Fanta et al.[30] designed the end-to-end neural network named single-tunneled GRU (SiTGRU); its performance in anomaly detection was significantly better than standard recurrent neural networks. To achieve a better performance, the author removed the reset gate from the standard GRUs and designed a single gate GRUs architecture. The experiment on a benchmark dataset showed a better performance of their architecture compared with methods designed on standard GRU and standard LSTM architectures.

2. Bayesian networks

The collection of probability and graph theory that contains a directed acyclic graph is called a Bayesian network. It is a graphical model that can also handle high dimensional data and uses the relation of probabilistic theory to relate the variables of interest. It is widely applicable for intrusion detection,[31] spam filters, pattern recognition, etc.

3. Support vector machine (SVM)

SVM is a supervised technique of machine learning that is highly applicable for classification, with each data item being plotted in an n-dimensional plane. Provided with the labeled data, the SVM assigns new upcoming data to one of two categories. If no label data is available, then supervised learning is not possible, and we have to choose another classifier, i.e., unsupervised learning. Wang et al.[32] detected the abnormality using two stages. The first stage is the normality estimation stage in which the autoencoder (AE) was trained iteratively to recognize the normal events globally from the complete unlabeled samples by a self-adaptive reconstruction loss thresholding mechanism. The second stage is normality modeling in which the normal activities detected in first stage are passed to the one-class SVM to further check for the presence of abnormality.

4. Rule-based classification

A rule-based classifier is a supervised learning technique used for the classification of the data instances. It utilizes IF-THEN rules for the purpose of classification. Rules for the detection of normal behavior are learned if any data instance fails to fall within the rule; such data instances are considered to be anomalies. The development of a better security system for a network continues due to the limitation of the present intrusion detection system. One of the major problems is the dimensionality of the data. Hence, for a better performance of intrusion detection system (IDS), the author in Ref. 24 proposed a system based on a rule-based technique that merges fuzzy logic and weighted fuzzy C-means clustering (WFCM). First, the size of the input dataset is reduced by the algorithm WFCM, and then the fuzzy logic algorithm is provided the reduced dataset as input. Finally, data instances are checked to determine if they are normal or abnormal using the rule.

Nearest-neighbor. One of the most popular and high-performance classifiers in supervised learning is the nearest neighbor. It has both the normal and abnormal training classes. The classification of the newly arrived data instance is done by the manipulation of the distance between the data instance and its nearest neighbor. The calculated distance is the anomaly score of that data instance. The KNN classifier uses this concept using the K number of nearest points. Gharaei et al.[33] detected the outlier using this technique in two-dimensional synthetic and multi-dimensional real datasets.

Clustering. Clustering is an unsupervised machine learning technique that groups similar data instances, resulting in the data instances in one group being similar to each other and different from the other data instances in the other group. The advantage of clustering methods is that it has less computation overhead and hence is considerably faster than other methods such as distance-based methods. The serious disadvantage of this method is that the accuracy of this method might be low when the number of provided data instances is too low. Due to the ongoing high demand for anomaly detection systems in cloud computing, the author in Ref. 34 fused K-means clustering with the SVM algorithm. The UNSW-NB15 dataset was used for the experiment with nine different types of attacks used for detection. The K-means algorithm was used for clustering the network flows, and consequently 64 clusters were formed and later are separated into normal/abnormal categories using the SVM classifier with an accuracy of 88.6%.

Due to increasing crime, such as fighting, looting, vandalism, robbery, etc., the demand for abnormal behavior detection systems has increased tremendously. Consequently, in this paper, we provide a comprehensive survey of abnormality detection and the algorithms that can aid in detecting the abnormality. In the past decade, numerous books, review articles, and surveys ranging from brief to comprehensive have been published in anomaly detection,[12,35–45] but our survey is relevant and different from others in many aspects. For instance, Nandhini Abirami et al.[46] only targeted the deep CNN and deep GAN in their survey. Similarly, Omarov et al.[12] discussed only anomaly detection techniques, and no approaches were examined in depth. In Ref. 40, only a description is provided regarding datasets that can assist in anomaly detection. The authors in Ref. 45 provided a brief systematic survey of motion and anomaly detection in video

surveillance. However, our survey is detailed and systematically ordered. The summary of the contributions of this paper follows.

1. Most existing works discuss only the machine learning-based techniques for abnormal behavior detection, but in this paper, we cover all of the machine-learning and non-machine-learning techniques.
2. After conducting a vast survey of the research articles published until 2022, we provide some of the recent gaps that exist in anomaly detection in surveillance videos.
3. We provide the description, pros, and cons of various popular approaches that have been employed recently for abnormality detection.
4. We proved a complete robust overview of GAN for anomaly detection and localization.
5. We offer an in-depth comparative study along with cons and pros for abnormal detection approaches.
6. We provide a quantitative comparison between the various recent popular approaches at the frame level in UCSD dataset.

2 Related Works

One of the most researched topics in computer science is anomaly detection. In the past decade, numerous books, review articles, and surveys ranging from brief to comprehensive have been published in anomaly detection.[42–44] However, some research gaps exist in the field of anomaly detection. Some of them are listed below:

1. Most of the methods are applicable only to areas where the density of pedestrians is sparse.
2. No suitable methods are available to detect an anomaly in multicamera setups in real time in crowd scenes.
3. Anomaly detection is basically contextual dependent. As the context changes, the nature of the object differs. Hence, there should be a system that learns the different kinds of objects in different scenarios and then makes a decision about an object's abnormality.
4. The detection and localization of anomalies generally depend upon the scene and type of anomalies. The environment can affect the performance of abnormality detection. Therefore, a robust system must be designed to adapt to the changing environment that includes fog, rain, night, etc.
5. A large number of different abnormal activities can be generated such that the robust system can be designed to learn the abnormalities in real time. This can be done by making a network of surveillance cameras throughout an entire city. The activities captured by several cameras located at several places can be examined and shared with the other locations.

In Ref. 35, a review was carried out in anomaly detection in a surveillance video. The author elaborates the aspects of the surveillance target, the definition of anomaly and assumption of anomalous behavior, the types of sensors used, the process of extracting the feature, and finally the learning method used for classifying the behavior. A survey of a related topic was covered in Ref. 36. A short survey was carried out in Ref. 37 in which the majority of research papers published within 5 years for the detection of the anomaly was studied. It also provided the snapshot of various architectures and the datasets used by the researcher in the network for identifying whether an anomaly was present or not. In Ref. 38, a survey was done regarding the various steps taken in a surveillance video, and the outcome and the obstacle arising were examined for understanding the behavior and detection of the anomaly while merging standoff biometrics, tracking of the object, analysis of the motion and the behavior. It also provided a wide overview of the topics of motion and object detection, object classification, object tracking, extraction of motion information, and behavior analysis. Recently, a survey was carried out for the detection of motion in the sequences of an image.[39] It prioritized the various algorithms available for motion detection. The technique used by Zhou et al.[47] for motion detection was also elaborated in full detail. Similarly, an explanation of another technique used by Wu et al.[48] for detecting motion, along with its advantages, was also given. Furthermore, the comparison of this novel

technique with earlier related works was also done to prove its effectiveness. Finally, a snapshot of extensively used datasets, such as the Weizmann dataset, KTH dataset, CAVIAR dataset, UCF Sports Action dataset, and UCF YouTube Action dataset along with their frames was also provided. The survey in Ref. 40 provided a brief description of the various currently used datasets applied for anomaly detection, and the survey in Ref. 41 differentiates several crowd datasets used for the calculation of the crowd density. Nowadays, home-based health care facilities have received tremendous research. Therefore, a comprehensive survey study was carried out by the author in Ref. 49; it explored the dense sensing network for anomalous behavior detection and mainly focused on elderly care. The advantages and disadvantages of existing anomalous behavior detection systems based on dense sensing networks were also highlighted. An explanation and comparison of the several anomaly detection systems, such as dense sensing-based, wearable-based, and vision-based, were also conducted. A brief description of the anomaly types was also included in the survey. The application of sensor fusion in a dense sensing network was also discussed, and the challenges arising in anomaly detection systems based on dense sensing networks were also stated clearly. The author also noted that the uses of the models based on sensor fusion are more vigorous than the tradition methods, and moreover their application also increased the ability of the dense sensing networks.

The main phases of video surveillance are detection of the object, object tracking, and recognition of the object. Various challenges arise while employing these phases. Some of the challenges and their solutions, such as the detection of objects that encompass complex structures, the detection of abnormal events, occluded objects, deformed objects, and changes in the level of intensity, are discussed thoroughly in Ref. 50. Due to the extensive research in anomaly detection, the author in Ref. 45 provided a brief systematic survey of motion and anomaly detection in video surveillance and discussed related works in a similar area. In addition, a simple and well-explained methodology for behavior analysis was also discussed. For security reasons, the use of anomaly detection systems by transit agencies has increased sharply. Consequently, an extensive survey of the recognition of human behavior in surveillance video was performed in Ref. 51. This survey aimed to explain the various state-of-the-art techniques that can be used to detect anomalies in transit surveillance. The anomaly method used can be for single people, interactions among several people, people and vehicle interactions, and people-facility/location interactions. In Ref. 52, the author performed a comprehensive survey of the recent work performed in visual analytics of anomalous behavior detection of the user and classified their behavior as social interaction (sharing of ideas and views between people), travel (locomotion of people from one place to another), network communication (sharing of data between machines using a network), and financial transaction (flow of money for the purpose of buying and selling). For every classification, similar methods for interactive analysis, visualization techniques (for egocentric and collective behaviors), and types of data are determined.

The most important thing when designing an anomaly detection system is to select the suitable optimization techniques and algorithms to enhance the performance and accuracy of the system. Some prefer using decision trees to model the appearance motion features. The approaches using decision trees[53] have the benefits of normalization and scaling of data not being required, no considerable impact of missing values, better visualization, and absence of irrelevant features, so these issues will not affect the decision trees. However, such approaches are prone to overfitting and require longer time to train the decision trees. Similarly, approaches that prefer KNN[33] have better performances for anomaly detection, but they are extremely slow for bigger datasets, perform poorly for a large number of features, and are sensitive to outliers.

Shreedarshan and Selvi[54] first generated a model for the detection of the optical flow in an image. Optical flows along with streak lines were implemented for the representation of motion. Later these motions were analyzed using particle swarm optimization (PSO). An experiment was done for anomaly detection in a crowd using the publicly available dataset of the University of Minnesota. For the detection and localization of abnormal activities in a crowd, Raghavendra et al.[55] used the concept of the social force model. PSO was used to optimize the interaction forces among the particles in the frame. Later, random sample consensus and a segmentation algorithm were implemented for detecting and localizing the global anomaly in each frame of the crowded scene. The main advantages of PSO are that it is easy to use and robust over parameter control compared with other mathematical algorithms and heuristic optimization techniques.

The main disadvantages are that, in this algorithm, it is easy to fall into local optimum in high-dimensional space, and it has a low convergence rate in the iterative process.

To solve the problem of online video anomaly detection, Leyva et al.[18] used binary features from the video. Subsequently, their model required very little processing time and worked in real time more effectively than models that use double precision features. First, they computed the foreground and temporal gradients from the input frame. Second, a fast accelerated segmentation test detector was used to detect the spatio temporal interest points from the temporal gradients. Afterward, binary wavelets differences was used to encode the corresponding support regions. Finally, Gaussian mixture modelling (GMM) was used to model the distance and detect the anomalies. Similarly, Fan et al.[14] designed end-to-end network called Gaussian mixture fully convolutional variational autoencoder (GMFC-VAE) for abnormality detection. Image patches of the RGB frame and dense flows are passed to separate network to learn the spatial and temporal pattern. During the testing, the separate latent representation of RGB frame patches and dense flow patches are obtained from two GMFC-VAEs that depict the conditional probability of the testing patches to belong to each of the components of GMM. Late fusion is done to these probabilities, and an energy-based technique is used to detect both the appearance and motion anomalies. In terms of performance, the GMM-based model has a better performance in comparison with K-means[34] algorithms; however, in comparison with neural network-based models, they are somewhat slow, are more sensitive to outliers, need sufficient data for each cluster, and require the number of clusters to be specified ahead of time.

Liu et al.[8] used the collaborative normality learning framework for video anomaly detection. Their proposed approach contains an AE and a channel attention-based MIL regression module. The anomaly is detected using two phases. In the first phase, AE is trained using the normal frames to learn the spatio/temporal patterns of normal activities. In the second phase, only the pre-trained encoder is used to extract the features from the test video frames. Afterward, a regression module is used to compute the anomaly score for each frame. The average value of anomalous frames is set to be larger than the maximum value of the normal frames. All frames in the test clip may not be abnormal; therefore, two sub-bags are used based on the anomaly score. Only those frames having a score lower than the average value are fed to the AE of the first phase. Though regression-based models are simple and effective, they perform poorly with non-linear data and irrelevant and highly correlated features. Moreover, they are not powerful enough and can be easily outperformed by other algorithms.

Hasan et al.[56] proposed two methods for anomaly detection in videos. For the first method, the author used the conventional handcrafted-based methods for extracting the spatio/temporal features and trained the end-to-end AE on those features. Second, the author designed a fully connected convolution AE and used it directly to extract and learn the spatio/temporal patterns. The main intuition behind the two methods was that the learned AE model was able to reconstruct the motion pattern in regular videos with a low reconstruction error, whereas motion pattern in irregular videos were reconstructed with a high reconstruction error. Though a better accuracy rate can be achieved through the handcrafted-based technology, manually extracting the spatio/temporal features from a large, high-dimensional dataset is an extremely difficult task.

Hu et al.[57] used the faster R-CNN to detect and localize pedestrians and vehicles; then a histogram of the large-scale optical flow descriptor was extracted from each detected objects to describe the object behavior. Finally, multiple instance SVM was trained to identify whether the object behavior was normal or abnormal. Sun et al.[20] designed an end-to-end, deep one-class model by fusing the CNN with the one-class SVM and used it to detect abnormal behavior in videos. The parameter of the model was optimized using the loss function derived from the one-class SVM. Though the approaches using SVM[20,57] have better accuracy due to their ability to handle high-dimensional data, such approaches are slow for the larger datasets. They also perform poorly for overlapped classes, and the selection of hyperparameters and kernel can be tricky.

Yong et al.[58] detected abnormal activities in real time in a crowd video at 140 fps using a spatio/temporal architecture. Their architecture has two components for learning the spatial representation and temporal representation of the spatial features. Their model was evaluated using the three benchmark datasets: Avenue, Subway, and UCSD. Ionescu et al.[59] detected the abnormality in the videos using the technique called unmasking, which has previously been preferred

for authorship verification in text documents. The performance of their proposed model was extremely high and worked in real time by processing at 20 fps. To detect and localize the anomalies, Li et al.[60] designed the cascade classifier spatialtemporal cascade autoencoder (ST-CaAE), a two-stream framework that enabled feeding the gradient and optical flow cuboids as input to the spatial-temporal adversarial autoencoder (ST-AAE) and ST-CAE. This classifier works in two phases: ST-AAE and ST-CAE. In the first phase, ST-AAE was built by fusing the 3D CNN with an adversarial AE to extract spatio/temporal features. Only the test cuboid with a latent representation that does not match with the prior distribution was considered to be anomalous. During the second phase, ST-CAE recognizes the anomalous patches in each anomalous cuboid using the reconstruction-based technique. Skip connections are only applied in the ST-CAE to fuse the low-level features with the high-level features and to avoid the feature loss across every layer. The main advantage of CNN over other approaches is that they only provide a better accuracy when the number of training images is large.

Ullah et al.[19] detected and localized anomalies using the recurrent conditional random field (R-CRF), which is designed by integrating the RNN and CRF. Initially, all video frames are broken into the fixed size blocks from where the spatio temporal features are extracted. Thus, extracted features are utilized as *a priori* for training the R-CRF. R-CRF was trained using the Gaussian kernel-based integration model (GKIM) features, and the performance of anomaly detection was evaluated in three benchmark datasets. To cope with a stable background and efficient feature extraction of various scales, Wang et al.[4] used a multi-path structure. Their proposed approach uses ConvGRU with non-local blocks to capture the motion information of the objects of different resolutions. To avoid the interference of the noisy pixels in the frame sequence, noise tolerance was used. This kind of setup allowed their model to train efficiently, and it was robust in detecting abnormalities in videos. The main advantage of RNN-based approaches is that they can remember every piece of information occurring through the time period analyzed; consequently, they can be used to capture the motion features of moving objects. But the training procedure of RNN is extremely difficult, and there is always the risk of gradient vanishing and the exploding problem. Also, they cannot capture very long sequences of information if the activation functions such as tanh and rectified linear units (ReLU) are used.

Song et al.[61] designed the end-to-end adversarial network called Ada-Net, which is the fusion of AE and a GAN model to enhance the reconstruction ability of the AE. The performance of the reconstruction was increased using the attention model fitted on the decoder; the model's task was to dynamically select the informative parts of the encoded features for the decoder. Consequently, learning the critical patterns of the normal behavior was preserved. Experiments on the Subway, UCSD, CUHK Avenue, and ShanghaiTech datasets depict the high performance of Ada-Net. Ravanbakhsh et al.[62] detected the anomalies in a crowd scene using the generative model called GAN. The model was trained using normal data, and during the test phase, the abnormality was detected through the local difference between the generated images and the ground truth. The model was evaluated following both frame- and pixel-level protocols. Liu et al.[63] detected abnormalities using the future frame prediction technique. First, a predictor design using a U-Net architecture was trained to predict the future frame for the training data. During testing, only those frames that did not match with their predictions were considered to be abnormal. For the appearance features, the frame was considered to be normal if the intensity and gradient maps of the predicted frame was close to its ground truth frame. Similarly, for the motion features, the frame was considered to be normal only if the optical flow of the predicted frame was close to the optical flow of its ground truth frame. Recently, Huang et al.[5] used the self-supervised attentive GAN that contains the three modules: a self-attentive generator, vanilla discriminator, and auxiliary self-supervised discriminator for the anomaly detection. The function of the generator is to predict the future frame from the set of consecutive frames that are then rotated 0, 90, 180, and 270 deg. Thus, the rotated frames are fed to the self-supervised discriminator for the rotation degree detection task. The function of the vanilla discriminator is to do the binary classification, i.e., true or fake. Consequently, the application of auxiliary rotation detection loss and vanilla prediction errors helps to achieve more discriminative anomaly scores. The most expensive task is data labeling, which is not required in GAN because GAN works in an unsupervised way and can generate very sharp images through adversarial training. However, training GAN can become unstable and slow due to the presence of a dual network, i.e., the

generator and discriminator, that compete against each other. Additionally, GAN requires a large amount of training data to generate effective results.

In this section, we mainly focused on the CNN-based model and GAN. CNN are the networks that contain numerous filters that slide across the images and yield an activation at all slide positions. Consequently, a feature map is generated as the output. The advantage of CNN is that it is spatially invariant; consequently, it can detect any critical information in the image. This spatial invariance property equally applies to all dimension data, i.e., one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D). The 2D convolution-based network can extract only the spatial information from the images, whereas 3D convolution-based network can extract both the spatial and temporal information. GAN utilizes CNN or RNN to design the generator and discriminator models. So basically, GAN can use CNN, but a CNN is not a GAN. Nandhini Abirami et al.[46] performed a comprehensive survey of deep CNN and deep GAN considering the various factors such as their principles, variants, and applications. Their work elaborated on the current opportunities and future challenges in the emerging domains. From their work, D-GAN was verified to be able to tackle the problem of insufficient data and improve the quality of images generated. It has numerous applications when merged with other deep learning algorithms. In the following section, we describe CNN-based GAN along with their architecture and characteristics.

3 Overview of Anomaly Detection Method

The authors of Refs. 17 and 62 provided in Sec. 2 designed robust models using the GAN for the abnormal behavior detection. Generative models are powerful models that, under certain controlled conditions, can achieve superior performances. Their performance can be verified by observing Table 2, which shows that they achieved a state-of-the-art performance for anomaly detection. Consequently, following the work of the author in Ref. 3 as described in Sec. 1, a new future frame prediction model can be designed for abnormal behavior detection using the GAN. The GAN contains a generator and a discriminator; it is trained using the training set that contains only normal activities, whereas the testing set encompass both normal and abnormal activities. First, RGB/grayscale images are extracted from the videos of the training and test sets and resized to $256 \times 256 \times 3$, and the pixel value is normalized in the range $[-1, 1]$, with 256×256 being the dimensions of images and 3 representing the number of channels (3 = RGB and 1 = grayscale). Frames are stacked in a sequence [i.e., 1,2,4,8 etc.] and passed in batches to the GAN model for the desired number of epochs. The Adam optimizer with the learning rate of 0.001 is selected, and MSE is used as the loss function for a better performance. For the encoder part, two convolution layers with a stride and kernel of 2 and 3, respectively, are used to extract the appearance features, and three ConvLSTM layer with a stride and kernel of 1 and 3, respectively, are used for the motion feature extraction. For the input sequence frame $I_1, I_2, I_3, \ldots\ldots I_t$, the encoder generates the encoded feature map $\hat{x}_{t+1}$ and passes it to the decoder. The decoder can be simply designed using the two-deconvolution layer with a stride and kernel of 2 and 3, respectively, to upsample the feature map $\hat{x}_{t+1}$ to $\hat{I}_{t+1}$. Moreover, after each layer in the generator, a normalization layer is added, and the filter size in each layer is set to 256→128→64→32→64→128→256. Similarly, the discriminator can be designed using a four-convolution layer followed by a normalization layer. The stride and kernel in each layer are set to 2 and 3, respectively, and the filter size is varied from 256→128→64→32. Further, two extra layers, such as a flattened layer followed by a dense layer, can be added for the binary classification, i.e., real or fake. The main task of the generator $G(z)$ is to continue generating high quality frames from the original frames until the discriminator $D(x)$ is unable to discriminate between the original images (real) and the fake images generated by the $G(z)$. If $D(x)$ recognizes the images as fake, then the parameters of the model are updated and optimized. This process continues iterating until the $D(x)$ is unable to distinguish between the real and fake images. Both the $G(z)$ and $D(x)$ interact with each other such that they together learn to obtain the optimal network parameters. Once the model is trained, it can be saved and used for testing. During the test phase when the sequence of frames $I_1, I_2, I_3, \ldots\ldots I_t$ are provided, the task of the GAN model is to use these frames to predict the future frame $\hat{I}_{t+1}$ of I_{t+1}. The difference between I_{t+1}

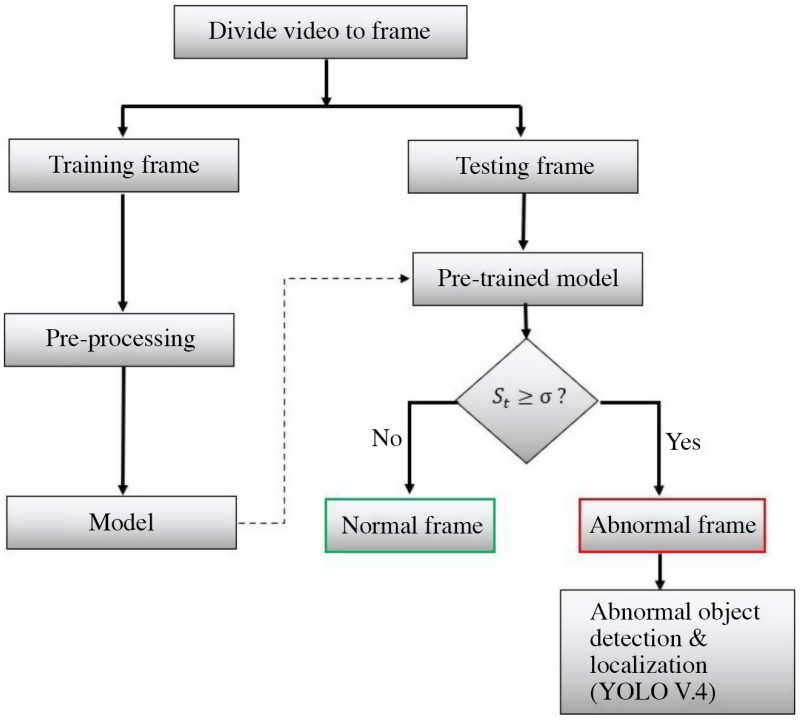

Fig. 3 General flowchart for the abnormal behavior detection using standard GAN model.

and $\hat{I}_{t+1}$ is used to compute the anomaly score (S_t). Only those frames with $S_t > \sigma$ (threshold) are considered to be abnormal. The value of σ is set in the range [0.6–0.9] depending upon the nature of dataset captured through the surveillance camera. Figure 3 provides the flowchart of the GAN method for anomaly detection and localization in the video.

To find whether the frame is normal or abnormal, we compute the anomaly score for each frame I_t. First, we measure the mean squared error between the predicted frame ($\hat{I}_t$) and its corresponding ground truth frame (I_t) using Eq. (1), where (H, W) are the dimensions of frame $I_t/\ (\hat{I}_t)$ and $I_t(i,j)/\hat{I}_t(i,j)$ depicts the pixel value at position (i,j) in frame $I_t/\hat{I}_t$. Subsequently, the peak signal-to-noise ratio (PSNR) with the value PSNR_t is calculated for every frame using Eq. (2), where MAX_{I_t} is the maximum possible value of I_t. Finally, the anomaly score (S_t) for the t'th frame I_t is measured by normalizing PSNR values in the range [0,1] as shown in Eq. (3), where $t' = 1,2,3,\ldots\ldots\ldots\ldots,T$ is the total number of frames in a testing sample. A larger value of S_t indicates that the frame is more likely to be an abnormal frame. The equations are given as

$$\text{MSE}(I_t, \hat{I}_t) = \frac{1}{HW}\sum_{i,j}\|I_t(i,j) - \hat{I}_t(i,j)\|_2^2, \tag{1}$$

$$\text{PSNR}_t = 10\log_{10}\frac{\text{MAX}_{I_t}^2}{\text{MSE}(I_t, \hat{I}_t)}, \tag{2}$$

$$S_t = 1 - \frac{P_t - \min_{t'} P_{t'}}{\max_{t'} P_{t'} - \min_{t'} P_{t'}}. \tag{3}$$

To reduce the computational complexity for anomaly localization, only the frames containing abnormal activities are passed to the object detection algorithm YOLO v.4. The reason for selecting this algorithm is its strength and ability to recognize almost any size of objects accurately in real time.[68] However, the application of YOLO is not mandatory, but it can be used for a better localization rate. As provided in Sec. 1, Gong et al.[15] used this algorithm to build the supervised LDA-Net and showed the powerful impact of this algorithm while detecting the videos

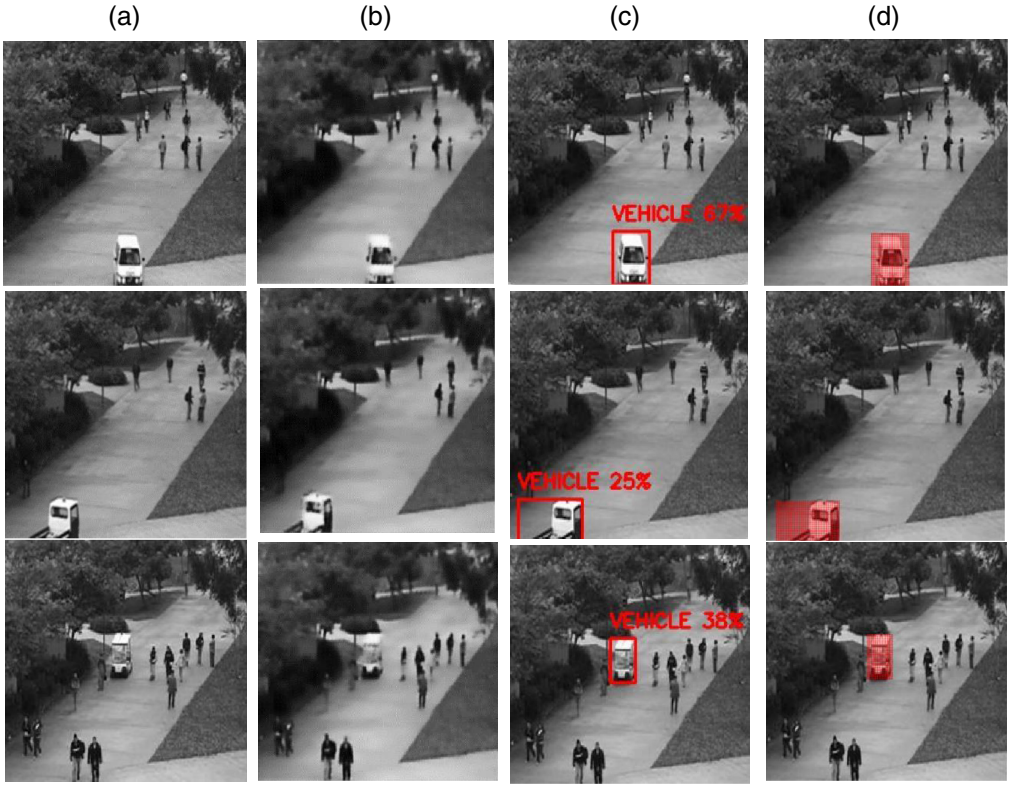

Fig. 4 Anomalies detected and localized on the UCSD Ped1 dataset: (a) original frame; (b) reconstructed frame; (c) and (d) anomaly localized using the YOLO V.4.

anomalies. Similarly, Yajing et al.[69] further demonstrated the effectiveness of YOLO by combining it with the convolutional autoencoder (Conv-Ae) and detected the abnormal activities in a semi-supervised fashion. The objects category detected from YOLO and the speed information were integrated with the reconstruction error of Conv-Ae for a better anomaly detection rate. Consequently, YOLO can be used for the localization of only the abnormal objects in the scene. Only those objects in the frame with a reconstruction error value of the pixel that is above σ_1, where the value of σ_1 ranges from $[0.1 \text{ to } \infty]$, are highlighted with the red bounding box.

Figure 4 demonstrates the results obtained if the GAN architecture as described in Sec. 3 is employed to detect the anomalies in the Ped1 dataset of UCSD. When the test samples are fed to the pretrained GAN, the anomalies are first detected by the GAN, and YOLO V.4 is used to localize the detected anomalies. The first, second, and third rows in Fig. 4 are the test video clip samples #19, #20, and #24, respectively, of the Ped1 dataset of UCSD, which encompass anomalies such as vehicles and small carts. The first column in the figure represents the original ground truth frame, and second column depicts the reconstructed frame by the GAN. The normal activities of the test frame are reconstructed with a low reconstruction error, whereas the abnormal objects in the frames are reconstructed with a high reconstruction error. Only those frames that are abnormal are fed to the YOLO. YOLO does the task of localizing the anomalies and generating the red bounding box only on the objects with reconstruction errors that are higher than σ_1. Subsequently, columns three and four represent the localized anomalies inside the red bounding box by YOLO.

Similarly, Li and Chang[64] modified the standard GAN and used it to detect the abnormal activities in a video. First, the frames are extracted from the videos; then for a better anomaly detection rate, a multi-scale patch structure is used to generate varying sizes of patches from these frames. The patches close to the camera are bigger in size, and the patches farther from the camera are smaller. Then the corresponding gradient and optical flow are extracted from the patches as shown in Fig. 6. For the anomaly detection and localization, a multivariate Gaussian fully convolution adversarial autoencoder (MGFC-AAE) was used. It contains the two networks:

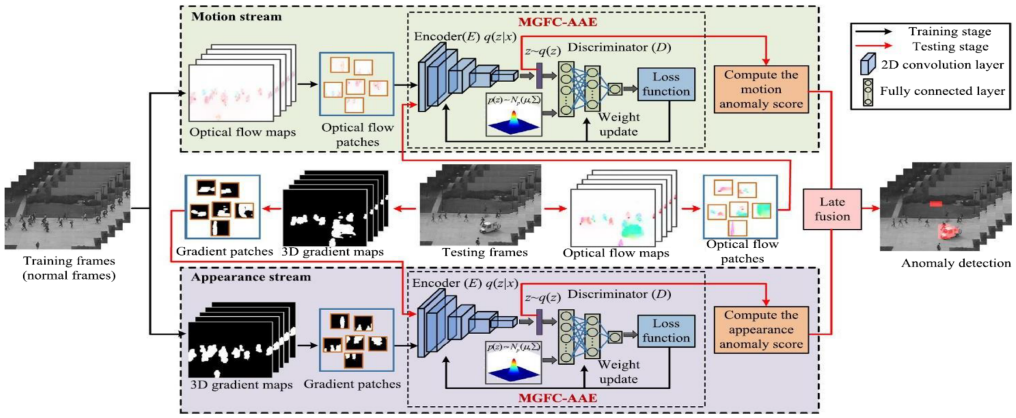

Fig. 5 Overview of anomaly detection methods.[64]

a generator and a discriminator. The generator lacks a decoder and contains only the encoder, which is designed using 2D CNN and encompasses three convolution layers, whereas the discriminator is designed using the three fully connected layers. MGFC-AAE has separate streams for the appearance and motion, and these streams share the same identical network architecture as shown in Fig. 5. The extracted gradient patches are passed from the appearance stream to extract the spatial pattern, whereas optical flow patches are fed to the motion stream to learn the temporal patterns. Once the model is trained, for the test sample, the corresponding gradient and optical flow patches are extracted and fed to the appearance stream and motion stream, respectively. To generate the appearance anomaly score S_a from the appearance stream and motion anomaly score S_m from the motion stream, an energy-based method in the form of a probability density function is used. Then, late fusion is used to achieve the total anomaly score S_t by combining S_a and S_m. A testing patch is only considered to be abnormal if $S_t > \theta$, where θ is the predefined threshold to find the sensitivity of the anomaly detection method. An overview of the method for anomaly detection can be seen in Figs. 5 and 6. The detection results of the method[64] on the UCSD dataset can be seen in Fig. 7, with the detected anomalies highlighted in red. The second and fourth rows depicts the results of detection in the Ped2 and Ped1 datasets, respectively, and the remaining rows represent the ground truth. Detected anomalies constitute bikers, cars, a wheelchair, and skaters.

The preprocessing steps prior to anomaly recognition, including motion detection, object classification, and object tracking, are known as core technologies. A complete description of the related literature, core technologies, and human behavior recognition is given in this paper. While performing the literature survey, we faced numerous difficulties. (1) Core technology, i.e., all of the available algorithms cannot be implemented to analyze every video, hardware-based

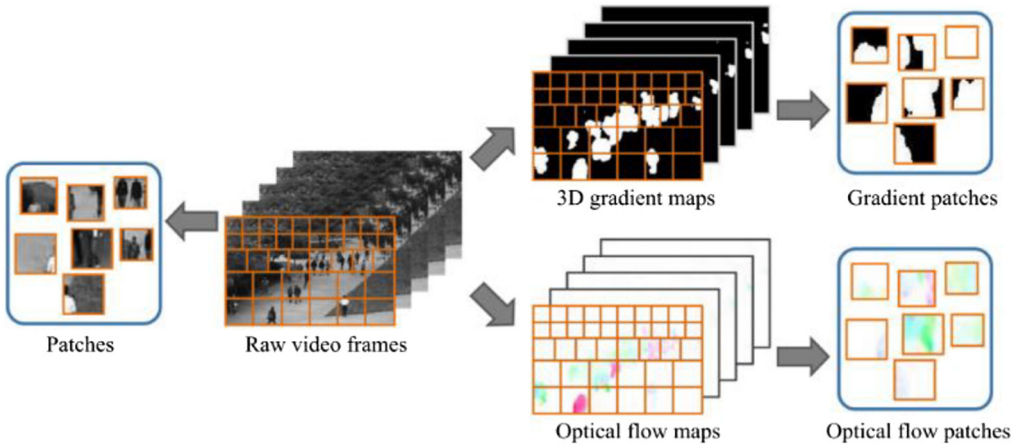

Fig. 6 Extraction of local patches.[64]

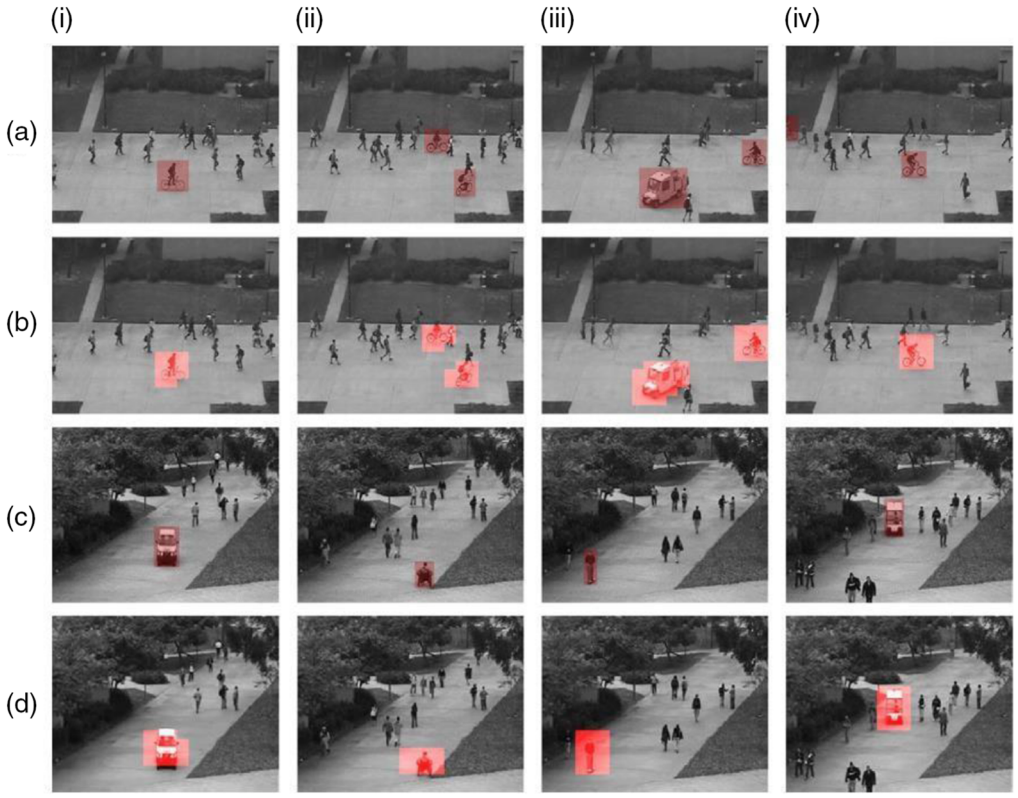

Fig. 7 UCSD dataset-based anomaly detection and localization results of Ref. 64: (a) Ped2 ground truth; (b) Ped2 detection results; (c) Ped1 ground truth; and (d) Ped1 detection results.

problem such as low processing power, and for online anomaly detection, optical flow demands special hardware, etc. (2) Dataset, i.e., only a few transit surveillance datasets are available due to privacy and security concerns. A lack of these datasets makes comparing the performance scores obtained from the various algorithm difficult. (3) Aerial surveillance, i.e., tracking methods that are used in transit surveillance, face issues with aerial video. (4) Commercial systems, i.e., commercial providers advertise anomaly detection systems that have high capability and accuracy for anomaly detection and recognition. Presently, no published work is available in the literature that can prove the providers' claims.

4 Comparative Study and Analysis

As discussed in the previous section, due to increasing crime and terrorist acts, the need for abnormal behavior systems has become a critical aspect within society. Therefore, numerous approaches have been proposed to detect abnormalities in surveillance videos. Unfortunately, only a few of them have performed in the expected zone. Here, in this section, we provide a snapshot of some of the recent popular approaches that can be used for abnormal behavior detection. The summary of these approaches along with their pros and cons can be observed in Table 1.

Once a model is designed, it is evaluated following the frame or pixel-level ground truth protocol on a publicly available benchmark datasets, such as CUHK Avenue,[9] ShanghaiTech,[10] UCSD,[11] etc. The training set in these datasets contains only the normal behavior, whereas the testing set encompass both normal as well as abnormal behaviors. (a) UCSD contains 6800 frames for training and 7200 frames for testing. Altogether, there are 40 different abnormal activities and anomalies such as bikers, skaters, wheelchairs, small carts, and pedestrians walking on a lawn. It encompasses two subsets: Ped1 and Ped2. In Ped1, pedestrians are moving toward and away from the camera, whereas in Ped2, pedestrians are moving parallel to the

Table 1 Summary of different anomaly detection methods.

Methodology	Discussion	Advantage	Disadvantages	Ref
Information theory and statistical analysis are used for detecting malicious attacks.	This approach uses Kullback Leibler Divergence for estimating deviation present between features, then the threshold value is calculated and used to characterized patterns of usage to either normal or abnormal.	Can be used in any smartphone for detecting anomalous behavior. Furthermore, it does not require high computational resources such as the machine learning method.	Can only detect smartphone anomalies.	70
ODM-ADS algorithm is used for detection of anomalies and calculation of robustness of statistical algorithms when poisoning attacks occur.	ODM-ADS is an adversarial statistical learning algorithm used because of its numerous advantages for the detection of anomalies. Datasets such as UNSW-NB15 and NSL-KDD are used for the experiment.	This algorithm is superior in processing time, false positive rate, rate of detection, and accuracy in comparison with seven peer algorithms	Capability and scalability of the proposed method in finding requirements and demands in real world fog-based environments for application in the real world was not performed yet.	71
Anomalous behavior in embedded system is detected using statistical approaches based on cumulative distribution functions.	Series and individual operation are continuously checked for the analysis of system internal timing. The deviation arising in the time required for the execution indicates malware.	This approach was able to achieve a low false positive rate and high detection rate.	N/A	72
MEDUSA framework is implemented for malware detection	This framework continuously looks for system behavior. Whenever the system shows some deviation, malware is detected.	Malware of any type, known or unknown, can be easily detected.	Implementing this framework in Windows OS was left as future work.	73
Statistical learning method to monitor device activity, and whenever the deviation in normal operation occurs, an anomaly is detected	Time series analysis methods are used for the anomaly detection. Experiment results reveal that machine learning and statistical methods can be utilized to recognize the anomalies in IoT devices.	The proposed approach is device and platform independent because the statics of the devices such as disk, CPU cycle usage, etc., can be easily gained from the IoT APIs.	N/A	74

Table 1 (*Continued*).

Methodology	Discussion	Advantage	Disadvantages	Ref
Pearson correlation and low variance filter are merged for the selection of only the appropriate features needed by the IDS.	First, a single dataset is formed, and then preprocessing is performed followed by the decision tree algorithm for the purpose of classification. For the experiment, the ISCX2017 dataset is used.	Number of features to be processed was decreased. Calculation time of the proposed system was reduced to 5.6 s from 71 s.	Can only detect a limited number of attacks.	53
Gaussian mixture model is used for outlier detection.	The data was produced by a Gaussian mixture distribution *a priori*, and the algorithm detects the outlier in the high dimension dataset.	A real dataset is used for the experiment and result reveals that this algorithm is appropriate for outlier detection in high dimension dataset.	In real applications, the generation of data from a Gaussian mixture distribution *a priori* does not hold; as a result, accuracy of the detection is reduced.	75
Clustering and data mining method with a sliding window is used for anomaly detection	Normal traffic data were used for comparing with the real-time data of the network to find any abnormal data of network traffic. A real dataset was used for the experiment.	Can detect anomalies in real-time network data and has a low false alarm rate with a high detection rate.	Data mining algorithm has to scan the whole database, which is time consuming.	76
AMDN used for anomaly detection.	In the first phase, motion and appearance features are learned. In phase 2, the output of phase 1 is provided to three different SVM algorithms. Then, a novel late fusion scheme detects anomaly.	Superior to handcrafted features and can learn features such as motion and appearance along with their correlation automatically.	For real-time processing, this method is not suitable due to the computational overhead in the test phase.	77
Rule-based classification technique is used to detect anomalous behavior in videos.	This method does not need a training sample that is labeled normal or abnormal behavior. Contains three phases of detection: moving object, tracking of object, and understanding behavior for activity identification.	Accuracy of this method highly relies on rules. Unlike other systems, new events in rule-based systems can be easily detected by changing the rules.	Overlapping of multiple objects and focus on more suspicious data were left as future work.	78

Table 2 Frame-level results comparison among some popular approaches in UCSD Ped1 dataset.

S.N.	Approaches	Frame-level		Ref.
		EER (%)	AUC (%)	
1	Ada-Net	15.8	90.4	61
2	SiTGRU	32.1	73.1	30
3	ConvAE	27.9	81.0	56
4	Unmasking	31.0	68.4	59
5	Hu et al.	22.0	71.0	57
6	Parallel spatial-temporal CNN	**6.29**	96.73	16
7	GAN	8	**97.4**	62
8	Binary feature	25.34	—	18
9	ST-CaAE	18.8	90.5	60
10	BR-GAN	22.5	84.7	17
11	GKIM	16.5	—	19
12	Deep one-class learning	15.60	91.40	20
13	Two-stage	29.2	77.8	32
14	ST-AE	12.5	89.9	58
15	FFP	23.5	83.1	63
16	MGFC-AAE	20	85	64
17	ISTL	29.8	75.2	13
18	GMFC-VA	11.3	94.9	14
19	DeepOC	23.4	83.5	65
20	Anomaly-Net	25.2	83.5	66
21	FSCN	25.2	82.4	67
22	DF-ConvLSTM-VAE	16.7	88.4	3

Note: Bold values represent the best values.

camera plane. (b) CUHK Avenue has 15,328 frames for training and 15,324 frames for testing. It contains 47 different abnormal activities, such as loitering, running, and throwing objects. (c) ShanghaiTech is the largest and most challenging dataset; it contains 330 videos for training and 107 videos for testing. A total of 130 abnormal activities are given in 13 various scenes.

In Table 2, we used the UCSD Ped1 dataset to compare the recent approaches designed for anomaly detection using performance metrics such as the equal error rate (EER) and area under the curve (AUC). A comparison is done at the frame level. A lower EER and higher AUC denote a better performance, and the best value is highlighted in bold. From the table, it can be clearly observed that GAN[62] has outstanding performance in achieving the best value of AUC (97.4%) in comparison with the other state-of-the-art methods, whereas for EER, it achieved the second-best value of 8 and lags by 1.71 behind the parallel spatial-temporal CNN,[16] which achieved the best EER value of 6.29.[16] Reference [16] is the most competitive model close to the GAN[62] for various reasons. One reason is its application of varying sizes of patches as in Ref. 64, which enables it to capture better appearance/ motion anomalies. Another reason is its deeper layers, which enables it to extract more discriminative spatio-temporal features using 3D CNN. It has

five convolution layers, five pooling layers, and three fully connected layers. By contrast, GAN[62] uses a shallow network and does not apply varying sizes of patches. In addition, other factors such as the selection of hyperparameters and loss function can also influence the detection accuracy. In addition, the performance of other GAN models such as in Refs. 7 and 64 achieved the state-of-the-art performance in the table. Therefore, it is proved that a better performance is can be achieved using the GAN in detecting abnormal activities in a video. Accuracy can be further improved using varying sizes of patches, including deeper layers with a residual connection, and tuning the hyperparameters.

5 Conclusions

Surveillance cameras nowadays are being used in every place to monitor the activities of the people, and it is not possible for a human being to stay 24/7 in front of camera and monitor people's activities. So, for security reasons, there is a need for fully automatic systems that can recognize anomalous activities with high accuracy; the demand for these systems has increased sharply in the past decades. To detect anomalies, there are numerous methods that have been adopted by researchers. This paper summarized a comprehensive survey of abnormal behavior detection. Description, pros, and cons of various machine and non-machine learning techniques that can be used for abnormality detection were discussed in depth. Similarly, more focus was given to one of the most powerful generative approaches: GAN. A detailed description was provided for designing a GAN for a better abnormality detection rate. Moreover, we supplied a comparative analysis between recent state-of-the-art methods based on their methodology, advantages, and disadvantages. Also, in the quantitative comparison, we compared the robust approaches at the frame level on the UCSD Ped1 dataset. During our comparison, the GAN model achieved the highest AUC and the second-highest EER. We also provided different suggestions on how to further increase the detection rate of GAN for abnormal behavior detection in surveillance videos.

References

1. D. Miljković, "Review of novelty detection methods," in *The 33rd Int. Convention MIPRO*, IEEE (2010).
2. M. Cho et al., "Unsupervised video anomaly detection via normalizing flows with implicit latent features," *Pattern Recognit.* **129**, 108703 (2022).
3. L. Wang et al., "Unsupervised anomaly video detection via a double-flow ConvLSTM variational autoencoder," *IEEE Access* **10**, 44278–44289 (2022).
4. X. Wang et al., "Robust unsupervised video anomaly detection by multipath frame prediction," *IEEE Trans. Neural Networks Learn. Syst.* **33**(6), 2301–2312 (2022).
5. C. Huang et al., "Self-supervised attentive generative adversarial networks for video anomaly detection," *IEEE Trans. Neural Networks Learn. Syst.* **33**, 1–15 (2022).
6. J. Luo et al., "SMD anomaly detection: a self-supervised texture–structure anomaly detection framework," *IEEE Trans. Instrum. Meas.* **71**, 5017611 (2022).
7. D. Zhang et al., "Weakly supervised video anomaly detection via transformer-enabled temporal relation learning," *IEEE Signal Process Lett.* **29**, 1197–1201 (2022).
8. Y. Liu et al., "Collaborative normality learning framework for weakly supervised video anomaly detection," *IEEE Trans. Circuits Syst. II Express Briefs* **69**(5), 2508–2512 (2022).
9. C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in matlab," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 2720–2727 (2013).
10. W. Liu et al., "Future frame prediction for anomaly detection–a new baseline," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 6536–6545 (2018).
11. V. Mahadevan et al., "Anomaly detection in crowded scenes," in *IEEE Computer Society Conf. Comput. Vision and Pattern Recognit.*, pp. 1975–1981 (2010).
12. B. Omarov et al., "State-of-the-art violence detection techniques in video surveillance security systems: a systematic review," *PeerJ Comput. Sci.* **8**, e920 (2022).

13. R. Nawaratne et al., "Spatiotemporal anomaly detection using deep learning for real-time video surveillance," *IEEE Trans. Ind. Inf.* **16**(1), 393–402 (2020).
14. Y. Fan et al., "Video anomaly detection and localization via Gaussian mixture fully convolutional variational autoencoder," *Comput. Vision Image Understanding* **195**, 102920 (2020).
15. M. Gong et al., "Local distinguishability aggrandizing network for human anomaly detection," *Neural Networks* **122**, 364–373 (2020).
16. Z. Hu et al., "Parallel spatial-temporal convolutional neural networks for anomaly detection and location in crowded scenes," *J. Vis. Commun. Image Represent.* **67**, 102765 (2020).
17. Z. Yang, J. Liu, and P. Wu, "Bidirectional retrospective generation adversarial network for anomaly detection in videos," *IEEE Access* **9**, 107842–107857 (2021).
18. R. Leyva, V. Sanchez, and C. T. Li, "Abnormal event detection in videos using binary features," in *Proc. IEEE Telecommun. And Signal Process.*, Barcelona, pp. 621–625 (2017).
19. H. Ullah et al., "Anomalous entities detection and localization in pedestrian flows," *Neurocomputing* **290**, 74–86 (2018).
20. J. Sun, J. Shao, and C. He, "Abnormal event detection for video surveillance using deep one-class learning," *Multimedia Tools Appl.* **78**(3) 3633–3647 (2019).
21. L. Zhang, Y. Chen, and S. Liao, "Algorithm optimization of anomaly detection based on data mining," in *10th Int. Conf. Meas. Technol. And Mechatron. Autom.*, pp. 402–404 (2018).
22. M. Chhabra et al., "Improving automated latent fingerprint detection and segmentation using deep convolutional neural network," *Neural Comput. Appl.* **34**, 1–24 (2022).
23. P. Sharma, M. Kumar, and H. Sharma, "Comprehensive analyses of image forgery detection methods from traditional to deep learning approaches: an evaluation," *Multimedia Tools Appl.* **34**, 1–34 (2022).
24. A. Garg et al., "Blockchain-based online education content ranking," *Educ. Inf. Technol.* **27**, 4793–4815 (2022).
25. A. Aggarwal et al., "COVID-19 risk prediction for diabetic patients using fuzzy inference system and machine learning approaches," *J. Healthc. Eng.* **2022**, 4096950 (2022).
26. S. Raheja et al., "Modeling and simulation of urban air quality with a 2-phase assessment technique," *Simul. Modell. Pract. Theory* **109**, 102281 (2021).
27. A. K. Singh et al., "Secure and energy efficient data transmission model for WSN," *Intell. Autom. Soft Comput.* **27**(3), 761–769 (2021).
28. M. Kumar et al., "Automatic brain tumor detection using machine learning and mixed supervision," in *Evolving Role of AI and IoMT in the Healthcare Market*, F. Al-Turjman et al., Eds., pp. 247–262, Springer, Cham (2021).
29. A. Goswami et al., "Sentiment analysis of statements on social media and electronic media using machine and deep learning classifiers," *Comput. Intell. Neurosci.* **2022**, 9194031 (2022).
30. H. Fanta, Z. Shao, and L. Ma, "SiTGRU: single-tunnelled gated recurrent unit for abnormality detection," *Inf. Sci.* **524**, 15–32 (2020).
31. W. Tylman, "Anomaly-based intrusion detection using Bayesian networks," in *Third Int. Conf. Dependability of Comput. Syst. DepCoS-RELCOMEX*, pp. 211–218 (2008).
32. S. Wang et al., "Detecting abnormality without knowing normality: a two-stage approach for unsupervised video abnormal event detection," in *Proc. 26th ACM Int. Conf. Multimedia*. Seoul, pp. 636–644 (2018).
33. R. H. Gharaei, R. Sharify, and H. Nezamabadi-Pour, "An efficient outlier detection method based on distance ratio of *k*- nearest neighbors," in *9th Iranian Joint Congr. Fuzzy and Intell. Syst.*, pp. 1–5 (2022).
34. I. Aljamal et al., "Hybrid intrusion detection system using machine learning techniques in cloud computing environments," in *IEEE 17th Int. Conf. Software Eng. Res., Manage. And Appl. (SERA)*, Honolulu, Hawaii, pp. 84–89 (2019).
35. A. A. Sodemann, M. P. Ross, and B. J. Borghetti, "A review of anomaly detection in automated surveillance," *IEEE Trans. Syst., Man, Cybern.—Part C: Appl. Rev.* **42**(6), 1257–1272 (2012).
36. K. Rohit, K. Mistree, and J. Lavji, "A review on abnormal crowd behavior detection," in *Int. Conf. Innov. In Inf. Embed. And Commun. Syst.* (2017).

37. M. Z. Zaheer et al., "A brief survey on contemporary methods for anomaly detection in videos," in *ICTC 2019* (2019).
38. T. Ko, "A survey on behavior analysis in video surveillance for homeland security applications," in *37th IEEE Appl. Imagery Pattern Recognit. Workshop* (2008).
39. I. Arora and M. Gangadharappa, "A survey of motion detection in image sequences," in *6th Int. Conf. Comput. For Sustain. Global Dev.* (2019).
40. N. Patil and P. K. Biswas, "A survey of video datasets for anomaly detection in automated surveillance," in *Sixth Int. Symp. Embed. Comput. And Syst. Design* (2016).
41. N. A. Nemade and V. V. Gohokar, "A survey of video datasets for crowd density estimation," in *Int. Conf. Global Trends in Signal Process., Inf. Comput. And Commun.* (2016).
42. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection for discrete sequences: a survey," *IEEE Trans. Knowl. Data Eng.* **24**(5), 823–839 (2012).
43. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Comput. Surv.* **41**(3), 15 (2009).
44. Z. Niu et al., "A survey of outlier detection methodologies and their applications," *Lect. Notes Comput. Sci.* **7002**, 380–387 (2011).
45. C. Savitha and D. Ramesh, "Motion detection in video surveillance: a systematic survey," in *2nd Int. Conf. on Invent. Syst. And Control*, Coimbatore, pp. 51–54 (2018).
46. R. Nandhini Abirami et al., "Deep CNN and deep GAN in computational visual perception-driven image analysis," *Complexity* **2021**, 5541134 (2021).
47. M. Zhou et al., "GraphMA: graph-based semi-supervised manifold alignment for indoor WLAN localization," *IEEE Sens. J.* **17**(21), 7086–7095 (2017).
48. Y. Wu, Y. Ye, and C. Zhao, "Coherent motion detection with collective density clustering," in *ACM Conf. Multimedia*, pp. 361–370 (2015).
49. S. Deep et al., "A survey on anomalous behavior detection for elderly care using dense-sensing networks," *IEEE Commun. Surv. Tutor.* **22**(1), 352–370 (2020).
50. K. Pragadeeswari and G. Yamuna, "Challenges and solutions in motion surveillance – a survey," in *Int. Conf. Commun. And Signal Process.*, Chennai, India, pp. 0960–0964 (2019).
51. J. Candamo et al., "Understanding transit scenes: a survey on human behavior-recognition algorithms," *IEEE Trans. Intell. Transp. Syst.* **11**(1), 206–224 (2010).
52. Y. Shi et al., "Visual analytics of anomalous user behaviors: a survey," *IEEE Trans. Big Data* **8**, 377–396 (2020).
53. F. A. Saputra et al., "Botnet detection in network system through hybrid low variance filter, correlation filter and supervised mining process," in *Thirteenth Int. Conf. Digital Inf. Manage. (ICDIM)*, Berlin, pp. 112–117 (2018).
54. K. Shreedarshan and S. S. Selvi, "An adaptive swarm optimization technique for anomaly detection in crowded scene," in *Int. Conf. Circuits, Controls, Commun. And Comput. (I4C)*, pp. 1–5 (2016).
55. R. Raghavendra et al., "Optimizing interaction force for global anomaly detection in crowded scenes," in *Proc. Int. Conf. Comput. Vision and Pattern Recognit., CVPRW'11*, IEEE, pp. 136–143 (2011).
56. M. Hasan et al., "Learning temporal regularity in video sequences," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. (CVPR)*, pp. 733–742 (2016).
57. X. Hu et al., "A weakly supervised framework for abnormal behavior detection and localization in crowded scenes," *Neurocomputing* **383**, 270–281 (2020).
58. S. C. Yong and H. T. Yong, "Abnormal event detection in videos using spatiotemporal autoencoder," in *Proc. Int. Symp. Neural Networks*, pp. 189–196 (2017).
59. R. T. Ionescu et al., "Unmasking the abnormal events in video," in *Proc. ICCV*, pp. 2914–2922) (2017).
60. N. Li, F. Chang, and C. Liu, "Spatial-temporal cascade autoencoder for video anomaly detection in crowded scenes," *IEEE Trans. Multimedia* **23**, 203–215 (2021).
61. H. Song et al., "Learning normal patterns via adversarial attention-based autoencoder for abnormal event detection in videos," *IEEE Trans. Multimedia* **22**(8), 2138–2148 (2020).
62. M. Ravanbakhsh et al., "Abnormal event detection in videos using generative adversarial nets," in *Proc. IEEE Int. Conf. Image Process.*, pp. 1–5 (2017).

63. W. Liu et al., "Future frame prediction for anomaly detection a new baseline," in *IEEE Conf. CVPR* (2018).
64. N. Li and F. Chang, "Video anomaly detection and localization via multivariate Gaussian fully convolution adversarial autoencoder," *Neurocomputing* **369**, 92–105 (2019).
65. P. Wu, J. Liu, and F. Shen, "A deep one-class neural network for anomalous event detection in complex scenes," *IEEE Trans. Neural Networks Learn. Syst.* **31**(7), 2609–2622 (2020).
66. J. T. Zhou et al., "AnomalyNet: an anomaly detection network for video surveillance," *IEEE Trans. Inf. Forensics Secur.* **14**(10), 2537–2550 (2019).
67. P. Wu et al., "Fast sparse coding networks for anomaly detection in videos," *Pattern Recognit.* **107**, 107515 (2020).
68. R. Garg and S. Singh, "Intelligent video surveillance based on YOLO: a comparative study," in *Int. Conf. Adv. In Comput. Commun. And Control*, pp. 1–6 (2021).
69. L. Yajing and D. Zhongjian, "Abnormal behavior detection in crowd scene using YOLO and Conv-AE," in *33rd Chin. Control and Decision Conf.*, pp. 1720–1725 (2021).
70. J. Prakash, S. Sankaran, and J. Jithish, "Attack detection based on statistical analysis of smartphone resource utilization," in *IEEE 16th India Council Int. Conf.*, Rajkot, India, pp. 1–4 (2019).
71. N. Moustafa et al., "Outlier irichlet mixture mechanism: adversarial statistical learning for anomaly detection in the fog," *IEEE Trans. Inf. Forensics Secur.* **14**(8), 1975–1987 (2019).
72. N. A. Carreón, A. Gilbreath, and R. Lysecky, "Statistical time-based intrusion detection in embedded systems," *Design, Autom. & Test in Eur. Conf. & Exhibit.*, Grenoble, pp. 562–567 (2020).
73. M. E. Ahmed, S. Nepal, and H. Kim, "MEDUSA: malware detection using statistical analysis of system's behavior," *IEEE 4th Int. Conf. Collaboration and Internet Comput.*, Philadelphia, Pennsylvania, pp. 272–278 (2018).
74. F. Li et al., "System statistics learning-based IoT security: feasibility and suitability," *IEEE Internet Things J.* **6**(4), 6396–6403 (2019).
75. W. Liu et al., "Outlier detection algorithm based on Gaussian mixture mode," in *IEEE Int. Conf. Power, Intell. Comput. And Syst.*, Shenyang, pp. 488–492 (2019).
76. A. R. Jakhale, "Design of anomaly packet detection framework by data mining algorithm for network flow," in *Int. Conf. Comput. Intell. In Data Sci.*, Chennai, pp. 1–6 (2017).
77. D. Xu et al., "Detecting anomalous events in videos by learning deep representations of appearance and motion," *Comput. Vision Image Understanding* **156**, 117–127 (2017)
78. S. Chaudhary, M. A. Khan, and C. Bhatnagar, "Multiple anomalous activity detection in videos," *Procedia Comput. Sci.* **125**, 336–345 (2018).

Biographies of the authors are not available.