Insights of Deep Learning-Based Video Anomaly Detection Approaches

Dipak Ramoliya and Amit Ganatra

Abstract Deep learning is a powerful computing strategy that has changed the landscape of computer vision. It has been used to tackle complicated cognitive tasks such as detecting abnormalities in videos. Anomalies in the video are events or objects in the footage that don't fit the typical, learned patterns. Using deep learning, it is possible to automatically and in real-time identify unusual actions and objects like fights, riots, traffic rule violations, abrupt rushes, and the presence of weapons in restricted areas or abandoned luggage. Despite the challenges posed by video anomaly detection, this review offers a comprehensive assessment of published deep learning algorithms for the task. Future research can build on this work by understanding the existing methods to create more effective solutions. First, the challenges of video anomaly identification are discussed as the benefits of deep learning in anomaly detection. Furthermore, several types of abnormalities were explored, followed by diverse methodologies for anomaly identification. Furthermore, significant aspects of anomaly detection using deep learning, including learning approaches, were presented. Finally, numerous datasets used in anomaly detection were examined, followed by a discussion of deep learning-based algorithms for spotting video anomalies.

Keywords Anomaly detection · Outlier detection · Deep learning · Computer vision

D. Ramoliya (✉)
Department of Computer Science and Engineering, Devang Patel Institute of Advance Technology and Research (DEPSTAR), Faculty of Technology and Engineering (FTE), Charotar University of Science and Technology (CHARUSAT), Changa 388421, India
e-mail: dipakramoliya.ce@charusat.ac.in

A. Ganatra
Department of Computer Engineering, Devang Patel Institute of Advance Technology and Research (DEPSTAR), Faculty of Technology and Engineering, Charotar University of Science and Technology (CHARUSAT), Changa, Anand 388421, India

Computer Science and Engineering Department, Faculty of Engineering and Technology, Parul University (PU), Waghodia, Vadodara 391760, India

G. Rajakumar et al. (eds.), *Intelligent Communication Technologies and Virtual Mobile Networks*, Lecture Notes on Data Engineering and Communications Technologies 171,
https://doi.org/10.1007/978-981-99-1767-9_48

1 Introduction

The usage of surveillance cameras has grown significantly in recent years to improve public safety. Unfortunately, the number and rate of production of surveillance data have outpaced the capacity of security forces to monitor this footage. As a result, there is a noticeable gap in the application of cameras used for surveillance, and the ratio of cameras to human monitors which is unacceptable. However, it is conceivable to create systems that automate the process of detecting anomalies using the computational power and resources available today. Such systems can handle issues in other relevant areas of computer vision and machine learning, such as activity recognition, and video annotation in addition to reducing (or perhaps eliminating) the requirement for user intervention in anomaly detection [1–3]. Figure 1 provides the general flow of anomaly detection.

The term "anomaly identification" refers to the process of uncovering an object or occurrence of any event that diverges from the norm, either inside a dataset or in relation to other datasets. Anomaly detection is becoming more and more important, as recognized in data mining, artificial intelligence, machine learning, computer vision, and deep learning communities. This is due to the rising demand in extensive domains, such as financial surveillance, health, medical risk, agriculture, safety, and security. Identifying abnormal events, such as traffic accidents, crimes, or unlawful activity, is a crucial duty in video surveillance. Compared to regular life, anomalous occurrences are extremely rare. Therefore, developing efficient computer vision algorithms for automatic video anomaly identification is crucial for saving time and manpower. The goal of any useful anomaly detection system is to provide an early warning of any activity that breaks with typical patterns, as well as pinpoint when this deviation first occurred [4, 5].

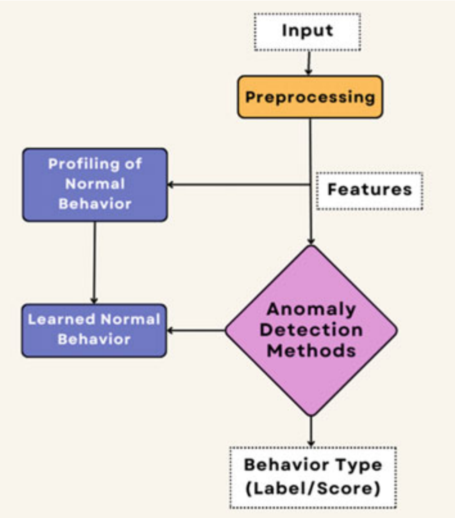
Input
Preprocessing
Profiling of Normal Behavior
Features
Learned Normal Behavior
Anomaly Detection Methods
Behavior Type (Label/Score)

Fig. 1 General flow of anomaly detection

2 Background Theory and Related Work

2.1 Different Types of Video Anomalies

(1) Anomalies both locally and globally: Local anomalies are activities that differ greatly from what is happening in the area around them. For example, a vehicle driving in the wrong direction. Individuals who act strangely here are different from their neighbors. Global anomalies are activities that happen worldwide and are abnormal or unexpected. Global anomalies are caused by a group or collective anomalies. If a group of data points shows normal behavior when examined individually, but anomalous behavior when looked at as a group, that set is called a collective or group anomaly [1, 6].

(2) Anomalies in point and interaction: Some data points are very different from the rest, and these are called point anomalies. To put it another way, a point anomaly is a form of random abnormality that has the potential to be misunderstood as deviant conduct, such as loitering. Interaction anomalies, such as a fight between two people, are examples of abnormal interactions between separate entities [7].

(3) Anomalies that are contextual or conditional: These are data items that don't fit in with the rest of the dataset, creating anomalies. Contextual data, such as time and place, is used to determine the context [8]. Behavioral characteristics, such as normal activities, are used to determine the anomaly. This category includes most video oddities such as violence, riots, stampedes, and so on. Spatial and temporal abnormalities can also be considered contextual anomalies. Contextual abnormalities are generally best represented by spatio-temporal anomalies [1, 7].

3 Video Anomalies Detection Using Deep Learning-Based Approaches

3.1 Trajectory-Based Methods

To understand the trajectory of the item or topic of interest, the researchers followed it between frames. They then used this information to deduce any abnormal behavior the object may have been carrying out [9]. The accuracy of trajectory-based approaches is highly dependent on object recognition and tracking. Tracking effectiveness can be negatively impacted by factors such as high crowd density, poor video quality, rapid camera movement, and occlusion. As a result, the trajectory approaches were discovered to provide better accuracy in sparsely packed environments, rather than in moderately or highly congested environments. Furthermore, while determining the anomaly, these approaches lack contextual information [10].

3.2 Global Pattern-Based Methods

There are a growing number of studies on convenience stores and grocery stores that use video footage to understand how these stores work. These studies use low- or medium-level features to examine the footage as a whole [11]. They rely on global pattern-based approaches, which are successful in both moderately and densely populated areas. Space–time gradients, motion energy, and the direction of light flux are all utilized in these methods. Using a crowd behavior-based social attribute-aware force model, we can identify unusual occurrences [8]. Statistical hypothesis tests and Gaussian Noise Model approximation are applied to discover abnormal events [12]. A model for anomalous crowd behavior is given using kinetic energy derived using the assistance of optical flow and crowd dispersion index [13]. Detecting anomalous behavior in crowds can be difficult. Several models have been created to try and make this process easier, including ones that use motion maps or hierarchical feature representation [14]. However, locating abnormalities can be time-consuming using these methods.

3.3 Grid Pattern-Based Methods

Grid pattern techniques are frequently employed because they can shorten the amount of time required for analysis. This is accomplished by restricting the number of characteristics that are gathered from a particular spatio-temporal region. This means that anomalies in each cell, or grid, are studied separately without respect to their neighbors [15]. Furthermore, similarities are retrieved from the frames' split blocks rather than interpreting the frames as a singular entity. These methods do not need the use of STIPs or saliency detection. It is possible to improve video anomaly detection by examining numerous attributes or grid patterns. Spatial anomaly detection using keypoint detection [16], temporal anomaly detection using background removal using the Gaussian Mixture Model (GMM) [17], and sparse combination learning [17] is just a few of the methods that have been used. Detecting spatial and temporal abnormalities in busy settings can be tricky. Some researchers have explored using maps that show both spatial and temporal anomalies [3]. Another approach is to look at object speed, texture, and size in order to identify changes in a particular area. Grid pattern-based approaches are similar to local region-based methods in that they use a probabilistic model to identify anomalies. Markov Chains, Gaussian mixture models, and Bag-of-words [9] are used to compile a set of compact characteristics for use in creating a grid with cells of varying sizes. Recently, grid-based video anomaly detection algorithms have begun to incorporate deep learning techniques, such as deep Gaussian mixture models and spatio-temporal convolution neural networks.

3.4 Representation Learning Models

Represent learning is the process of learning effective representations of raw video data while counting the fact critical previous information about the situation. It reduces the impact of the dimensionality curse by translating very high-dimensional video data into manageable d-dimensional vectors, making computational operations easier [15]. Classifiers and predictors built on representation learning models could be applied in contexts as diverse as action recognition, anomaly detection, and object recognition. Additionally, this method aids in the extraction of a generic, compact, efficient, and simple video descriptor or representation [11].

(1) Deep neural networks inspired by sparse coding: One of the most extensively utilized strategies for anomaly identification is normal distribution learning. The training set of video data serves as a dictionary for the sparse coding-based video anomaly detection algorithm. It then discovers the aberrant actions that the learned dictionary is unable to rebuild [8]. Adaptive sparse representations combine a combined sparse model for multiple-item anomaly detection with nonlinearity for class separation. These approaches seldom took into account the encoding of both spatial and temporal links between different blocks and frames. The accuracy of anomaly detection is improved when an intra-frame classification approach is used [11].

(2) Models based on reconstruction: Video anomaly detection reconstruction models can only learn what is typical behavior or activity from the training video dataset. Anomalous or aberrant behaviors deviate from the acquired model, resulting in poor reconstruction. Auto-encoders and Principal Component Analysis (PCA) are among the techniques employed. Because of its hierarchical feature extraction technique, Deep auto-encoder (Deep-AE) may provide superior feature representations. A deep CAE was trained with an end-to-end framework to maintain spatio-temporal properties independent of encoding dynamics [9]. It is noise-proof, can detect visual anomalies, and can even pick them out by itself. A sparse AE-based system sufficient for dynamically identifying and localizing video anomalies is described, which reduces the memory required at the time of execution and the false positive rate [12]. The concept, which is founded on the basis of subspace clustering, structured AE, has the ability to make use of nonlinear transformations in order to seamlessly translate the input to the output while still maintaining the local and global subspace structures [13].

(3) Slow feature analysis: The slowness concept underpins slow feature analysis (SFA) [9]. Human activity recognition uses SFA-based video feature representations at first. For video anomaly detection, D-IncSFA integrates the process of extraction of features and detection of anomaly into a single step [17].

3.5 Discriminative Models

The techniques which attempt for finding out the distinguishing traits between classes of data are discriminative modeling. After these models have been trained using supervised learning techniques, they are successful with balanced datasets. However, these techniques are rarely applied for video anomaly identification because of the scarcity of well-balanced video anomaly datasets and the ambiguity surrounding what exactly constitutes an anomalous action or entity. There is a need for a framework to identify video abnormalities using conventional density estimation techniques [16]. However, this has not been extensively explored and a commonly acknowledged definition for the anomaly does not exist. Discriminative models could be more effective in this case, but they are not widely used due to the lack of a consensus on what an anomaly is.

3.6 Predictive Models

A video can be thought of as a spatio-temporal signal, with each frame creating its own unique pattern. Because of their ability to incorporate both geographical and time-based information, predictive models (also known as spatio-temporal models) are frequently employed for video anomaly identification (or motion data) [18].

3.7 Generative Models

Joint probability $P(X, Y)$ can be found via generative modeling, and from there the conditional posterior probability $P(X/Y)$ can be calculated. In general, generative models model each class's actual distribution, while discriminative models model the line of separation between classes [18]. There is no restriction on how the likelihood should be represented for deep generative models; they are still able to learn using the maximum likelihood principle. In order to overcome difficulties caused by limited data and imbalanced data, deep generative models are gaining popularity for application in video anomaly detection. However, more research is needed before these models can be adopted for video anomaly detection.

3.8 Deep Hybrid Models

Auto-encoders are used to learn hidden representations, which are then used to find outliers in deep hybrid models for anomaly detection. The transfer learning models are tremendously successful in extracting the features from the pre-trained models which inspire the use of these models as in hybrid models. Outlier detection using deep learning algorithms is a growing field. However, current hybrid models that use coupled feature extractor training do not perform well. This is due to the absence of a trainable goal specifically for anomaly detection [19]. Researchers have proposed employing singular anomaly detection goals such deep one-class classification or single-class neural networks to counteract this shortcoming [20].

3.9 One-Class Neural Networks

There is increasing enthusiasm in one-class classification methods for anomaly detection, and Chalapathy et al. [21] present an approach that is based on kernel-based deep networks. This technique blends the deep networks' ability to take out a richer data representation with a single-class objective of creating a secure perimeter around normal data. The OC-NN objective dictates the data format in the hidden layer, which is optimized for anomaly detection, making this technique unique. Deep neural networks (DNNs) are a sort of neural network that is more advanced than ordinary neural networks. They are trained to extract common patterns of variation in data. This is done by mapping common factors of variation using Singular Value Decomposition (SVDD). This technique was applied to the MNIST and CIFAR-10 datasets, and it was found that data instances moved to the center of the sphere improved performance [21, 22].

3.9.1 Statistical Techniques

There is a statistical signal processing method known as the Hilbert transform which can be used to obtain a real-valued signal's analytic representation [14]. This attribute is highly promising for detecting abnormalities in health-related time series datasets in real time. The method discovers real-time anomalies by sequentially integrating wavelet analysis, neural networks, and the Hilbert transform. To completely comprehend the potential and usefulness of statistical approaches and deep learning methodologies for anomaly detection, more research is required [19].

4 Performance Evaluatıon Methodologıes

4.1 Error Matrix

Video anomaly detection, along with other areas of computer vision, shares the same aim of classifying data points as normal or abnormal, although using quite different approaches to accomplish this. Video anomaly detection is therefore a binary classification problem that can be addressed. Additionally, any of the fundamental components of performance analysis, such as True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN), can be used to describe each binary classifier choice.

4.2 Receiver Operating Characteristic Curve

To evaluate the efficacy of detection at varying false positive rates (FPRs), a receiver operating characteristic (ROC) plot is constructed with the TPR (sensitivity) on the Y axis and the FPR (probability of false alarm) on the X axis (or thresholds). The Area Under the Curve (AUC) of the Receiver Operating Characteristic curve is used to evaluate the effectiveness of an anomaly detector across a variety of evaluation metrics (AU-ROC). The AU-ROC value ought to be maximized as closely to one as possible (between zero and one) [23].

4.3 Precision-Recall Curve

A plot of precision and recall is known as a precision-recall (PR) plot. With regards to solving the anomaly detection problem, the Area Under the Precision-Recall curve (AU-PR) is superior to the Area Under the Receiver Operating Characteristics (AU-ROC). This is a result of anomaly detection's data imbalance issue, where TNs are larger than TPs. A further feature of the PR curve is its weighting of optimistic and out-of-the-ordinary forecasts. The AU-PR should have the highest possible value between 0 and 1 [23].

4.4 Equal Error Rate

The percentage of incorrectly labeled frames is denoted by the acronym "EER" (for "equal error rate") where both the TPR and FNR are equivalent. When looking for video anomalies, it performs admirably [21].

4.5 Detection Rate

The ratio of detected anomalies to all abnormalities included in the data, expressed as a percentage, is known as the detection rate (DR). The detection rate, or accuracy rate, as determined by the Equal Error Rate, is useful for pinpointing video anomalies.

$$\mathrm{DR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

4.6 Reconstruction Error

The reconstruction error error_reconst(t) for a given frame at the instant in time t is calculated using the equation.

$$\text{error\_reconst(t)} = \sum_{(x,y)} e(x, y, t)$$

where $e(x, y, t)$ is the pixel level reconstruction error for intensity level I at a position (x, y) in the specific frame at time instant t.

Higher reconstruction error scores indicate a higher possibility of spotting a video anomaly [21].

4.7 Anomaly Score

An abnormality indexes, the scoreano(t), is a numeric value between 0 and 1 that describes the degree of an anomaly.

$$\text{Scoreano(t)} = \frac{\text{error\_reconst(t)} - \text{error\_reconst}_{\text{min}}\text{min(t)}}{\text{error\_reconst}_{\text{max}}\text{min(t)}}$$

Higher anomaly score values reflect a higher level of anomaly.

4.8 Regularity Score

Regularity score Scorereg(t), which is determined using the opposite of the anomaly score is calculated using the following equation

$$\text{Scorereg}(t) = 1 - \ \text{Scoreano}(t)$$

The lower the level of anomaly, the greater the regularity score.

Table 1 Datasets

Dataset name	Total video clips
UCSD dataset	98
Avenue dataset	37
ShanghaiTech Campus dataset	437
UCF-Crime dataset	1900
LV dataset	30

5 Datasets

5.1 UCSD Dataset

This dataset (in Table 1) was captured in 2010 at UC San Diego. The data was captured by a high, stationary camera that looked down on the walkways below. Two separate cases, Peds1 and Peds2, are represented in the dataset. In every scene, there is a distinct train test, with normal samples only in the train and abnormal ones in the test. The Peds1 dataset includes 34 training video clips and 36 testing video clips, while the Peds2 dataset only includes 16 training video clips and 12 testing video clips [11].

5.2 Avenue Dataset

Another extensively used dataset is The Avenue dataset, which was published in 2013 and was photographed in CUHK campus avenue. It is comprised of brief video clips that were taken outdoors. The camera faces a sidewalk that runs alongside a building. The majority of anomalies include actions, such as running away, stopping briefly, or tossing something into the air [8].

5.3 ShanghaiTech Campus Dataset

The dataset was made public in 2017. In order to build one of the largest datasets for video abnormalities, CCTV cameras were installed all around the ShanghaiTech campus. There are a broad variety of settings and camera perspectives included in the 13 included scenarios. The 330 normal event movies have a resolution of 480 × 856 pixels, while the 107 mixed resolution recordings have a size of 720 × 576 pixels [12].

5.4 UCF-Crime Dataset

The UCF-Crime dataset was published in 2018. It is a huge collection of online videos shot with a large number of different cameras, producing a variety of situations. It consists of 1900 uncut surveillance videos, 950 of which are of ordinary happenings, and the remaining 950 are from 13 preset categories of criminal activity. Arrest, Assault, Accident, Abuse, Burglary, Explosion, Fighting, Arson, Robbery, Shooting, Stealing, Shoplifting, and Vandalism are a few examples of the events that are covered [6, 7, 24].

5.5 ARENA Dataset

ARENA stands for "Architecture for the REcognition of threats to mobile assets using Networks of multiple Affordable sensors." This dataset was built as a part of the PETS2014 challenge. The dataset was captured by deploying four cameras (frame rate: 30 FPS, resolution: 1280 × 960 pixels) at the University of Reading's crossing path and parking lot without any overlap [9].

5.6 LV Dataset

The live videos (LV) dataset is a comprehensive collection of real-world videos that were recorded by surveillance cameras under challenging environmental conditions such as varying lighting and camera movements. It was made available to the public in 2017 and comprised of thirty video snippets that were culled from various online places in order to illustrate thirty distinct scenarios. The entire time span covered by this dataset is 3.93 h [13]. Figure 2 is the sample of anomalies.

6 Limitations and Open Challenges

6.1 The Necessity of Improved Datasets

Given that this research area is very young, there are significantly fewer publicly accessible datasets regarding anomaly detection from videos. Additionally, the adoption of supported learning-based models is hindered by the data imbalance between anomalous occurrences (Positive samples) and normal events (Negative samples). Good benchmarks are needed to assess the effectiveness of the algorithms used to locate and recognize video anomalies [1, 25].

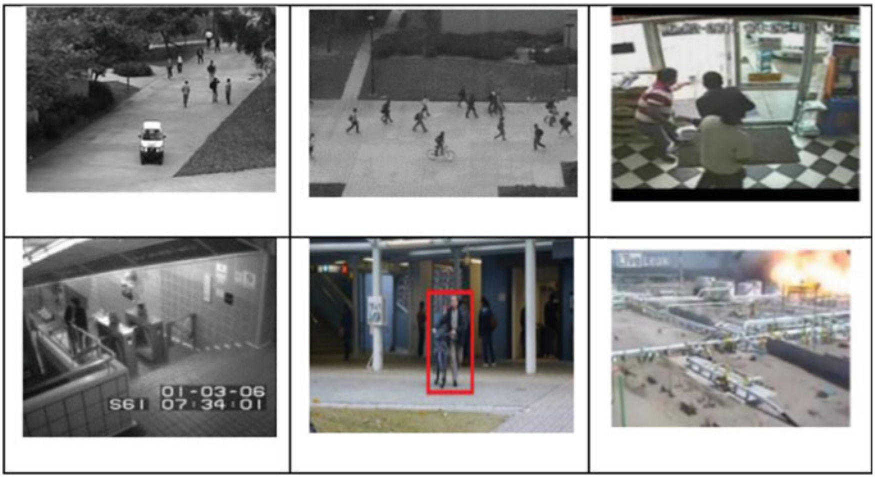

Fig. 2 Sample of anomalies from different datasets

6.2 Reduction in Computational Complexity

The feature representation process used in video anomaly detection is typically computationally expensive and takes a long time, which is a significant impediment to its use in real-world applications. The currently used algorithms for video anomaly detection have a high cost of space and time complexities. Consequently, these techniques are not appropriate for real-world applications [25].

6.3 Methodology Incompleteness

The technique used by the current methods for anomaly detection in the video is insufficient, indicating that no technique is capable of finding every kind of anomaly. As a result, there is a requirement for an effective anomaly detection platform for the video that can solve the issue of incompleteness by utilizing hybrid models [25].

6.4 Accuracy and Processing Time Trade-Off

It is challenging to find the best balance between accuracy in detection and execution time in video anomaly detection systems by using a limited number of highly descriptive features to achieve competitive online performance. Furthermore, the high computational complexity and lengthy processing times are required by deep learning-based algorithms to accomplish the high accuracy of identification and localization of video anomalies. Therefore, there is always an adjustment between the required detection exactness and the processing time along with computing complexity [6, 7].

6.5 Environmental Factors

The changes in object viewpoint and the distance between the object of interest and the monitoring camera are two sensitive areas of video anomaly detection methods. The mentioned issue has been discussed using grid patterns, although there remains room for development. There is a pressing need for efficient approaches to deal with environmental changes such as shifting illumination and backdrops, occlusion problems, noisy input, and the working state of the surveillance camera [4, 5].

7 Conclusion

In this article, the approaches for the detection of video anomaly detection using deep learning have been discussed and it shows the great promise in detecting abnormal events in surveillance footage. Various architectures such as auto-encoder, RNN, and 3D CNN have been proposed and evaluated for their performance on different datasets. The results have shown that these methods can achieve high accuracy in detecting anomalies, but there is still room for improvement in terms of computational efficiency and generalizability to different types of videos. In the future, other architectures such as transformer-based models, and to explore the use of unsupervised learning techniques for anomaly detection can be explored. Overall, the insights gained from this paper demonstrate the potential of deep learning for video anomaly detection and pave the way for further research in this area.

References

1. Nayak R, Pati UC, Das SK (2021) A comprehensive review on deep learning-based methods for video anomaly detection. Image Vis Comput 106:104078
2. Wang L et al (2022) Unsupervised anomaly video detection via a double-flow ConvLSTM variational autoencoder. IEEE Access 10:44278–44289
3. Luo W et al (2019) Video anomaly detection with sparse coding inspired deep neural networks. IEEE Trans Patt Anal Mach Intell 43(3):1070–1084
4. Wang H-G, Li X, Zhang T (2018) Generative adversarial network based novelty detection usingminimized reconstruction error. Front Inform Technol Electr Eng 19(1):116–125
5. Franklin RJ, Dabbagol V (2021) Anomaly detection in videos for video surveillance applications using neural networks. In: 2020 fourth international conference on inventive systems and control (ICISC). IEEE

6. Li N et al (2015) Anomaly detection in video surveillance via gaussian process. Int J Patt Recogn Artif Intell 29(06):1555011
7. Sultani W, Chen C, Shah M (2018) Real-world anomaly detection in surveillance videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition
8. Lu C, Shi J, Jia J (2013) Abnormal event detection at 150 fps in Matlab. In: Proceedings of the IEEE international conference on computer vision 2013, pp 2720–2727
9. Xu K, Jiang X, Sun T (2018) Anomaly detection based on stacked sparse coding with intraframe classification strategy. IEEE Trans Multimedia 20(5):1062–1074
10. Munir M et al (2019) A comparative analysis of traditional and deep learning-based anomaly detection methods for streaming data. In: 2019 18th IEEE international conference on machine learning and applications (ICMLA). IEEE
11. Mahadevan V, Li W, Bhalodia V, Vasconcelos N (2010) Anomaly detection in crowded scenes. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, IEEE, 2010, pp 1975–1981
12. Kumari P, Bedi AK, Saini M (2021) Multimedia datasets for anomaly detection: a survey. arXiv preprint: arXiv:2112.05410
13. Piciarelli C, Micheloni C, Foresti GL (2008) Trajectory-based anomalous event detection. IEEE Trans Circ Syst Video Technol 18(11):1544–1554
14. Kiran BR, Thomas DM, Parakkal R (2018) An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. J Imag 4(2):36
15. Zhao Y et al (2017) Spatio-temporal autoencoder for video anomaly detection. In: Proceedings of the 25th ACM international conference on Multimedia
16. Wu P, Liu J, Shen F (2019) A deep one-class neural network for anomalous event detection in complex scenes. IEEE Trans Neural Netw Learn Syst 31(7):2609–2622
17. Leyva R, Sanchez V, Li C-T (2017) Video anomaly detection with compact feature sets for online performance. IEEE Trans Image Process 26(7):3463–3478
18. Liu W, Luo W, Lian D, Gao S (2018) Future frame prediction for anomaly detection—a new baseline. In: Proceedings of the IEEE conference on computer vision and pattern recognition 2018, pp 6536–6545
19. Leyva R, Sanchez V, Li C-T (2017) The lv dataset: a realistic surveillance video dataset for abnormal event detection. In: Proceedings of the 5th international workshop on biometrics and forensics (IWBF), IEEE 2017, pp 1–6
20. Del Giorno A, Andrew Bagnell J, Hebert M (2016) A discriminative framework for anomaly detection in large videos. In: European conference on computer vision. Springer, Cham
21. Sabokrou M et al (2017) Deep-cascade: cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes. IEEE Trans Image Process 26(4):1992–2004
22. Zhou JT et al (2019) Anomalynet: an anomaly detection network for video surveillance. IEEE Trans Inform Forensics Secur 14(10):2537–2550
23. Chalapathy R, Chawla S (2019) Deep learning for anomaly detection: a survey. arXiv preprint: arXiv:1901.03407
24. Ramachandra B, Jones M (2020) Street scene: a new dataset and evaluation protocol for video anomaly detection. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2020
25. Berroukham A et al (2023) Deep learning-based methods for anomaly detection in video surveillance: a review. Bull Electr Eng Inform 12(1):314–327
26. Ullah W et al (2021) CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. Multimedia Tools Appl 80(11):16979–16995