Multimedia Tools and Applications (2023) 82:12635–12651
https://doi.org/10.1007/s11042-022-13954-1

Analysis of anomaly detection in surveillance video: recent trends and future vision

Rohit Raja[1] · Prakash Chandra Sharma[2] · Md Rashid Mahmood[3] · Dinesh Kumar Saini[4]

Received: 13 May 2022 / Revised: 16 August 2022 / Accepted: 12 September 2022 /
Published online: 27 September 2022

Abstract

Video Surveillance (VS) systems are popular. For enhancing the safety of public lives as well as assets, it is utilized in public places like marketplaces, hospitals, streets, education institutions, banks, shopping malls, city administrative offices, together with smart cities. The main purpose of security applications is the well-timed and also accurate detection of video anomalies. Anomalous activities along with anomalous entities are the video anomalies, which are stated as the irregular or abnormal patterns on the video that doesn't match the normal trained patterns. Automatic detection of Anomalous activities, say traffic rule infringements, riots, fighting, and stampede in addition to anomalous entities, say, weapons at the sensitive place together with deserted luggage ought to be done. The Anomaly Detection (AD) in VS is reviewed in the paper. This survey concentrates on the Deep Learning (DL) application in finding the exact count, involved individuals and the occurred activity on a larger crowd at every climate condition. The fundamental DL implementation technology concerned in disparate crowd Video Analysis (VA) is discussed. Moreover, it presented the available datasets as well as metrics for performance evaluation and also described the examples of prevailing VS systems utilized in the real

✉ Prakash Chandra Sharma
prakashsharma12@gmail.com

Rohit Raja
drrohitraja1982@gmail.com

Md Rashid Mahmood
er.mrashid@gmail.com

Dinesh Kumar Saini
dineshkumar.saini@jaipur.manipal.edu

1 Department of Information Technology, Guru Ghasidas Vishwavidyalaya, Bilaspur (CG), India

2 Department of Information Technology, Manipal University Jaipur, Jaipur, India

3 Department of Electronics and Communication Engineering, Guru Nanak Institutions Technical Campus, Hyderabad, India

4 Department of Computer & Communication Engineering, Manipal University Jaipur, Jaipur, India

life. Lastly, the challenges together with propitious directions for additional research are outlined. Pattern recognition has been the subject of a great deal of study during the previous half-century. There isn't a single technique that can be utilised for all kinds of applications, whether in bioinformatics or data mining or speech recognition or remote sensing or multimedia or text detection or localization or any other area. Methodologies for object recognition are the primary focus of this paper. All aspects of object recognition, including local and global feature-based algorithms, as well as various pattern-recognition approaches, are examined here. Please note that we have attempted to describe the findings of many technologies and the future extent of this paper's particular technique. We used the datasets' properties and other evaluation parameters found in an easily accessible web database. Research in pattern recognition and object recognition can greatly benefit from this study, which identifies the research gaps and limits in this subject.

Keywords Anomaly detection · Video surveillance · Deep learning · Machine learning · Feature extraction · Real-time video detection

1 Introduction

A vital part is played by abnormal activity detection in surveillance applications. Automatically capturing the video could be implemented for capturing the abnormal behaviour of humans without the intervention of a system [43]. It is still among the difficult and famous topics of research owing to its complex nature, unlimited possibilities of abnormal cases, together with the limited quantity of normal examples available [71]. Nowadays, from security-sensitive places, namely borders together with military bases to migrate terrorist activities, to private homes intended for burglary prevention, surveillance cameras are everywhere. For a longer time period owing to monotony as well as fatigue, manual operators could not assure vigilant monitoring on account of the huge quantity of data involved. Conventionally, human operators perform the monitoring task. Video feeds as of numerous cameras must be visually inspected by those operators simultaneously [70]. The accelerated deployment of surveillance cameras has been caused by technological advances and decreased costs, both in public along with private facilities [50]. When implemented in crowded scenarios wherein occlusions and clutter occur, the detection rate of anomalies reduces quickly which is the chief issue of this approach [51].

A crucial part is played by abnormal event detection on VS and also smart camera systems. In the literature, the prevailing methods are not generally object-aware, in which disparate objects aren't distinguished in processing [72]. For detecting the anomaly, trajectory features are extracted, or semantic analysis is created by the methods on the initial category centered on detection or tracking outcomes. For example, the issues of track matching along with dynamic event detection are addressed on a frames series [35]. Severe challenges and difficulties are posed by the processing of camera (surveillance) information on crowded events. The utilization of advanced machine learning like DL approaches is one among the best approaches to process this information as well as attain the goal-oriented pattern [23]. However, the issue of AD is significantly open for interpretation. Research endeavors are scattered in the interpretation of the issue, assumptions, and goals [52, 53]. A more concentrated review of recent publications on AD on automated surveillance is created by this study. Therefore, addressing those specific challenges incorporated with this topic in-depth is possible. The broader issue formulations as well as assumptions implemented in AD on automated surveillance studies will be concentrated by this review, instead of primarily offering a review of specific pattern-classification methods. The basic architecture of AD in the surveillance video is exhibited in Fig. 1.

2 Literature survey

This paper is arranged as: section 2.1 explains the Feature Extraction (FE) as well as encoding techniques briefly. Section 2.2 discusses the AD in real-time videos. Section 2.3 illustrates the ML techniques utilized in AD. Section 2.4 elucidates the DL for AD.

2.1 Feature extraction and encoding

Extracting, finding out, and matching the feature or object in one frame of the video is called FE. The FE and encoding techniques used in AD in the surveillance video are elucidated in the section below.

Chandrakar ***et al.*** [4] recommended a Gaussian Process Regression (GPR) centered Video Anomaly Detections (VAD) in addition to localization with hierarchical feature depiction. GPR constructed and modeled a codebook of interaction templates. An inference technique for gauging the possibility of an observed interaction was developed. These local probability scores and globally steady anomaly masks were integrated. Then, as of which anomalies could be identified succinctly. For modeling the relationship of the close by STIP for AD, GPR was employed. Simulations were done on ‘4’ widespread datasets. The top-notch methods were trounced by this method with a lower computational burden. However, complex causality was not handled by this method.

Roberto Leyva ***et al.*** [25] generated Gaussian Mixture Model (GMM) for FE of VAD. For detecting abnormal events, an inference mechanism was used, which estimated the compact feature set via GMM, Markov Chains, together with Bag-of-Words. For augmenting detection

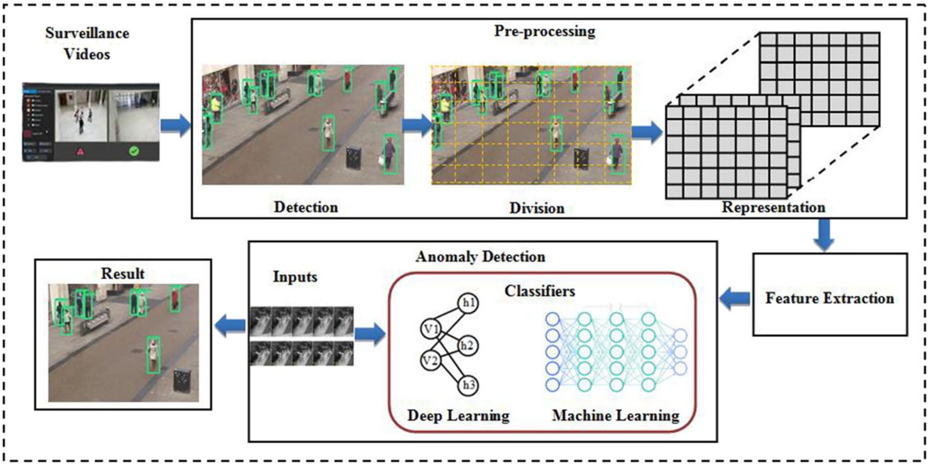

Fig. 1 Basic architecture of AD in surveillance video

accuracy, the joints responses of the models on the local Spatios-temporal neighborhood were also considered. Popular prevailing datasets were taken, which comprised an extensive range of practical videos that were captured by means of surveillance cameras. Other online techniques were outperformed. Contrasted with top-notch non-online methods, the framework attained a competitive detection performance. Nevertheless, long processing times were needed.

Xinwen Gao *et al.* [19] rendered a particle filter-centered algorithm aimed at feature series extracted as of videos. The complete process comprised feature series generation in addition to particle filter tracking. Centered upon the optical flow, an L2-norm extractor was modeled for representing the video's features. After that, these feature series were kept on track by the particle filter. The shift of features series in addition to a larger error on the PF tracking could be caused by the abnormal events' occurrence. This allowed the computers to comprehend and state the anomalies occurrences. UMN dataset was considered. 90% accuracy was reached by the algorithm in frame-level detection. Nevertheless, the particle filter's generalization capability was much weaker contrasted with the deep Neural Networks (NN) on fitting the data.

Qiang Li *et al.* [27] analyzed a Fussy C-Means (FCM) for AD in VS utilizing multi-FE. For detecting a global anomaly, kinetic energy was defined and the first derivative was gauged, and then, an inclusive anomaly score of every test frame was derived. Appearance anomaly, location anomaly, in addition to velocity anomaly was the '3' sorts of the local anomaly that was defined. Disparate sorts of features were extracted; lastly, fused into a united framework. For classifying abnormal instances as of normal instances, an enhanced Normality Sensitive Hashing technique was rendered. Global as well as local anomalies were detected with a comparative performance. However, the technique was not effectual for instances extraction.

Muhammad Zaigham Zaheer *et al.* [68] formed a completely connected Convolution 3D (C3D) aimed at FE of anomaly videos. Pseudo labels were generated utilizing binary clustering of Spatios-temporal video features for carrying out the self-reasoning-centered training. It aided in lessening the noise presented on the labels of anomalous videos. The chief network as well as the clustering was encouraged to complement one another in attaining the object of more accurate AD. dUCF-crime, ShanghaiTech in addition to UCSD Ped2 were publicly available real-world AD datasets, which were used here. The framework's superiority was shown over the current top-notch methods. However, the processing time was augmented for larger fragments.

Tan Xiao *et al.* [54] commenced a Sparse Semi-nonnegative Matrix Factorizations aimed at FE for detecting anomalies on surveillance video. The probability design was studied at the model level. The spatial together with temporal con-textual information were considered. It was unsupervised and also does not need any human-labeled training data. The framework accurately detected as well as localized anomalies on surveillance video with more expressive features as well as a more complicated design. Numerous benchmark video datasets were taken for AD. This framework's superiority over the top-notch approaches was demonstrated and also its effectiveness was validated. However, the technique was less in computational effectiveness.

Bjorn Barz *et al.* [2] recommended a Principal Components Analysis for extracting features as of anomaly videos. Maximally Divergent Intervals was developed for unsupervised detection of coherent spatial areas in addition to time intervals characterized via higher Kullback-Leibler divergences against prevailing techniques to detect secluded anomalous data points. Experiments were done on synthetic as well as real data as of climate analysis, VS, and also text forensics. The technique was extensively applicable as well as a valuable tool to find interesting events in disparate data types. Nevertheless, when it comes to real data, it might perform less dependably. Entropy varied more extensively over time contrasted with this synthetic benchmark.

Dongyue Chen *et al.* [9] recommended a U-Net-centered FE of surveillance video. A framework was developed centered on the bidirectional prediction for ameliorating the AD performance in surveillance video. The same target frame was predicted via the forward in addition to the backward prediction sub-networks, correspondingly. Next, centered upon the real target frame together with its bi-directional forecast frame, the loss function was built. In addition, centered upon the sliding window that concentrated on the fore-grounds of the prediction error map, an anomaly score estimation technique was presented. The model trounced utmost competing models on disparate VS datasets. However, the optimal setting might not be suitable for all the probable datasets.

Tong Li *et al.* [32] suggested a '2'-streams deep spatial-temporal auto-encoder aimed at VS abnormal events detection. Appearance characteristics were extracted by the spatial stream together with the motion patterns were extracted via the temporal streams. Next, this design fused the spatial stream along with the temporal stream in extracting spatial-temporal characteristics in detecting anomalies centered upon the policy joint reconstruction error. The optical flow was introduced for enhancing the ability to extract continuity betwixt adjoining frames and inter-frame motion information as the optical flows was invariant to appearances, say, light or color. It was not suitable aimed at real-time surveillance.

Yeqi Liu *et al.* [36] introduced a Reference Frame Kanades-Lucas-Tomasis (RF-KLT) aimed at motion FE in VS. Next, a dimension diminution technique of time series was described to establish a feature dataset with clear boundaries betwixt classes. For building the feature classifier, ML algorithms were used. Real-time, robust, in addition to the cost-free AD of aerators, were realized by the expert system in the actual video as well as the augmented video dataset. Nevertheless, a large quantity of training data as well as higher-performance computing power was needed.

Fernando P. dos Santos *et al.* [14] developed CNN aimed at FE of disparate VAD domains. This generalization analysis signified the theoretical approach as well as helpful practice as a passageway to comprehending which datasets permitted the better transfer of knowledge. Better anomaly detectors were attained aimed at video frames. It allowed examination of transfer learning's positive as well as negative aspects. However, a larger training process was required for a larger quantity of data. However, the learning rate was augmented.

JunYi Lim *et al.* [34] developed a deep multiple-level feature pyramid network to ameliorate the FE in surveillance video. In unconstrained surroundings, containing handguns for depiction learning, a dataset was built. A compilation of 250 recorded videos in addition to over 2500 distinct labeled frames was considered for building a dataset. As per the comparative receptive field, the base features were ameliorated by means of concatenating shallow, medium, in addition to deep features as of the backbone. Focal Loss was integrated as its classification loss for enhancing detection of smaller-scale handguns. The recommended model attained 87.42% accuracy.

Peng Wu *et al.* [61] recommended a Fast Sparse Coding Networks (FSCN) for higher-level FE in anomaly videos. For extracting Spatial-Temporal Fusions Features (STFF) on hidden layers, a '2'-stream NN was developed. For building a normal dictionary, an FSCN was utilized with the STFF. The FSCN produced sparse co-efficients inside a forward pass that was simple as well as computationally efficient by means of leveraging the predictor in producing approximate sparse coefficients. FSCN was quicker at the test stage contrasted with conventional sparse coding-centered methods. The method attained the top-notch level. However, the system encompassed less convergence.

Fuqiang Zhou *et al.* [76] analyzed an LSTM aimed at abnormal event detection on surveillance video via hybrid Auto-Encoder (AE). The learned illustration of the LSTM encoder-decoder was learned as of the encoder. It was vital for the decoder. Hybrid AE architecture was explored. It not just extracted better Spatios-temporal context, however, enhanced the extrapolate ability of the equivalent decoder through the shortcut connection. Compared to top-notch AD methods, the hybrid model performed better in qualitative together with quantitative ways on standard datasets. Nevertheless, the system didn't perform with larger dimensional data.

As per our knowledge, this paper is contributing to the following areas:

1. The literature review on anomaly detection in surveillance video methods has been integrated in this survey study.
2. This work examines both local and global feature extraction approaches, as well as all five categories (classification, clustering, regression) of classification and clustering methods. Pattern recognition methods (sequence labelling and parsing) use this technique.
3. Pattern recognition and feature extraction techniques are studied side by side, with a focus on research gaps, future directions, and dataset details.
4. We've tried to include the most recent research in that topic in this publication.

2.2 Real-time processing in video analysis

A variety of realistic anomalies can well be captured by the Surveillance videos. The AD in real-time VA is discussed in Table 1. Numerous methods are involved in real-time VA.

Research in the field of 2D object recognition and pattern recognition is on the rise. There is a thorough examination of 2D object and pattern detection methods in this work. We've attempted to summarise the performance of various object and pattern recognition techniques in various contexts. The majority of the papers included in the survey are supported and/or funded by the government [1, 48]. It demonstrates the work's authenticity. Numerous efforts have been made to extract, classify and cluster data features. A number of authors have focused on specific 2D object recognition issues, such as illumination or noise or scaling. However, there are still a number of unexplored regions in which scholars can devote their attention. In the realm of robotics, CAD/CAM systems, medical data mining, remote sensing, and so on, there is a need for improved algorithms that can increase recognition rates. The storing of feature vectors is another hard issue. The 2D image's feature vectors use up a lot of memory, and that may not be the best approach to store the data, given that memory space is a critical consideration in any computer vision application [6, 20].

In the future, we hope to develop a new feature representation approach that uses less memory than the current one. The amount of memory required to store a computer representation of a 2D image is considerable. Another tough task is reducing data dimensionality, which can help speed up data transfer and minimise memory usage. More than 95% of pattern recognition algorithms are accurate, but sequence labelling and syntactic pattern identification, especially for 2D data, have received less attention. In addition, researchers can focus on increasing computational speed. Because robots and GPU-based technologies rely on object identification, it is imperative that the calculation speed be increased [10, 44].

Table 1 AD in real-time VA

Author	Title	Method	Tool	Data set
Rashmika Nawaratne *et al.* [42]	Spatiotemporal AD using DL for Real-time VS	Incremental Spatio-Temporal Learner (ISTL)	TensorFlow framework	CUHK Avenue
Tao Xiang and Shaogang Gong [63]	Video Behavior Profiling for AD	Dynamic Bayesian Network (DBN).	–	Publicly available dataset
Lucas A. Thomaz *et al.* [55]	AD in Moving-Camera Video Sequences Using Principal Subspace Analysis	Principal Subspace Analysis	–	VDAO database
Rikard Laxhammar and Göran Falkman [24]	Online Learning and Sequential AD in Trajectories	Sequential Hausdorff Nearest-Neighbors Conformal Anomaly Detector (SHNN-CAD)	Macbook Pro 2.66 GHz Intel Core 2 Duo processors	Publicly available trajectory datasets.
Ilker Bozcan *et al.* [3]	GridNet: Image-Agnostic Conditional AD for Indoor Surveillance	GridNet	–	Real-world datasets
Yuan Yuan *et al.* [67]	AD in Traffic Scenes via Spatial-Aware Motion Reconstruction	Bayesian model	MATLAB execution on a machine with Intel i5–3470 3.2GHz CPU as well as 4 GB RAM.	car accident dataset in addition to QMUL Junction dataset
Chunyu Chen *et al.* [8]	Detection of Anomalous Crowd Behavior Based on the Acceleration Feature	Gradient-Based Acceleration	–	UMN dataset in addition to the PETS2009 dataset
Anima Pramanik *et al.* [46]	A real-time VS system for traffic pre-events detection	GMM	–	CamSeq01 and ISLab-PVD
Renzhi Wu *et al.* [62]	Improving VAD performance by mining useful data from unseen video frames	Semi-supervised learning	NVIDIA GeForce GTX 1080 GPUs as well as Tensorflow	PED1
Ramna Maqsood *et al.* [40]	Anomaly recognition from surveillance videos using 3D Convolution Neural Network (CNN)	3-dimensional convolutional networks (3D ConvNets)	Ge-Force GTX 1080ti	UCF Crime dataset

2.3 Machine learning models for anomaly detection

A pathway is paved by Artificial intelligence to make computers think like a human. By means of adding training as well as learning components, ML makes the way more even. This section describes the ML classifiers used in AD.

Sahu *et al.* [49] formed a K-means algorithm intended for effectual crowd AD. Superior accuracy over many DL-centered as well as handcrafted feature-centered approaches was rendered. A lower-power FPGA implementation was presented. The features were extracted over non-over-lapping pixels. This allowed gating inputs to several modules, which bought about high power effectiveness. The maximal energy needed per pixel was 2.43nJ as well as 126.65 M-pixels could be processed per second. Noisy data in addition to outliers were not handled.

Claudio Piciarelli *et al.* [45] generated a single-class Support Vectors Machine (SVM) intended for trajectory-centered anomalous events detection. Particularly, AD was addressed by trajectory analysis, which was an approach with numerous application fields, utmost markedly VS together with traffic monitoring. The approach was centered upon SVM clustering. Aimed at the recognition of anomalous trajectories, the detection SVM capabilities were utilized. In lack of a priori information on the outliers' distribution, specific attention was rendered to trajectory classification. The approach's validity was proved. Nevertheless, it didn't prove that, when classifying unseen patterns, the system encompassed good generalization performances.

Xinfeng Zhang *et al.* [73] analyzed a K-Nearest Neighbor (KNN) aimed at VAD along with localization. A technique was presented to detect anomalies over time as well as space via judging whether the similarities betwixt the testing samples and the retrieved K-NN samples followed the pattern distributions of homogeneous intra-class similarities. It was unsupervised '1'-class learning and did not need any clustering or prior assumption. As the probability was utilized to judge as well as the computation of probability wasn't affected via motion distortions arising as of standpoint distortion that gained a benefit over the prevailing solutions, such a system could adapt to the complete scene. Nevertheless, as this approach required doing specialized training for disparate scenes, the tremendously slower training speed restricted its actual deployment.

Sondos Fadl *et al.* [15] recommended a Gaussian RBF multiple-class-SVM (RBF-MSVM) aimed at surveillance video forgery detection. An inter-frame forgeries detection system was rendered utilizing a 2D-CNN of spatiotemporal information as well as fusion for FE. Frame deletion, insertion, in addition to duplication was included. The system's efficiency was shown to detect every inter-frame forgery. However, because of a lack of motion modeling, it wasn't suitable directly for videos.

Shaoci Xie *et al.* [64] posited Video crowd detection together with abnormal behavior model detection centered upon the ML. Learning disparate data mining as well as endeavoring to enhance detection accuracy was the goal. An IDS-AD model was rendered for the absence of user behavior AD. The illustration of user behavior patterns together with behavior profiles was improved. A similarity task was adopted. Experimentation centered upon UNIX user shells commands data exhibited that the detection design encompassed higher detection's performances. However, estimating the optimal sequence length intended for particular users was not easy.

2.4 Deep learning models in surveillance

The DL to detect real-world anomalies on surveillance videos is discussed in this section. Many approaches were developed intended for video action classification because of the

Table 2 DL classifiers for AD in surveillance video

Author	Techniques	Dataset	Results	disadvantages
Yuanyuan Li et al. [29]	Convolution Long shorts term memory (CLSTM)	Ped2	AUC - 96.5	Not appropriate for detecting non-obvious anomalies.
YUMNA ZAHID et al. [69]	IBaggedFCNet	UCF-Crime dataset	AUC- 92.06	Significantly slower than other systems.
Yanshan Li et al. [28]	Fuzzy theory	Publicly available dataset	Accuracy- 93.4	Increase the false rate for complex situations.
Wenqing Chu et al. [11]	Deep 3-dimensional convolutionals network (C3D)	Avenue, Subway, and UCSD	Frame AUC- 82.1 Pixel AUC- 93.7	Poor knowledge about temporal order information in the training videos.
Tian Wang et al. [59]	Generative NNs	UCSD, Avenue, UMN, in addition to PETS	Accuracy- 98.8	The model parameters oscillate, destabilize in addition to never converge.
Zhaoyan Li et al. [30]	Multi-scale 3D CNN	UCSDSped2		The method was to augment the total learning parameters of the network.
Thittaporn Ganokratanaa et al. [18]	Deep Spatiotemporal Translations Network (DSTN)	UCSD pedestrians, UMN, as well as CUHK Avenue	EER- 21.8 AUC- 83.1 SSIM- 96	Degrade the performance for complex scene detection.
Lingping Dong et al. [13]	CNNs	Publicly available dataset	Accuracy – 95%	Reduce the system accuracy for camera redirecting detection.
Peng Wu et al. [60]	Deep One-Class NN	UCSD	AUC- 70.6 EER- 35.1	Deep OC struggles with smaller objects that happen at the far end of the camera as well as objects that show apparently normal movement.
Cem Direkoglu et al. [12]	CNN	UMN and PETS2009	Accuracy- 98.39	High computational complexity.
Ke Xu et al. [66]	Adaptive Intra-frame Classifications Network (AICN)	UCSD Ped1 datasets	AUC- 95.1% EER-9.4%	Manual segmentation of sub regions.
Weixin Luo et al. [37]	special stacked Recurrent NN (sRNN)	MNIST	AUC- 90.11%	Not robust in complex or crowded scenes.
Ragedhaksha et al. [47]	CNN	COCO train database	Accuracy- 78.9%	Increase the processing time for complex features.
Romany F. Mansour et al. [39]	Quicker RCNN with Deep Reinforcements Learning Model	UCSD anomaly database	Accuracy- 85.30%	Degrade the performance for complex features.

Table 2 (continued)

Author	Techniques	Dataset	Results	disadvantages
Mr.M.Murugesan et al. [41]	Multi-layer perception recurrent NN (MLP-RNN)	Real world dataset	Accuracy- 98.30% Sensitivity - 98.36%	Less efficiency.
Yi Hao et al. [21]	3D CNN	ShanghaiTechs, CUHK Avenue, as well as UCSD Ped2	AUC- 96.9% EER- 8%	Trained with small datasets.
Yunpeng Chang et al. [7]	Deep K-means cluster	UCSD pedestrian dataset	AUC- 96.7%	Optical flow may not be optimal aimed at learning regularity since they aren't designed specifically aimed at AD.
Weixin Luo et al. [38]	Graph convolutional networks	ShanghaiTech Campus and CUHK Avenue	AUC- 87.3%	Not robust to some larger-scale VAD datasets because of sub-optimal features as well as classifiers.
Yaxiang Fan et al. [16]	Fully Convolutional Network (FCN)	UCSD and Avenue	Accuracy- 94.9% EER- 11.3%	Intricate conditions like the changing scenes have difficulties to study the proper appraisal of the distribution parameters due to lack of enormous similar normal events.
Tiwari et al. [56]	Mask RCNN	UCSDped2	Accuracy- 94%	Trained with particular real-time scenarios.

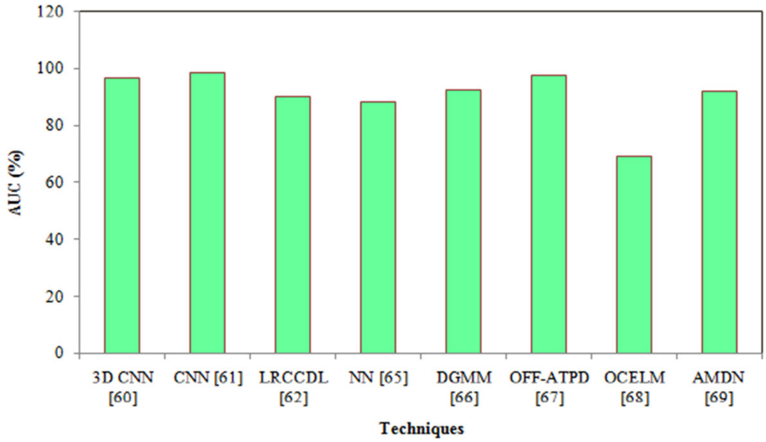

Fig. 2 AUC of DL classifiers

successful demonstration of DL aimed at image classification. The DL classifiers aimed at anemology detection are discussed in Table 2.

For the above-discussed methods, the accuracy analysis is conducted centered on disparate assessment criteria, say Area Under the Curves (AUC) along with Equal Errors rate (EER) is discussed here. The point on the ROC curve is the EER wherein the False Positives Rates (FPR, in which normal behavior is regarded abnormal) is equivalent to the False-Negative Rates (FNR, in which the abnormal behavior is recognized as normal). The measure of the AD's accuracy is the AUC (the ROC). Centered upon some performance metrics, the performance metrics are examined. For comprehending the DL as well as ML classifiers' performance in AD in VS, these comparison figures were utilized.

The AUC of DL classifiers is exhibited in Fig. 2. 3D-CNN [22] renders 96.3% of AUC. CNN [26] encompasses 98.6%, which has higher AUC compared to 3D-CNN [22]. 90.01% as well as 88% of AUC is encompassed by the Low-Rank and Compacts Coefficient Dictionary Learning (LRCCDL) [31] and NN [33]. The Deep Gaussian mixtures model (DGMM) [17]

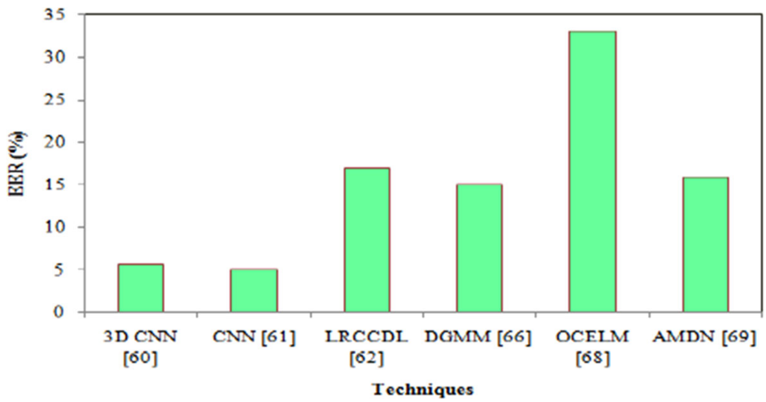

Fig. 3 EER of DL classifiers

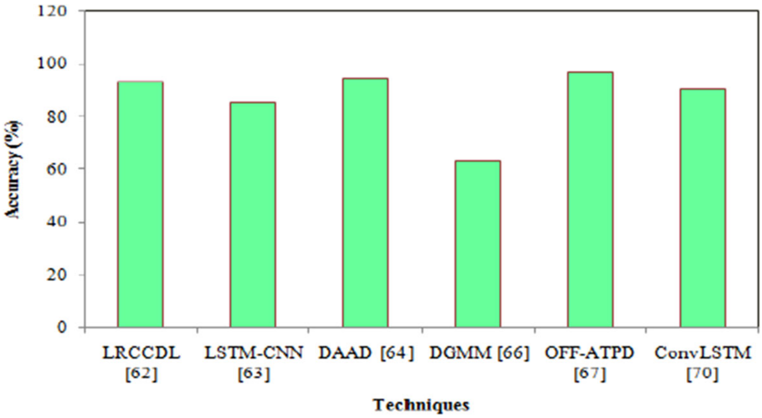

Fig. 4 Accuracy of DL classifiers

renders 92.5% of AUC. Offline Anomalous Traffics Patterns Detection (OFF-ATPD) [74] encompasses 97.25% of AUC that encompasses a higher true positive rate. One-class Extremes Learning Machine (OCELM) [58] as well as Appearance and Motions DeepNet (AMDN) [65] encompass 68.9% and 92.1% of AUC.

The EER of DL methods is depicted in Fig. 3. 3D CNN [22] encompasses a 5.59% of error rate. CNN [26] as well as LRCCDL [31] encompass 5% and 17.05%, which has higher EER. After that, 15.1% of EER is attained by DGMM [17], 33% of EER by OCELM [58], and 16% of EER by AMDN [65].

DL classifiers' accuracy is exhibited in Fig. 4. LRCCDL [31] renders 93.21% of accuracy. 94.67% of accuracy is rendered by LSTM-CNN [57] and Deep Adversarial Anomaly Detections (DAAD) [75]. 63% and 96.93% of accuracy is attained by DGMM [17] and OFF-ATPD [74]. ConvLSTM [5] encompasses 90.7% of accuracy.

ML classifiers' accuracy is exhibited in Fig. 5. 93.75% of accuracy is obtained by the K-means classifier [49]. 94.3% of accuracy is rendered by the SVM [45]. KNN [73] encompasses 88.3% of accuracy. Lastly, RBF-MSVM [15] renders 98.9%, and also it has higher accuracy compared to all other methods.

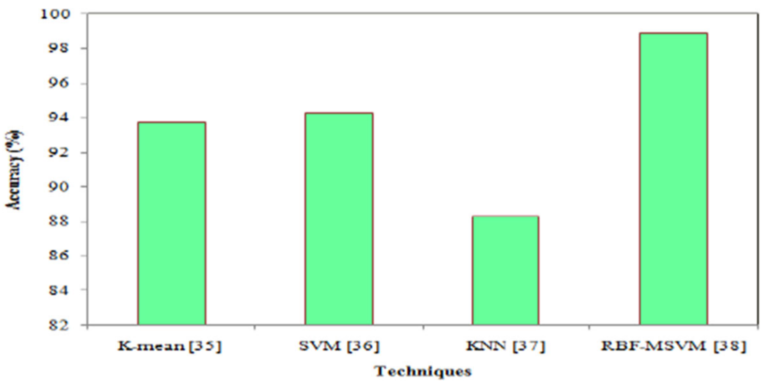

Fig. 5 Accuracy of ML classifiers

Discussion By means of examining the prevailing techniques, the shortcomings are identified. Real-life issues encompass subsequent objectives,

- Bad weather conditions
- Time complexity
- Real-life dynamics
- Occlusions
- Overlapping of objects

The issues were handled by the prevailing separately. Every objective was not handled as features by any methods in a single proposal. The technique ought to be able to render solutions to all these issues for handling effectual intelligent crowd VA in real-time. An effectual economic solution on a time-bounded manner was not produced by the customary methods. The execution of DL-cantered solutions for the quick processing of big data is allowed with the availability of higher-performance computational resources, say GPU. By means of including good features as well as eradicating unwanted features, prevailing DL architectures or models can well be combined. The lack of widespread utilization of automated AD solutions was on account of the absence of techniques that can well be applied to an extensive gamut of targets. This is since every solution encompasses merely a particular use and also may fail when exposed to the disparate targets as well as behaviour common on realistic scenarios. In the individual or crowd surveillance area, this absence of cross-target research is of utmost significance. This mixture of targets is extremely common as a practical scenario as well as extremely hard owing to the difference in FE techniques implemented to each. Additional study is required for stating the techniques' applicability to a broader gamut of surveillance targets in differing environments.

3 Conclusion

A review exploring current research in AD in automated surveillance athwart main aspects of the issue domain, approach, together with technique is presented in this paper. An extensive variety of applications is covered in the reviewed papers. The techniques, tools, as well as dataset identified, were listed in table form. First, the focus is given to VS analysis from a common standpoint, and then, it goes to crowd analysis. Crowd size is large as well as dynamic in real-life scenarios, thus, crowd analysis is hard. It is hard to identify each entity along with its behaviour. It discussed the methods that analyse crowd behaviour. For providing an efficient solution, the issues identified in prevailing techniques were enlisted as future directions. The disparate methods were successfully validated. For addressing issues taking place in the deployment of surveillance systems, some works are needed. Issues, like the ability to generalize in new environments, scaling to wide-region distribution, in addition to robustness to unexpected events will require to be addressed before the techniques can well be implemented in real-life scenarios as the studies in the literature characteristically assess the methods in merely a controlled conditions or single environment.

Funding The authors have no financial or proprietary interests in any material discussed in this article.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Conflicts of interest/competing interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Al Farid F, Hashim N, Abdullah J, Bhuiyan MR, Shahida Mohd Isa WN, Uddin J, Haque MA, Husen MN (2022) A Structured and Methodological Review on Vision-Based Hand Gesture Recognition System. *J Imaging* 8:153. https://doi.org/10.3390/jimaging8060153
2. Barz B, Rodner E, Garcia YG, Denzler J (2018) Detecting regions of maximal divergence for spatio-temporal anomaly detection. IEEE Trans Pattern Anal Mach Intell 41(5):1088–1101
3. Bozcan I, Le Fevre J, Pham HX, Kayacan E (2021) GridNet image-agnostic conditional anomaly detection for indoor surveillance. IEEE Robot Autom Lett 6(2):1638–1645
4. Chandrakar R, Raja R, Miri R, Patra RK, Sinha U (2021) Computer Succored Vaticination of Multi-Object Detection and Histogram Enhancement in Low Vision. Int J of Biometrics. Special Issue: Investigation of Robustness in Image Enhancement and Preprocessing Techniques for Biometrics and Computer Vision Applications 1:1–20
5. Chandrakar R, Raja R, Miri R (2021) Animal detection based on deep convolutional neural networks with genetic segmentation. Multimed Tools Appl. https://doi.org/10.1007/s11042-021-11290-4
6. Chandrakar R, Miri R, Kushwaha A (2022) Enhanced the moving object detection and object tracking for traffic surveillance using RBF-FDLNN and CBF algorithm. Expert Syst Appl 191:116306, ISSN: 0957-4174. https://doi.org/10.1016/j.eswa.2021.116306
7. Chang Y, Zhigang T, Xie W, Luo B, Zhang S, Sui H, Yuan J (2022) Video anomaly detection with spatio-temporal dissociation. Pattern Recogn 122:1–12
8. Chen C, Yu S, Bi X (2015) Detection of anomalous crowd behavior based on the acceleration feature. IEEE Sensors J 15(12):7252–7261
9. Chen D, Wang P, Yue L, Zhang Y, Jia T (2020) Anomaly detection in surveillance video based on bidirectional prediction. Image Vis Comput 98:1–8
10. Choudhary S, Lakhwani K, Kumar S (2022) Three Dimensional Objects Recognition & Pattern Recognition Technique; related challenges: a review. Multimed Tool Appl 23(1):1–44
11. Chu W, Xue H, Yao C, Deng C (2017) Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos. J Latex Class Files 14(8):1–11
12. Direkoglu C (2020) Abnormal crowd behavior detection using motion information images and convolutional neural networks. IEEE Access 8:80409–80416
13. Dong L, Zhang Y, Wen C, Wu H (2016) Camera anomaly detection based on morphological analysis and deep learning. IEEE International Conference on Digital Signal Processing (DSP), 16–18 Oct, Beijing, China
14. dos Santos FP, Ribeiro LSF, Ponti MA (2019) Generalization of feature embeddings transferred from different video anomaly detection domains. J Vis Commun Image Represent. https://doi.org/10.1016/j.jvcir.2019.02.035
15. Fadl S, Han Q, Li Q (2020) CNN spatiotemporal features and fusion for surveillance video forgery detection. Signal Process Image Commun 90:116066. https://doi.org/10.1016/j.image.2020.116066
16. Fan Y, Wen G, Li D, Qiu S, Levine MD, Xiao F (2020) Video anomaly detection and localization via gaussian mixture fully convolutional variational autoencoder. Comput Vis Image Underst 195:1–12
17. Feng Y, Yuan Y, Lu X (2016) Learning deep event models for crowd anomaly detection. Neurocomputing 219:548–556
18. Ganokratanaa T, Aramvith S, Sebe N (2020) Unsupervised anomaly detection and localization based on deep spatiotemporal translation network. IEEE Access 8:50312–50329
19. Gao X, Xu G, Li S, Wu Y, Dancigs E, Juan D (2019) Particle filter-based prediction for anomaly detection in automatic surveillance. IEEE Access 7:107550–107559
20. Gupta D, Varshney N, Kumar A (2022) Detection of behavioral patterns employing a hybrid approach of computational techniques. Comput Mater Contin 72(1):2015–2031
21. Hao Y, Li J, Wang N, Wang X, Gao X (2022) Spatiotemporal consistency enhanced network for video anomaly detection. Pattern Recogn 121:1–11

22. Hu Z-p, Zhang L, Li S-f, Sun D-g (2020) Parallel spatial-temporal convolutional neural networks for anomaly detection and location in crowded scenes. J Vis Commun Image Represent 67:102765. https://doi.org/10.1016/j.jvcir.2020.102765
23. Khaleghi A, Moin MS (2018) Improved anomaly detection in surveillance videos based on a deep learning method. 8th Conference of AI & Robotics and 10th RoboCup Iranopen International Symposium (IRANOPEN), 10–10 April, Qazvin, Iran
24. Laxhammar R, Falkman G (2014) Online learning and sequential anomaly detection in trajectories. IEEE Trans Pattern Anal Mach Intell 36(6):1158–1173
25. Leyva R, Sanchez V, Li C-T (2015) Video anomaly detection with compact feature sets for online performance. Journal of Latex Class Files 14(8):1–16
26. Li N, Chang F (2019) Video anomaly detection and localization via multivariate Gaussian fully convolution adversarial autoencoder. Neurocomputing 369:92–105. https://doi.org/10.1016/j.neucom.2019.08.044
27. Li Q, Li W (2016) A novel framework for anomaly detection in video surveillance using multi-feature extraction. 9th international symposium on computational intelligence and design, 10–11 Dec, Hangzhou, China
28. Li Y, Guo T, Xia R, Xie W (2018) Road traffic anomaly detection based on fuzzy theory. IEEE Access 6: 40281–40288
29. Li Y, Cai Y, Liu J, Lang S, Zhang X (2019) Spatio-temporal unity networking for video anomaly detection. IEEE Access 7:172425–172432
30. Li Z, Li Y, Gao Z (2020) Spatiotemporal representation learning for video anomaly detection. IEEE Access 8:25531–25542
31. Li A, Miao Z, Cen Y, Zhang X-P, Zhang L, Chen S (2020) Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning. Pattern Recogn 108:1–16
32. Li T, Chen X, Zhu F, Zhang Z, Yan H (2021) Two-stream deep spatial-temporal auto-encoder for surveillance video abnormal event detection. Neurocomputing 439:256–270
33. Li B, Leroux S, Simoens P (2021) Decoupled appearance and motion learning for efficient anomaly detection in surveillance video. Comput Vis Image Underst 210(4):1–8
34. Lim JY, Al Jobayer MI, Baskaran VM, Lim JMY, See J, Wong KS (2021) Deep multi-level feature pyramids application for non-canonical firearm detection in video surveillance. Eng Appl Artif Intell 97:1–18
35. Liu C, Wang G, Ning W, Lin X, Li L, Liu Z (2010) Anomaly detection in surveillance video using motion direction statistics. IEEE 17th international conference on image processing, September 26–29, Hong Kong
36. Liu Y, Yu H, Gong C, Chen Y (2020) A real time expert system for anomaly detection of aerators based on computer vision and surveillance cameras. J Vis Commun Image Represent 68:102767. https://doi.org/10.1016/j.jvcir.2020.102767
37. Luo W, Liu W, Lian D, Tang J, Duan L, Peng X, Gao S (2019) Video anomaly detection with sparse coding inspired deep neural networks. IEEE Trans Pattern Anal Mach Intell 43(3):1070–1084
38. Luo W, Liu W, Gao S (2021) Normal graph spatial temporal graph convolutional networks based prediction network for skeleton based video anomaly detection. Neurocomputing 444:332–337
39. Mansour RF, Gutierrez JE, Gamarra M, Villanueva JA, Leal N (2021) Intelligent video anomaly detection and classification using faster RCNN with deep reinforcement learning model. Image Vis Comput 112:1–9
40. Maqsood R, Bajwa UI, Saleem G, Raza RH, Anwar MW (2021) Anomaly recognition from surveillance videos using 3D convolution neural network. Multimed Tools Appl 80:18693–18716
41. Murugesan M, Thilagamani Dr S (2020) Efficient anomaly detection in surveillance videos based on multi layer perception recurrent neural network. Microprocess Microsyst 79:103303. https://doi.org/10.1016/j.micpro.2020.103303
42. Nawaratne R, Alahakoon D, De Silva D, Yu X (2019) Spatiotemporal anomaly detection using deep learning for real-time video surveillance. IEEE Trans Industr Inform 16(1):393–402
43. Ovhal KB, Patange SS, Shinde RS, Tarange VK, Kotkar VA (2017) Analysis of anomaly detection techniques in video surveillance. International conference on intelligent sustainable systems, 7–8 Dec, Palladam, India
44. Pandey S, Miri R, Sinha GR (2022) AFD Filter and E2n2 Classifier for Improving Visualization of Crop Image and Crop Classification in Remote Sensing Image. Int J Remote Sens 43(1). https://doi.org/10.1080/01431161.2021.2000062
45. Piciarelli C, Micheloni C, Foresti GL (2008) Trajectory-based anomalous event detection. IEEE Trans Circuits Syst Video Technol 18(11):1544–1554
46. Pramanik A, Sarkar S, Maiti J (2021) A real-time video surveillance system for traffic pre-events detection. Accid Anal Prev 154:1–21
47. Ragedhaksha, Darshini, Shahil, Nehru A (2021) Deep learning-based real-world object detection and improved anomaly detection for surveillance videos. Mater Today Proc. https://doi.org/10.1016/j.matpr.2021.07.064

48. Reccetti M, Marfia G, Zanichelli M (2010 Article No.: 28) The art and craft of making the Tortellini: playing with a digital gesture recognizer for preparing pasta culinary recipes. Comput Entertain 8(4):1–20. https://doi.org/10.1145/1921141.1921148
49. Sahu AK, Sharma S, Tanveer M, Raja R (2021) Internet of Things attack detection using hybrid Deep Learning Model. Comput Commun 176:146–154, ISSN 0140–3664. https://doi.org/10.1016/j.comcom.2021.05.024
50. SanMiguel JC, Martínez JM, Caro-Campos L (2017) Object-size invariant anomaly detection in videosurveillance. International Carnahan Conference on Security Technology (ICCST), 23–26 Oct, Madrid, Spain
51. Singh H, Hand EM, Alexis K (2020) Anomalous motion detection on highway using deep learning. ArXiv:200608143
52. Sodemann AA, Ross MP, Borghetti BJ (2012) A review of anomaly detection in automated surveillance. IEEE Trans Syst Man Cybern Part C Appl Rev 42(6):1257–1272
53. Sun J, Wang X, Xiong N, Shao J (2016) Learning sparse representation with variational auto-encoder for anomaly detection. IEEE Access 4:1–10
54. Tan X, Zhang C, Zha H (2015) Learning to detect anomalies in surveillance video. IEEE Signal Process Lett 22(9):1477–1481
55. Thomaz LA, Jardim E, da Silva AF, da Silva EAB, Netto SL, Krim H (2017) Anomaly detection in moving-camera video sequences using principal subspace analysis. IEEE Trans Circuits Syst 65(3):1003–1015
56. Tiwari L, Raja R, Awasthi V, Miri R, Sinha GR, Alkinani MH, Polat K (2021) Detection of lung nodule and cancer using novel Mask-3 FCM and TWEDLNN algorithms. Measurement 172:108882, ISSN 0263-2241. https://doi.org/10.1016/j.measurement.2020.108882
57. Ullah A, Ahmad J, Muhammad K, Sajjad M, Baik SW (2017) Action recognition in video sequences using deep bi-directional LSTM with CNN features. IEEE Access 6:1155–1166
58. Wang S, Zhu E, Yin J, Porikli F (2017) Video anomaly detection and localization by local motion based joint video representation and OCELM. Neurocomputing 277:161–175. https://doi.org/10.1016/j.neucom.2016.08.156
59. Wang T, Qiao M, Lin Z, Li C, Snoussi H, Liu Z, Choi C (2018) Generative neural networks for anomaly detection in crowded scenes. J Latex Class Files 14(8):1–11
60. 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
61. Wu P, Liu J, Li M, Sun Y, Shen F (2020) Fast sparse coding networks for anomaly detection in videos. Pattern Recogn 107:107515. https://doi.org/10.1016/j.patcog.2020.107515
62. Wu R, Li S, Chen C, Hao A (2021) Improving video anomaly detection performance by mining useful data from unseen video frames. Neurocomputing 462:523–533
63. Xiang T, Gong S (2008) Video behavior profiling for anomaly detection. IEEE Trans Pattern Anal Mach Intell 30(5):893–908
64. Xie S, Zhang X, Cai J (2018) Video crowd detection and abnormal behavior model detection based on machine learning method. Neural Comput & Applic 31:175–184. https://doi.org/10.1007/s00521-018-3692-x
65. Xu D, Yan Y, Ricci E, Sebe N (2016) Detecting anomalous events in videos by learning deep representations of appearance and motion. Comput Vis Image Underst 156:117–127. https://doi.org/10.1016/j.cviu.2016.10.010
66. Xu K, Sun T, Jiang X (2019) Video anomaly detection and localization based on an adaptive intra-frame classification network. IEEE Trans Multimedia 22(2):394–406
67. Yuan Y, Wang D, Wang Q (2016) Anomaly detection in traffic scenes via spatial-aware motion reconstruction. IEEE Trans Intell Transp Syst 18:1198–1209. https://doi.org/10.1109/TITS.2016.2601655
68. Zaheer MZ, Mahmood A, Shin H, Lee S-I (2020) A self-reasoning framework for anomaly detection using video-level labels. IEEE Signal Process Lett 27:1705–1709
69. Zahid Y, Tahir MA, Durrani NM, Bouridane A (2020) IBaggedFCNet an ensemble framework for anomaly detection in surveillance videos. IEEE Access 8:220620–220630
70. Zaidi S, Jagadeesh B, Sudheesh KV, Audre Arlene A (2017) Video anomaly detection and classification for human activity recognition. International Conference on Current Trends in Computer, Electrical, Electronics and Communication, 8–9 Sept, Mysore, India
71. Zaigham Zaheer M, Lee JH, Lee S-Ik, Seo B-S (2019) A brief survey on contemporary methods for anomaly detection in videos. International Conference on Information and Communication Technology Convergence (ICTC), 16–18 Oct, Jeju, Korea (South)
72. Zang X, Li G, Li Z, Li N, Wang W (2016) An object-aware anomaly detection and localization in surveillance videos. IEEE second international conference on multimedia big data, 20–22 April, Taipei, Taiwan

73. Zhang X, Yang S, Zhang J, Zhang W (2020) Video anomaly detection and localization using motion-field shape description and homogeneity testing. Pattern Recogn 105:1–13
74. Zhang X, Zheng Y, Zhao Z, Liu Y, Blumenstein M, Li J (2021) Deep learning detection of anomalous patterns from bus trajectories for traffic insight analysis. Knowl-Based Syst 217:1–13
75. Zhang X, Jie M, Zhang X, Liu H, Zong L, Li Y (2022) Deep anomaly detection with self-supervised learning and adversarial training. Pattern Recogn 121:1–14
76. Zhou F, Lin W, Li Z, Zuo W, Tan H (2019) Unsupervised learning approach for abnormal event detection in surveillance video by hybrid autoencoder. Neural Process Lett 52:961–975. https://doi.org/10.1007/s11063-019-10113-w