



Anomaly detection using edge computing in video surveillance system: review

Devashree R. Patrikar¹ · Mayur Rajaram Parate¹

Received: 3 August 2021 / Revised: 13 February 2022 / Accepted: 21 February 2022 / Published online: 29 March 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract

The current concept of smart cities influences urban planners and researchers to provide modern, secured and sustainable infrastructure and gives a decent quality of life to its residents. To fulfill this need, video surveillance cameras have been deployed to enhance the safety and well-being of the citizens. Despite technical developments in modern science, abnormal event detection in surveillance video systems is challenging and requires exhaustive human efforts. In this paper, we focus on evolution of anomaly detection followed by survey of various methodologies developed to detect anomalies in intelligent video surveillance. Further, we revisit the surveys on anomaly detection in the last decade. We then present a systematic categorization of methodologies for anomaly detection. As the notion of anomaly depends on context, we identify different objects-of-interest and publicly available datasets in anomaly detection. Since anomaly detection is a time-critical application of computer vision, we explore the anomaly detection using edge devices and approaches explicitly designed for them. The confluence of edge computing and anomaly detection for real-time and intelligent surveillance applications is also explored. Further, we discuss the challenges and opportunities involved in anomaly detection using the edge devices.

Keywords Video surveillance · Anomaly detection · Edge computing · Machine learning

1 Introduction

Computer vision (CV) has evolved as a key technology in the last decade for numerous applications replacing human supervision. CV has the ability to gain a high-level understanding and derive information by processing and analyzing digital images or videos. These systems are also designed to automate various tasks that the human visual system does. There are numerous interdisciplinary fields where CV is used; Automatic Inspection, Modelling Objects, Controlling Processes, Navigation, Video Surveillance, etc.

Video surveillance is a key application of CV which is used in most public and private places for observation and monitoring. Nowadays, intelligent video surveillance systems are used to detect, track and gain a high-level understanding of objects without human supervision. Such intelligent video surveillance systems are used in homes, offices, hospitals,

malls, parking areas depending upon the preference of the user.

There are several computer vision-based studies that primarily discuss on aspects such as scene understanding and analysis [118,148], video analysis [74,129], anomaly/ abnormality detection methods [149], human-object detection and tracking [36], activity recognition [121], recognition of facial expressions [27], urban traffic monitoring [169], human behavior monitoring [96], detection of unusual events in surveillance scenes [82], etc. Out of these different aspects, anomaly detection in video surveillance scenes has been discussed further in our review. Anomalies can be contextual, point, or collective. Contextual anomalies are data instances that are considered anomalous when viewed against a certain context associated with the data instance [56]. Point anomalies are single data instances that are different with respect to others [9]. Finally, collective anomalies are data instances that are considered anomalous when viewed with other data instances, concerning the entire dataset [12]. Examples of an anomaly in video surveillance scenes are shown in Fig. 1; vehicles moving on the footpath, pedestrian walking on the lawn, incorrect parking of vehicle, etc.

✉ Devashree R. Patrikar
devashreepatrikar@gmail.com

Mayur Rajaram Parate
mparate@iiitn.ac.in

¹ Indian Institute of Information Technology, Nagpur, India

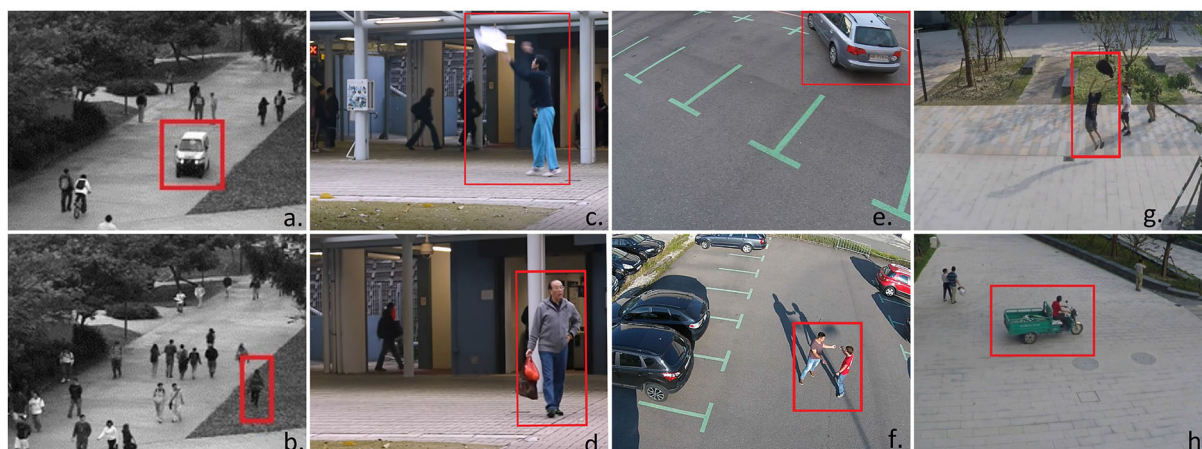


Fig. 1 Anomaly detection in video surveillance scenes. **a** A truck moving on the footpath (UCSD Dataset); **b** Pedestrian walking on the lawn (UCSD Dataset); **c** A person throwing an object (Avenue); **d** a per-

son carrying a suspicious bag (Avenue); **e** Incorrect parking of vehicle (MDVD); **f** people fighting (MDVD); **g** a person catching a bag (ShanghaiTech); **h** vehicles moving on the footpath (ShanghaiTech)

Intelligent video surveillance systems track unusual suspicious behavior and raise alarms without human intervention [96]. The general overview of the anomaly detection is shown in Fig. 2. In this process, visual sensors in the surveillance environment collect the data. This raw visual data is then subjected to pre-processing and feature extraction [31]. The resulting data is provided to a modeling algorithm, in which a learning method models the behavior of surveillance targets and determines whether the behavior is abnormal or not. For anomaly detection, various machine learning tools use cloud computing for data processing and storage [32]. Cloud computing requires large bandwidth and has longer response time because of inevitable network latency [126,155]. Anomaly detection in video surveillance is a delay-sensitive application and requires low latency. Cloud computing in combination with edge computing provides a better solution for real-time intelligent video surveillance [116].

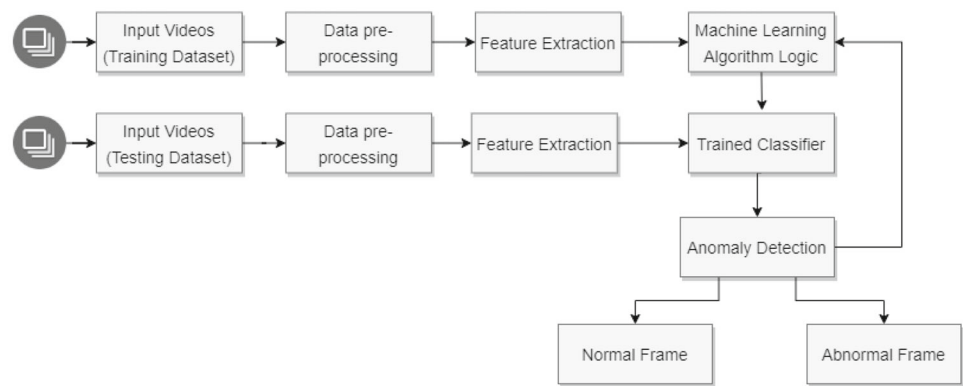
The research efforts in anomaly detection for video surveillance are not only scattered in the learning methods but also approaches. Followed initially, the researchers broadly focused on the use of different handcrafted spatiotemporal features and conventional image processing methods. Recently, more advanced methods like object-level information and machine learning methods for tracking, classification, and clustering have been used to detect anomalies in video scenes. In this survey, we aim to bring together all these methods and approaches to provide a better view of different anomaly detection schemes.

Further, the choice of surveillance target varies according to the application of the system. The reviews done so far have disparity in the surveillance targets. We have categorized the surveillance targets primarily focusing on five types:

automobile, individual, crowd, object, and event. Anomaly detection is a time-sensitive application thus, network latency and operational delays make cloud computing inefficient for delay-sensitive applications such as anomaly detection. Thus, this survey discusses the application of Edge Computing (EC) with cloud computing which enhances the response time for anomaly detection. This survey also presents recent techniques in anomaly detection using edge computing in video surveillance. None of the previous surveys address the convergence of anomaly detection in video surveillance and edge computing. In this study, we provide a review on anomaly detection in video surveillance and also its evolution with edge computing. This review will also address the challenges and opportunities involved in anomaly detection using edge computing.

The research contributions of this review article are as follows:

1. Presented review attempts to connect the disparity in the problem formulations and suggested solutions for the anomaly detection.
2. Review presents a detailed categorization of anomaly detection algorithms based on surveillance targets and methodologies.
3. Discussion of anomaly detection techniques in the context of application area, surveillance targets, learning methods, and modeling techniques.
4. We explore anomaly detection techniques used in vehicle parking, automobile traffic, public places, industrial and home surveillance. The emphasis in these surveillance scenarios is on humans (individual/crowd), objects, automobiles, events, and their interactions.

Fig. 2 General block diagram of anomaly detection

5. The review also focuses on modern-age edge computing technology employed to detect anomalies in video surveillance applications and further discusses the challenges and opportunities involved.

Further, to the best of our knowledge anomaly detection using edge computing paradigm in video surveillance systems is less explored and not surveyed.

We present this survey from the aforementioned perspectives and organize it into eleven sections; Sect. 2 presents the evolution of anomaly detection techniques, Sect. 3 focuses on the prior published surveys, Sect. 4 presents categorization based on different surveillance targets in corresponding application areas. Section 5 explores and categorizes methodologies employed in anomaly detection, and Sect. 6 elaborates about popular datasets and evaluation parameters. Section 7 talks about the adoption of edge computing, its overview, challenges, and opportunities in video surveillance and anomaly detection, Sect. 8 presents empowering anomaly detection with edge devices using machine learning models, Sect. 9 discusses the challenges and future opportunities regarding anomaly on the edge, and lastly, observation followed by the conclusion is discussed in Sects. 10 and 11, respectively.

2 Evolution of anomaly detection techniques

Over a period of time, different researchers used various methods for the purpose of anomaly detection. Handcrafted methods for outlier detection were the most commonly used approach in earlier times. These methods include Histogram of Oriented Gradient (HOG), Histogram of Optical Flow (HOF) [45], Trajectory based [55], etc.

With the advancements in machine learning, spatiotemporal features were automatically learned by machine learning algorithms using neural networks. The issue with neural networks was the problem of vanishing gradient. In 2011, [33] proposed the Deep Sparse Rectifier Neural Network using

ReLU as the activation function to solve the problem of vanishing gradient. This was one of the major breakthroughs in the history of neural networks. With this milestone, convolution neural networks (CNNs) were designed to learn spatiotemporal features automatically. Since 2012, CNNs have become the primary choice for many image-processing problems and are extensively used for anomaly detection [43]. This neural network computational model uses a variation of multilayer perceptrons and contains one or more convolutional layers that can be either entirely connected or pooled. These convolutional layers create feature maps that record a region of the image which is ultimately broken into rectangles and sent out for nonlinear processing [117].

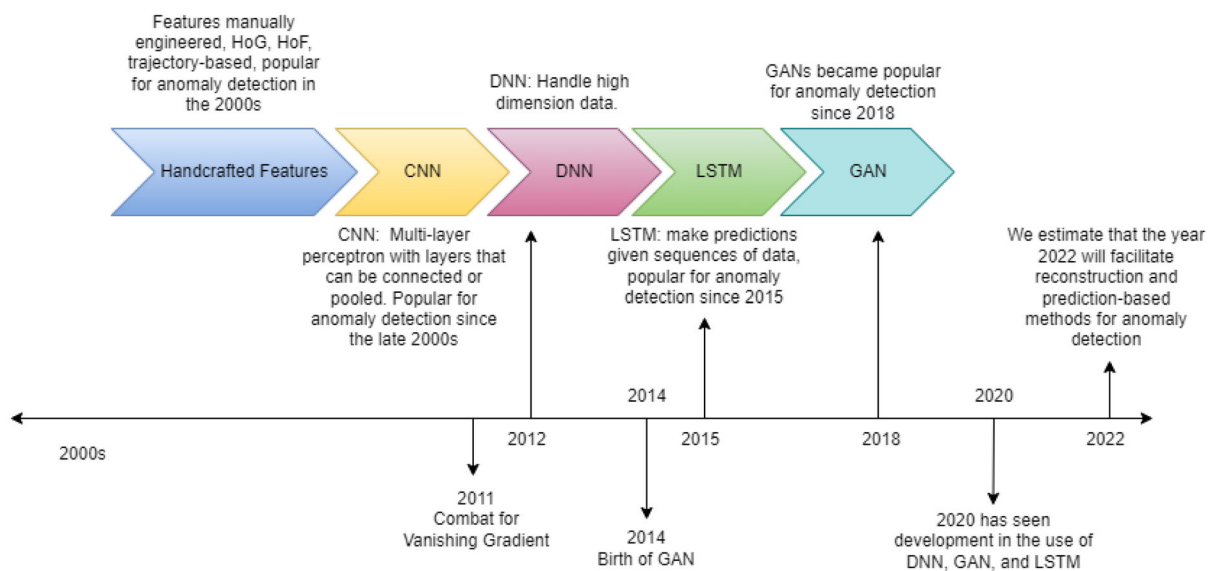
To extract more features and increase the accuracy of prediction, Deep CNN networks are employed which have multiple hidden layers [37]. The deep learning approaches typically fall within a family of encoder–decoder models: an encoder that learns to generate an internal representation of the input data, and a decoder that attempts to reconstruct the original input based on this internal representation. While the exact techniques for encoding and decoding vary across models, the overall benefit they offer is the ability to learn the distribution of normal input data and construct a measure of anomaly.

Table 1 shows the Evolution of Anomaly Detection techniques over recent years. Figure 3 shows the Timeline for Anomaly Detection Techniques.

As far as sequential data is concerned, the results of the CNN network are not the same. Temporal information or data that comes in sequences is well processed by a Recurrent Neural Network (RNN) [164]. When we compare CNNs with RNNs, CNNs are faster as they are designed to handle images, while RNNs are designed to handle text and audio. While RNNs can be trained to handle images, it is difficult for them to separate contrasting features that are closer together [141]. With the fall-outs of RNN, long short-term memory (LSTM) is implemented for anomaly detection [138]. LSTM networks are a special type of RNN that includes a memory cell that can maintain information in memory for a long

Table 1 Evolution of anomaly detection techniques over the recent years

Year	Handcrafted methods	CNN	DNN	LSTM	GAN
2021		[18,53,102,112]	[10,37,52,152]	[4,10,106,138,139]	[111,113,156,165] [98]
2020		[13,17,19,128] [67]	[31,50,109,129,167]	[13,105]	[25,30,44,54]
2019		[82,83,89,135]	[75,127,148]	[28,57,84,170]	[125,147]
2018	[153]	[69,87,88,90] [35,48]	[20,65,81,108,163] [122]		[73]
2017			[39,107,117]	[136]	
2016	[21,22,40]				
2015	[12,45]	[32]			
2014	[158]				

**Fig. 3** Timeline for evolution of anomaly detection techniques

period. LSTM is used to process and make predictions given sequences of data and is a very useful tool in anomaly detection [170].

In recent years, DNNs have attained great success in handling high-dimensional data, especially images. However, generating realistic images containing enormous features for different tasks like image detection, classification and reconstruction continue to be difficult tasks. In 2014, Ian Goodfellow et al. designed a model using the concept of generative modeling which has the potential to learn any kind of data distribution in an unsupervised manner [34]. Generative adversarial network (GAN) is composed of two sub-networks: the generator and discriminator where the generator generates authentic images [165]. The generated and real images are shuffled and given to the discriminator. The network losses are then determined from the prediction of the discriminator. The two sub-networks essentially fight with

each other through what is called adversarial training and hence these networks are called generative adversarial networks [156]. Table 1 shows that GANs have become popular in the research community since 2018 and their implementation for anomaly detection has achieved optimum results [111].

Handcrafted methods were extensively used in the late 2000s. With the advancements in neural networks and the combat for vanishing gradient, CNN becomes popular within the research community and still dominates the field of anomaly detection. Later deep convolutional networks showed optimum results for data with higher parameters. Currently, LSTMs and GANs are largely used for abnormal event detection.

Another technique studied in the literature is the One-Class Classification (OCC), which tries to identify objects of a specific class [90]. SVM-based one-class classification

(OCC) is a popular formulation that identifies the smallest hyper-sphere consisting of all the data points [94]. As the information regarding negative class is unavailable, OCC focuses on maximizing the boundary with respect to the origin [148]. Predominantly used for outlier detection, anomaly detection, and novelty detection, OCC can be further categorized as: boundary methods [18,19] which define a specific margin, generative approaches [37] which mainly use generative adversarial networks, and Discriminative approaches [90] which depend on the loss function.

3 Recent surveys for anomalies in different context

In recent years, remarkable surveys have been done in the respective fields of video surveillance and anomaly detection; however, the convergence of both the fields is less reviewed. The context of an anomaly in this review can broadly be categorized into anomalies in road traffic and human or crowd behavior [1,63,68,96,124,159]. The advances in vehicle detection using monocular, stereo vision, and active sensor–vision fusion for the on-road vehicle are surveyed in [123]. Automobile detection and tracking are surveyed in [133]. The performance dependency of a vehicle surveillance system on traffic conditions is also discussed and a general architecture for the hierarchical and networked vehicle surveillance is presented. The techniques for recognizing vehicles based on attributes such as color, logos, license plates are discussed in [119]. The anomaly detection methodologies in road traffic are surveyed in [56]. As the anomaly detection schemes cannot be applied universally across all traffic scenarios, the paper categorizes the methods according to features, object representation, approaches, and models.

Unlike anomaly detection in-vehicle surveillance, anomalies in human or crowd behavior are much more complex. Approaches to understanding human behavior are surveyed in [63,96,124] based on human tracking, human–computer interactions, activity tracking, and rehabilitation. The recognition of complex human behavior and various anomaly detection techniques are discussed in [68]. Further, the use of moving object trajectory-clustering [159], and trajectory-based surveillance [1] to detect abnormal events are observed in the literature. In [124], the learning methods and classification algorithms are discussed considering crowd and individuals as separate surveillance targets to detect the anomaly. However, the occlusions and visual disparity in the crowded scenes reduce the accuracy in detecting the anomalies. [63] focuses on the aforementioned aspects and learns the motion pattern of the crowd to detect abnormal behavior. [135] showcases various techniques based on convolutional neural networks (CNNs) for anomaly detection in crowd behavior, whereas [99] opts for the analysis of a single scene in videos. With the recent advancements in deep learning (DL), [101] presents a real-time analysis of crowd anomaly detection in video surveillance and [85] provides a review of the techniques used in deep learning for anomaly detection. The recent surveys on anomaly detection and automated video surveillance are listed in Table 2.

4 Categorization of anomalies according to surveillance targets

Surveillance targets are those entities on which the anomaly detection method aims to detect anomalies. In the context of anomaly, the surveillance targets can be categorized as: the individual, crowd, automobile traffic, object and event,

Table 2 Recent surveys for anomalies in different context

Year	Existing work	Broad topics
2021	Khosro Rezaee et al. [101]	DL-based real-time crowd anomaly detection
2021	Rashmiranjan Nayak et al. [85]	DL-based methods for video anomaly detection
2020	Bharathkumar Ramachandra et al. [99]	A survey of single-scene video anomaly detection
2019	Gaurav Tripathi et al. [135]	CNN for crowd behavior
2019	Ahmed et al. [1]	Trajectory-based surveillance
2018	Shobha et al. [119]	Vehicle detection, Recognition and tracking
2016	Yuan et al. [159]	Moving object trajectory clustering
2016	Xiaoli Li et al. [68]	Anomaly detection techniques
2015	Li et al. [63]	Crowded scene analysis
2014	Tian et al. [133]	Vehicle surveillance
2013	Sivaraman et al. [123]	Vehicle detection, Tracking,
2012	Popoola et al. [96]	Abnormal human behavior recognition
2012	Angela A. Sodemann [124]	Human behavior detection

the interaction between humans and objects, etc. The correlation of surveillance, surveillance targets, and associated anomalies is illustrated in Fig. 4.

4.1 Individual

Anomaly detection is deployed to ensure the safety of individuals in public places like hospitals, offices, public places, or at home. It recognizes patterns of human behavior based on sequential actions and detects abnormalities [60]. Visual rhythms are extracted from videos to generate a Histogram of Oriented Gradients (HOG) and train a classifier to detect abnormal actions like running, walking, hopping, jumping, waving a hand [134]. Several approaches have been proposed to detect anomalies in behavior involving breach of security [97], running [170], lawbreaking actions like robbery [59], and fall of elderly people [72].

4.2 Crowd

This review distinguishes between individuals and crowd as shown in Fig. 4. Both of these targets consist of a single entity, the methods used to identify abnormalities are distinct for individuals and crowd [62,135]. Any change in motion vector or density or kinetic energy indicates an anomalous crowd motion [107,108]. In [158], this change in motion is captured by Structural Context Descriptor, [21] incorporates a feature descriptor based on optical flow information, whereas [45] uses Histogram of Swarms (HoS) where both motion and appearance information is captured by the descriptor. In [68],

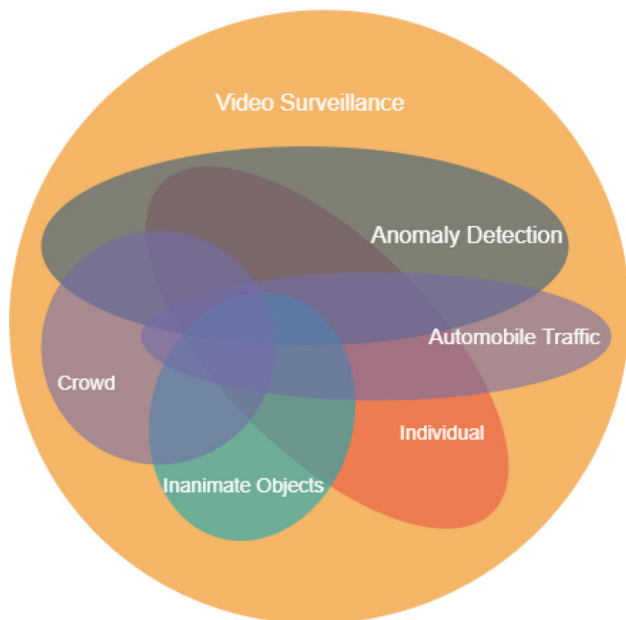


Fig. 4 Correlation of surveillance, surveillance targets, and associated anomalies

behavior such as people suddenly running in different directions or the same direction is considered anomalous. A crowd cannot only be a crowd of individuals but a fleet of taxis as well; [7] allows the scene understanding and monitoring on a fleet of taxis.

4.3 Automobiles and traffic

The automobile and traffic surveillance intends to monitor and understand automobile traffic, traffic density, traffic law violations, and safety issues like; accident or parking occupancy. In smart cities, automobiles become important surveillance targets and extensively surveyed for traffic monitoring, lane congestion, and behavior understanding [7,11,47,48,56,144,169]. In metro cities, [87] allows drivers to find a vacant parking area. Moreover, for better accessibility, security, and comfort of the citizens, studies also focus on traffic law violations which include vehicles parked in an incorrect place [49], predicting anomalous driving behavior, abnormal license plate detection [69], detection of road accidents [122] and detection of collision-prone behavior of vehicles [105].

4.4 Inanimate objects and events

The target in this category is divided into events and inanimate objects. Some of the examples of abnormal events are; an outbreak of fire, which is a common calamity in industries [82] and needs automatic detection and quick response. Similarly, it is challenging to detect smoke in the foggy environment; [83] presents smoke detection in a foggy environment which plays a key role in disaster management. Further, there are defects in the manufacturing system that are tedious for humans to examine considering them as anomalies; [65] proposes a scheme for detecting manufacturing defects in industries.

4.5 Interaction between humans and objects

In this category, anomaly detection schemes are associated with the interaction between humans and objects. Both individuals and objects together give the potential benefits of detecting interaction between them such as an individual carrying a suspicious baggage [84], individual throwing a chair [3]. Some studies attempt to account for both pedestrians and vehicles in the same scene such as cyclists driving on a footpath, pedestrians walking on the road [84,120,147,153]. In [107], abnormal behavior is identified by objects like a skateboarder, a vehicle, or a wheelchair on the footpath.

5 Anomaly detection methodologies in video surveillance

To detect anomalies in automated surveillance, advanced detection schemes have been developed over a decade [9,12]. In this survey, we categorize them broadly into learning-based and modeling-based approaches and further sub-categorize for a clear understanding as shown in Fig. 5.

5.1 Learning

Any event or behavior that deviates from the normal is called an anomaly. The learning algorithms learn anomalies or normal situations based on the labeled and unlabeled training data. Depending upon the data and approach, the learning methods for anomaly detection can be classified broadly as

1. Supervised Learning,
2. Unsupervised Learning and
3. Semi-supervised Learning.

The anomaly detection classification based on learning is shown in Table 3.

5.1.1 Supervised learning

Supervised learning uses labeled dataset to train the algorithm to predict or classify the outcomes. It gives a categorical output or probabilistic output for the different categories. The training data is processed to form different class formulations; single class, two-class or multi-class. When the training data contains data samples either of normal situations or anomalous situations only, it is called single class formulation [124]. In a single class approach, if the detector is trained on normal events, then the events that fall outside the learned class are classified as anomalous [4]. Various approaches to classify and model anomalies with such training data use a 3D convolutional neural network model [43], stacked sparse coding (SSC) [149], adaptive iterative hard-thresholding algorithm [170].

Apart from single and two-class formulation, an approach where multiple classes of events are learned is called multi-class formulation. In this approach before anomaly detection, certain rules are defined regarding behavior classification. Anomaly detection is then performed using these set of rules [59,110]. However, this approach has a drawback that the events that are learned can only be reliably recognized and the events that do not span the learned domain are incorrectly

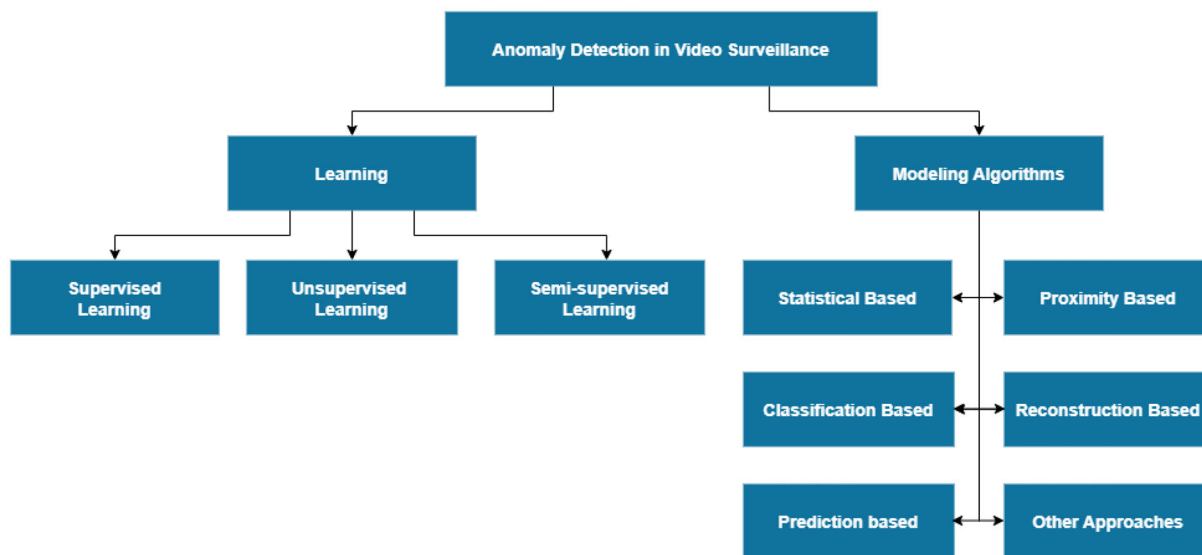


Fig. 5 Anomaly classification

Table 3 Anomaly detection classification

Type of learning	Reference
Supervised	[4,43,59,110,149,170]
Unsupervised	[17–20,22,28,37,55,58,84,111,113,125,128,148,165] [5,95,100,104]
Semi-supervised	[10,25,70,112,135,147,168]

classified. Thus, the multi-class approach may not provide optimum results outside a scripted environment.

5.1.2 Unsupervised learning

In unsupervised learning, given a set of unlabeled data, we discover patterns in data by cohesive grouping, association, or frequent occurrence in the data. In this approach, we consider both normal and abnormal training data samples do not have any label. An algorithm discovers patterns and groups them together with an assumption that the training data consist of mostly normal events and occurs frequently, while rare events are termed as anomalous [124,148]. Deep CNNs are widely used for anomaly detection in video surveillance applications to capture the local spatiotemporal patterns [17,20]. Because of the non-deterministic nature of anomalous events and insufficient training data, it is challenging to automatically detect anomalies in surveillance videos. To address these issues, [125] presents an adversarial attention-based auto-encoder network to detect anomalies. Such generative adversarial networks (GANs) aim to learn the spatiotemporal patterns and train the auto-encoder by using the de-noising reconstruction error and adversarial learning strategy to detect anomalies without supervision [22,111,113,128,165].

To distinguish between new anomalies and normality that evolve, a prediction-based approach of long short-term memory (LSTM) is adopted which is principally used for sequence data [28]. [84] makes use of incremental spatiotemporal learner (ISTL) to remain updated about the changing nature of anomalies by utilizing active learning with fuzzy aggregation. For action recognition in surveillance scenes, [104] proposes a Gaussian mixture model called universal attribute modelling (UAM) using an unsupervised learning approach. The UAM is also been used for facial expression recognition where it captures the attributes of all expressions [95].

One-class classifiers [18,19] have evolved as state of the art for anomaly detection in the research community in recent years. One-class classification (OCC) is a technique where the model is trained using only one class that is positive class [37,148]. [90] proposes a feature extractor network and a CNN-based OCC to detect anomalies using binary cross-entropy as the loss function.

Along with the methods mentioned prior, clustering-based [58], histogram of magnitude and momentum (HoMM) [5], trajectory-based [55], and support vector data description (SVDD) [100] methods are widely used for anomaly detection.

5.1.3 Semi-supervised learning

Semi-supervised learning falls between supervised learning and unsupervised learning. It combines a small amount of

labeled data with a large amount of unlabeled data during training. Semi-supervised learning is used where less variety of labeled training datasets is available. In some applications, GANs are used, where different combinations of labeled and unlabeled data are used in the process of training to obtain misclassified events [168]. [25] uses Dual discriminator GAN, while [147] uses SaliencyGAN to distinguish real samples from the abnormal ones. In some applications, Laplacian support vector machine (LapSVM) utilizes unlabeled samples to learn a more accurate classifier [70]. It is observed that there is a considerable improvement in learning accuracy when unlabeled data is used in conjunction with a small number of labeled data [10,112,135].

5.2 Modeling algorithms for anomaly detection

In the last 10 years, the modeling algorithms for anomaly detection have undergone a drastic transition from traditional handcrafted methods to deep neural networks (DNNs). The detailed categorization of anomaly detection techniques is discussed below, and the technique and highlights of each method are listed in Table 4.

5.2.1 Statistical based

In a statistical-based approach, the parameters of the model are learned to estimate anomalous activity. The aim is to model the distribution of normal-activity data. The expected outcome under the probabilistic model will have a higher likelihood for normal activities and a lower likelihood for abnormal activities [6]. Statistical approaches can further be classified as parametric and nonparametric methods. Parametric methods assume that the normal-activity data can be represented by some kind of probability density function [56]. Some methods use Gaussian mixture model (GMM) which works only if the data satisfies the probabilistic assumptions implicated by the model [12]. The nonparametric statistical model is determined dynamically from the data. Examples of nonparametric models are histogram of gradients (HoG)-based [45] models, histogram of magnitude and momentum (HoMM)-based models [5], Bayesian models [42,86]. Recently, efficient way to detect and localize anomalies in surveillance videos is to use fully convolutional networks (FCNs) [108], spatiotemporal features [67] and structural context descriptor (SCD) [158].

5.2.2 Proximity based

When the video frame is sparsely crowded it is easier to detect anomalies, but it becomes a tedious job to find irregularities in a densely crowded frame. The proximity-based technique utilizes the distance between the object and its surroundings to detect anomalies. In [21], a distance-based approach is

Table 4 Categorization of anomaly detection techniques

Approach	Ref	Technique	Highlights	
Statistical based	[12]	Gaussian process regression	Hierarchical feature representation, Gaussian process regression. Spatiotemporal interest points (STIP) is used to detect local and global video anomaly detection; Dataset: Subway, UCSD, Behave, QMUL Junction; Parameters: AUC, EER (AUC: area under curve), EER: equal error rate)	
	[45]	Histogram-based model	HOG, Histograms of oriented swarms (HOS) KLT interest point tracking are used to detect anomalous event in crowd; Aims at achieving high accuracy and low computational cost; Dataset used: UCSD, UMN; Parameters: ROI, EER, DR; (ROI: region of interest, DR: detection rate)	
	[5]	Histogram based	Crowd anomaly detection using histogram of magnitude and momentum (HoMM); background removal, feature extraction (optical flow), anomaly detection (using K-means clustering); Dataset: UCSD, UMN; Parameters: AUC, EER, RD (rate reduction) true positive rate (TPR), false positive rate (FPR)	
	[86]	Bayesian model	Bayesian nonparametric (BNP) approach, hidden Markov model (HMM) and Bayesian nonparametric factor analysis is employed for data segmentation and pattern discovery of abnormal events. Dataset: MIT parameters: energy distribution	
	[42]	Bayesian model	Bayesian nonparametric dynamic topic model is used. Hierarchical Dirichlet process (HDP) is used to detect anomaly. Dataset: QMUL-junction; Parameters: ROC, AUC (ROC: receiver operating characteristics)	
	[108]	Gaussian classifier	Fully convolutional neural network, Gaussian classifier is used. It extracts distinctive features of video regions to detect anomaly. Dataset: UCSD, Subway; Parameters: ROC, EER, AUC; Aims: to increase the speed and accuracy.	
	[67]	Gaussian model	3D CNN model used to extract spatiotemporal characteristics, Scene background modelled by Gaussian model Dataset: Subway, UCSD ped2, Parameters: AUC, EER	
	[158]	Histogram-based model	Structural context descriptor (SCD) selective histogram of optical flow (SHOF), 3D discrete cosine transform (DCT) object tracker, Spatiotemporal analysis for abnormal detection is crowd using Energy Function Dataset: UMN and UCSD; Parameters: ROC, AUC, TPR, FPR	
	Proximity based	[21]	Histogram	Histogram of optical flow and motion entropy (HOFME) is used to detect the pixel level features diverse anomalous events in crowd anomaly scenarios as compared with conventional features. Nearest neighbor threshold is used by HOFME. Dataset: UCSD, Subway, Badminton; Parameters: AUC, EER
		[71]	Accumulated relative density (ARD)	ARD method is used for large-scale traffic data and detect outliers; Dataset: self-deployed parameter: detection success rate (DSR)
[41]		Density based	A weighted neighborhood density estimation is used to detect anomalies. Hierarchical context-based local kernel regression. Dataset: KDD, Shuttle, Parameters: precision, recall	
Classification based	[58]	Density based	Trajectory-based DBSCAN clustering; Foreground segmentation by using active contouring; Dataset: UCF Web, Collective Motion, Violent Flows Parameters: MAE (mean absolute error) and F-score	
	[58]	Trajectory extraction flow analysis	Foreground segmentation by using active contouring density-based DBSCAN clustering. Dataset: UCF Web, collective motion, and violent flows parameters: MAE (mean absolute error) and F-score	
	[80]	Adaptive sparse representation	Trajectory-based video anomaly detection using joint sparsity model. Dataset: CAVIAR. Parameters: ROC	
	[143]	One-class classification	OCC-based anomaly detection techniques using SVM. Pixel level features are extracted. Dataset: PETS2009 and UMN; Parameters: ROC	
	[88]	Harr-Cascade, HOG feature extraction	Smart Surveillance as an Edge Network using Harr-Cascade, SVM, lightweight-CNN. Provides fast object detection and tracking. Dataset: VOC07, ImageNet; Parameters: accuracy	
	[153]	Stacked Sparse Coding	Intraframe classification strategy; Dataset: UCSD, Avenue, Subway; Parameters: EER, AUC, accuracy	

Table 4 continued

Approach	Ref	Technique	Highlights
	[117]	Sequential deep trajectory descriptor	Dense trajectories are projected into 2D plane, long term motion CNN-RNN network is employed. Dataset: KTH, HMDB51, UCF101; Parameters: accuracy
	[32]	Autoencoders, data reduction	Feature learning with deep learning, autoencoder is placed on the edge, decoder part is placed on the cloud Dataset: HAR, MHEALTH; Parameters: accuracy
	[47]	Kanade–Lucas–Tomasi (KLT) tracker	k-means clustering, connected graphs and traffic flow theory are used estimate real-time parameters from aerial videos; Dataset: Self-Deployed; Parameters: speed, density, volume of vehicles
	[53]	Deep neural network	Face mask identification; video restoration, face detection using DNN; edge computing-based dataset: bus drive monitoring and public dataset; Parameters: accuracy
	[112]	3D convolutional neural network (3D CNN)	Spatiotemporal features are extracted using temporal 3D CNN; Dataset:UCF-Crime dataset; Parameters: AUC;
Reconstruction	[61]	Hyperspectral image (HSI) Analysis	Discriminative reconstruction method for HSI anomaly detection with spectral learning (SLDR). Loss function of SLDR model is generated. Dataset: ABU, San Diego Parameters: ROC, AUC
	[150]	Hyperspectral imagery (HSI)	Low-rank sparse matrix decomposition (LRaSMD)-based dictionary reconstruction is used for anomaly detection. Dataset: STONE, AVIRIS Parameters: ROC
	[20]	3D convolutional network (C3D)	C3D network is used to perform feature extraction and detect anomaly using sparse coding, DL. Dataset: Avenue, Subway, UCSD ; Parameters: AUC, EER
Reconstruction based	[148]	Deep OC neural network	One stage model is used to learn compact features and train a DeepOC (Deep One Class)classifier. Dataset: UCSD, Avenue, Live Video; Parameters: ROC
	[44]	Generative adversarial network (GAN)	Anomaly detection using generative adversarial network for hyperspectral images (HADGAN). Dataset: ABU, San Diego, HYDICE; Parameters: ROC, computing time
	[125]	Adversarial attention based, auto-encoder GAN	Normal patterns are learnt through adversarial attention-based auto-encoder and anomaly is detected. Dataset: ShanghaiTech, Avenue, UCSD, Subway; Parameters: AUC EER
	[128]	Adversarial 3D Conv,Autoencoder	Spatiotemporal patterns are learnt using adversarial 3D Conv, Autoencoder to detect abnormal events in videos; Dataset: Subway, UCSD, Avenue, ShanghaiTech; Parameters: AUC/EER
	[157]	Sparse reconstruction	Sparsity-based reconstruction method is used with low rank property to determine abnormal events. Datasets: UCSD, Avenue; Parameters: ROC, AUC, EER
	[162]	Sparsity-based method	Abnormal event detection in traffic surveillance using low-rank sparse representation (CLSR). Dataset: UCSD, Subway, Avenue; Parameters: AUC, EER
	[111]	GAN	Singular value decomposition (SVD) loss function for reconstruction; Dataset: UCSD, ShanghaiTech, Avenue Parameters: AUC, EER
	[165]	GAN	Auto-encoder/generator = dense residual networks + self-attention; Discriminator= introduces self attention Dataset: CUHK Avenue, Shanghai Tech Parameters: AUC
	[113]	GAN	Proposed GAN: Multi-scale U-Net unsupervised anomaly detection Dataset: UCSD, CHUK Avenue, ShanghaiTech Parameters: AUC
	[30]	GAN	Deep spatiotemporal translation network based on GAN; edge wrapping: to reduce the noise Dataset: UCSD, UMN, CUHK Avenue
	[25]	Dual discriminator-based GAN	Semi-supervised; dual discriminator-based GAN to generate and distinguish motion data; Dataset: CUHK, UCSD, Shanghai; Parameters: AUC, EER
	[156]	GAN	Bidirectional retrospective GAN with frame prediction model; 3D CNN to capture temporal relations between frames; Dataset: Avenue, CUHK, UCSD, Shanghai; Parameters: AUC

Table 4 continued

Approach	Ref	Technique	Highlights
Prediction Based	[98]	GAN	Combination of prediction based and reconstruction based methods to construct a U-shaped GAN. Dataset: UCSD, CUHK Avenue, ShanghaiTech; Parameters: AUC
	[73]	Video prediction framework	Spatial/motion constraints are used for future frame prediction for normal events and identifies abnormal events; Dataset: CUHK, UCSD, ShanghaiTech; Parameters: AUC
	[84]	Incremental spatiotemporal learner	ISTL, an unsupervised deep learning approach with fuzzy aggregation is used to distinguish between anomalies that evolve over time in assistance with spatiotemporal autoencoder, ConvLSTM. Dataset: CUHK, Avenue, UCSD; Parameters: AUC, EER
	[57]	LSTM, Cross Entropy	Recognizing industrial equipment in manufacturing system using edge computing; Big Data, smart meter Dataset: Self deployed ; Parameter: Accuracy
	[13]	Optical flow, GMM, HoF	Detection accuracy, less computational time Dataset: UMN, UCSD, Subway, LV; Parameter: AUC, EER;
	[170]	Spatiotemporal feature extraction	CNN, Adaptive ISTA and LSTM Dataset: CUHK Avenue, UCSD, UMN ; Parameter:AUC/EER
	[102]	Predictive convolutional attentive block	Integration of reconstruction and predictive approach to predict masked information Dataset: MVTEC AD, Avenue, ShanghaiTech Parameter:AUC
	[168]	Temporal attention network	Event score prediction for each video; TANet to learn features and temporal values Dataset: UCSD; Parameters: accuracy, specificity, sensitivity
	[28]	Bi-directional LSTM (BD-LSTM)	Frame level prediction using BD-LSTM and CNN Dataset: UCSD; Parameters: accuracy, AUC;
	[136]	Deep BD-LSTM	Action recognition using Deep BD-LSTM; Redundancy reduction; Capable of analyzing long videos Dataset: UCF-101, YouTube 11 Actions, HMDB51; Parameters: accuracy;
Other Approaches	[138]	LSTM	Feature extraction using lightweight CNN; Residual attention-based LSTM for abnormal event detection; Dataset: UCF-Crime, UMN, Avenue; Parameters: accuracy;
	[10]	BD-LSTM	Deep learning for segmentation; BD-LSTM to analyze anomalies in UAV real-world aerial datasets; Dataset: real-world aerial dataset; Parameters: mean precision, mean, recall, F1 mean
	[64]	Fuzzy theory	Anomaly detection in road traffic using fuzzy theory. The Gaussian distribution model is trained. Dataset: SNA2014-Nomal; Parameters: accuracy, false detection rate
	[80]	Sparse reconstruction	Joint sparsity model for abnormality detection, multi-object anomaly detection for real world scenarios. Dataset: CAVIAR Parameters:ROC
	[66]	Sparsity based	Video surveillance of traffic. Background subtraction. Non-convex optimization; generalized shrinkage thresholding operator (GSTO), Joint estimation Dataset: I2R, CDnet2014 Parameters: F-measure
	[36]	Frequency domain (power spectral density; correlation)	Detection of vehicle anomaly using edge computing high-frequency correlation, sensors, reduces computation overhead, privacy; Dataset: Open Source Platform; Parameters: FPR, TPR, ROC
	[132]	Prediction model	Particle filtering used to detect anomaly in frames; of particle filtering less processing time; Dataset: UCSD, LIVE Parameters: EER
	[142]	Magnitude of motion detection (MoMD)	MoMD based on spatiotemporal interest points (STIP); segmentation of video into key frames; elimination of redundant frames; CNN; YOLO; Dataset: UA-DETRAC, crossroad in Beijing city; Parameters: precision, recall, F-measure

used that assumes normal data has dense neighborhoods and anomalies are identified by their proximity to their neighbors. Further, density-based approaches identify distinctive groups or clusters depending upon the density and sparsity, to detect the anomaly [41,58,71].

5.2.3 Classification based

Another commonly used method of anomaly detection is a classification based which aims to distinguish between events by determining the margin of separation [109]. In [80], support vector machine (SVM) uses a classic kernel to learn a feature space to detect the anomaly. Further, a nonlinear one-class SVM is trained with histogram of optical flow (HOF) orientation to encode the moving information of each video frame [143]. Aiming at intelligent human object surveillance scheme, Harr-cascade and HOG+SVM are applied together to enable a real-time human-object identification [88]. Some approaches utilize object trajectories estimated by object tracking algorithms [92,93,153] to understand the nature of the object in the scene and detect anomalies. Trajectory-based descriptors are also widely used to capture long-term motion information and to estimate the dynamic information of foreground objects for action recognition [117]. Moreover, PCA [32], K-means [47], video restoration [53], 3D neural network [112] are observed in the literature for detection of anomaly.

5.2.4 Reconstruction based

In reconstruction-based techniques, the anomalies are estimated based on reconstruction error. In this technique, every normal sample is reconstructed accurately using a limited set of basis functions, whereas abnormal data is observed to have larger reconstruction loss [56]. Depending on the model type, different loss functions and basis functions are used. Some of the methods use hyperspectral image (HSI) [61,150] and 3D convolution network [20] to estimate reconstruction loss. A deep neural network (DeepOC) in [148] can simultaneously train a classifier and learn compact feature representations. This framework uses the reconstruction error between the ground truth and the predicted future frame to detect anomalous events. Another set of methods use generative adversarial network (GAN) [54,111,113,165] to learn the reconstruction loss function [44]. GAN-based auto-encoder proposed in [125] produces reconstruction error and detects anomalous events by distinguishing them from the normal events [111]. Further, an adversarial learning strategy and denoising reconstruction error are used to train a 3D convolutional auto-encoder to discriminate abnormal events [128]

Another paradigm to detect anomalous events is by exploiting the low-rank property of video sequences. Depending on low-rank approximation, a weighted sparse recon-

struction method is estimated to describe the abnormality in the testing samples [157,162]. Various researches include GAN using self attention [165], U shaped GAN [113], deep spatiotemporal translation [30], dual discriminator [25], Bidirectional Retrospective Generation Adversarial Network (BR-GAN) [156]. Moreover, [98] uses the combination of prediction-based and reconstruction-based methods to construct a U-shaped GAN.

5.2.5 Prediction based

Prediction-based approach uses known results to train a model [102,168] and predicts the probability of the target variable based on the estimated significance from the set of input variables. In [28], the difference between the actual and predicted spatiotemporal characteristics of the feature descriptor is calculated to detect the anomaly [73]. Nawaratne et al. proposed incremental spatiotemporal learning (ISTL) with fuzzy aggregation to distinguish anomalies that evolve [84]. Long short-term memory (LSTM) is very powerful as they store past information to estimate future predictions. LSTM networks are used to learn the temporal representation to remember the history of the motion information to achieve better predictions [57]. To enhance the approach, [13] integrates autoencoder and LSTM in a convolutional framework to detect video anomalies. Another technique of learning spatiotemporal characteristics is estimating an adaptive iterative hard-thresholding algorithm (ISTA) where a recurrent neural network is used to learn sparse representation and dictionary to detect anomalies [170]. Lately, Bidirectional LSTM (BD-LSTM) has been extensively used to provide better prediction accuracy [106]. Waseem Ullah et al. study a deep bidirectional LSTM [136], attention-based residual block concept [138] and multi-layer LSTM [139] to reduce redundancy and time complexity in videos. [10] uses BD-LSTM to analyze anomalies in UAV real-world aerial datasets.

5.2.6 Other approaches

To handle complex issues in traffic surveillance, [64] estimates a fuzzy theory and proposes a traffic anomaly detection algorithm. To perform the state evaluation, virtual detection lines are used to design the fuzzy traffic flow, pixel statistics are used to design fuzzy traffic density, and vehicle trajectory is used to design the fuzzy motion of the target. To recognize abnormal events in traffic such as accidents, unsafe driving behavior, on-street crime, traffic violations, [80] proposes an adaptive sparsity model to detect such anomalous events. Similarly, [66] estimates sparsity-based background subtraction method and shrinkage operators. Other approaches also include high-frequency correlation sensors [36], particle filtering [132], redundancy removal [142] to detect vehicle anomaly.

6 Evaluation parameters and datasets

Detecting the abnormal events from videos is very challenging due to the ambiguous nature of anomalies, lack of enough training data, and circumstances at which events took place. Along with these other factors include illumination conditions (day/night), scene of surveillance, number of entities in the scene, variation in environmental conditions and most importantly working status of capturing cameras. There are many publicly available datasets for training and validating anomaly detection. Table 5 gives the detailed list of datasets categorized according to surveillance targets. Depending upon the application, researchers have chosen specific datasets and evaluation parameters to work with, as the impact and the result of each dataset are different. The popular evaluation parameters observed are area under the curve (AUC), receiver operating characteristics (ROC), equal error rate (EER), precision, recall, and detection rate (RD). It is observed that there AUC and EER are effectively used for quantifying the anomalies in UCSD, Avenue, Subway, and ShanghaiTech datasets. Precision and recall are both extensively used in information extraction providing maximum precision of 95.6% [170] on Avenue dataset and maximum recall of 100% [170] for UCSD dataset. Moreover, maximum AUC of 96.9% and minimum EER of 8.8% are observed on UCSD dataset [148]. We present the breakthroughs and attainments of evaluation parameters by various algorithms on different datasets in Table 6.

6.1 UCSD

UCSD is an extremely popular dataset, widely used by anomaly detection researchers. Videos in the UCSD dataset include events captured from various crowd scenes that range from sparse to dense. The dataset represents different situations like; walking on the road, walking on the grass, vehicular movement on the footpath, unexpected behavior like skateboarding, etc. Ped1 consists of 34 training video samples and 36 testing video samples and Ped2 has 16 training video samples and 12 testing video samples [79]. From Table 6, we observe that UCSD has been deployed using DNNs, CNNs, LSTMs, and GANs throughout the years. The year 2021 has seen the use of GAN and LSTM modules for anomaly detection at a great extent. UCSD has acquired the highest accuracy of 94.8% for LSTM [106] and 98.5% for DeepGAN [30] using Ped1 and while LSTM and GAN offering an accuracy of 96.5% [106] and 97.1% [98] respectively using Ped2.

6.2 Avenue

Publicly released in 2013, Avenue is another popular dataset used for abnormal event detection [77]. The 16 training

and 21 testing video clips are captured in CUHK Campus Avenue with 30652 (15328 training, 15324 testing) frames in total. The anomalies in the dataset include a random person running, any abandoned object, person walking with a suspicious object. As given in Table 6, LSTM outperforms other machine learning methods with an accuracy of 98% [138], followed by GAN with 89.82% [111] accuracy.

6.3 ShanghaiTech

The above datasets are restricted to a sparse crowd. The ShanghaiTech dataset offers is a large-scale crowd counting dataset consisting of 1198 (Part-A: 482; Part-B:716 images) crowd images where each individual is annotated with one point close to the center of the head [166]. As compared to a sparse crowd, anomaly detection in the crowded scene becomes onerous and it is noticeable in the results manifesting the highest accuracy as 78.43% [111] using the GAN model (Table 6).

6.4 Subway

Subway dataset captures the view of underground train station both at entrance and exit where anomalous situations include walking in the wrong way (people entering the exit gate), loitering, suspicious interactions, avoiding turnstiles [21]. Table 6 implies that the majority of the machine learning algorithms including GAN, DNN, CNN, and handcrafted features show optimum results for accuracy greater than 90%.

6.5 Mini-Drone video dataset

The anomaly detection dataset discussed heretofore was using a stationary camera. Mini-Drone Video Dataset (MDVD) offers drone-based surveillance with privacy protection filters that offer an aerial view of a car parking lot. With a resolution of 960×540 MDVD consists of 21 different video sequences for assessment [8]. This UAV-based surveillance dataset has anomalies that include bad parking, car stealing, people quarreling, suspicious activities [19]. The quantitative results obtained by [18] using this challenging dataset are impressive with an AUC of 0.93.

6.6 Other datasets

Other datasets that are often found in the literature are Badminton [21] Behave [12] and QMUL Junction [42], Mind's Eye [22] and Vanaheim dataset [22]. These datasets include normal videos and abnormal videos for training and testing purposes depending upon the application. UCSD dataset includes individuals and vehicles, while DAVIS dataset is composed of various objects (human, vehicles, animals) to obtain the class diversity [164]. The Uturn dataset is a video

Table 5 List of Datasets

Dataset	Usage in literature	Surveillance target
UCSD	[12,121,128,148,149,170] [20,73,84,107,108,125] [21,110,157,158]	Automobile, Individual, Crowd (Public Places)
Avenue, CUHK	[13,84,148,149,170] [20,73,125]	Individual, Crowd, Object (Public Places)
UMN	[107,110,121,157,170]	Individual, Crowd (Public Places)
Subway	[12,45,125,128,148,149] [20–22,108,110]	Individual, Crowd (Entrance and Exit of Subway Stations)
Uturn	[110]	Objects (Vehicles), Individuals (Pedestrians)
Vanaheim	[22]	Individual, Crowd (Metro Stations)
Mind's Eye	[22]	Individual, Crowd, Objects (Vehicles) (Parking Area)
UVSD, DAVIS	[164]	Individuals, Objects (Vehicles)
Shanghai	[73,76,125,128]	Individual, Crowd, Objects (Vehicles) (Public Places)
Badminton	[21]	Individual, Crowd (Badminton Game)
Behave	[12]	Individual, Crowd Public Places)
ABU dataset	[44,61]	Objects (Aerial View)
MDVD	[18,19]	Objects (Aerial View)
IEEE SP Cup-2020	[16,17]	Objects (Lobby, Parking Area)
AU-AIR	[9]	Objects (Aerial View)
Facial expression 2013 (FER-2013)	[2]	Individuals (Facial Expressions)
ISLD-A	[3]	Individuals, Objects (Human Action Recognition)
Self-deployed	[43,57,65,69,87,167] [47,49,59,70,122] [23,64,71,161] [78,151]	Individuals, Objects, Crowd, Automobiles
STONE	[150]	Objects (Grassy Scene)
AVIRIS	[150]	Objects (Aerial View)
SkyEye	[105]	Objects (Aerial View)
UCF101	[104]	Individual (Action Videos)
HMDB51	[104]	Individual (Action Videos)
THUMOS14	[35,104]	Individual (Action Videos)
BP4D	[95]	Individual (Facial Expressions)
AFEW	[95]	Individual (Facial Expressions)
CAVIAR	[80]	Individuals, Objects (Vehicles)
PASCAL VOC	[88,89,147,169]	Objects (Vehicles)
DS1, DS2	[82]	Events (Fire/Smoke/Fog)
SFpark	[7]	Object (Taxis)
LISA 2010	[169]	Objects (Vehicles)
ImageNet	[83]	Events (Fire/Smoke/Fog)
QMUL junction	[12,42]	Objects (Vehicles)
HockeyFight	[4]	Individuals
ViolentFlow	[4]	Individuals

Table 6 Breakthroughs and attainments on different datasets (FL=frame level; PL = pixel level, NA = not applicable, Ent=Entrance)

Dataset	Year	Ref	AUC	EER	Method
UCSD (Ped1)	2021	[106]	94.8%	NA	LSTM
	2021	[113]	85.3%	23.6%	GAN
	2021	[111]	73.26%	28.75%	GAN
	2021	[156]	84.7%	22.5%	GAN
	2020	[128]	90.2%	11.6%	Adversarial 3D CNN Autoencoder
	2020	[5]	82.31%	21.43%	HoMM
	2020	[30]	FL=98.5%; PL=77.4%	FL=5.2%; PL=27.3%	Deep GAN
	2019	[84]	75.2%	29.8%	Deep Learning ISTL Autoencoder
	2019	[170]	FL=83.5% ; PL=45.2 %	25.2%	LSTM
	2019	[125]	90.5%	11.9%	GAN
	2019	[148]	FL=83.5; PL=63.1%	23.4%	DeepOC
	2018	[20]	90.6%	16.2%	Deep Learning
	2018	[149]	FL= 94.1% ; PL = 74.8%	FL = 11.4; PL=28.8%	SVM
	2017	[107]	99.6%	FL = 9.1; PL= 15.18	Deep Cascade
	2016	[21]	72.7%	33.1%	HOF
UCSD (Ped2)	2021	[106]	96.5%	NA	LSTM
	2021	[113]	95.7%	12%	GAN
	2021	[111]	76.98%	23.46%	GAN
	2021	[98]	97.1%	NA	GAN
	2021	[156]	94.6%	7.6%	GAN
	2020	[128]	91.1%	10.9%	Adversarial 3D CNN Autoencoder
	2020	[5]	94.16%	13.25%	HoMM
	2020	[30]	FL = 95.5%; PL=83.1%	FL=9.4%; PL=21.8%	Deep GAN
	2020	[67]	NA	FL = 10.5%; PL=13.8%	CNN
	2020	[25]	95.6%	NA	GAN
	2019	[84]	91.1%	8.9%	Deep Learning ISTL Autoencoder
	2019	[170]	FL = 94.4% ; PL=52.8%	10.3%	LSTM
	2019	[125]	90.7%	11.3%	GAN
	2019	[148]	FL=96.9%; PL=95.0%	8.8%	DeepOC
	2018	[20]	90.2%	17.3%	Deep Learning
2018	[149]	PL=89.2%	PL=16.7%	SVM	
2018	[108]	NA	FL= 11%; PL=15%	Fully CNN	
2017	[107]	NA	FL=8.2; PL= 19	Deep Cascade	
2016	[21]	87.5%	20.0%	HOF	
Avenue	2021	[165]	89.2%	NA	GAN
	2021	[113]	86.9%	20.2%	GAN
	2021	[138]	98%	NA	LSTM
	2021	[98]	85.8%	NA	GAN
	2021	[156]	88.6%	NA	GAN
	2021	[111]	89.82 %	21.55%	GAN
	2020	[128]	88.9%	18.2%	Adversarial 3D Autoencoder
	2020	[30]	87.9%	20.2%	DeepGAN
	2020	[53]	84.9 %	NA	DNN
2019	[84]	76.8%	29.2%	DL ISTL Autoencoder	

Table 6 continued

Dataset	Year	Ref	AUC	EER	Method
Avenue	2019	[170]	FL=86.1%; PL=94.1%	22%	LSTM
	2019	[125]	89.2%	17.6	GAN
	2019	[148]	88.6%	18.5%	DeepOC
	2018	[20]	FL=82.1; PL=93.7%	NA	Deep Learning
ShanghaiTech	2021	[165]	75.7%	NA	GAN
	2021	[113]	73%	32.3%	GAN
	2021	[111]	78.43%	25.16%	GAN
	2021	[98]	73.7%	NA	GAN
	2021	[156]	74.5%	31.6%	GAN
	2020	[25]	73.7%	32%	GAN
	2020	[128]	74.6%	27.6%	Adversarial 3D Convolution Autoencoder
	2019	[125]	70%	36.5%	GAN
Subway	2020	[128]	Ent=90.5%; Exit=94.8%	Ent=22.7%; Exit=9.6%	Adversarial 3D
	2020	[67]	Ent=NA; Exit=NA	Ent=10.1%; Exit=15.9%	3D CNN
	2019	[125]	Ent=90.2%; Exit=94.6%	Ent=22.67%; Exit=9.3%	GAN
	2019	[148]	Ent=91.1%; Exit=89.5%	Ent=NA; Exit=NA	DeepOC
	2018	[108]	Ent=90.4%; Exit=90.2%	Ent=17%; Exit=16%	FCNN
	2016	[21]	Ent=81.6%; Exit=84.9%	Ent=22.8%; Exit=17.8%	HOF

of a road crossing with trams, vehicles, and pedestrians on the scene. The abnormal activity videos cover illegal U-turns and trams [110]. Vanaheim Dataset consists of videos containing people passing turnstiles while entering/exiting stations recorded in metro stations [22]. The abnormal events encountered were a person loitering, a group of people suddenly stopping, a person jumping over turnstiles. Some authors have also used live videos for the implementation of their respective methods [148]. Anomalous events from live videos like an accident, kidnapping, robbery, and crime (a man being murdered) are seen in the literature. Various algorithms have been developed to tackle challenges in video surveillance in different datasets. Table 5 gives a detailed list of datasets with surveillance targets.

7 Overview of edge computing

Edge computing has seen remarkable growth in recent years proving its potential in data processing at the edge of the network, thereby reducing latency and saving cost [114]. Feature of computing the data where it is generated remarkably eases the growth of latency-sensitive applications [116], thereby reducing data traffic in the network and saving the bandwidth and system cost. These benefits of edge computing facilitate remarkable progress, especially in time-critical real-time applications such as anomaly detection [140]. With the advancement in the terminal or edge devices, few contributions are observed in detecting anomalies at the edge

or terminal devices. Schneible et al. present a federated learning approach in which autoencoders are deployed on edge devices to identify anomalies. Utilizing a centralized server as a back-end processing system, the local models are updated and redistributed to the edge devices [115]. Despite the rapid development in learning methods and edge computing devices, there is a gap between software and hardware implementations [131].

The general architectural overview of the edge computing paradigm is shown in Fig. 6. The top-level entities are cloud storage and computing devices which comprise data centers and servers. The middle level represents fog computing. Any device with compute capability, memory, and network connectivity is called a fog node. Examples of fog devices are switches, routers, servers, controllers. The bottom-most part of the pyramid includes edge devices like sensors, actuators, smartphones, and mobile phones. These terminal devices participate in processing a particular task using a user access encryption [46].

7.1 Edge computing paradigms

7.1.1 Low latency computing and decentralized cloud

As far as anomaly detection using the cloud is concerned, the data is captured on the device and is processed away from the device leading to a delay. Moreover, if the cloud centers are geographically distant, the time response is hampered further. Edge computing has the capability of processing

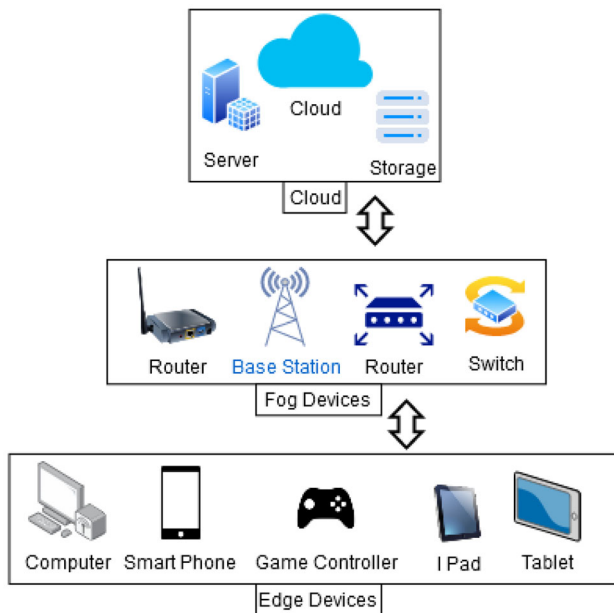


Fig. 6 Architectural overview of edge computing

the data where it is produced thereby reducing the latency [88,155]. Other conventional methods focus on improving either transmission delay or processing delay, but not both. Service delay puts forth a solution that reduces both [103].

7.1.2 Security and accessibility

In edge computing, a significantly large number of nodes (edge devices) participate in processing tasks and each device requires a user access encryption [46]. Also, the data that is processed needs to be secured as it is handled by many devices during the process of offloading [36].

7.1.3 Quality of service

The quality of service delivered by the edge nodes is determined by the throughput where the aim is to ensure that the nodes achieve high throughput while delivering workloads. The overall framework should ensure that the nodes are not overloaded with work; however, if they are overloaded in the peak hours, the tasks should be partitioned and scheduled accordingly [103,130].

7.1.4 Distributed computing

Edge computing uses the technique of dividing computationally expensive tasks to other nodes available in the network thereby reducing response time. The transfer of these intensive tasks to a separate processor such as a cluster, cloud-let, or grid is called computation offloading. It is used to accelerate applications by dividing the tasks between the nodes such

as mobile devices. Mobile devices have physical limitations and are restricted in memory, battery, and processing. This is the reason that many computationally heavy applications do not run on such devices. To cope with this problem, the anomaly detection task is migrated to various edge devices according to the computing capabilities of respective devices. Xu et al. tried to optimize running performance, responsive time, and privacy by deploying task offloading for video surveillance in edge computing enabled Internet of Vehicles [151]. Similarly, a Fog-enabled real-time traffic management system uses a resource management offloading system to minimize the average response time of the traffic management server [24,144,154]. The resources are efficiently managed with the help of distributed computing or task offloading [144,145,167].

8 Anomaly detection at edge devices using machine learning

In recent years, the research community has witnessed extensive development and growth in the field of edge computing and anomaly detection respectively. Fortunately, the latest quantum leap in anomaly detection using machine learning elucidates on the edge application scenarios, providing proficiency in big data management, cost-saving, delay management, etc. The confluence of anomaly detection and edge computing can further create momentum, empowering the development of imminent applications. To fully unleash the underlying capabilities of edge computing, we present a detailed survey of the recent integration of anomaly detection techniques with edge computing. We also summarize them in Table 7 and categorize them based on the methodology used.

8.1 Handcrafted features on edge

Handcrafted Features such as feature-based classification approaches are noticeable candidates for edge application. As discussed earlier, the features that are manually engineered by data scientists using techniques such as histogram of oriented gradients (HOG), histogram of optical flow (HOF), etc., are called handcrafted features. For human-object detection, [153] uses HOG to extract spatiotemporal features and support vector machine (SVM) with kernelized correlation filters (KCF) for classification. On similar lines, a Kerman algorithm which is a combination of kernelized Kalman filter, Kalman filter (KF), and background subtraction (BS) is proposed by [89] to achieve enhanced performance on edge. To implement a low source enabled edge computing device, [2] proposes a lightweight quantized neural network based on the FPGA Platform. Here, quantization refers to techniques for performing computations and storing tensors at lower bit widths in order to yield a high throughput. The author

Table 7 Anomaly detection at the edge FP=false positive; FN=false negative; FPR=false positive rate; TPR=true positive rate; ROC=receiver operating characteristics; MAE=maximum absolute error

Method	Ref/Year	Features	Learning	Anomaly criteria	Dataset	Parameters
Hand-crafted Features	[153] (2018)	Low computational cost with high accuracy, performance	HOG,SVM,KCF	Human object tracking	Self deployed	Speed, Performance
	[89] (2019)	Real-time, good accuracy, with limited resources	LCNN, Kerman (KCF, KF, BS) Objects	Tracking human	VOC07, VOC12	Accuracy Precision
	[2] (2021)	Accuracy	Binary Neural Network; FPGA system design	Recognition of facial emotions	JAFFE	Accuracy
CNN	[88] (2018)	Processing requires less memory	Harr-Cascade, SVM, L-CNN	Smart surveillance	ImageNet	Accuracy precision, recall, FP, FN F-measure
	[83] (2019)	Processing requires less memory	LCNN	Smoke detection in foggy surveillance	ImageNet	FP, FN, Accuracy precision, recall, F-measure
DNN	[152] (2021)	Real-time low complexity	CNN	Detecting pedestrian	Multi pedestrian	Accuracy
	[142] (2022)	Reduced Bandwidth, Security, Real-time	Video motion magnitude	Vehicle anomaly	UA-DETRAC, crossroad in Beijing city	Precision Recall F-measure
	[91] (2021)	Real-time feasible	Spatiotemporal feature extraction, trajectory association	Pedestrian identification	UCSD Ped1 UMN	Accuracy
	[65] (2018)	Improved efficiency	CNN, DL	Industry manufacture inspection	Self deployed	FPR, TPR, ROC
	[49] (2020)	Smart parking	CNN, HOG	Parking surveillance	Self deployed	Accuracy
DNN	[11] (2019)	Balances computational power and workload	DIVS	Vehicle classification	Self deployed	Efficiency
	[36] (2019)	Reduced computation , overhead privacy	high-frequency correlation, sensors	Vehicle Anomaly	Open Source Platform	FPR, TPR, ROC
	[51] (2020)	Resource optimization	Deep learning, Video analytics	Video optimization	VIRAT 2.0 Ground	F1 Score accuracy
DNN	[53] (2020)	Accuracy time efficiency	DNN	Face Mask Identification	Bus Drive Monitoring, public dataset	Accuracy
	[167] (2020)	Reduce network occupancy and system response delay	Intelligent Edge Surveillance	Cloud, DL, Edge	Self deployed	Accuracy Model loss
LSTM	[57] (2019)	Parallel computing to improve efficiency	LSTM, Cross entropy	Industrial electrical equipment	Self deployed	Accuracy
GAN	[137] (2021)	Effective utilization of resources	BD-LSTM, L-CNN in videos	Crime scene	UCF-Crime,RWF-2000	Accuracy
	[147] (2019)	Increases computing performance	SaliencyGAN Deep SOD CNN Adversarial learning semi-supervised	Object detection/Tracking	PASCALS	MAE F-Measure precision recall

implements binary neural network (BNN) on FPGA to track facial expressions of passengers classified into six categories, namely: fear, happy, sad, anger, surprise, and disgust.

8.2 Convolutional neural network on edge

A Convolutional neural network (CNN) is a type of artificial neural network which is specifically designed to process pixel data. Many researchers have introduced a lightweight convolutional neural network (L-CNN) to train the network using Harr-Cascade and HOG feature [88]. Such models are pretrained using VOC07, MobileNet V2 [83]. To provide real-time object detection, YOLO—a neural network algorithm comprising convolution layers—is used for faster object detection. Further, [152] evaluates a fast and lightweight YOLO named FL-YOLO comprising depth-wise separable convolution layers. To migrate the tasks to the edge some kind of pre-processing is required. [142] proposes a technique to discard redundant frames and divide the video into segments of interest-based into spatiotemporal interest points (STIP), thereby decreasing the total number of video frames. [18] uses a pretrained CNN using the technique of one-class classification to extract the features from UAV videos. These videos are obtained from a very complicated dataset called MDVD where the videos are captured through drones in a parking area. Further [91] proposes a CPU-only edge device to detect complex anomalies in video scenes by extracting spatiotemporal features and for trajectory association and re-identification. In order to migrate the tasks to the edge devices, the author uses Raspberry Pi 4B as the edge node and 433MHz LoRa Module as the link to connect the edge node with the MobileNet architecture.

8.3 Deep neural networks on edge

Compared to the convolutional neural network (CNN), deep neural network (DNN) is a fully connected neural network where each input layer is connected to hidden layers with activation functions and bias. DNNs are widely used in smart industries to detect manufacturing anomalies [65]. DNNs are a popular choice for smart city [160] including intelligent parking systems [49], vehicle classification and traffic flow prediction [11], to prevent vehicle accidents [36].

As the processing capacity of the edge devices is less there needs to be some kind of video pre-processing to make the videos compatible on the edge. For this purpose, video analytics system is used to increase the performance and efficiency of the application with proper utilization of resources. [51] addresses the curbs of current video analytics systems which are inefficient to process the video analytics of multiple video streams and proposes a DNN-based GPU-enabled edge server to process large data streams. The module was trained using VIRAT 2.0 Ground dataset and the

results show reduced accuracy degradation in various scenarios. On similar grounds, [127] conducts video analytics on video streams by using a mobile edge computing scheme. The aim is to utilize available resources by task partitioning and pre-processing the video data using a DNN model. [52] proposes a solution for real-time videos by designing a Front-CNN consisting of a Shallow 3D CNN and pre-trained 2D CNN as the Back-CNN. This end-to-end trainable architecture is capable of learning both spatiotemporal information of videos thereby achieving state-of-the-art performance. The shallow lightweighted module is trained on NVIDIA Jetson Nano Developer as an edge-computing device to show its real-time video recognition, especially in edge-computing habitat.

Due to the outbreak of the COVID-19 pandemic, the World Health Organization (WHO) has recommended preventive measures one of them is mask-wearing. To help prevent the pandemic, real-time detection of face masks becomes an urgent need. Keeping this most effective non-pharmaceutical measure of mask-wearing in practice, [53] designed a framework that detects face masks based on real-time edge computing devices. The framework consists of three stages: video restoration, face detection, and mask detection using Intel's Neural Compute Stick 2 (NCS) as the edge computing device and tested efficiency and accuracy of the model on the Bus Drive Monitoring dataset and public dataset.

The deep learning platform well integrates with the edge devices providing benefits of reduced network occupancy, low time response [167] and reduced communication overhead [144].

8.4 Long short-term memory on edge

Different from DNNs and CNNs which are profound in abstracting spatiotemporal features in the data, long short-term memory (LSTM) is thoughtfully designed to process sequential data. LSTM networks are well-suited to make a prediction, classification, and processing of data based on time series [57]. [137] utilizes a two-stream neural network strategy where stream one performs instant anomaly detection at the edge using lightweight CNN and then stream two performs detailed anomaly detection at the cloud using bidirectional LSTM (BD-LSTM) to detect anomalous events like abuse and assault. LSTMs efficiently deal with the problem vanishing gradient and hence have a promising future for anomaly-related problems [15].

8.5 Generative adversarial network on edge

The above-mentioned methods though performing well in some scenarios are not as accurate as the reconstruction-based learning approaches [54]. Generative adversarial

network popularly called GANs are two neural networks that contest with each other to attain optimum results. In practice, the generative network learns to map a data distribution from a latent space, while the discriminative network distinguishes the data produced by the generator from the true data distribution. GANs are well suited for anomaly detection as the aim is to distinguish anomalies from normal events. In [147], GAN in combination with Tesla K80 GPU with 12G memory, a fog device is used to analyze anomalies. Though originally proposed for unsupervised learning, GANs are also proved useful for semi-supervised learning.

9 Anomaly detection on the edge: challenges and future research directions

Despite the fact that the confluence of Edge Computing and Anomaly Detection has divulged great potentials and brought about the rapid development of many applications, there still exist enormous hurdles in attaining steady, vigorous, empirical usage which calls for the uninterrupted endeavor in this area from many aspects. In the next section, we discuss some pivotal research challenges and favorable future research directions.

9.1 Training of model

The execution of edge-enabled-anomaly-detection applications immensely depends on the way learning models perform in an edge environment, where the model training is a supreme process. It is well established that the training of the model demands plenty of computation and thus consumes enormous computer resources, principally for DNNs [31]. The servers at the edge are not economical to entirely lay hold of the task of training the model. To resolve this issue, distributed learning [11], task offloading [144], resource sharing [144], federated learning [115] are propitious solutions.

9.2 Application improvement

Though edge computing and machine learning integration has accomplished exceptional improvement in numerous applications, there are a few perilous time-sensitive applications like VR gaming, desiring for a quantum leap. Such applications require less delay and vigorous computation. In such scenarios, video analytics is performed at the edge nodes and the outcomes are handed over to the cloud in real time [51].

Another perspective to enhance the application is federated learning. Joseph Schneible et al. take the advantage of the annually escalating potential of edge devices and performs local analytics using federated learning to detect anomaly [114]. For the detection of anomalous scenes in

residential surveillance videos, [91] uses a CPU-only edge device to apprehend object-level speculations. Even though the fundamental foundation of feasibility is established, practical application is still far-flung and needs improvement.

9.3 Annotation of anomaly

The interpretation of anomaly is different in different scenarios. For example, riding a motorbike at a speed of 60 km/hr would be normal on a highway but abnormal on a crowded street. Applying machine learning to anomaly detection requires a good understanding of the problem, especially in situations with unstructured data. Videos and images encoded as a sequence of pixels carry little interpretation and render the old algorithms useless until the data becomes structured. Further, a founding principle of any good machine learning algorithm is that a large dataset is necessary for training the model [85]. Inferences can be made only when predictions can be validated. Anomaly detection benefits from even larger amounts of data because the assumption is that anomalies are rare [85]. Researchers use benchmarks to compare the performance of one algorithm to another. Different kinds of models use different bench-marking datasets like image classification has MNIST and IMAGENET [88]. However, in anomaly detection, none of the datasets has yet become a standard as the applications are varied. Despite the varied applications of video surveillance, we observed that UCSD, Avenue, and Subway are the widely used datasets that provide an accuracy of over 90% when used with deep neural networks [30,128,138].

9.4 Hardware–software optimization

Along with the enhancement of model training and application improvement, hardware–software escalation is challenging yet propitious. Although the current hardware–software environment is specially designed for the cloud, efforts have been made to optimize them with the edge-computing platform [116,140].

From the hardware architecture approach, [26] proposed streaming hardware accelerator using TSMC 65nm technology to enhance the performance of CNN in the edge-based application. Furthermore, [2] studies FPGA system design to support edge-computing-based platforms. For the real-time video analysis, CNN layers are used in confluence with NVIDIA Jetson TX1 is the edge device [152]. Similarly, [53] uses Intel Neural Compute Stick 2 (NCS) as the edge device. To accomplish higher processing speed for real-time applications, [91] utilizes Movidius Neural Compute Stick (MNCS) as a hardware accelerator.

From the software approach, some of the firms have proposed their private platforms regarding edge-computing. AWS IoT Greengrass, an open-source edge and cloud service

assists you to deploy intelligent software-based applications (<https://aws.amazon.com/cn/greengrass/>). Meanwhile MXNet [14], Tensorflow Lite (<https://www.tensorflow.org/lite>) are frameworks especially designed for edge environments. Despite the current frameworks, there is a need to amalgamate them to attain a high-yielding, high-throughput practical system for edge-computing applications.

10 Observations

After studying different paradigms of anomaly detection in video surveillance systems, we observe that some approaches intend to neglect background and focus only on foreground features for anomaly detection [29]. We think that background information would be useful to model environmental conditions like rainy, sunny, or snowy weather that can cause anomalies [88,89]. Anomaly detection using CNN [142], Deep Learning [148], LSTM [106] and GAN [30,54] has a promising future as they offer higher accuracy in most of the surveillance scenarios. Further, the performance depends on the density of the crowd, as the crowd increases the performance of the anomaly detection model decreases and it works best when the crowd is sparse [62,107]. Moreover, only benchmark data-set-based comparison may not be relevant for all real-life situations, as they are not enough to consider all real-life scenarios [146].

Out of the modeling algorithms discussed above, the prediction-based approach used to process the data that evolves is befitting for anomaly detection [15]. This approach learns the temporal characteristics of the data and preserves the representations to estimate future probabilities providing refined solutions [57]. The prediction-based model using LSTM deals with the problem of vanishing gradient and is suitable for processing time-series data that is suitable in anomaly detection [136]. Another methodology that imparts optimum results for anomaly detection is reconstruction-based modeling where GAN models are predominantly used to detect anomalies as they learn representations faster than CNN models [38], whereas GANs generally face an issue while dealing with a large number of distributions [54]. With advanced techniques like self attention GAN [165], U shaped GAN [113] and BR-GAN [156] this issue can be overcome.

For delay-sensitive applications like intelligent surveillance and anomaly detection, edge computing is a promising approach [145]. It offers more privacy and security as the data is processed on the device itself [142]. With continuous improvement in edge devices and task offloading the workload is divided thereby improving the overall efficiency [151]. The confluence of anomaly detection with edge computing will bring new possibilities in the field of computer vision [91].

11 Conclusion

As described herein, we explored and categorized various anomaly detection techniques applied for different video surveillance scenarios. As the context anomaly is subjective, we considered surveillance scenarios with pedestrian, crowd, traffic, industries, and public places. We presented the evolution of the anomaly detection techniques over the years and the surveys done so far. We emphasized the learning techniques, methodologies, approaches, and scenarios for anomaly detection. This survey intends to provide detailed insight and associated diversities in anomaly detection techniques. Out of the discussed modeling algorithms, we estimate that the reconstruction and prediction-based techniques will be dominant in the years to come. We investigate how the recent advances in anomaly detection can be used with edge computing to provide real-time, time-efficient video surveillance applications. In the future anomaly detection in combination with edge computing, using LSTMs and GANs has great potential in the future. We also establish that anomaly detection using edge devices is less explored and has a large scope of improvement to achieve state-of-the-art anomaly detection and intelligence surveillance.

Funding Not applicable

Data Availability All the data and material corresponding to the manuscript will be made publicly available.

Declarations

Conflicts of interest Both the authors have checked the manuscript and have agreed to the submission in International Journal of Multimedia Information Retrieval. There is no conflict of interest between the authors.

Code availability Not applicable as the manuscript presents a survey of anomaly detection in video surveillance and focus on anomaly detection using edge devices.

References

1. Ahmed SA, Dogra DP, Kar S, Roy PP (2018) Trajectory-based surveillance analysis: a survey. *IEEE Trans Circuits Syst Video Technol* 29(7):1985–1997
2. Ajay B, Rao M (2021) Binary neural network based real time emotion detection on an edge computing device to detect passenger anomaly, In: 2021 34th International conference on VLSI design and 2021 20th international conference on embedded systems (VLSID) pp 175–180, IEEE
3. Angelini F, Yan J, Naqvi SM (2019) Privacy-preserving online human behaviour anomaly detection based on body movements and objects positions. In: ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP) pp 8444–8448, IEEE

4. Asad M, Yang J, He J, Shamsolmoali P, He X (2021) Multi-frame feature-fusion-based model for violence detection. *Visual Comput* 37(6):1415–1431
5. Bansod SD, Nandedkar AV (2020) Crowd anomaly detection and localization using histogram of magnitude and momentum. *Vis Comput* 36(3):609–620
6. Bergman L, Hoshen Y (2020) Classification-based anomaly detection for general data. *arXiv preprint arXiv:2005.02359*
7. Bock F, Di Martino S, Origlia A (2019) Smart parking: Using a crowd of taxis to sense on-street parking space availability. *IEEE Transactions on Intelligent Transportation Systems* 21(2):496–508
8. Bonetto M, Korshunov P, Ramponi G, Ebrahimi T (2015) Privacy in mini-drone based video surveillance. In 2015 11th IEEE international conference and workshops on automatic face and gesture recognition (FG) IEEE vol 4, pp 1–6
9. Bozcan I, Kayacan E (2021) Context-dependent anomaly detection for low altitude traffic surveillance, *arXiv preprint arXiv:2104.06781*
10. Chakravarthy AS, Sinha S, Narang P, Mandal M Dronesegnet: robust aerial semantic segmentation for uav-based iot applications
11. Chen J, Li K, Deng Q, Li K, Philip SY (2019) Distributed deep learning model for intelligent video surveillance systems with edge computing. *IEEE Trans Ind Inf.* <https://doi.org/10.1109/TII.2019.2909473>
12. Cheng K-W, Chen Y-T, Fang W-H (2015) Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation. *IEEE Trans Image Process* 24(12):5288–5301
13. Cheng H, Liu X, Wang H, Fang Y, Wang M, Zhao X (2020) Securead: a secure video anomaly detection framework on convolutional neural network in edge computing environment. *IEEE Trans Cloud Comput*
14. Chen T, Li M, Li Y, Lin M, Wang N, Wang M, Xiao T, Xu B, Zhang C, Zhang Z (2015) Mxnet: a flexible and efficient machine learning library for heterogeneous distributed systems. *arXiv preprint arXiv:1512.01274*
15. Choi K, Yi J, Park C, Yoon S (2021) Deep learning for anomaly detection in time-series data: review, analysis, and guidelines. *IEEE Access* 9:120043–120065
16. Chowdhury SS, Islam KM, Noor R (2020) Anomaly detection in unsupervised surveillance setting using ensemble of multimodal data with adversarial defense. *arXiv preprint arXiv:2007.10812*
17. Chowdhury SS, Islam KM, Noor R (2020) Unsupervised abnormality detection using heterogeneous autonomous systems. *arXiv preprint arXiv:2006.03733*
18. Chriki A, Touati H, Snoussi H, Kamoun F (2021) Deep learning and handcrafted features for one-class anomaly detection in uav video. *Multimed Tools Appl* 80(2):2599–2620
19. Chriki A, Touati H, Snoussi H, Kamoun F (2020) Uav-based surveillance system: an anomaly detection approach. In: 2020 IEEE Symposium on computers and communications (ISCC)
20. Chu W, Xue H, Yao C, Cai D (2018) Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos. *IEEE Trans Multimed* 21(1):246–255
21. Colque RVHM, Caetano C, de Andrade MTL, Schwartz WR (2016) Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos. *IEEE Trans Circuits Syst Video Technol* 27(3):673–682
22. Coşar S, Donatiello G, Bogorny V, Garate C, Alvares LO, Brémond F (2016) Toward abnormal trajectory and event detection in video surveillance. *IEEE Trans Circuits Syst Video Technol* 27(3):683–695
23. Deng X, Liu Y, Zhu C, Zhang H (2021) Air-ground surveillance sensor network based on edge computing for target tracking. *Comput Commun* 166:254–261
24. Dinh TQ, La QD, Quek TQ, Shin H (2018) Learning for computation offloading in mobile edge computing. *IEEE Trans Commun* 66(12):6353–6367
25. Dong F, Zhang Y, Nie X (2020) Dual discriminator generative adversarial network for video anomaly detection. *IEEE Access* 8:88170–88176
26. Du L, Du Y, Li Y, Su J, Kuan Y-C, Liu C-C, Chang M-CF (2017) A reconfigurable streaming deep convolutional neural network accelerator for internet of things. *IEEE Trans Circuits Syst I Regul Pap* 65(1):198–208
27. Elgarrai Z, El Meslouhi O, Kardouchi M, Allali H (2016) Robust facial expression recognition system based on hidden markov models. *Int J Multimed Inf Retr* 5(4):229–236
28. Ergen T, Kozat SS (2019) Unsupervised anomaly detection with lstm neural networks. *IEEE Trans Neural Netw Learning Syst* 31(8):3127–3141
29. Farooq MU, Khan NA, Ali MS (2017) Unsupervised video surveillance for anomaly detection of street traffic. *Int J Adv Comput Sci Appl(IJACSA)* 12(8):270–275
30. Ganokratanaa T, Aramvith S, Sebe N (2020) Unsupervised anomaly detection and localization based on deep spatiotemporal translation network. *IEEE Access* 8:50312–50329. <https://doi.org/10.1109/ACCESS.2020.2979869>
31. Georgiou T, Liu Y, Chen W, Lew M (2020) A survey of traditional and deep learning-based feature descriptors for high dimensional data in computer vision. *Int J Multimed Inf Retr* 9(3):135–170
32. Ghosh AM, Grolinger K (2020) Edge-cloud computing for internet of things data analytics: embedding intelligence in the edge with deep learning. *IEEE Trans Ind Inf* 17(3):2191–2200
33. Glorot X, Bordes A, Bengio Y (2011) Deep sparse rectifier neural networks. In: Proceedings of the fourteenth international conference on artificial intelligence and statistics, JMLR Workshop and conference proceedings, pp 315–323. <http://proceedings.mlr.press/v15/glorot11a>
34. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. *Adv Neural Inf Process Syst* vol 27
35. Guo D, Li W, Fang X (2018) Fully convolutional network for multiscale temporal action proposals. *IEEE Trans Multimed* 20(12):3428–3438
36. Guo F, Wang Z, Du S, Li H, Zhu H, Pei Q, Cao Z, Zhao J (2019) Detecting vehicle anomaly in the edge via sensor consistency and frequency characteristic. *IEEE Trans Veh Technol* 68(6):5618–5628
37. Hamdi S, Bouindour S, Snoussi H, Wang T, Abid M (2021) End-to-end deep one-class learning for anomaly detection in uav video stream. *J Imaging* 7(5):90
38. Han X, Chen X, Liu L-P (2020) Gan ensemble for anomaly detection. *arXiv preprint arXiv:2012.07988*, vol 7, no 8
39. Hou J, Wu X, Sun Y, Jia Y (2017) Content-attention representation by factorized action-scene network for action recognition. *IEEE Trans Multimed* 20(6):1537–1547
40. Hu P, Ning H, Qiu T, Zhang Y, Luo X (2016) Fog computing based face identification and resolution scheme in internet of things. *IEEE Trans Ind Inf* 13(4):1910–1920
41. Hu W, Gao J, Li B, Wu O, Du J, Maybank S (2018) Anomaly detection using local kernel density estimation and context-based regression. *IEEE Trans Knowl Data Eng* 32(2):218–233
42. Isupova O, Kuzin D, Mihaylova L (2016) Anomaly detection in video with bayesian nonparametrics. *arXiv preprint arXiv:1606.08455*. <https://doi.org/10.48550/arXiv.1606.08455>
43. Ji S, Xu W, Yang M, Yu K (2012) 3d convolutional neural networks for human action recognition. *IEEE Trans Pattern Anal Mach Intell* 35(1):221–231
44. Jiang T, Li Y, Xie W, Du Q (2020) Discriminative reconstruction constrained generative adversarial network for hyperspectral

- anomaly detection. *IEEE Trans Geosci Rem Sens* 58(7):4666–4679
45. Kaltsa V, Briassoulis A, Kompatsiaris I, Hadjileontiadis LJ, Strintzis MG (2015) Swarm intelligence for detecting interesting events in crowded environments. *IEEE Tran Image Process* 24(7):2153–2166
 46. Kang J, Yu R, Huang X, Zhang Y (2017) Privacy-preserved pseudonym scheme for fog computing supported internet of vehicles. *IEEE Trans Intell Transp Syst* 19(8):2627–2637. <https://doi.org/10.1109/TITS.2017.2764095>
 47. Ke R, Li Z, Kim S, Ash J, Cui Z, Wang Y (2016) Real-time bidirectional traffic flow parameter estimation from aerial videos. *IEEE Transactions on Intelligent Transportation Systems* 18(4):890–901
 48. Ke R, Li Z, Tang J, Pan Z, Wang Y (2018) Real-time traffic flow parameter estimation from uav video based on ensemble classifier and optical flow. *IEEE Trans Intell Transp Syst* 20(1):54–64
 49. Ke R, Zhuang Y, Pu Z, Wang Y (2020) A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on IoT devices. *IEEE Trans Intell Transp Syst* 22(8):4962–4974
 50. Khan SD, Basalamah S (2020) Scale and density invariant head detection deep model for crowd counting in pedestrian crowds. *Visual Comput* 37(8):2127–37
 51. Kim W-J, Youn C-H (2020) Lightweight online profiling-based configuration adaptation for video analytics system in edge computing. *IEEE Access* 8:116881–116899
 52. Kim J-H, Kim N, Won CS (2021) Deep edge computing for videos. *IEEE Access* 9:123348–123357
 53. Kong X, Wang K, Wang S, Wang X, Jiang X, Guo Y, Shen G, Chen X, Ni Q (2021) Real-time mask identification for covid-19: an edge computing-based deep learning framework. *IEEE Internet Things J* 8(21):15929–38
 54. Kumar MP, Jayagopal P (2020) Generative adversarial networks: a survey on applications and challenges. *Int J Multimed Inf Retr* 10(1):1–24
 55. Kumar D, Bezdek JC, Rajasegarar S, Leckie C, Palaniswami M (2017) A visual-numeric approach to clustering and anomaly detection for trajectory data. *Visual Comput* 33(3):265–281
 56. Kumaran SK, Dogra SK, Roy PP (2019) Anomaly detection in road traffic using visual surveillance: A survey, [arXiv:1901.08292](https://arxiv.org/abs/1901.08292)
 57. Lai C-F, Chien W-C, Yang LT, Qiang W (2019) Lstm and edge computing for big data feature recognition of industrial electrical equipment. *IEEE Trans Ind Inf* 15(4):2469–2477
 58. Lamba S, Nain N (2019) Segmentation of crowd flow by trajectory clustering in active contours. *Vis Comput* pp 1–12
 59. Lao W, Han J, De With PH (2009) Automatic video-based human motion analyzer for consumer surveillance system. *IEEE Transactions on Consumer Electronics* 55(2):591–598
 60. Lee SW, Kim YS, Bien Z (2009) A non-supervised learning framework of human behavior patterns based on sequential actions. *IEEE Transactions on Knowledge and Data Engineering* 22(4):479–492
 61. Lei J, Fang S, Xie W, Li Y, Chang C-I (2020) Discriminative reconstruction for hyperspectral anomaly detection with spectral learning. *IEEE Trans Geosci Rem Sens* 58(10):7406–7417
 62. Li W, Mahadevan V, Vasconcelos N (2013) Anomaly detection and localization in crowded scenes. *IEEE transactions on pattern analysis and machine intelligence* 36(1):18–32
 63. Li T, Chang H, Wang M, Ni B, Hong R, Yan S (2014) Crowded scene analysis: a survey. *IEEE Trans Circuits Syst Video Technol* 25(3):367–386
 64. Li Y, Guo T, Xia R, Xie W (2018) Road traffic anomaly detection based on fuzzy theory. *IEEE Access* 6:40281–40288
 65. Li L, Ota K, Dong M (2018) Deep learning for smart industry: efficient manufacture inspection system with fog computing. *IEEE Trans Ind Inf* 14(10):4665–4673
 66. Li L, Wang Z, Hu Q, Dong Y (2020) Adaptive non-convex sparsity based background subtraction for intelligent video surveillance. *IEEE Trans Ind Inf* 17(6):4168–78
 67. Li Z, Li Y, Gao Z (2020) Spatiotemporal representation learning for video anomaly detection. *IEEE Access* 8:25531–25542
 68. Li X, Cai Z-m (2016) Anomaly detection techniques in surveillance videos. In: 2016 9th International congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI), pp. 54–59, *IEEE*
 69. Liu C, Chang F (2018) Hybrid cascade structure for license plate detection in large visual surveillance scenes. *IEEE Trans Intell Transp Syst* 20(6):2122–2135
 70. Liu P, Yang P, Wang C, Huang K, Tan T (2016) A semi-supervised method for surveillance-based visual location recognition. *IEEE Trans Cybern* 47(11):3719–3732
 71. Liu SW, Ngan HY, Ng MK, Simske SJ (2018) Accumulated relative density outlier detection for large scale traffic data. *Electron Imaging* 2018(9):1–239
 72. Liu J, Xia Y, Tang Z (2021) Privacy-preserving video fall detection using visual shielding information. *The Visual Computer* 37(2):359–370
 73. Liu W, Luo W, Lian D, Gao S (2018) Future frame prediction for anomaly detection—a new baseline. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6536–6545
 74. Long C, Cao Y, Jiang T, Zhang Q (2017) Edge computing framework for cooperative video processing in multimedia IoT systems. *IEEE Trans Multimed* 20(5):1126–1139
 75. 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
 76. Luo W, Liu W, Gao S (2017) A revisit of sparse coding based anomaly detection in stacked rnn framework. In: Proceedings of the IEEE international conference on computer vision pp. 341–349
 77. 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 pp. 2720–2727
 78. Ma X, Wang S, Zhang S, Yang P, Lin C, Shen XS (2019) Cost-efficient resource provisioning for dynamic requests in cloud assisted mobile edge computing. *IEEE Trans Cloud Comput* 9(3):968–980
 79. Mahadevan V, Li W, Bhalodia V, Vasconcelos N (2010) Anomaly detection in crowded scenes. In 2010 IEEE Computer society conference on computer vision and pattern recognition, *IEEE* pp. 1975–1981
 80. Mo X, Monga V, Bala R, Fan Z (2013) Adaptive sparse representations for video anomaly detection. *IEEE Trans Circuits Syst Video Technol* 24(4):631–645
 81. Muhammad K, Ahmad J, Lv Z, Bellavista P, Yang P, Baik SW (2018) Efficient deep cnn-based fire detection and localization in video surveillance applications. *IEEE Trans Syst Man Cybern Syst* 49(7):1419–1434
 82. Muhammad K, Khan S, Elhoseny M, Ahmed SH, Baik SW (2019) Efficient fire detection for uncertain surveillance environment. *IEEE Trans Ind Inf* 15(5):3113–3122
 83. Muhammad K, Khan S, Palade V, Mehmood I, De Albuquerque VHC (2019) Edge intelligence-assisted smoke detection in foggy surveillance environments. *IEEE Trans Ind Inf* 16(2):1067–1075
 84. Nawaratne R, Alahakoon D, De Silva D, Yu X (2019) Spatiotemporal anomaly detection using deep learning for real-time video surveillance. *IEEE Trans Ind Inf* 16(1):393–402

85. Nayak R, Pati UC, Das SK (2021) A comprehensive review on deep learning-based methods for video anomaly detection. *Image and Vision Computing* 106:104078
86. Nguyen V, Phung D, Pham D-S, Venkatesh S (2015) Bayesian nonparametric approaches to abnormality detection in video surveillance. *Ann Data Sci* 2(1):21–41
87. Nieto RM, García-Martín Á, Hauptmann AG, Martínez JM (2018) Automatic vacant parking places management system using multicamera vehicle detection. *IEEE Trans Intell Transp Syst* 20(3):1069–1080
88. Nikouei SY, Chen Y, Song S, Xu R, Choi B-Y, Faughnan T (2018) Smart surveillance as an edge network service: srom harr-cascade, svm to a lightweight cnn. In: 2018 IEEE 4th international conference on collaboration and internet computing (cic) pp. 256–265, IEEE
89. Nikouei SY, Chen Y, Song S, Choi B-Y, Faughnan TR (2019) Toward intelligent surveillance as an edge network service (isense) using lightweight detection and tracking algorithms. *IEEE Trans Serv Comput.* 14(6):1624–37
90. Oza P, Patel VM (2018) One-class convolutional neural network. *IEEE Signal Process Lett* 26(2):277–281
91. Parate MR, Bhurchandi KM, Kothari AG (2021) Anomaly detection in residential video surveillance on edge devices in iot framework. arXiv preprint [arXiv:2107.04767](https://arxiv.org/abs/2107.04767)
92. Parate MR, Bhurchandi KM (2017) Structurally enhanced correlation tracking. *KSII Trans Internet & Inf Syst* 11(10):4929–4947
93. Parate MR, Satpute VR, Bhurchandi KM (2018) Global-patch-hybrid template-based arbitrary object tracking with integral channel features. *Appl Intell* 48(2):300–314
94. Perera P, Oza P, Patel VM (2021) One-class classification: a survey. arXiv preprint [arXiv:2101.03064](https://arxiv.org/abs/2101.03064)
95. Perveen N, Roy D, Mohan CK (2018) Spontaneous expression recognition using universal attribute model. *IEEE Trans Image Process* 27(11):5575–5584
96. Popoola OP, Wang K (2012) Video-based abnormal human behavior recognition-a review. *IEEE Trans Syst Man Cybern. Part C (Appl Rev)* 42(6):865–878
97. Puvvadi UL, Di Benedetto K, Patil A, Kang K-D, Park Y (2015) Cost-effective security support in real-time video surveillance. *IEEE Transactions on Industrial Informatics* 11(6):1457–1465
98. Qiang Y, Fei S, Jiao Y (2021) Anomaly detection based on latent feature training in surveillance scenarios. *IEEE Access* 9:68108–68117
99. Ramachandra B, Jones M, Vatsavai RR (2020) A survey of single-scene video anomaly detection. *IEEE Trans Patt Anal Mach Intell.* <https://doi.org/10.1109/TPAMI.2020.3040591>
100. Ren W, Li G, Sun B, Huang K (2015) Unsupervised kernel learning for abnormal events detection. *Vis Comput* 31(3):245–255
101. Rezaee K, Rezakhani SM, Khosravi MR, Moghimi MK (2021) A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. *Personal Ubiquitous Comput.* <https://doi.org/10.1007/s00779-021-01586-5>
102. Ristea N-C, Madan N, Ionescu RT, Nasrollahi K, Khan FS, Moeslund TB, Shah M (2021) Self-supervised predictive convolutional attentive block for anomaly detection. arXiv preprint [arXiv:2111.09099](https://arxiv.org/abs/2111.09099)
103. Rodrigues TG, Suto K, Nishiyama H, Kato N (2016) Hybrid method for minimizing service delay in edge cloud computing through vm migration and transmission power control. *IEEE Trans Comput* 66(5):810–819
104. Roy D, Murty KSR, Mohan CK (2018) Unsupervised universal attribute modeling for action recognition. *IEEE Trans Multimed* 21(7):1672–1680
105. Roy D, Ishizaka T, Mohan CK, Fukuda A (2020) Detection of collision-prone vehicle behavior at intersections using siamese interaction lstm. *IEEE Trans Intell Transp Syst.* <https://doi.org/10.1109/TITS.2020.3031984>
106. Sabih M, Vishwakarma DK (2021) Crowd anomaly detection with lstms using optical features and domain knowledge for improved inferring. *The Visual Computer*, pp 1–12
107. Sabokrou M, Fayyaz M, Fathy M, Klette R (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
108. Sabokrou M, Fayyaz M, Fathy M, Moayed Z, Klette R (2018) Deep-anomaly: fully convolutional neural network for fast anomaly detection in crowded scenes. *Comput Vis Image Understanding* 172:88–97
109. Sabokrou M, Fathy M, Zhao G, Adeli E (2020) Deep end-to-end one-class classifier. *IEEE Trans Neural Netw Learning Syst* 32(2):675–684
110. Saligrama V, Chen Z (2012) Video anomaly detection based on local statistical aggregates. In 2012 IEEE Conference on computer vision and pattern recognition pp 2112–2119, IEEE,
111. Samuel DJ, Cuzzolin F (2021) Svd-gan for real-time unsupervised video anomaly detection
112. Sarker MI, Losada-Gutiérrez C, Marrón-Romera M, Fuentes-Jiménez D, Luengo-Sánchez S (2021) Semi-supervised anomaly detection in video-surveillance scenes in the wild. *Sensors* 21(12):3993
113. Saypadith S, Onoye T (2021) An approach to detect anomaly in video using deep generative network. *IEEE Access* 9:150903–150910
114. Schneible J, Lu A (2017) Anomaly detection on the edge. In MILCOM 2017-2017 IEEE Military communications conference (MILCOM), IEEE pp 678–682
115. Schneible J, Lu A (2017) Anomaly detection on the edge. In: MILCOM 2017 - 2017 IEEE Military communications conference (MILCOM), pp 678–682
116. Shi W, Cao J, Zhang Q, Li Y, Xu L (2016) Edge computing: vision and challenges. *IEEE Internet Things J* 3(5):637–646. <https://doi.org/10.1109/JIOT.2016.2579198>
117. Shi Y, Tian Y, Wang Y, Huang T (2017) Sequential deep trajectory descriptor for action recognition with three-stream cnn. *IEEE Trans Multimed* 19(7):1510–1520
118. Shirahama K, Grzegorzec M, Uehara K (2015) Weakly supervised detection of video events using hidden conditional random fields. *Int J Multimed Inf Retr* 4(1):17–32
119. Shobha B, Deepu R (2018) A review on video based vehicle detection, recognition and tracking. In: 2018 3rd International conference on computational systems and information technology for sustainable solutions (CSITSS) pp 183–186, IEEE
120. Shojaei G, Razzazi F (2019) Semi-supervised domain adaptation for pedestrian detection in video surveillance based on maximum independence assumption. *Int J Multimed Inf Retr* 8(4):241–252
121. Singh D, Mohan CK (2017) Graph formulation of video activities for abnormal activity recognition. *Pattern Recognit* 65:265–272
122. Singh D, Mohan CK (2018) Deep spatio-temporal representation for detection of road accidents using stacked autoencoder. *IEEE Trans Intell Transp Syst* 20(3):879–887. <https://doi.org/10.1109/TITS.2018.2835308>
123. Sivaraman S, Trivedi MM (2013) Looking at vehicles on the road: a survey of vision-based vehicle detection, tracking, and behavior analysis. *IEEE Trans Intell Transp Syst* 14(4):1773–1795
124. Sodemann AA, Ross MP, Borghetti BJ (2012) A review of anomaly detection in automated surveillance. In: *IEEE Trans Syst Man Cybern, Part C (Applications and Reviews)* 42(6):1257–1272
125. Song H, Sun C, Wu X, Chen M, Jia Y (2019) Learning normal patterns via adversarial attention-based autoencoder for abnormal event detection in videos. *IEEE Trans Multimed* 22(8):2138–2148

126. Srivastava S, Singh SP (2016) A survey on latency reduction approaches for performance optimization in cloud computing. In: 2016 Second International conference on computational intelligence & communication technology (CICCT), pp 111–115, IEEE
127. Sun H, Yu Y, Sha K, Lou B (2019) mvideo: edge computing based mobile video processing systems. *IEEE Access* 8:11615–11623
128. Sun C, Jia Y, Song H, Wu Y (2020) Adversarial 3d convolutional auto-encoder for abnormal event detection in videos. *IEEE Trans Multimed* 23:3292–305
129. Suresha M, Kuppa S, Raghukumar D (2020) A study on deep learning spatiotemporal models and feature extraction techniques for video understanding. *Int J Multimed Inf Retr* 9(2):81–101
130. Suto K, Miyanabe K, Nishiyama H, Kato N, Ujikawa H, Suzuki K-I (2015) Qoe-guaranteed and power-efficient network operation for cloud radio access network with power over fiber. *IEEE Trans Comput Soc Syst* 2(4):127–136
131. Sze V, Chen Y, Yang T, Emer JS (2017) Efficient processing of deep neural networks: a tutorial and survey. *Proc IEEE* 105(12):2295–2329
132. Tariq S, Farooq H, Jaleel A, Wasif SM et al (2021) Anomaly detection with particle filtering for online video surveillance. *IEEE Access* 9:19457–19468
133. Tian B, Morris BT, Tang M, Liu Y, Yao Y, Gou C, Shen D, Tang S (2014) Hierarchical and networked vehicle surveillance in its: a survey. *IEEE Trans Intell Transp Syst* 16(2):557–580
134. Torres BS, Pedrini H (2018) Detection of complex video events through visual rhythm. *The Visual Computer* 34(2):145–165
135. Tripathi G, Singh K, Vishwakarma DK (2019) Convolutional neural networks for crowd behaviour analysis: a survey. *Vis Comput* 35(5):753–776
136. 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
137. Ullah W, Ullah A, Hussain T, Muhammad K, Heidari AA, Del Ser J, Baik SW, De Albuquerque VHC (2021) Artificial intelligence of things-assisted two-stream neural network for anomaly detection in surveillance big video data. *Future Gener Comput Syst* 129:286–297
138. Ullah W, Ullah A, Hussain T, Khan ZA, Baik SW (2021) An efficient anomaly recognition framework using an attention residual lstm in surveillance videos. *Sensors* 21(8):2811
139. Ullah W, Ullah A, Haq IU, Muhammad K, Sajjad M, Baik SW (2021) Cnn features with bi-directional lstm for real-time anomaly detection in surveillance networks. *Multimed Tools Appl* 80(11):16979–16995
140. Varghese B, Wang N, Barbhuiya S, Kilpatrick P, Nikolopoulos DS (2016) Challenges and opportunities in edge computing. In: 2016 IEEE International conference on smart cloud (Smart-Cloud), IEEE pp. 20–26
141. Vosta S, Yow K-C (2022) A cnn-rnn combined structure for real-world violence detection in surveillance cameras. *Appl Sci* 12(3):1021
142. Wan S, Ding S, Chen C (2022) Edge computing enabled video segmentation for real-time traffic monitoring in internet of vehicles. *Pattern Recognit* 121:108146
143. Wang T, Snoussi H (2014) Detection of abnormal visual events via global optical flow orientation histogram. *IEEE Trans Inf Forensics Secur* 9(6):988–998
144. Wang X, Ning Z, Wang L (2018) Offloading in internet of vehicles: A fog-enabled real-time traffic management system. *IEEE Transactions on Industrial Informatics* 14(10):4568–4578
145. Wang Y, Wang K, Huang H, Miyazaki T, Guo S (2018) Traffic and computation co-offloading with reinforcement learning in fog computing for industrial applications. *IEEE Trans Ind Inf* 15(2):976–986
146. Wang T, Qiao M, Zhu A, Niu Y, Li C, Snoussi H (2018) Abnormal event detection via covariance matrix for optical flow based feature. *Multimed Tools Appl* 77(13):17375–17395
147. Wang C, Dong S, Zhao X, Papanastasiou G, Zhang H, Yang G (2019) Saliencygan: deep learning semisupervised salient object detection in the fog of iot. *IEEE Trans Ind Inf* 16(4):2667–2676
148. Wu P, Liu J, Shen F (2019) A deep one-class neural network for anomalous event detection in complex scenes. *IEEE Trans Neural Netw Learning Syst* 31(7):2609–2622
149. Xu K, Jiang X, Sun T (2018) Anomaly detection based on stacked sparse coding with intraframe classification strategy. *IEEE Trans Multimed* 20(5):1062–1074
150. Xu Y, Du B, Zhang L, Chang S (2019) A low-rank and sparse matrix decomposition-based dictionary reconstruction and anomaly extraction framework for hyperspectral anomaly detection. *IEEE Geosci Remote Sens Lett* 17(7):1248–1252
151. Xu X, Wu Q, Qi L, Dou W, Tsai S-B, Bhuiyan MZA (2020) Trust-aware service offloading for video surveillance in edge computing enabled internet of vehicles. *IEEE Trans Intell Trans Syst* 22(3):1787–1796
152. Xu Z, Li J, Zhang M (2021) A surveillance video real-time analysis system based on edge-cloud and fl-yolo cooperation in coal mine. *IEEE Access* 9:68482–68497
153. Xu R, Nikouei SY, Chen Y, Polunchenko A, Song S, Deng C, Faughnan TR (2018) Real-time human objects tracking for smart surveillance at the edge. In: 2018 IEEE International conference on communications (ICC), pp. 1–6, IEEE
154. Yang L, Cao J, Liang G, Han X (2015) Cost aware service placement and load dispatching in mobile cloud systems. *IEEE Trans Comput* 65(5):1440–1452
155. Yang P, Lyu F, Wu W, Zhang N, Yu L, Shen XS (2019) Edge coordinated query configuration for low-latency and accurate video analytics. *IEEE Trans Ind Inf* 16(7):4855–4864
156. Yang Z, Liu J, Wu P (2021) Bidirectional retrospective generation adversarial network for anomaly detection in videos. *IEEE Access* 9:107842–107857
157. Yu B, Liu Y, Sun Q (2016) A content-adaptively sparse reconstruction method for abnormal events detection with low-rank property. *IEEE Trans Syst Man Cybern Syst* 47(4):704–716
158. Yuan Y, Fang J, Wang Q (2014) Online anomaly detection in crowd scenes via structure analysis. *IEEE Trans Cybern* 45(3):548–561
159. Yuan G, Sun P, Zhao J, Li D, Wang C (2017) A review of moving object trajectory clustering algorithms. *Artif Intell Rev* 47(1):123–144
160. Zahra A, Ghafoor M, Munir K, Ullah A, Ul Abideen Z (2021) Application of region-based video surveillance in smart cities using deep learning. *Multimed Tools Appl*. <https://doi.org/10.1007/s11042-021-11468-w>
161. Zhai X, Liu K, Nash W, Castineira D (2020) Smart autopilot drone system for surface surveillance and anomaly detection via customizable deep neural network. In: International petroleum technology conference, OnePetro
162. Zhang Z, Mei X, Xiao B (2015) Abnormal event detection via compact low-rank sparse learning. *IEEE Intell Syst* 31(2):29–36
163. Zhang Q, Yang LT, Yan Z, Chen Z, Li P (2018) An efficient deep learning model to predict cloud workload for industry informatics. *IEEE Trans Ind Inf* 14(7):3170–3178
164. Zhang J, Xu C, Gao Z, Rodrigues JJ, Albuquerque V (2020) Industrial pervasive edge computing-based intelligence iot for surveillance saliency detection. *IEEE Transactions on Industrial Informatics* 17(7):5012–5020
165. Zhang W, Wang G, Huang M, Wang H, Wen S (2021) Generative adversarial networks for abnormal event detection in videos based on self-attention mechanism. *IEEE Access* 9:124847–124860

166. Zhang Y, Zhou D, Chen S, Gao S, Ma Y (2016) Single-image crowd counting via multi-column convolutional neural network. In: Proceedings of the IEEE conference on computer vision and pattern recognition pp 589–597
167. Zhao Y, Yin Y, Gui G (2020) Lightweight deep learning based intelligent edge surveillance techniques. *IEEE Trans Cogn Commun Netw* 6(4):1146–1154
168. Zheng X, Zhang Y, Zheng Y, Luo F, Lu X (2021) Abnormal event detection by a weakly supervised temporal attention network. *CAAI Trans Intell Technol*
169. Zhou Y, Liu L, Shao L, Mellor M (2017) Fast automatic vehicle annotation for urban traffic surveillance. *IEEE Trans Intell Transp Syst* 19(6):1973–1984
170. Zhou JT, Du J, Zhu H, Peng X, Liu Y, Goh RSM (2019) Anomaly-net: an anomaly detection network for video surveillance. *IEEE Trans Inf Forensics Secur* 14(10):2537–2550

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.