Anomaly Detection in Road Traffic Using Visual Surveillance: A Survey

K. K. SANTHOSH and D. P. DOGRA, Indian Institute of Technology Bhubaneswar, India
P. P. ROY, Indian Institute of Technology, Roorkee, India

Computer vision has evolved in the last decade as a key technology for numerous applications replacing human supervision. Timely detection of traffic violations and abnormal behavior of pedestrians at public places through computer vision and visual surveillance can be highly effective for maintaining traffic order in cities. However, despite a handful of computer vision–based techniques proposed in recent times to understand the traffic violations or other types of on-road anomalies, no methodological survey is available that provides a detailed insight into the classification techniques, learning methods, datasets, and application contexts. Thus, this study aims to investigate the recent visual surveillance–related research on anomaly detection in public places, particularly on road. The study analyzes various vision-guided anomaly detection techniques using a generic framework such that the key technical components can be easily understood. Our survey includes definitions of related terminologies and concepts, judicious classifications of the vision-guided anomaly detection approaches, detailed analysis of anomaly detection methods including deep learning–based methods, descriptions of the relevant datasets with environmental conditions, and types of anomalies. The study also reveals vital gaps in the available datasets and anomaly detection capability in various contexts, and thus gives future directions to the computer vision–guided anomaly detection research. As anomaly detection is an important step in automatic road traffic surveillance, this survey can be a useful resource for interested researchers working on solving various issues of Intelligent Transportation Systems (ITS).

CCS Concepts: • **Computer systems organization** → *Embedded and cyber-physical systems*; • **Computing methodologies** → *Artificial intelligence*; *Machine learning*;

Additional Key Words and Phrases: Learning methods, classification, road traffic analysis

ACM Reference format:
K. K. Santhosh, D. P. Dogra, and P. P. Roy. 2020. Anomaly Detection in Road Traffic Using Visual Surveillance: A Survey. *ACM Comput. Surv.* 53, 6, Article 119 (December 2020), 26 pages.
https://doi.org/10.1145/3417989

1 INTRODUCTION

With the widespread use of surveillance cameras in public places, computer vision–based scene understanding has gained a lot of popularity amongst the Computer Vision (CV) research community. Visual data contains rich information as compared to other information sources such as GPS, mobile location, radar signals, and so forth. Thus, it can play a vital role in detecting/predicting

Authors' addresses: K. K. Santhosh and D. P. Dogra, Indian Institute of Technology Bhubaneswar, School of Electrical Sciences, Bhubaneswar, Odisha, 752050, India; emails: {sk47, dpdogra}@iitbbs.ac.in; P. P. Roy, Indian Institute of Technology, Roorkee, Department of Computer Science and Engineering, Roorkee, India; email: proy.fcs@iitr.ac.in.

0360-0300/2020/12-ART119 $15.00
https://doi.org/10.1145/3417989

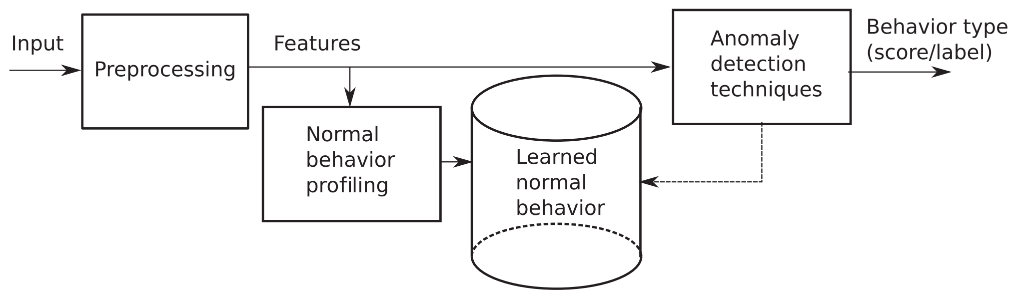

Fig. 1. Overview of a typical anomaly detection scheme. Preprocessing block extracts features/data in the form of descriptors. The normal behavior is represented in abstract form in terms of rules, models, or data repository. Specific anomaly detection techniques are used for detecting anomalies using an anomaly scoring or labeling mechanism.

congestions, accidents, and other anomalies apart from collecting statistical information about the status of road traffic.

Several computer vision–based studies have been conducted focusing on data acquisition [1], feature extraction [2], scene learning [3], activity learning [4], behavioral understanding [5], and so forth. These studies primarily discuss aspects such as scene analysis, video processing techniques, anomaly detection methods, vehicle detection and tracking, multi-camera–based techniques and challenges, activity recognition, traffic monitoring, human behavior analysis, emergency management, event detection, and so forth.

Anomaly detection is a sub-domain of behavior understanding [1] from surveillance scenes. Anomalies are typically aberrations of scene entities (vehicles, human, or the environment) from the normal behavior. With the availability of video feeds from public places, there has been a surge in research outputs on video analysis and anomaly detection [2, 5–7]. Typically, anomaly detection methods learn the normal behavior via training. Anything deviating significantly from the normal behavior can be termed as anomalous. Vehicle presence on walkways, a sudden dispersal of people within a gathering, a person falling suddenly while walking, jaywalking, signal bypassing at a traffic junction, or U-turn of vehicles during red signals are a few examples of anomalies. Anomaly detection frameworks typically use supervised, semi-supervised or unsupervised learning. In this survey, we mainly explore anomaly detection techniques used in road traffic scenarios focusing on *entities* such as vehicles, pedestrian, environment, and their interactions.

We have noted that the scopes of the study should cover the nature of input data and their representations, employed learning methods, types of anomalies, suitability of the techniques in application contexts, and anomaly detection mechanisms. We present this survey from the above perspectives. A typical anomaly detection framework is presented in Figure 1. Usually, anomaly detection systems work by learning the normal data patterns to build a normal profile. Once the normal patterns are learned, anomalies can be detected with the help of established approaches [8, 9]. Output of the system can be a score typically in the form of a metric or a label that notifies whether the data is anomalous or not. Some examples of anomaly detection results are shown in Figure 2.

1.1 Recent Surveys

During the last decade, a few interesting surveys have been published in this field of research. The authors of [3] have explored object detection, tracking, scene modeling, and activity analysis using video trajectories. The study presented in [14] covers vehicle detection, tracking, behavior

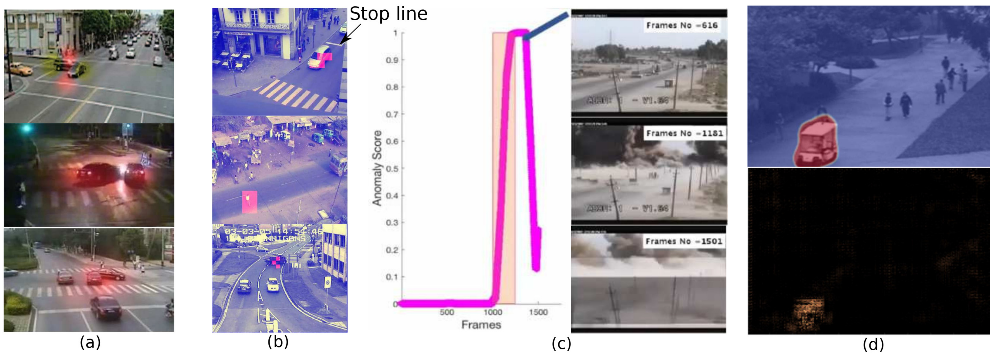

Fig. 2. Visual snapshots of some of the state-of-the-art anomaly detection techniques to present an overview about the survey. (a) Accident detection using Motion Interaction Field (MIF) [10]. (b) Anomaly detection using topic-based models [11]. The top row shows a vehicle that crossed the stop line, the middle row represents a jaywalking scenario, and the bottom row represents a vehicle taking an unusual turn. (c) Real world anomaly detection using multiple instance learning (MIL) [12]. The anomaly detection is measured using an anomaly score in explosion scene. (d) Presence of a vehicle on a walkway detected using spatio-temporal adversarial networks (STAN) [13]. The top row represents the anomaly visualization from the generator and the bottom row represents the anomaly visualization from the discriminator.

understanding, and incident detection from the purview of intelligent transportation systems (ITS). The authors of [15] have conducted an in-depth study of traffic analysis frameworks under different taxonomies with pointers at integrating information from multiple sensors. The review presented in [2] is possibly the first work covering anomaly detection techniques. It covers sensors, entities, feature extraction methods, learning methods, and scene modeling to detect anomalies. An object-based approach from the perspective of vehicle mounted sensors has been presented in [5] with a focus on object detection, tracking, and behavior analysis. The Mmulti-camera study presented in [16] covers the researches related to surveillance in multi-camera setups. The authors of [17] discuss events, which are considered as a subset of anomalous events, requiring immediate attention, occurring unintentionally, abruptly, and unexpectedly. The research presented in [18] discusses safety, security, and law enforcement–related applications from the computer vision perspective. The review presented in [4] discusses the elements of human activity and behavioral understanding frameworks. The authors of [19] present the researches on human behavioral understanding through actions and interactions of human entities. Intelligent video systems covering the analytics aspect has been studied in [20]. Surveillance systems with specific application areas have been presented in [21]. The authors of [1] systematically divide road traffic analysis into four layers, namely, image acquisition, dynamic and static attribute extraction, behavioral understanding, and ITS services. Datasets used for anomaly detections have been covered in [22]. Traffic monitoring using different types of sensors has been discussed in [23]. Algorithms used for spatio-temporal point detections and their applications in vision domain have been covered in [24]. Traffic entities have been studied from the perspective of safety in [6]. The authors of [25] explore studies on video trajectory–based analysis and applications. The authors of [26] discuss various ways of handling emergency situations by assessing the risks, preparedness, response, recovery, and mitigation using the extracted information from the visual features with the help of various learning mechanisms. In [7], the authors have presented anomalous human behavior recognition work with a focus on behavior representation and modeling, feature extraction techniques, classification and behavior modeling frameworks, performance evaluation techniques, and datasets with examples of video surveillance systems. Table 1 summarizes the

Table 1. Surveys on Computer Vision–Based Methods in Surveillance

Ref.	Focus	Explored research areas
Morris (2008) [3]	Video trajectory–based scene analysis	Scene modeling methods; applications of scene modeling; path learning approaches; activity analysis.
Tian (2011) [14]	Video processing techniques applied for traffic monitoring	Traffic parameter collection; traffic incident detection, vehicle detection methods; vehicle tracking and algorithms, model-based classification, region, deformable template, and feature study; traffic incident detection and behavior understanding.
Buch (2011) [15]	Video analytics system for urban traffic	Applications of video analytics; analytics system components; foreground segmentation techniques; top-down and bottom-up vehicle classification techniques; tracking methods; classification of traffic analytic systems.
Sodemann (2012) [2]	Anomaly detection	Study on sensors; learning methods; classification algorithms.
Sivaraman (2013) [5]	Vision-based vehicle detection, tracking, and behavior analysis	Sensors classification; vehicle detection; vehicle tracking; behavior analysis; future direction of vehicle detection, tracking, their on-road behavior and public benchmarks.
Wang (2013) [16]	Multi-camera–based surveillance	Multi-camera calibration; topology computation; multi-camera object tracking; object re-identification; multi-camera activity analysis; cooperative video surveillance using active and static cameras; background modeling and object tracking with active cameras.
Suriani (2013) [17]	Abrupt event detection	Human-centered, vehicle-centered, and small area–centered studies; methods of detection.
Loce (2013) [18]	Traffic management	Vehicle mounted camera-based safety applications; efficiency studies; security management; law enforcement methods.
Vishwakarma (2013) [4]	Human activity recognition and behavior analysis	Application areas; object detection methods; object tracking methods; action recognition techniques; human behavior understanding; dataset description.
Borges (2013) [19]	Human behavior analysis	Human detection methods; action recognition approaches; interaction recognition; datasets description.
Liu (2013) [20]	Intelligent video systems and analytics	Video systems; analytics; employed methods; applications areas.
Zablocki (2013) [21]	Characteristics of intelligent video surveillance systems	System classification; anomaly detection, identification and warning/alarming systems; vehicle detection, traffic and parking lot analysis systems; object counting systems; integrated camera view handling systems; privacy preserving systems; cloud-based systems.
Tian (2015) [1]	Vehicle surveillance	Dynamic and static attribute extraction; behavior understanding; image acquisition; ITS service study.
Patil (2016) [22]	Video datasets for anomaly detection	Dataset classification as traffic, subway, panic-driven, pedestrian, abnormal activity, campus, train, sea, crowd.
Datondji (2016) [23]	Traffic monitoring at intersections	Camera-based classification; Vehicle sensing; Challenges; Vehicle detection methods; vehicle representation and tracking approaches; vehicle tracking algorithms; monitoring systems; vehicle tracking; vehicle behavior analysis.
Li (2017) [24]	Spatio-temporal interest point (STIP) detection algorithms	STIPs algorithms; detection challenges; applications.
Shirazi (2017) [6]	Intersections analysis from safety perspective	Vehicular behavior; driver behavior; pedestrian behavior; safety assessment; intersection safety systems.
Ahmed (2018) [25]	Trajectory-based analysis	Trajectory analysis; clustering algorithms; event detection methods and learning procedures; localization of abnormal events; video summarization and synopsis generation.
Lopez-Fuentes (2018) [26]	Emergency management using computer vision	Emergency classification; monitoring objective; acquisition methods; feature extraction algorithms; semantic information extraction using machine learning.
Mabrouk (2018) [7]	Abnormal behavior recognition	Behavior representation; anomalous behavior recognition methods; performance evaluation; existing surveillance systems.

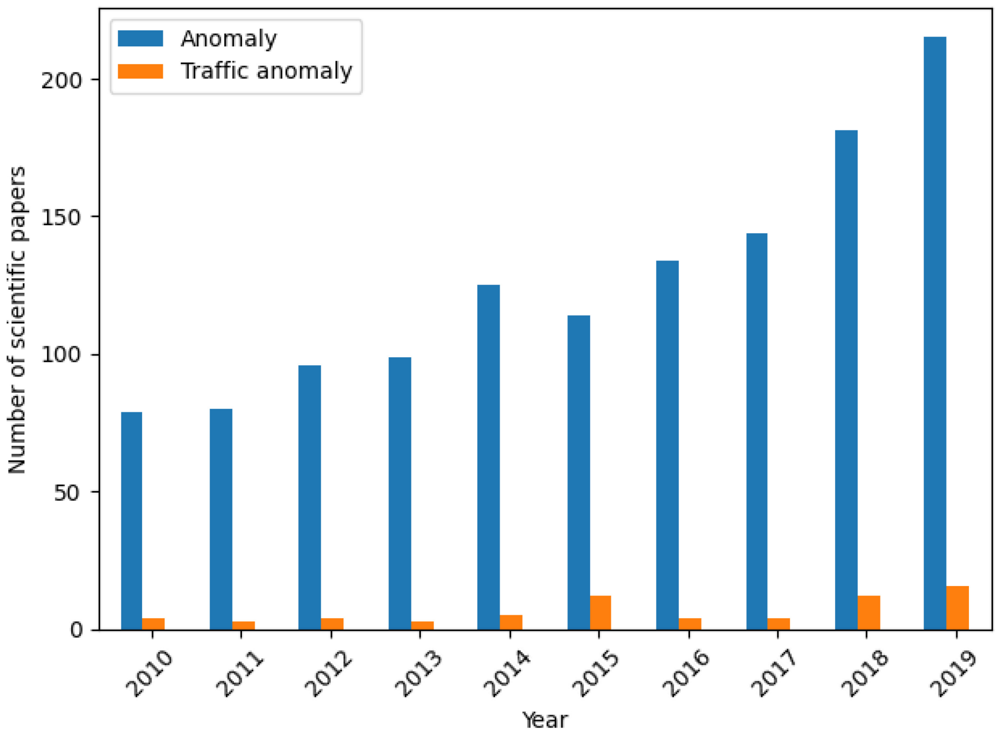

Fig. 3. The number of articles with scopus index published in the last decade.

major computer vision–based studies done during the last decade. In our survey, we particularly focus on the studies on anomaly detection that are relevant on road traffic scenarios.

The number of scientific papers on road traffic anomalies is relatively small as compared to the computer vision–based anomaly detection studies as can be seen from the trends shown in Figure 3. This survey can be a good starting point for a researcher of this domain.

1.2 Contributions of this Survey

Anomalies are contextual in nature. The assumptions used in anomaly detection cannot be applied universally across different traffic scenarios. We analyze the capabilities of anomaly detection methods used in road traffic surveillance from the perspective of data. In the process, we categorize the methods according to scene representation, employed features, used models, and approaches. The study introduces the relevant technologies and provides an end-to-end perspective of the anomaly detection method with several examples of the learning mechanisms, employed detection methods, applied anomaly scenes, types of anomalies detected, and so forth. The following are the contributions of the article:

- (i) To the best of our knowledge, this is the first comprehensive survey that focuses specifically on anomaly detection applied on road traffic with challenges and solutions.
- (ii) It also presents a structured study of various traffic anomaly detection methods by representing the system through a generic framework, thus enhancing its readership base.
- (iii) This study collates all the relevant road traffic anomaly datasets and highlights the gaps that can be exploited by future researchers.

(iv) It also does a critical analysis of the current methods by highlighting the weaknesses that will essentially help shape the future anomaly detection studies.
(v) Finally, the recent advancement in machine learning (ML) and hardware has enabled the CV community to come up with various techniques to detect traffic anomalies. This study summarizes the shift in focus from classical machine learning to deep learning.

The rest of the article is organized as follows. First, the background and the terminologies used in the article are introduced in Section 2.1. Anomaly detection–related visual scene learning methods are presented in Section 2.2. Anomaly detection approaches and classification are elaborated in Section 2.3. Features used for anomaly detection and application areas are presented in Sections 2.4 and 2.7, respectively. A critical analysis of the existing methods followed by discussions on the challenges and future possibilities of anomaly detection are presented in Section 3. Section 4 discusses the recent shift towards deep learning–based methods from classical ML and its relevance in anomaly detection–related research. We conclude the article in Section 5.

2 COMPUTER VISION GUIDED ANOMALY DETECTION STUDIES

2.1 Background and Terminologies

Features are assumed as data in the present context and are represented in the form of feature descriptors. Data typically occupy a position in a multi-dimensional space depending on the feature descriptor length.

Anomalies are data patterns that do not conform to a well-defined notion of normal behavior [27]. There have been other synonyms of anomalies such as outliers, a novelty in various application areas [28]. In this article, we use anomaly or outlier in the subsequent part.

2.1.1 Anomaly Classification. Traditionally, anomalies are classified as *point* anomalies [29–31], *contextual* anomalies [32, 33], and *collective* anomalies [34, 35]. Data correspond to a point anomaly if they are far away from the usual distribution. For example, a non-moving car on a busy road can be termed as a point anomaly. Contextual anomalies correspond to data that may be termed normal in a different context. For example, in slow moving traffic, if a biker rides faster as compared to others, we may term it as anomaly. Conversely, in a less dense road it may be a normal behavior. A group of data instances together may cause anomaly even though individually they may be normal. For example, a group of people dispersing within a short span of time can be termed as collective anomaly.

In the context of visual surveillance, it is common to see anomalies classified as *local* and *global* anomalies [11, 36, 37]. Global anomalies can be present in a frame or a segment of the video without specifying where exactly it has happened [36]. Local anomalies usually happen within in a specific area of the scene, but may be missed by global anomaly detection algorithms [11, 37]. Some methods can detect both global and local anomalies [35, 38–40].

2.1.2 Challenges and Scope of Study. The key challenges in anomaly detection are as follows: (i) defining a representative normal region, (ii) boundaries between the normal and anomalous regions may not be crisp or well defined, (iii) the notion of anomaly is not same in all application contexts, (iv) limited availability of data for training and validation, (v) data is often noisy due to inaccurate sensing, and (vi) normal behavior evolves over time.

We have done this survey based on the studies conducted on videos captured through a static camera. The key reasons are as follows: (i) We want to discuss anomaly from a first person's view. This study can be a good starting point for new researchers working in the area of traffic surveillance. (ii) Static cameras are frequently used for public place surveillance. Hence, a survey on this can help the solution developers to plan for efficient implementations. (iii) The challenges

Table 2. Broad Categorization Based on Learning Methods

Learning method	**Assessment**
Supervised [13, 42–46]	**Adv**: Useful when the number of classes are known *a priori* and decision boundary is clear between classes. **Disadv**: Manual labeling; retraining required when defined classes change over time; difficulty in classifying unforeseen anomalies; chance of overfitting when training data is less in neural network–based systems.
Unsupervised [29, 47–50]	**Adv**: No requirement for labeling data; useful in detecting time-sensitive anomalies. **Disadv:** Anomaly detection is difficult when nature of data changes with time; works on the assumption that anomalies are rare as compared to normal data.
Semi-supervised [11, 51–53]	**Adv**: Only weak labeling of data is required; useful when the anomalous behavior evolves over time. **Disadv**: Manual intervention may be required in case of uncertainties; the criteria of uncertainty needs to be clearly defined in the solution design.
Hybrid [13, 37, 54]	**Adv**: Can exploit advantages of different learning methods. **Disadv**: Detection of anomalies can involve complex steps.

involved in multi-camera anomaly detection [16] include camera calibration, inference of topology, and multi-camera tracking. Therefore, the focus often shifts from the actual task, e.g., anomaly detection. Anomaly detection using multiple cameras includes additional challenges and the frameworks can be completely different [41]. (iv) It can help the future researchers to understand the technical gaps of the detection mechanisms.

2.2 Learning Methods

Learning the normal behavior is not only relevant for anomaly detection, but also for diverse use cases. Pattern analysis [79], classification [74], prediction [80], and behavior analysis [81] are a few among them.

Learning methods can be classified as *supervised*, *unsupervised*, *semi-supervised*, or *hybrid*. In supervised learning, the normal profile is built using labeled data [12]. It is typically applied for classification and regression-related applications. In unsupervised learning, normal profile is structured from the relationships between elements of the unlabeled dataset [79]. Semi-supervised learning primarily uses unlabeled data with some supervision with a small amount of labeled data for specifying example classes known *a priori* [52]. If learning happens through interactive labeling of data as and when the label info is available, such learning is called active learning [82, 83]. Such methods are used when unlabeled data are abundant and manual labeling is expensive. Hybrid methods [13] employ a combination of the above methods for understanding different characteristics of the data. Some of the important works are summarized in Table 2.

Learned models are not only used in feature extraction, but are also used in object detection [84], classification [85], activity recognition [86], segmentation [87], anomaly detection [88], and so forth. Table 3 presents some important learning methods used in anomaly detection.

2.3 Anomaly Detection Approaches

As anomalies usually correspond to deviation from the normal behavior, the way the problem is formulated and the underlying features decide type of anomalies. Thus, the approaches do not limit the capability to identify the type of an anomaly. Furthermore, the anomaly detection approaches can be classified as depicted in Figure 4.

2.3.1 Statistical Model–Based. In this approach, statistical methods are used in general to learn the parameters of the model as they try to fit the data using a stochastic process. Anomalous

Table 3. Examples of Learning Methods Used in Anomaly Detection

Learning Method	Assessment
Supervised:	
Hidden Markov Model (HMM) [55]: A supervised statistical Markov model where the system modeled is assumed to be a Markov process with hidden states [56].	**Pros**: Highly suitable for sequence learning and prediction. **Cons**: Anomaly detection capability depends on feature and model states.
Support Vector Machine (SVM) [57] : A representation of data points in space, mapped such that separate categories are divided by a clear separation between them [36].	**Pros**: Useful in anomaly classification and even works for unforeseen anomaly detection. **Cons**: Anomaly detection depends on the chosen feature.
Gaussian Regression (GR) [58]: A generic supervised learning method designed to solve regression and probabilistic classification problems [35, 59].	**Pros**: Data uncertainty is incorporated in the model; It has the flexibility to use prior knowledge. **Cons**: Performance is highly dependent on the feature representation and choice of the kernel.
Convolutional Neural Networks (CNNs) [60]: A class of deep neural networks, applied usually to analyze visual imagery using convolution [54].	**Pros**: Automatic extraction of semantic features from inputs. **Cons**: Requirement of large number of labeled training data.
Multiple Instance Learning (MIL) [61]: The learning from a set of labeled bags, each containing many instances instead of using individually labeled instances [12, 37].	**Pros**: Useful in the detection of unknown event instances. **Cons**: Anomaly detection depends highly on the problem formulation and the representation of instances.
Long short-term memory (LSTM) networks [62]: A special kind of recurrent neural network (RNN) used in time series applications [54, 63].	**Pros**: Neural network variant for sequence prediction without vanishing gradient problem; real-time prediction and localization of anomalies. **Cons**: Anomaly prediction on video sequence depends on the training data volume.
Fast Region–based–CNN (Fast R-CNN) [64]: A higher variation of neural Deep Neural Networks (DNNs) that works efficiently in object classification over conventional CNNs [44].	**Pros**: Faster feature extraction than CNN due to reduced search space. **Cons**: Requirement of large number of labeled input data.
Unsupervised:	
Latent Dirichlet Allocation (LDA) [65]: A topic model using statistical analysis to retrieve underlying topic distribution in documents [31].	**Pros**: Behavior characterization possible without the need to have labeled data and there is no issue of overfitting. **Cons**: Needs the number of topics in advance; anomalies cannot be detected in real-time as there is a requirement for the availability of visual words.
Probabilistic Latent Semantic Analysis (pLSA) [66]: A model for representing co-occurrence information under a probabilistic framework [67].	**Pros**: Useful in offline analysis of traffic data. **Cons**: Needs the number of topics in advance; overfitting is possible on small data.
Hierarchical Dirichlet Process (HDP) [68]: A nonparametric Bayesian approach, built based on LDA, to cluster data [39].	**Pros**: Number of topics need not be known *a priori* and can be learned from data. **Cons**: Cannot perform real-time anomaly detection.
Gaussian Mixture Model (GMM) [69]: A probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters [70, 71] .	**Pros**: Highly effective when the data distribution is Gaussian. **Cons**: Performance depends on the used feature descriptor.
Density-based spatial clustering of applications with noise (DBSCAN) [72]: A density-based nonparametric clustering algorithm used extensively for modeling and learning data patterns [73].	**Pros**: User need not specify the number of clusters in advance. **Cons**: Difficulty in identifying a rare pattern from anomaly.
Fisher kernel method [74]: A function to measure the similarity of two objects on the basis of sets of measurements for each object and a statistical model [51].	**Pros**: Has advantages of both generative and discriminative models while measuring dissimilarity. **Cons**: Training requires both normal and anomalous data instances.
Principal Component Analysis (PCA) [75]: A statistical procedure of orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables [76].	**Pros**: Effective in linear transformation of high dimensional data in a lower dimensional space. **Cons**: Preparation of training data requires separate classifier.
Generative Adversarial Networks (GANs) [77]: Two neural networks (generator and discriminator) contesting with each other to outsmart each other [78].	**Pros**: Adaptive learning algorithm. **Cons**: Anomaly detection capability depends on the formulation of the problem in a GAN setup.
Semi-supervised:	
Autoencoder Model [53]: Learns the model by attempting to reconstruct the input through minimization of the reconstruction error. User feedback is allowed/frame during training.	**Pros**: User feedback reduces false alarm. **Cons**: Accuracy not as good as state-of-the-art.
Hybrid:	
GAN-LSTM [13]: Fake frames required for adversarial learning used in [54] are generated using bidirectional Conv-LSTM.	**Pros**: Takes advantages of both Conv-LSTM and GAN for adaptive learning. **Cons**: Anomaly detection purely depends on the formulation of the setup.

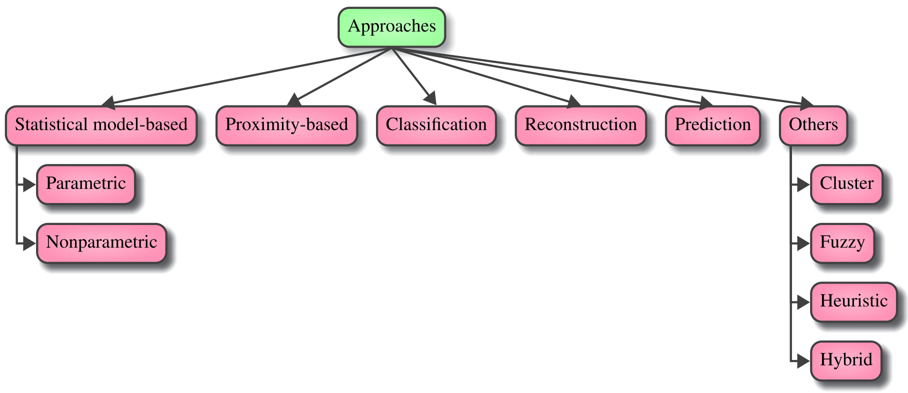

Fig. 4. Classification of the anomaly detection methods based on different approaches.

samples are the data points not generated by the assumed stochastic model. These models can be parametric or nonparametric.

Parametric model–based: Parametric models assume that the normal data is generated through parametric distribution and probability density function. Examples are Gaussian mixture models [70], regression models [35], and so forth. Deep Neural Networks (DNNs) can also be categorized under parametric models, where the parameters are the weights and biases of the neural networks [13, 54]. However, some researchers consider them as classification approaches [9], while a combination of approaches (statistical, classification, information theoretic, reconstruction based) are used in the anomaly detection. Neural network–based methods typically adopt an information-theoretic approach to reduce cross entropy between expected and the predicted outputs in the model learning [89]. Although a combination of methods have been used in anomaly detection, it would be more appropriate to include them under parametric approaches, since the detection mechanisms primarily revolve around the learning of the model parameters.

Nonparametric model–based: In nonparametric statistical models, the structure is not defined *a priori*, but instead determined dynamically from the data. Examples are histogram-based [90], Dirichlet process mixture models (DPMMs) [47], Bayesian network–based models [83], and so forth. Bayesian network estimates the posterior probability of observing a class label from a set of normal class labels and the anomaly class labels, given a test data instance [83]. The class label with the biggest posterior is regarded as predicted class for the given test instance. Typically, topic model–based anomaly detection methods use Bayesian nonparametric approaches [67, 91].

2.3.2 Proximity-Based. In proximity-based approaches, anomalies are decided by how close they are to their neighbors. Distance or relative density may be used as a criterion for anomaly detection. In distance-based approaches, the assumption is that normal data have dense neighborhood [92]. Density-based approaches compare the density around a point with the density around its local neighbors. The relative density of a point compared to its neighbors is computed as an outlier score [49].

2.3.3 Classification-Based. Classification-based anomaly detection methods assume that a classifier can distinguish between normal and anomalous classes in a given feature space. Class-based anomaly detection techniques can be divided into two categories: one class and multi-class. Multi-class classification-based anomaly detection techniques assume that the training data contain labeled instances of normal and anomalous classes. A data point is assumed anomalous if it falls in the anomalous class [46]. One-class classification (OCC)–based anomaly detection techniques assume that all training data have one label [34, 36, 93]. Such techniques learn a discriminative boundary around the normal instances using a OCC algorithm. Support Vector Machines (SVMs) can be used for anomaly detection in the one-class setting extensively in visual surveillance [27, 36]. Another one-class example is a rule-based approach that learns rules to capture a system's normal behavior [94]. A test instance that is not covered by any such rule, is considered as an anomaly.

2.3.4 Prediction-Based. Prediction-based approaches detect anomaly by calculating the variation between predicted and actual spatio-temporal characteristics of the feature descriptor [95]. HMM and LSTM models rely on such approaches for anomaly detection [54, 56].

2.3.5 Reconstruction-Based. In reconstruction-based techniques, the assumption is, normal data can be embedded into a lower dimensional subspace in which normal instances and anomalies appear differently. Anomaly is measured based on the data reconstruction error. Some of the examples are sparse coding [96, 97], autoencoder [98], and principal component analysis (PCA)–based approaches [49].

2.3.6 Other Approaches. There are two types of clustering approaches. One relies on an assumption that the normal data lie in a cluster, while anomaly data do not get associated with any cluster [73]. The latter type is based on an assumption that normal data instances belong to big and dense clusters, while anomalies belong to either little or small clusters. Fuzzy inference systems take a fuzzy data point and use the rules related to membership and strength at which data point fires to decide whether the data is anomalous or not [99, 100]. Heuristic methods intuitively decide about the feature values, spatial location, and contextual information to decide on anomalies [10, 101, 102]. However, many practical systems do not entirely depend on one technology; rather, hybrid approaches are used for anomaly detection [76, 103, 104].

Table 4 presents some important works from the above mentioned classification.

2.4 Features Used in Anomaly Detection

As mentioned earlier, anomaly detection is essentially done by applying a specific technique on the extracted feature. However, in visual surveillance, primary data is a video which is a sequence of frames. Hence, it is essential to extract the relevant features from the videos as these features become input to the specific technique used in anomaly detection. Broadly, the features can be classified as object-based and non-object–based. The classification is represented in Figure 5. Using object-based features, anomalies can be detected by extracting the objects [106, 117] or trajectories [105, 115, 118]. Objects or trajectories represented in the form of feature descriptors become the data for anomaly detection. In the latter approach, low-level descriptors for pixel or pixel group features, intensities, optical flows, or resultant features from spatio-temporal cubes (STCs) [29, 59] have been used for anomaly detection. Some methods use hybrid features for anomaly detections [98, 108].

The choice of feature plays a key role in the capability of detecting specific anomalies. In some methods, preprocessing essentially involves extracting the foreground information and applying specific techniques for finding objects from the foreground [30, 113]. Also, histograms extracted from the pixel-level features can become inputs to anomaly detection methods [34, 92, 112]. Some methods use detected objects or object trajectories as inputs to the anomaly detection methods [105, 115, 119]. Particle Swarm Optimization [120] is used in [88] to obtain optimized motion descriptor from a set of particles having individual motion characteristics. DNNs extract features automatically and use them for anomaly detection [13, 110].

Table 4. Examples of Anomaly Detection Approaches

Ref.	Approach	Specific techniques	Assessment
[70]	Parametric	Gaussian mixture model	**Pros**: Since real-life data typically follow some statistical distributions, outlier calculation is easier due to the samples low probability of occurrences in the distributions. **Cons**: Outlier calculation is often challenging for high dimensional data. Also, these techniques are not robust as mean and standard deviations are sensitive to outliers.
[83]	Nonparametric	Bayesian nonparametric	**Pros**: Robust due to multiple abnormality detection models; allows user interaction to decide anomalies. **Cons**: Requires long duration videos for effective detection.
[49]	Proximity	Relative-density	**Pros**: Suitable when normal data is highly concentrated around a region. **Cons**: Works based on the assumption that effectiveness depends on the anomaly scoring mechanisms and the distance measures used in calculating the relative density.
[36]	Classification	Support vector machine	**Pros**: Can obtain high accuracy for known anomalies; useful even for anomaly classification. **Cons**: Detection of unforeseen anomalies are difficult.
[98]	Reconstruction	Autoencoder	**Pros**: Good at finding unforeseen anomalies; dependent on the data representation. **Cons**: Requirement for high performance systems such as GPUs; requirement of large volume of training data; detection capability highly depends on the choice of reconstruction loss function while training the Autoencoder; overfitting often causes high false alarms.
[54]	Prediction	LSTM	**Pros**: Highly efficient for video data as it is time-series data; useful for real-time prediction of anomalies. **Cons**: Large volume of training data required; accuracy highly depends on the information contained in the feature.
[73]	Clustering-based	DBSCAN clustering	**Pros**: Unsupervised method. **Cons**: Assumes clusters with less data points are potential anomalies.
[100]	Fuzzy logic–based	Fuzzy theory	**Pros**: Capability to handle uncertainty in data. **Cons**: Effective low traffic density; accuracy depends on the detection accuracy.
[101]	Heuristic	Probabilistic rule–based model	**Pros**: Useful when generic model fails to detect certain kinds of anomaly. **Cons**: Can work only for a certain kind of anomaly detection.
[103]	Hybrid	Clustering, fuzzy logic, and autoencoders	**Pros**: Tries to incorporate advantages of different approaches. **Cons**: Effectiveness depends on the problem formulation and the features used.

Features are typically in the form of vectors, corresponding to the data. The method proposed in [98] uses histograms of oriented gradients (HOGs), histograms of optical flows (HOFs), improved trajectory features [114], and automatic features extracted using DNNs. A mixture of dynamic textures has been used in [30]. Histograms of oriented swarm accelerations (HOSA) coupled with HOGs have been used in learning [88]. The authors of [115] have used 3D-tube representation of trajectories as features using the contextual proximity of neighboring trajectory for learning normal trajectory. In [121], a Fisher vector corresponding to each trajectory obtained using optical flow of the object and its position, has been used. HOF and motion entropy (HOFME) have been used in [92]. In DNN-based systems, high-level features are automatically extracted. Some of the important work using the various aforementioned features are summarized in Table 5.

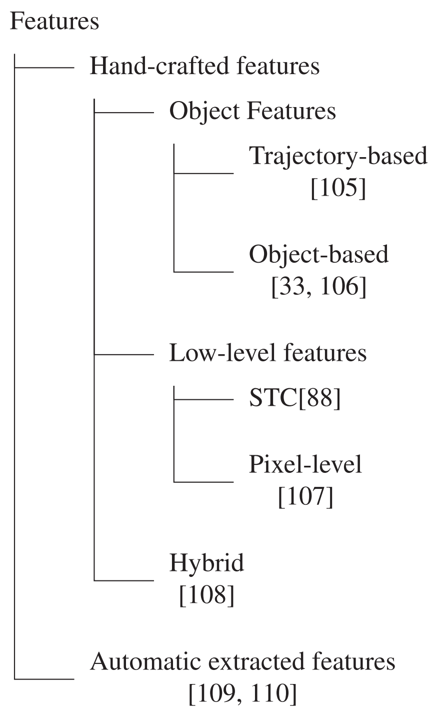

Fig. 5. Overall classification of features used in anomaly detection.

2.5 Detection Methods

Developing an anomaly detection framework typically involves two stages. In the first stage, the normal characteristics of the scene are learned by training a model with the features from the normal videos. Later, features from the test videos are given to the trained model. Based on the selected anomaly criteria, test videos are categorized as normal or anomalous. However, the exact detection mechanisms and the criteria of anomaly are different across these methods. Thus, it is difficult to group them only based on the detection mechanisms. An illustration of the anomaly detection scheme proposed in [122] has been shown in Figure 6. The method detects and localizes anomalies by combining the Temporal CNN pattern (TCP) maps extracted using a deep learning architecture and the optical flow maps.

Furthermore, our study reveals that the anomaly detection mechanisms are diverse in nature and difficult to be compared directly. To summarize the contributions in the area of anomaly detection, we discuss some of the key detection mechanisms in this section.

The authors of [123] use a threshold on maximum likelihood score of test trajectory features on HMM states to decide the anomaly. The LDA-based [31] method uses threshold on posterior probability on class assignment to determine anomaly. In [104], a dictionary model has been used to represent the trajectory classes, and the reconstruction errors of test trajectories have been used for deciding the anomalies. Statistical confidence levels have been used for anomaly detection in the method proposed in [124], which learns activity zones using clustering from the trajectory features. Anomaly detection in [10] has been done using an abnormality score derived from the MIF intensity using an autoregression model. The authors of [125] have constructed motion subspaces using low rank approximation from the motion matrix of the training videos. A threshold on the low rank optimization error has been used for deciding anomaly. The method proposed

Table 5. Representative Work Based on Used Features

Ref.	Features	Learning	Anomaly Detection Mechanisms	Observations
Yang (2013) [37]	Sub-trajectories	MIL	Nearest neighborhood distance obtained from MIL model on the features from test trajectory is compared with a normal distance score to detect anomaly.	**Highlights**: Local anomaly detection capability. **Limitations**: Wrong trajectory partitioning can cause an anomaly miss; trajectories are dependent on tracking algorithm.
Jeong (2014) [31]	Trajectories and pixel velocities	Hybrid (LDA + GMM)	Anomaly if the features extracted from input video have low probability score as compared with dominant traffic flow when applied on the learned model.	**Highlights**: Model remembers the dominant traffic flow allowing it to detect unforeseen motion easily; rare false alarms. **Limitations**: Presence of miss detection and false alarms.
Zhu (2014) [40]	Histogram of optical flow features (HOF)	Sparse coding	Anomaly threshold on reconstruction cost on the features from the test video frames used for global anomaly detection. Local anomaly detection is done using pixel-level ground truth masks.	**Highlights**: Detect both local and global anomalies; can detect unforeseen anomalies. **Limitations**: Region-of-Interest is manually marked for dictionary learning in local anomaly detection; nonadaptive.
Maousavi (2015) [91]	Histogram of Oriented Tracklets (HOT)	LDA	Detection using a threshold on likelihood score when HOT obtained from the test video is applied on the learned model.	**Highlights**: Anomaly detection and localization; training data need not contain anomalies. **Limitations**: Performance depends on the temporal window size; does not work with dense trajectories.
Cheng (2015) [35]	Spatio-temporal interest points (STIPs) [111]	Gaussian regression	Local anomalies are detected using a k-NN-based likelihood threshold with respect to the visual vocabulary of STIP codebook. Global anomalies are detected using global negative log likelihood threshold.	**Highlights**: Global and local anomaly detection. **Limitations**: The method assumes that anomalies are caused by movements, hence cannot detect anomalies caused by stopped vehicles.
Mendel (2016) [54]	CNN-derived features	Conv-LSTM	Detection using a threshold on regularity score (derived from mean square error between predicted and actual output) when the test video is applied on the model.	**Highlights**: Useful for detecting unforeseen anomalies. **Limitations**: No local anomaly detection; false alarms may be present.
Zhang (2016) [112]	Histogram of optical flow	Clustering	Detection using a threshold applied on anomaly score derived using Hamming distance from the learned cluster center.	**Highlights**: Anomaly localization using locality sensitive hashing filters. **Limitations**: May not be suitable for real-time detection due to the usage of PSO-based search for optimal solution.
Lan (2016) [113]	HOG	Heuristic method	Detection using relative difference in the extracted speed-based HOG of test video from the normal videos.	**Highlights**: Abandoned object detection on road. **Limitations**: False alarms may be present.
Hasan (2016) [98]	Handcrafted HOG + HOF [114] and semantic features	Dual Autoencoder model	Detection based on a threshold applied on regularity score when the features from test video are applied on the learned model.	**Highlights**: Global anomaly detection. **Limitations**: High false alarms; non-anomalous training data preparation from real-life videos can be difficult; requirement for GPUs; localization through visual inspection of heat map.
Hinami (2017) [44]	Deep features from CNN	Multi-test Fast R-CNN	Detection using a threshold on anomaly score based on Nearest Neighbor (NN)+OCSVM+KDE when the test video is applied on the learned model.	**Highlights**: Environment-dependent anomaly detection; abnormal event recounting concept introduced. **Limitations**: False alarms; failures in recounting in some cases.
Wen (2017) [71]	Object velocity and direction	GMM	Detection using a threshold on probability score obtained once the object features are applied on the learned model.	**Highlights**: Specifically for speed variation detection; real-time performance. **Limitations**: Limited applications.
Lin (2017) [115]	Droplet vector	SVM	Test trajectory's droplet vector is used to derive a measure (in terms of droplet size and direction) and is applied on the OCSVM trained using normal trajectories to obtain anomaly.	**Highlights**: Contextual anomalies using trajectories. **Limitations**: Offline detection.
Colque (2017) [92]	HOFME	Histogram-based model	Detection using Nearest neighbor–based anomaly threshold when HOFME from the test video applied on the trained model.	**Highlights**: A novel feature descriptor HOFME. **Limitations**: Presence of false alarms.
Lee (2018) [13]	Real and Fake frames	GAN	Detection using abnormality score derived using the losses of the generator and discriminator when the test frames are applied on the learned GAN model.	**Highlights**: Detect anomalies containing complex motion and frequent occlusions; adaptive model. **Limitations**: Corrupted frames may lead to false alarms.
Nguyen (2019) [116]	Deep features	Conv-AE + U-Net	Patch-based anomaly score derived from the input and optical flow frame reconstruction used for deciding anomaly.	**Highlights**: Noise resilient; no manual feature extraction. **Limitations**: Presence of false alarms.

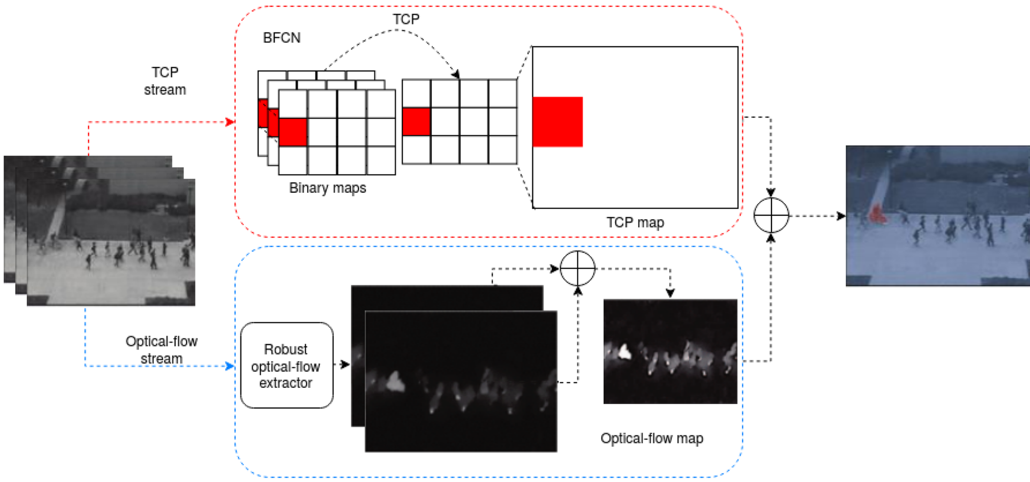

Fig. 6. Depiction of the anomaly detection scheme proposed in [122]. CNN-based binary map is extracted from the input frame sequence. Temporal CNN pattern (TCP) is calculated from the map. Motion features from the TCP and optical flow methods are combined to detect anomalies.

in [35] uses Gaussian process regression using 3D interest point features and STIP to create a likelihood score for deciding anomaly. Appearance, motion, and joint features extracted using a stacked autoencoder have been used in [48] with the help of one-class SVM. The authors of [67] propose a mixture of methods using pLSA, LDA, Fully Sparse Topic Models (FSTM), and Nonparametric Model (NPM) by considering quantized optical flow directions in square cells of a frame as visual words. The topic models such as pLSA, LDA, and FSTM use likelihood measure for anomaly decisions. The NPM-based method uses sparse reconstruction cost for determining anomaly. The abnormality detection algorithm in [83] works by projecting the optical flow–based feature vector from the video clips to the residual space using singular value decomposition. Obtained residual error has been used for anomaly detection. The authors of [11] use the HOG-HOF descriptors as visual words on a pLSA model. The detection uses a projection algorithm in which a test video clip is considered abnormal if the number of anomalous words are more than a predefined threshold.

In [45], the normal and abnormal video classes are trained using optical flow–based spatial-temporal volumes of interest (SVOI) using CNN classifier. Anomalies are detected based on the classifier results. Generative Adversarial Networks (GANs) have been used in [95] to a predict future frame from a continuous set of preceding frames with the help of deep features and optical flow features. The difference between the predicted future frame and the actual frame determines whether a frame is normal or abnormal. The authors of [140] use Region Association Graph (RAG) to learn the activity regions of a scene from the dynamic trajectory features. Anomaly of a test trajectory is decided based on the classification results obtained from the trained SVM. Reconstruction loss on the dictionary model has been used in [141] and [51] using the cuboid feature from MIF and Fisher kernel of trajectory, respectively. The authors of [39] use multiple HDP and code books to represent each region using superpixel-generated spatio-temporal volumes. A confidence score derived from reconstruction loss of the video clip has been used to decide anomaly. Deep MIL model–based [12] uses classification based on the ranking loss to decide anomaly.

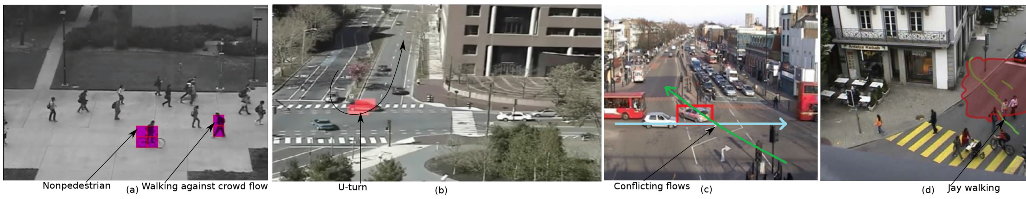

Fig. 7. Depiction of some examples of anomaly detections. (a) Detection of nonpedestrians on road using [92] on Ped2 dataset. It also detects a person walking against the normal crowd flow as anomaly. (b) Detection of illegal U-turn using [31] on MIT traffic dataset. (c) Detection of conflicting flows using [51] on QMUL dataset. (d) Detection of jay walking using [39] applied on Idiap dataset.

2.6 Datasets

A study of some publicly available datasets commonly used for evaluating anomaly detection, is presented in this section. We do not claim that these datasets are only used to evaluate anomaly-related research. These datasets have already been used for other surveillance-related tasks. These have been summarized in Table 6. However, we reveal some crucial gaps in these datasets in the context of road traffic anomaly detection: (i) The number of anomaly detection datasets depicting vehicle violations is low [127, 130, 133, 138] as compared to human-related anomalies [12, 63, 98, 124, 129, 131, 134, 136]. (ii) The number of datasets that present anomalies under different lighting and climatic conditions is very small [10, 42, 135]. (iii) Many of these datasets used in anomaly detection research are local datasets and are not publicly available [104, 125, 142]. (iv) The datasets maintained across do not have any standard format. Although some of these datasets are available with proper annotation and description [98, 129, 131, 134], some are not available [42, 127, 130, 132, 133, 138]. (v) There is a dearth of long duration videos that can potentially capture anomalies occurring during day and night. We hope that the findings can help future researchers to look into these aspects and address them in the future.

2.7 Applied Areas

In this section, we discuss the research work that has been carried out so far focusing on scenes and types of anomalies. Typical scenes are road segments, junctions, parking areas, highways, pedestrian paths, and so forth. A few of the important research works have been summarized in Table 7. We mainly highlight the underlying techniques, applicable scenes, anomaly types, and datasets. Some examples of anomaly detections applied on these datasets are shown in Figure 7.

3 CRITICAL ANALYSIS

This discussion is purely in the context of visual surveillance. Although most of the papers discussed in this survey address anomaly detection, we have observed four key issues with these methods: (i) Benchmark dataset–based comparisons are used to show the effectiveness against the state-of-the-art [43, 78, 93]. It has been observed that performance has gradually improved from statistical approaches [134] to reconstruction-based approaches [63, 78] to prediction-based approaches [95, 143] as can be observed from the Table 8. Although benchmarks may be relevant for comparisons, they may not contain all real-life situations. For example, though anomaly detection works fine on Avenue [98] dataset, it gives higher false alarms when applied on a real dataset QMUL [138] using two of the proposed methods [43, 144]. Therefore, we believe the methods need to be relevant for real-life scenarios and should be applicable to long duration videos. (ii) Secondly, due to the aforementioned trend, a very limited amount of research [12, 42, 46] has been carried out for developing generic techniques applicable to a variety of datasets. (iii) There has been hardly

Table 6. Publicly Available Datasets

Dataset	Scene	Source	Anomaly Types	Assessment
CAVIAR (2004) [126]	Passage	https://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/	Point and collective	**Highlights**: Presence of various anomaly scenarios relevant for road traffic. **Gaps**: Not suitable for illumination agnostic anomaly detection.
i-LIDS (2007) [127]	Road	http://www.eecs.qmul.ac.uk/~andrea/avss2007_d.html	Point	**Highlights**: Video under different illumination conditions. **Gaps**: Only one type of anomaly specified in the dataset.
NGSIM (2007) [128]	Highway	https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm	Point, contextual	**Highlights**: Trajectory dataset; presence of many real-life anomalies; long duration video. **Gaps**: Dataset needs anomaly annotation.
PETS2009 (2009) [129]	Road	http://www.cvg.reading.ac.uk/PETS2009/a.html	Point and collective.	**Highlights**: Many scenarios of people motion from multiple cameras. **Gaps**: Staged scenarios.
Idiap (2009) [130]	Intersection	https://www.idiap.ch/~odobez/RESSOURCES/DataRelease-TrafficJunction.php	Point and contextual	**Highlights**: Long duration videos; many anomalous scenarios. **Gaps**: Lack of anomaly annotation.
UMN (2009) [131]	Public area	https://www.crcv.ucf.edu/projects/Abnormal_Crowd/	Collective	**Highlights**: Specifically suited for collective anomaly. **Gaps**: Staged scenarios.
U-turn (2009) [132]	Intersection	https://sites.google.com/view/ybenezeth/cvpr2009	Point	**Highlights**: Unique dataset containing illegal U-turn. **Gaps**: Not annotated; short duration videos.
MIT (2009) [133]	Intersection	https://www.ee.cuhk.edu.hk/~xgwang/MITtraffic.html	Point and contextual	**Highlights**: Presence of many real-life anomalies; long duration video. **Gaps**: Dataset needs anomaly annotation.
UCSD (2010) [134]	Walkways	http://www.svcl.ucsd.edu/projects/anomaly/dataset.html	Point	**Highlights**: Well described anomaly scenarios. **Gaps**: Very few anomaly scenarios.
Belleview (2010) [135]	Intersection	http://vision.eecs.yorku.ca/research/anomalous-behaviour-data/	Point and contextual	**Highlights**: A challenging dataset containing transition from day to night. **Gaps**: Short duration video; very few anomalies.
Behave (2010) [136]	Road	http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/INTERACTIONS/	Point, collective, and contextual.	**Highlights**: People interaction in various scenarios. **Gaps**: Staged scenarios.
MIT Trajectory (2011) [137]	Car parking area	https://www.ee.cuhk.edu.hk/~xgwang/MITtrajsingle.html	Point	**Highlights**: Trajectories from 5-day recording; large number of trajectories. **Gaps**: Many of the trajectories are truncated; the video is missing.
QMUL (2012) [138]	2 intersections and a roundabout	https://personal.ie.cuhk.edu.hk/~ccloy/downloads_qmul_junction.html.	Point and contextual	**Highlights**: Presence of many real-life anomalies. **Gaps**: Anomaly annotations not in all the videos.
ARENA (2014) [124]	Parking area	http://www.cvg.reading.ac.uk/PETS2014/a.html	Point	**Highlights**: Various anomalous scenarios from multiple sensors available. **Gaps**: All the anomalies not annotated.
Avenue (2016) [98]	Walkway	http://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html	Point	**Highlights**: Well maintained and annotated dataset. **Gaps**: Staged scenarios.
ShanghaiTech (2017) [63]	Road, walkways	https://svip-lab.github.io/dataset/campus_dataset.html	Point, collective, and contextual	**Highlights**: Many real-life scenarios; large number of videos . **Gaps**: Less focus on vehicles.
NVDIA CITY (2017) [139]	Road, intersections	https://www.aicitychallenge.org/	Point and contextual	**Highlights**: Challenging anomalies. **Gaps**: Presence of low-quality videos.
SJTU Trajectories (2017) [115]	Intersection	http://min.sjtu.edu.cn/lwydemo/Trajectory%20analysis.htm	Point and contextual	**Highlights**: Well annotated dataset. **Gaps**: Lack of timing information; contains synthetic scene.
UCF-Crime (2018) [12]	Road, outdoor, and indoor areas	https://webpages.uncc.edu/cchen62/dataset.html	Point	**Highlights**: Presence of 13 real-life anomalies. **Gaps**: Anomaly detailing missing.
IITH Accident (2018) [42]	Road, intersections	https://sites.google.com/site/dineshsinghindian/iith_accident-dataset	Point	**Highlights**: Numerous real-life accident scenarios under various lighting conditions. **Gaps**: Lack of anomaly annotation.

Table 7. Representative Work on Scope of Applied Areas

Ref.	Model	Scene	Anomalies	Dataset
Jeong (2014) [31]	LDA + GMM	Junctions, walkway, roads, public gathering area	Illegal U-turn, vehicle in opposite direction, disordering in the traffic signal, over speed on a pavement, unusual crowds speed, a car stops on a railway	UCSD, UMN, MIT, QMUL and In-house datasets
Mo (2014) [104]	Sparsity model	Junction, road, parking area	Man suddenly falls on floor, vehicle almost hits a pedestrian, car violates the stop sign rule, car fails to yield to oncoming car while turning left, driver backs his car in front of stop sign	i-LIDS, CAVIAR, and In-house dataset, namely, XEROX
Patino (2014) [124]	Heuristic-based	Parking lot, road, intersection	U-turn, vehicle stopping at pedestrian path, person stopping outside zebra passages, person crossing outside zebra passages, loitering and vehicle/person stopping for longer duration	ARENA, CAVIAR, and MIT trajectory dataset
Akoz (2014) [123]	HMM + SVM	Intersection	Collision, nearby passes	NGSIM and CAVIAR
Yun (2014) [10]	MIF	Junction	Accident detection	Car accident
Xia (2015) [125]	Low rank approximation	Road, intersection	Accident detection	In-house dataset
Cheng (2015) [35]	Gaussian regression	Road, walkways, subway, intersection	Non-pedestrians appearing in walkway, chase, fight, run together, traffic interruption, jaywalk, illegal U-turn, strange driving	UCSD (Ped1), Behave and QMUL
Xu (2015) [48]	SVM	Walkways	Non-pedestrians appearing on walkways	UCSD (Ped1 and Ped2)
Nguyen (2015) [83]	HMM + PCA + BN	Junctions	Street fight, loitering, truck—unusual stopping, big truck blocking camera	MIT
Pathak (2015) [11]	pLSA	Junction, highway, roadways	Car stops after the stop-line, jaywalk, vehicle abruptly crossing the road	ldiap, highway (In-house), and i-LIDS
Zhou (2016) [45]	CNN	Junction, walkways, dispersing crowd	U-turn, unexpected presence of vehicles	UCSD, UMN, and U-turn
Liu (2017) [95]	GAN	Roadways, walkways, junction	Throwing objects, loitering and running, non-pedestrians on walkways, presence of people at unexpected area of road	Avenue, UCSD (Ped1 and Ped2), and ShanghaiTech.
Chebiyyam (2017) [140]	SVM + RAG	Parking lot, walkways	Object encircling a particular region, target switching between two or more regions for a sustained period of time	MIT Parking trajectory, Avenue, and a custom dataset.
Yun (2017) [141]	Sparsity model.	Junction, roadways, public gathering area	Car accidents, crowd riots, and uncontrolled fighting	BEHAVE, UMN, and Car accident
Wang (2018) [51]	Sparse dictionary model	Junction, roadways	Car deviating from normal pattern, conflicting patterns, vehicle suddenly interrupting normal pattern, jaywalk, vehicle retrograde, pedestrian near collisions with vehicle	i-LIDS and QMUL
Kaltsa (2018) [39]	HDP	Intersections	Jaywalking, illegal U-turns, wrong vehicle direction, traffic break	QMUL, ldiap, and U-turn
Sultani (2018) [12]	Deep-MIL	Intersection, roadways, walkways	Abuse, arrest, arson, assault, accident, burglary, fighting, robbery	UMN, UCSD (Ped1, Ped2), Avenue, Subway, and Local datasets

Table 8. Performance Evaluation on UCSD (Ped2) Dataset

Ref.	AUC
Mahadevan (2010) [134]	82.9%
Hasan (2016) [98]	85.0%
Ionescu (2017) [146]	82.2%
Luo (2017) [147]	88.1%
Hinami (2017) [44]	92.2%
Luo (2017) [63]	92.2%
Zhao (2017) [148]	91.2%
Ravanbakhsh (2017) [78]	93.5%
Sun (2017) [149]	94.1%
Liu (2018) [95]	95.4%
Ye (2019) [143]	96.8%

Note: Most commonly used evaluation metric is Receiver Operation Characteristic (ROC) by changing the regularity scores threshold gradually. The Area Under Curve (AUC) is accumulated to a scalar for performance evaluation. The higher the AUC, the better the anomaly detection performance.

any illuminating independent research [10, 42] except for accident-type anomaly detection. The problem is not entirely due to the limitations of the learning models. It is equally dependent on the dataset types and lack of illumination independent feature extraction. Possibly with the emergence of DNN-based modeling, we hope to address these issues in the future. An object-based approach might yield better results than histogram-based approaches as humans do not think of pixels and their motion in detecting anomalies, but with mere object motion observations. Researchers can make datasets containing segments of the same scene at varying illumination conditions. (iv) Some approaches remove the background and focus on foreground features for anomaly detections [113]. We believe background information should not be ignored as anomalies also depend on environmental conditions. For example, the chance of accidents on a rainy day is higher than that on a sunny day. Obstructions on roads due to various factors should be taken into consideration while preparing datasets. Very little work has happened on this front [113, 145].

Relevance of Anomaly Detection in Weak Environment: The majority of the existing works have focused on anomaly detection in standard lighting conditions using benchmark datasets [39, 45, 48]. Very few studies have addressed anomaly detection under challenging lighting conditions [42, 135]. For example, the research presented in [42] can detect road accidents in low light conditions. Another study performed in [135] can detect traffic violations at an intersection. However, such studies are more relevant as they can substantially help the authorities in alarming accidents, crimes, hit and run cases, drunken driving, and so forth. Especially at night or in hazy conditions, it is more difficult to do surveillance manually.

3.1 Challenges and Possibilities

Some of the stringent challenges on video-based anomaly detection are as follows:

- **Illumination:** Even though a handful of anomaly detection methods have already been proposed, the number of methods that can handle illumination variations, are limited [67, 70, 125]. This is due to the incapabilities of illuminations agnostic feature extraction from the videos. The criteria or methods used under different illumination conditions can be different for real-life applications.

- **Pose and Perspective:** Often camera angles focusing on the surveillance area can have a substantial impact on the performance of anomaly detection as the appearance of a vehicle may change depending on its distance from the cameras [1]. Although object detection accuracy has increased manifolds using DNN-based methods, there are still challenges in tracking smaller objects. Humans can detect objects at different poses with ease, while machine learning may face difficulties in detecting and tracking the same object under pose variations.
- **Heterogeneous object handling:** Anomaly detection frameworks are largely based on modeling the scene and its entities [12, 31, 35, 37, 54, 56, 59, 67]. However, modeling heterogeneous objects in a scene or learning the movement of heterogeneous objects in a scene can be difficult at times.
- **Sparse vs. Dense:** The methods used for detecting anomalies in sparse and dense conditions are different. Although some of the methods [43, 144] are good at locating anomalies in sparse conditions, dense scene-based methods can generate many false negatives.
- **Curtailed tracks:** Since many anomaly detections are based on vehicle trajectories [25, 37, 108], underlying tracking algorithms are supposed to perform accurately. Even though tracking accuracies have increased in the last decade, many of the existing tracking algorithms do not work under different scenarios [1, 150]. Tracking under occlusion is also another challenge, though humans can easily track them visually.
- **Lack of real-life datasets:** There is a need for real-life datasets to see the effectiveness of anomaly detection techniques.

There are ample scopes and requirements for anomaly detection research based on the gaps discussed earlier. With the advancements in machine learning techniques and affordable hardware, computer vision–based behavior analysis, anomaly detection, and anomaly prediction can leapfrog in the coming years. Deep learning–based hybrid frameworks can handle diverse traffic scenarios. This can also help to build fully automatic traffic analysis frameworks capable of reporting events of interest to the stakeholders.

4 DISCUSSIONS: TOWARDS THE FUTURE

During the last decade, there has been a paradigm shift in machine learning, especially toward DNN-based approaches. It may be observed that some anomaly detection problems have already been solved using deep learning methods [12, 13, 48, 54, 78, 95, 98]. DNN-based studies such as [151] have been successful in extracting illumination independent features. However, there is a lack of studies relevant to traffic anomaly applications. Classical ML often fails, especially for object detection due to pose and perspective of cameras. However, DNN-based solutions such as [84, 152] have been highly accurate in object detection despite their higher computational cost. As object tracking is an important step in many anomaly detection systems, pure deep learning–based methods have not been successful to accurately track objects, especially in dense scenarios. Methods such as [153] employ DNNs for object detection and Kalman filter for object association to make the tracks. However, this also suffers from tracking failures resulting in truncated trajectories in dense and occluded scenarios even though it uses YOLO [84]. During practical implementation of traffic anomaly detection systems using DNNs, access to high computing power can be a challenge. Although most of the industries are supporting academic research through free grants and access to cloud computing resources, unless the hardware prices come down, research outreach may be limited. Figure 8 highlights the strengths and weaknesses of both approaches. However, we cannot completely discard the conventional ML-guided methods as they provide the theoretical foundation for the learning process and have been used together with deep leaning methods

	Classical ML	Deep Learning
Strengths	i) Learning is possible even with small dataset. ii) Does not require high computing power. iii) Working mechanism easily interpretable. iv) Works well with structured data.	i) Works well with unstructured data (e.g. Audio, video, time series). ii) Automatic feature extraction. iii) Provides end-to-end solution.
Weakness	i) Domain expertize required in handcrafting the features. ii) Problem needs to be broken down do find solutions for the sub-problems and the results need to be combined.	i) Requirement of large datasets in learning. ii) Working mechanism is difficult to interpret. iii) More time to train the model. iv) High computing power required.

Fig. 8. Classical ML vs. deep learning systems.

for anomaly detection. Moreover, they are easily implementable using traditional computing platforms.

The present study is relevant for the following reasons: (i) The vehicle volume on roads has increased manifolds in the last decade. Manual monitoring is not only error prone but also can be time-consuming due to the high volume of surveillance data. (ii) With the advancements in machine learning, it has been possible to automatically detect anomalous behaviors using intelligent algorithms. However, our study on the existing literature finds some gaps for developing holistic solutions to detect traffic anomalies. (iii) Timely detection of anomalies is essential in traffic monitoring and control, as misdetections can lead to chaos. (iv) Our study presents the anomaly detection through the lens of features as basic inputs allowing the readers to easily understand the relevance of learning methods, applied scenarios, employed techniques, and the types of anomalies. We believe this survey can provide valuable insight to the researchers aspiring to leverage the effort towards road traffic anomaly detection through visual computing.

5 CONCLUSION

In this article, we have revisited important computer vision–based survey papers. Then we explored various anomaly detection techniques that can be applied for road network entities involving vehicles, people, and their interactions with the environment. We treat anomaly detection by taking data as the primary unit detailing the learning techniques, features used in learning, approaches employed for anomaly detection, and applied scenarios for anomaly detection. We believe the survey presented in this article will help the perspective researchers who intend to carry forward the research in this direction. Some of the immediate scopes can be exploiting the possibilities of implementing anomaly detection in indoor situations, where the visuals may be heavily affected due to occlusion, change in illumination, and geometrical constraints.

REFERENCES

[1] B. Tian, B. T. Morris, M. Tang, Y. Liu, Y. Yao, C. Gou, D. Shen, and S. Tang. 2017. Hierarchical and networked vehicle surveillance in ITS: A survey. *IEEE Transactions on Intelligent Transportation Systems* 18, 1 (2017), 25–48.

[2] A. A. Sodemann, M. P. Ross, and B. J. Borghetti. 2012. A review of anomaly detection in automated surveillance. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (Nov. 2012), 1257–1272.

[3] B. T. Morris and M. M. Trivedi. 2008. A survey of vision-based trajectory learning and analysis for surveillance. *IEEE Transactions on Circuits and Systems for Video Technology* 18, 8 (2008), 1114–1127.

[4] S. Vishwakarma and A. Agrawal. 2013. A survey on activity recognition and behavior understanding in video surveillance. *The Visual Computer* 29, 10 (2013), 983–1009.

[5] S. Sivaraman and M. M. Trivedi. 2013. Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis. *IEEE Transactions on Intelligent Transportation Systems* 14, 4 (2013), 1773–1795.

[6] M. S. Shirazi and B. T. Morris. 2017. Looking at intersections: A survey of intersection monitoring, behavior and safety analysis of recent studies. *IEEE Transactions on Intelligent Transportation Systems* 18, 1 (2017), 4–24.

[7] A. B. Mabrouk and E. Zagrouba. 2018. Abnormal behavior recognition for intelligent video surveillance systems: A review. *Expert Systems with Applications* 91 (2018), 480–491.

[8] S. Omar, A. Ngadi, and H. H. Jebur. 2013. Machine learning techniques for anomaly detection: An overview. *International Journal of Computer Applications* 79, 2 (2013).

[9] X. Li and Z.-M. Cai. 2016. Anomaly detection techniques in surveillance videos. In *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics*. IEEE, 54–59.

[10] K. Yun, H. Jeong, K. M. Yi, S. W. Kim, and J. Y. Choi. 2014. Motion interaction field for accident detection in traffic surveillance video. In *International Conference on Pattern Recognition*. IEEE, 3062–3067.

[11] D. Pathak, A. Sharang, and A. Mukerjee. 2015. Anomaly localization in topic-based analysis of surveillance videos. In *Winter Conference on Applications of Computer Vision*. IEEE, 389–395.

[12] W. Sultani, C. Chen, and M. Shah. 2018. Real-world anomaly detection in surveillance videos. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 6479–6488.

[13] S. Lee, H. G. Kim, and Y. M. Ro. 2018. STAN: Spatio-temporal adversarial networks for abnormal event detection. In *International Conference on Acoustics, Speech and Signal Processing*. IEEE, 1323–1327.

[14] B. Tian, Q. Yao, Y. Gu, K. Wang, and Y. Li. 2011. Video processing techniques for traffic flow monitoring: A survey. In *Conference on Intelligent Transportation Systems*. IEEE, 1103–1108.

[15] N. Buch, S. A. Velastin, and J. Orwell. 2011. A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems* 12, 3 (2011), 920–939.

[16] X. Wang. 2013. Intelligent multi-camera video surveillance: A review. *Pattern Recognition Letters* 34, 1 (2013), 3–19.

[17] N. S. Suriani, A. Hussain, and M. A. Zulkifley. 2013. Sudden event recognition: A survey. *Sensors* 13, 8 (2013), 9966–9998.

[18] R. P. Loce, E. A. Bernal, W. Wu, and R. Bala. 2013. Computer vision in roadway transportation systems: A survey. *Journal of Electronic Imaging* 22, 4 (2013), 41–121.

[19] P. V. K. Borges, N. Conci, and A. Cavallaro. 2013. Video-based human behavior understanding: A survey. *IEEE Transactions on Circuits and Systems for Video Technology* 23, 11 (2013), 1993–2008.

[20] H. Liu, S. Chen, and N. Kubota. 2013. Intelligent video systems and analytics: A survey. *IEEE Transactions on Industrial Informatics* 9, 3 (2013), 1222–1233.

[21] M. Zabłocki, K. Gościewska, D. Frejlichowski, and R. Hofman. 2014. Intelligent video surveillance systems for public spaces—a survey. *Journal of Theoretical and Applied Computer Science* 8, 4 (2014), 13–27.

[22] N. Patil and P. K. Biswas. 2016. A survey of video datasets for anomaly detection in automated surveillance. In *International Symposium on Embedded Computing and System Design*. IEEE, 43–48.

[23] S. R. E. Datondji, Y. Dupuis, P. Subirats, and P. Vasseur. 2016. A survey of vision-based traffic monitoring of road intersections. *IEEE Transactions on Intelligent Transportation Systems* 17, 10 (2016), 2681–2698.

[24] Y. Li, R. Xia, Q. Huang, W. Xie, and X. Li. 2017. Survey of spatio-temporal interest point detection algorithms in video. *IEEE Access* 5 (2017), 10323–10331.

[25] S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy. 2018. Trajectory-based surveillance analysis: A survey. *IEEE Transactions on Circuits and Systems for Video Technology* 29, 7 (2018), 1985–1997.

[26] L. Lopez-Fuentes, J. van de Weijer, M. González-Hidalgo, H. Skinnemoen, and A. D. Bagdanov. 2018. Review on computer vision techniques in emergency situations. *Multimedia Tools and Applications* 77, 13 (July 2018), 17069–17107.

[27] V. Chandola, A. Banerjee, and V. Kumar. 2009. Anomaly detection: A survey. *ACM Computing Surveys* 41, 3 (2009), 1–58.

[28] M. Gupta, J. Gao, C. C. Aggarwal, and J. Han. 2014. Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and Data Engineering* 26, 9 (2014), 2250–2267.

[29] M. J. Roshtkhari and M. D. Levine. 2013. An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions. *Computer Vision and Image Understanding* 117, 10 (2013), 1436–1452.
[30] W. Li, V. Mahadevan, and N. Vasconcelos. 2014. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 1 (2014), 18–32.
[31] H. Jeong, Y. Yoo, K. M. Yi, and J. Y. Choi. 2014. Two-stage online inference model for traffic pattern analysis and anomaly detection. *Machine Vision and Applications* 25, 6 (2014), 1501–1517.
[32] X. Song, M. Wu, C. Jermaine, and S. Ranka. 2007. Conditional anomaly detection. *IEEE Transactions on Knowledge and Data Engineering* 19, 5 (2007), 631–645.
[33] Y. Yuan, J. Fang, and Q. Wang. 2015. Online anomaly detection in crowd scenes via structure analysis. *IEEE Transactions on Cybernetics* 45, 3 (2015), 548–561.
[34] T. Wang and H. Snoussi. 2014. Detection of abnormal visual events via global optical flow orientation histogram. *IEEE Transactions on Information Forensics and Security* 9, 6 (2014), 988–998.
[35] K. Cheng, Y. Chen, and W. Fang. 2015. Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation. *IEEE Transactions on Image Processing* 24, 12 (2015), 5288–5301.
[36] N. Patil and P. K. Biswas. 2016. Global abnormal events detection in surveillance video—A hierarchical approach. In *International Symposium on Embedded Computing and System Design*. IEEE, 217–222.
[37] W. Yang, Y. Gao, and L. Cao. 2013. TRASMIL: A local anomaly detection framework based on trajectory segmentation and multi-instance learning. *Computer Vision and Image Understanding* 117, 10 (2013), 1273–1286.
[38] A. C. B. Abdallah, M. Gouiffès, and L. Lacassagne. 2016. A modular system for global and local abnormal event detection and categorization in videos. *Machine Vision and Applications* 27, 4 (2016), 463–481.
[39] V. Kaltsa, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis. 2018. Multiple hierarchical Dirichlet processes for anomaly detection in traffic. *Computer Vision and Image Understanding* 169 (2018), 28–39.
[40] X. Zhu, J. Liu, J. Wang, C. Li, and H. Lu. 2014. Sparse representation for robust abnormality detection in crowded scenes. *Pattern Recognition* 47, 5 (2014), 1791–1799.
[41] P. Babaei. 2015. Vehicles behavior analysis for abnormality detection by multi-view monitoring. *International Research Journal of Applied and Basic Sciences* 9, 11 (2015), 1929–1936.
[42] D. Singh and C. K. Mohan. 2018. Deep spatio-temporal representation for detection of road accidents using stacked autoencoder. *IEEE Transactions on Intelligent Transportation Systems* 20, 3 (2018), 879–887.
[43] C. Lu, J. Shi, and J. Jia. 2013. Abnormal event detection at 150 FPS in MATLAB. In *International Conference on Computer Vision*. IEEE, 2720–2727.
[44] R. Hinami, T. Mei, and Shin'ichi S.2017. Joint detection and recounting of abnormal events by learning deep generic knowledge.. In *International Conference on Computer Vision*. IEEE, 3619–3627.
[45] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei, and Z. Zhang. 2016. Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes. *Signal Processing: Image Communication* 47 (2016), 358–368.
[46] Y. Chen, Y. Yu, and T. Li. 2016. A vision based traffic accident detection method using extreme learning machine. In *International Conference on Advanced Robotics and Mechatronics*. IEEE, 567–572.
[47] H. Y. T. Ngan. 2015. Outlier detection in traffic data based on the Dirichlet process mixture model. *Intelligent Transport Systems* 9, 7 (2015), 773–781.
[48] D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe. 2015. Learning deep representations of appearance and motion for anomalous event detection. *Arxiv Preprint Arxiv:1510.01553*.
[49] S. W. T. T. Liu, H. Y. T. Ngan, M. K. Ng, and S. J. Simske. 2018. Accumulated relative density outlier detection for large scale traffic data. *Electronic Imaging* 2018, 9 (2018), 1–10.
[50] K. K. Santhosh, D. P. Dogra, and P. P. Roy. 2018. Temporal unknown incremental clustering model for analysis of traffic surveillance videos. *IEEE Transactions on Intelligent Transportation Systems* 20, 5 (2018), 1762–1773.
[51] J. Wang, L. Xia, X. Hu, and Y. Xiao. 2019. Abnormal event detection with semi-supervised sparse topic model. *Neural Computing and Applications* 31, 5 (2019), 1607–1617.
[52] P. Liu, P. Yang, C. Wang, K. Huang, and T. Tan. 2017. A semi-supervised method for surveillance-based visual location recognition. *IEEE Transactions on Cybernetics* 47, 11 (2017), 3719–3732.
[53] S. Bhakat and G. Ramakrishnan. 2019. Anomaly detection in surveillance videos. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*. Association for Computing Machinery, 252–255.
[54] J. R. Medel and A. Savakis. 2016. Anomaly detection in video using predictive convolutional long short-term memory networks. *Arxiv Preprint Arxiv:1612.00390*.
[55] L. E. Baum and T. Petrie. 1966. Statistical inference for probabilistic functions of finite state Markov chains. *The Annals of Mathematical Statistics* 37, 6 (1966), 1554–1563.
[56] T. Wang, M. Qiao, Y. Deng, Y. Zhou, H. Wang, Q. Lyu, and H. Snoussi. 2018. Abnormal event detection based on analysis of movement information of video sequence. *Optik-International Journal for Light and Electron Optics* 152 (2018), 50–60.

[57] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. 1998. Support vector machines. *IEEE Intelligent Systems and their applications* 13, 4 (1998), 18–28.
[58] C. E. Rasmussen. 2003. Gaussian processes in machine learning. In *Summer School on Machine Learning*. Springer, 63–71.
[59] M. Sabokrou, M. Fathy, M. Hoseini, and R. Klette. 2015. Real-time anomaly detection and localization in crowded scenes. In *Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 56–62.
[60] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. 2016. *Deep Learning*. Vol. 1. MIT Press, Cambridge.
[61] B. Babenko, M.-H. Yang, and S. Belongie. 2009. Visual tracking with online multiple instance learning. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 983–990.
[62] S. Hochreiter and J. Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.
[63] W. Luo, W. Liu, and S. Gao. 2017. A revisit of sparse coding based anomaly detection in stacked RNN framework. In *International Conference on Computer Vision*. IEEE, 341–349.
[64] R. Girshick. 2015. Fast r-cnn. In *International Conference on Computer Vision*. IEEE, 1440–1448.
[65] D. M. Blei, A. Y. Ng, and M. I. Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3, (Jan. 2003), 993–1022.
[66] T. Hofmann. 1999. Probabilistic latent semantic indexing. In *Conference on Research and Development in Information Retrieval*. ACM, 50–57.
[67] R. Kaviani, P. Ahmadi, and I. Gholampour. 2015. Automatic accident detection using topic models. In *Iranian Conference on Electrical Engineering*. IEEE, 444–449.
[68] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei. 2004. Sharing clusters among related groups: Hierarchical Dirichlet processes. In *Advances in Neural Information Processing Systems*. MIT Press, 1385–1392.
[69] S. Bengio. 2006. Statistical machine learning from data: Gaussian mixture models. *Lectures* (2006).
[70] Y. Li, W. Liu, and Q. Huang. 2016. Traffic anomaly detection based on image descriptor in videos. *Multimedia Tools and Applications* 75, 5 (2016), 2487–2505.
[71] J. Wen, Z. Lai, Z. Ming, W. K. Wong, and Z. Zhong. 2017. Directional Gaussian model for automatic speeding event detection. *IEEE Transactions on Information Forensics and Security* 12, 10 (2017), 2292–2307.
[72] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, Vol. 96. 226–231.
[73] R. Ranjith, J. J. Athanesious, and V. Vaidehi. 2015. Anomaly detection using DBSCAN clustering technique for traffic video surveillance. In *International Conference on Advanced Computing*. IEEE, 1–6.
[74] F. Perronnin, J. Sánchez, and T. Mensink. 2010. Improving the Fisher kernel for large-scale image classification. In *European Conference on Computer Vision*. Springer, 143–156.
[75] I. Jolliffe. 2011. Principal component analysis. In *International Encyclopedia of Statistical Science*. Springer, 1094–1096.
[76] L.-L. Wang, H. Y. T. Ngan, and N. H. C. Yung. 2018. Automatic incident classification for large-scale traffic data by adaptive boosting SVM. *Information Sciences* 467 (2018), 59–73.
[77] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. Generative adversarial nets. In *Advances in Neural Information Processing Systems*. MIT Press, 2672–2680.
[78] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe. 2017. Abnormal event detection in videos using generative adversarial nets. In *International Conference on Image Processing*. IEEE, 1577–1581.
[79] R. Emonet, J. Varadarajan, and J. M. Odobez. 2014. Temporal analysis of motif mixtures using Dirichlet processes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 1 (2014), 140–156.
[80] B. T. Morris and M. M. Trivedi. 2011. Trajectory learning for activity understanding: Unsupervised, multilevel, and long-term adaptive approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 11 (2011), 2287–2301.
[81] K. K. Santhosh, D. P. Dogra, P. P. Roy, and B. B. Chaudhuri. 2019. Trajectory-based scene understanding using Dirichlet process mixture model. *IEEE Transactions on Cybernetics* (2019), 1–14.
[82] J. Varadarajan, R. Subramanian, N. Ahuja, P. Moulin, and J. Odobez. 2017. Active online anomaly detection using dirichlet process mixture model and Gaussian process classification. In *Winter Conference on Applications of Computer Vision*. IEEE, 615–623.
[83] V. Nguyen, D. Phung, D.-S. Pham, and S. Venkatesh. 2015. Bayesian nonparametric approaches to abnormality detection in video surveillance. *Annals of Data Science* 2, 1 (2015), 21–41.
[84] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. 2016. You only look once: Unified, real-time object detection. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 779–788.
[85] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. 2014. Large-scale video classification with convolutional neural networks. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 1725–1732.
[86] J. C. Nascimento, M. A. T. Figueiredo, and J. S. Marques. 2013. Activity recognition using a mixture of vector fields. *IEEE Transactions on Image Processing* 22, 5 (2013), 1712–1725.

[87] A. Khoreva, F. Galasso, M. Hein, and B. Schiele. 2015. Classifier based graph construction for video segmentation. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 951–960.

[88] V. Kaltsa, A. Briassouli, I. Kompatsiaris, and M. G. Strintzis. 2014. Swarm-based motion features for anomaly detection in crowds. In *International Conference on Image Processing*. IEEE, 2353–2357.

[89] D. P. Kingma and J. Ba. 2014. Adam: A method for stochastic optimization. *Arxiv Preprint Arxiv:1412.6980*.

[90] Y. Zhang, H. Lu, L. Zhang, and X. Ruan. 2016. Combining motion and appearance cues for anomaly detection. *Pattern Recognition* 51 (2016), 443–452.

[91] H. Mousavi, S. Mohammadi, A. Perina, R. Chellali, and V. Mur. 2015. Analyzing tracklets for the detection of abnormal crowd behavior. In *Winter Conference on Applications of Computer Vision*. IEEE, 148–155.

[92] R. V. H. M. Colque, C. Caetano, M. T. L. de Andrade, and W. R. Schwartz. 2017. Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 3 (2017), 673–682.

[93] D. Xu, Y. Yan, E. Ricci, and N. Sebe. 2017. Detecting anomalous events in videos by learning deep representations of appearance and motion. *Computer Vision and Image Understanding* 156 (2017), 117–127.

[94] V. Saligrama and Z. Chen. 2012. Video anomaly detection based on local statistical aggregates. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 2112–2119.

[95] W. Liu, W. Luo, D. Lian, and S. Gao. 2018. Future frame prediction for anomaly detection—A new baseline. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 6536–6545.

[96] Z. Zhang, X. Mei, and B. Xiao. 2016. Abnormal event detection via compact low-rank sparse learning. *IEEE Intelligent Systems* 31, 2 (2016), 29–36.

[97] B. Yu, Y. Liu, and Q. Sun. 2017. A content-adaptively sparse reconstruction method for abnormal events detection with low-rank property. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 47, 4 (2017), 704–716.

[98] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis. 2016. Learning temporal regularity in video sequences. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 733–742.

[99] D. Wijayasekara, O. Linda, M. Manic, and C. G. Rieger. 2014. Mining building energy management system data using fuzzy anomaly detection and linguistic descriptions. *IEEE Transactions on Industrial Informatics* 10, 3 (2014), 1829–1840.

[100] Y. Li, T. Guo, R. Xia, and W. Xie. 2018. Road traffic anomaly detection based on fuzzy theory. *IEEE Access* 6 (2018), 40281–40288.

[101] M.-C. Chang, Y. Wei, N. Song, and S. Lyu. 2018. Video analytics in smart transportation for the AIC '18 challenge. In *Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 61–68.

[102] S. C. Lee and R. Nevatia. 2014. Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system. *Machine Vision and Applications* 25, 1 (2014), 133–143.

[103] Z. Chen, Y. Tian, W. Zeng, and T. Huang. 2015. Detecting abnormal behaviors in surveillance videos based on fuzzy clustering and multiple auto-encoders. In *International Conference on Multimedia and Expo*. IEEE, 1–6.

[104] X. Mo, V. Monga, R. Bala, and Z. Fan. 2014. Adaptive sparse representations for video anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology* 24, 4 (2014), 631–645.

[105] N. B. Ghrab, E. Fendri, and M. Hammami. 2016. Abnormal events detection based on trajectory clustering. In *International Conference on Computer Graphics, Imaging and Visualization*. IEEE, 301–306.

[106] A. Kumar and C. J. Prabhakar. 2017. Vehicle abnormality detection and classification using model based tracking. *International Journal of Advanced Research in Computer Science* 8, 5 (2017).

[107] O. Isupova, D. Kuzin, and L. Mihaylova. 2016. Anomaly detection in video with Bayesian nonparametrics. In *ICMLW*.

[108] S. Coşar, G. Donatiello, V. Bogorny, C. Garate, L. O. Alvares, and F. Brémond. 2017. Toward abnormal trajectory and event detection in video surveillance. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 3 (2017), 683–695.

[109] Z. Fang, F. Fei, Y. Fang, C. Lee, N. Xiong, L. Shu, and S. Chen. 2016. Abnormal event detection in crowded scenes based on deep learning. *Multimedia Tools and Applications* 75, 22 (2016), 14617–14639.

[110] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, and R. Klette. 2018. Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes. *Computer Vision and Image Understanding* 172 (2018), 88–97.

[111] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie. 2005. Behavior recognition via sparse spatio-temporal features. In *International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*. IEEE, 65–72.

[112] Y. Zhang, H. Lu, L. Zhang, X. Ruan, and S. Sakai. 2016. Video anomaly detection based on locality sensitive hashing filters. *Pattern Recognition* 59 (2016), 302–311.

[113] J. Lan, Y. Jiang, G. Fan, D. Yu, and Q. Zhang. 2016. Real-time automatic obstacle detection method for traffic surveillance in urban traffic. *Journal of Signal Processing Systems* 82, 3 (2016), 357–371.

[114] H. Wang and C. Schmid. 2013. Action recognition with improved trajectories. In *International Conference on Computer Vision*. IEEE, 3551–3558.

[115] W. Lin, Y. Zhou, H. Xu, J. Yan, M. Xu, J. Wu, and Z. Liu. 2017. A tube-and-droplet-based approach for representing and analyzing motion trajectories. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 8 (2017), 1489–1503.
[116] T. N. Nguyen and J. Meunier. 2019. Anomaly detection in video sequence with appearance-motion correspondence. In *International Conference on Computer Vision*. IEEE, 1273–1283.
[117] M. K. Lim, S. Tang, and C. S. Chan. 2014. iSurveillance: Intelligent framework for multiple events detection in surveillance videos. *Expert Systems with Applications* 41, 10 (2014), 4704–4715.
[118] F. Mehboob, M. Abbas, R. Jiang, A. Rauf, S. A. Khan, and S. Rehman. 2017. Trajectory based vehicle counting and anomalous event visualization in smart cities. *Cluster Computing* (2017), 1–10.
[119] S. Zhou, W. Shen, D. Zeng, and Z. Zhang. 2015. Unusual event detection in crowded scenes by trajectory analysis. In *International Conference on Acoustics, Speech and Signal Processing*. IEEE, 1300–1304.
[120] J. Kennedy. 2011. Particle swarm optimization. In *Encyclopedia of Machine Learning*. Springer, 760–766.
[121] P. Giannakeris, V. Kaltsa, K. Avgerinakis, A. Briassouli, S. Vrochidis, and I. Kompatsiaris. 2018. Speed estimation and abnormality detection from surveillance cameras. In *Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, 93–99.
[122] H. Xia, T. Li, W. Liu, X. Zhong, and J. Yuan. 2019. Abnormal event detection method in surveillance video based on temporal CNN and sparse optical flow. In *International Conference on Computing and Data Engineering*. ACM, 90–94.
[123] Ö. Aköz and M. E. Karsligil. 2014. Traffic event classification at intersections based on the severity of abnormality. *Machine Vision and Applications* 25, 3 (2014), 613–632.
[124] L. Patino and J. Ferryman. 2014. Multiresolution semantic activity characterisation and abnormality discovery in videos. *Applied Soft Computing* 25 (2014), 485–495.
[125] S. Xia, J. Xiong, Y. Liu, and G. Li. 2015. Vision-based traffic accident detection using matrix approximation. In *Asian Control Conference*. IEEE, 1–5.
[126] CAVIAR. Video Dataset. Retrieved on July 15, 2020 from http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/.
[127] AVSS2007. i-LIDS. Retrieved on July 15, 2020 from http://www.eecs.qmul.ac.uk/ andrea/avss2007_d.html.
[128] NGSIM. Next Generation Simulation Community. Retrieved on July 15, 2020 from https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm.
[129] A. Ellis, A. Shahrokni, and J. M. Ferryman. 2009. PETS2009 and Winter-PETS 2009 results: A combined evaluation. In *International Workshop on Performance Evaluation of Tracking and Surveillance*. IEEE, 1–8.
[130] J. Varadarajan and J.-M. Odobez. 2009. Topic models for scene analysis and abnormality detection. In *International Conference on Computer Vision Workshops*. IEEE, 1338–1345.
[131] R. Mehran, A. Oyama, and M. Shah. 2009. Abnormal crowd behavior detection using social force model. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 935–942.
[132] Y. Benezeth, P.-M. Jodoin, V. Saligrama, and C. Rosenberger. 2009. Abnormal events detection based on spatio-temporal co-occurences. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 2458–2465.
[133] X. Wang, X. Ma, and W. E. L. Grimson. 2009. Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31, 3 (2009), 539–555.
[134] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos. 2010. Anomaly detection in crowded scenes. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 1975–1981.
[135] A. Zaharescu and R. Wildes. 2010. Anomalous behaviour detection using spatiotemporal oriented energies, subset inclusion histogram comparison and event-driven processing. In *European Conference on Computer Vision*. Springer, 563–576.
[136] S. Blunsden and R. B. Fisher. 2010. The BEHAVE video dataset: Ground truthed video for multi-person behavior classification. *Annals of the British Machine Vision Association* 4, 1–12 (2010), 4.
[137] X. Wang, K. T. Ma, G.-W. Ng, and W. E. L. Grimson. 2011. Trajectory analysis and semantic region modeling using nonparametric hierarchical Bayesian models. *International Journal of Computer Vision* 95, 3 (2011), 287–312.
[138] T. Hospedales, S. Gong, and T. Xiang. 2012. Video behaviour mining using a dynamic topic model. *International Journal of Computer Vision* 98, 3 (2012), 303–323.
[139] NVIDIA. NVIDIA AI CITY. Retrieved on July 15, 2020 from https://www.aicitychallenge.org/.
[140] M. Chebiyyam, R. D. Reddy, D. P. Dogra, H. Bhaskar, and L. Mihaylova. 2017. Motion anomaly detection and trajectory analysis in visual surveillance. *Multimedia Tools and Applications* (2017), 1–26.
[141] K. Yun, Y. Yoo, and J. Y. Choi. 2017. Motion interaction field for detection of abnormal interactions. *Machine Vision and Applications* 28, 1–2 (2017), 157–171.
[142] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz. 2008. Robust real-time unusual event detection using multiple fixed-location monitors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 3 (2008), 555–560.

[143] M. Ye, X. Peng, W. Gan, W. Wu, and Y. Qiao. 2019. AnoPCN: Video anomaly detection via deep predictive coding network. In *International Conference on Multimedia*. ACM, 1805–1813.
[144] Y. S. Chong and Y. H. Tay. 2017. Abnormal event detection in videos using spatiotemporal autoencoder. In *International Symposium on Neural Networks*. Springer, 189–196.
[145] D. Danilescu, A. Lodin, L. Grama, and C. Rusu. 2015. Road anomalies detection using basic morphological algorithms. *Carpathian Journal of Electronic and Computer Engineering* 8, 2 (2015), 15.
[146] T. R. Ionescu, S. Smeureanu, B. Alexe, and M. Popescu. 2017. Unmasking the abnormal events in video. In *IEEE International Conference on Computer Vision*. IEEE, 2895–2903.
[147] W. Luo, W. Liu, and S. Gao. 2017. Remembering history with convolutional LSTM for anomaly detection. In *International Conference on Multimedia and Expo*. IEEE, 439–444.
[148] Y. Zhao, B. Deng, C. Shen, Y. Liu, H. Lu, and X.-S. Hua. 2017. Spatio-temporal autoencoder for video anomaly detection. In *International Conference on Multimedia*. ACM, 1933–1941.
[149] Q. Sun, H. Liu, and T. Harada. 2017. Online growing neural gas for anomaly detection in changing surveillance scenes. *Pattern Recognition* 64 (2017), 187–201.
[150] S. Ojha and S. Sakhare. 2015. Image processing techniques for object tracking in video surveillance—A survey. In *International Conference on Pervasive Computing*. IEEE, 1–6.
[151] C. Chen, Q. Chen, J. Xu, and V. Koltun. 2018. Learning to see in the dark. In *Conference on Computer Vision and Pattern Recognition*. IEEE, 3291–3300.
[152] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. 2016. Ssd: Single shot multibox detector. In *European Conference on Computer Vision*. Springer, 21–37.
[153] N. Wojke, A. Bewley, and D. Paulus. 2017. Simple online and realtime tracking with a deep association metric. In *International Conference on Image Processing*. IEEE, 3645–3649.

Received January 2020; revised July 2020; accepted July 2020