A Survey on Urban Anomaly: Description, Detection and Prediction

Urban anomalies may result in loss of life or property if not handled properly. It is of great value for citizens if anomalies can be automatically alerted in their early stage or even predicted before happening. With the recent rise of various sensors located all over cities, modern urban anomaly detection and prediction framework has been forming, which utilizes big urban data and machine learning algorithms to detect and predict urban anomalies automatically. In this survey, we make an exhaustive review on the state-of-the-art research efforts on the urban anomaly. We first give an overview of four main types of urban anomalies, traffic anomaly, unexpected crowds, environment anomaly, and individual anomaly. Next, we summarize various types of urban datasets obtained from diverse devices, trajectory, trip records, CDRs, urban sensors, social media, camera surveillance, and environment data. Subsequently, a comprehensive survey of issues on detecting and predicting techniques for urban anomalies is presented. Finally, open research challenges and future directions as presented as well.

ACM Reference Format:
. 2018. A Survey on Urban Anomaly: Description, Detection and Prediction. 1, 1 (November 2018), 21 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Urban anomalies are typically unusual events that occur in urban environments, such as traffic congestion and unexpected crowd gathering, which may pose tremendous threats to public safety and stability if not timely handled [122]. For example, on January 26, 2017, in Harbin, a big city in China, traffic congestion caused a serial rear-end collision accident where eight people were killed, and thirty-two people were injured. The government then admitted that they did not detect the traffic congestion timely, which caused they could not take immediate actions to prevent the happening of this tragic event. For policymakers and government, detecting anomalies at the early stage or even predicting anomalies before happening is of great value to prevent serious incidents from occurring. Detecting and predicting urban anomalies are also of great importance to improve the quality of life for citizens [38]. For example, traffic jams are the most headache problem for metropolises. A severe traffic jam can bring a great number of economic loss and ruin a good mood on that day. If most traffic jams happened in a city can be predicted, it can further be avoided by notifying people to change their travel routes or transport. In this way, it will save people a lot of time on the commute and improve their quality of life.

On the other hand, the recent rise of smart devices and various kinds of sensors located all over cities lead to an explosion of real-time, high-resolution data produced in the urban environment. These data provide a comprehensive view of urban dynamics, such as understanding city-scale people motion modes [42, 127], inferring land usage and region functions [96, 118] and modeling human activity patterns [28, 35]. By understanding urban dynamics, new approaches to solving urban problems are also becoming research hotspots [126], like discovering traffic problems with vehicle trajectories [18], predicting air quality based on historical monitor station records [128] and diagnosing urban noise problems with public service records [129].

Under this big urban data era, modern urban anomaly detection and prediction framework has been forming as well, which utilize data-driven intelligence to detect and predict urban anomalies automatically [14, 31, 67].

Author's address:

XXXX-XXXX/2018/11-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

Compared with traditional methods which rely on human observations and reports, the modern urban anomaly detection and prediction framework, based on big urban data, is low-cost and efficient [59, 126]. For example, many taxis are equipped with GPS devices which can report precise locations of the car in a second-level sample rate. From these location records, we then can quickly know the trajectories of taxis and further infer the traffic condition in specific areas and detect traffic anomaly by looking at the speed and routing choices of cars.

To better illustrate and formulate the process of extracting anomalous events from big urban data, we define three concepts and show their relationship in Fig. 1.

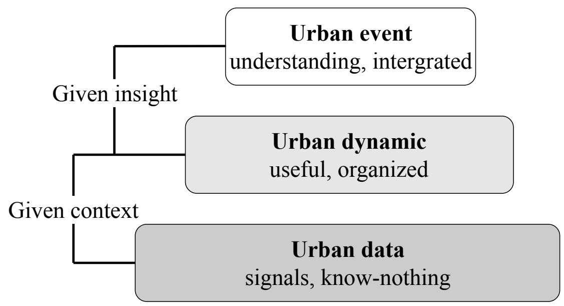

Fig. 1. The relationship among urban data, urban dynamics and urban events.

- **Urban data** are spatial-temporal data produced by mobile devices or distributed sensors in cities. These data are usually associated with timestamps and location tag. The urban data can be in different forms and contain signals about a location in a particular time or time interval. Directly from raw urban data, we usually know nothing.
- **Urban dynamics** refer to the people flows or traffic volume in urban areas. They are the fundamental elements of urban events and a basic description of the condition in a location. Urban dynamics cannot be observed directly but can be inferred from urban data. For example, Call Detail Record (CDR) data, as a kind of urban data, record a user's location when the user makes a call. We can use CDR to estimate the density of residents and trace the flow of people in a city [54].
- **Urban events** are all kinds of social or individual activities that happen in a city. They are the underlying causes of urban dynamics. It is a common fact that anomalous events cause unusual changes of urban dynamics, which is the basic assumption or hypothesis of modern urban anomaly detection and prediction framework.

We then use a soccer match as an example to illustrate the relationships among these three concepts. The holding of a soccer match is an urban event. That match will attract many soccer fans to gather in the stadium and further causes a rapid rise in traffic volume on nearby roads, which is how the urban dynamic changed by urban events. Meanwhile, many people may choose to take a taxi to come and go home after the match. The trip records of taxis are the urban data that we can directly observe. On the other hand, if we explore the taxi trip records, we can find an abnormally significant number of taxis arrive at or leave the stadium in a few hours, which reflects the abnormal change of urban dynamics.

In this way, detecting and predicting anomalous urban events from urban data consists of two steps, i) detecting abnormal changes in urban dynamics based on urban data and ii) identifying the underlying events based on urban dynamics. However, since anomalous urban events are complex and sometimes even not recorded, the second step is hard to conduct in practice. Many algorithms in studied literature solely stop at the first step and do not trace the events that cause the unusual fluctuation of urban dynamics, but we still include them in our survey. Besides, it is worth to mention that although there are many papers on events discovering from Twitter [37, 91], these works only focus on general events instead of urban anomalies. Thus, we do not discuss them in our survey.

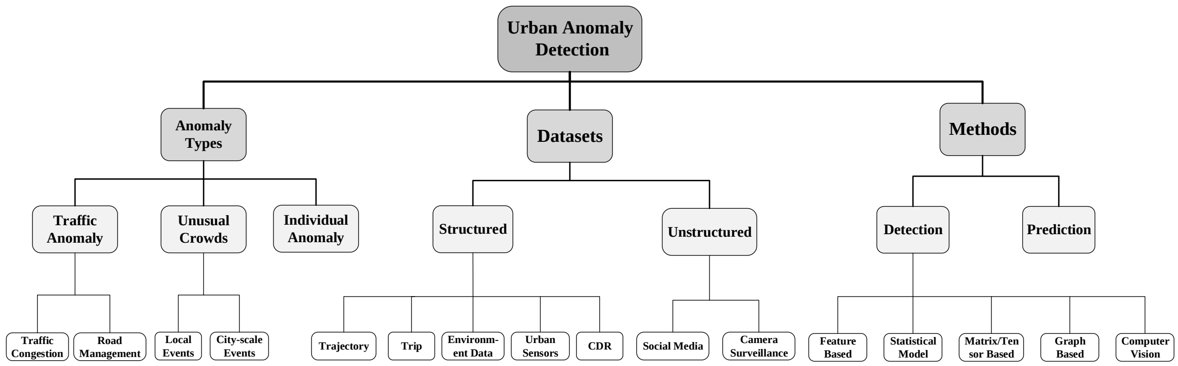

Fig. 2. Classification of urban anomalous events detecting literature.

In this survey, we discussed the following questions. What kinds of urban data are used? What kinds of anomalous events can be detected and predicted? What are the general detection and prediction methods? To answer these three questions, we systematically studied relevant researches in recent years and proposed our classification and summary. Fig. 2 shows the structure of our survey. We overview the related works from three aspects. Firstly, we introduce three types of urban anomalies studied in existing works in section II. We then describe the datasets that are commonly used to detect and predict anomalous events in section III, including structured data and unstructured data. After that, we summarize the algorithms proposed in the related literature based on their main thoughts in section IV. In section V, we conclude the open challenges in urban anomaly filed. In Table 1, we list the literature by the categories mentioned above.

2 URBAN ANOMALOUS EVENTS

The events happening in urban areas can be divided into two parts, normal events, and anomalous events. The normal events refer to regular activities that follow certain patterns or rules, while the anomalous events refer to incidental events that happen by accident. For example, crowds gathering at rush hours around a subway station is a normal event because it follows a regular pattern occurring periodically every weekday. On the other hand, although a pop concert held around can also cause a peak of subway traffic in this station, it should be considered as an anomalous event since it rarely happens and follows an irregular pattern. In this section, we discuss three kinds of urban anomalies that are most commonly studied, i.e., traffic anomaly, unexpected crowds, environment anomaly, and individual anomaly.

2.1 Traffic Anomaly

The traffic is of vital importance for the daily lives of citizens. Detecting and predicting traffic anomalies attracts a lot of researchers [2, 70, 76, 113]. Traffic anomalies mainly have two types. The first is traffic congestion which is usually caused by traffic incidents or traffic overload. The effect of traffic congestion is the slowdown in traffic speed or increase in traffic volume on specific roads, which only last for a short time. The other kind is road

Table 1. Main Features of Urban Anomaly Detection Works. (In Datasets columes, Tj. is Trajectory, Tp. is Trip, US. is Urban Sensor, SM. is Social Media, ER. is Event Records, EV. is Environment Data. In Anomaly Type columes, Tf. is Traffic Anomaly, Cd. is Unexpected Crowds, EA. is Environment Anomaly, Ind. is Individual Anomaly. In Method columes, F. is Feature Based, S. is Statistical Method, T. is Tensor/Matrix based Methods, G. is Graph based Methods, O. is Others, C. is Classification Methods, R. is Regression Methods.)

Literatures		Datasets							Anomaly Type				Method						
Name	Year	Tj.	Tp.	C.	US.	SM.	ER.	EV.	Tf.	Cd.	EA.	Ind.	Detection					Prediction	
													F.	S.	T.	G.	O.	C.	R.
Lee et al.[63]	2008	√										√	√						
Li et al.[66]	2009		√						√							√			
Ge et al.[40]	2010	√										√	√						
Yang et al.[115]	2011				√					√				√					
Yang et al.[116]	2011				√					√					√				
Liu et al.[70]	2011	√							√							√			
Pang et al.[79]	2011	√								√				√					
Ge et al.[39]	2011	√										√	√						
Zhang et al.[121]	2011	√										√	√						
Chen et al.[19]	2011	√										√		√					
Ceapa et al.[16]	2012		√						√					√					
Zhang et al.[123]	2012		√						√				√						
Chawla et al.[18]	2012	√							√						√				
Witayangkurn et al.[103]	2013			√						√				√					
Pan et al.[78]	2013	√							√				√						
Rozenshtein et al.[88]	2014		√			√				√						√			
Yang et al.[114]	2014				√				√						√				
Kinoshita et al.[57]	2015	√							√					√					
Chen et al.[22]	2015		√							√				√					
Zhang et al.[125]	2015	√								√				√					
Zheng et al.[130]	2015		√				√			√				√					
Dong et al.[34]	2015			√						√			√						
Wang et al.[101]	2016	√							√				√						
Wang et al.[100]	2016				√					√									
Huang et al.[52]	2016						√					√		√					
De et al.[31]	2016				√					√						√			
Banerjee et al.[7]	2016	√										√	√						
Wu et al.[105]	2017	√										√		√					
Chiang et al.[25]	2017	√							√				√						
Vahedian et al.[98]	2017	√								√				√					
Hu et al.[51]	2017		√			√				√				√					
Zhu et al.[131]	2017		√							√									
Khezerlou et al.[55]	2017	√								√				√					
Teng et al.[94]	2017		√			√				√						√			
Chen et al.[21]	2017		√							√					√				
Tomaras et al.[95]	2017		√						√										
Lin et al.[68]	2018		√							√					√				
Zhang et al.[122]	2018		√				√			√			√						
Chong et al.[27]	2004						√		√									√	
Oh et al.[76]	2005				√				√					√				√	
Adbel et al.[2]	2006				√				√					√				√	
Ahmed et al.[5]	2012				√				√					√				√	
Xu et al.[110]	2013				√				√				√					√	
Xu et al.[109]	2013				√				√					√				√	
Yu et al.[117]	2014				√				√					√				√	
Copeland et al.[29]	2015							√			√		√					√	
Madaio et al.[73]	2015							√			√		√					√	
Potash et al.[82]	2015							√			√		√					√	
Madaio et al.[72]	2016							√			√		√					√	
Chojnacki et al.[26]	2017							√			√		√					√	
Ren et al.[86]	2017																	√	
Wu et al.[107]	2017						√											√	
Singh et al.[93]	2018							√			√		√						
Abernethy et al.[3]	2018							√			√		√						√
Kumar et al.[62]	2018							√			√		√					√	
He et al.[47]	2018	√							√					√					√
Yuan et al.[119]	2018				√				√										√

management, like maintenance or close of roads. This kind of traffic anomaly usually causes a sharp drop in traffic volume, and the effect will keep for a longer time.

Many works treat the road network as multiple independent road segments and have been done to detect and predict the anomalous road segments [57, 66, 101, 114]. However, the effect of traffic anomalies is usually not limited on one road segment. If the road network in the urban area is regarded as a graph, a traffic anomaly is more likely an anomalous subgraph instead of an abnormal edge. Based on this consideration, some works regard a group of connected road segments as an unit[25, 78]. In [18, 70], the authors further explored the causal interactions among the anomalous road segments and identified the root cause of traffic anomalies.

2.2 Unexpected Crowds

Unexpected crowds in the urban area are one of the major threats to public safety. For example, on Dec. 31th, 2014, more than 300,000 people flowed into the Bund in Shanghai for the light show on New Year's Eve, which highly exceeds the expectation of organizers. The overcrowding led to a tragic stamped and caused 36 people killed and 49 injured in the end. Such accidents could be prevented if the gathering of people can be detected or predicted in its early stage [34, 78]. The increase in people density in a region has direct impacts on urban data, such as the sudden increase in the number of cellular subscribers for the base stations around that area and the unexpected rise in the number of exiting passengers of nearby subway stations [16]. These abnormal changes can help to discover the happening of unexpected crowds.

One significant difficulty of unexpected crowds detection and prediction is the limited number of recorded events. Thus, it is hard to evaluate the effectiveness of different detecting and predicting algorithms. In the relevant literature, the authors usually use festival celebration, pop concerts, and sports matches as abnormal events. These unusual events can be discriminated by their effect scope. Anomalous events such as concerts and matches only have a local influence, while other events like festivals and extreme weather usually cause unusual changes of urban dynamics in city-scale.

2.3 Environment Anomaly

Urban environment anomalies are also a significant category of anomalous urban events, which is highly related to public safety. For example, the fire incident in cities is a severe threat to people's lives and properties [72, 93], and the pollution of the water system is also a common issue that harms residents' health [26, 62].

Unlike other kinds of urban anomalies, environment anomaly is mainly caused by environment changes instead of large-scale human activities and usually does not show a significant sign before its happening. Instead of directly detecting and alerting the urban anomalous, the current works of environment anomaly focus on evaluating the risk or tracing the causes. Micheal et al. [72] evaluated whether a building has a risk of fire based on the building condition information. Alex et al. [26] explored the residential water contamination based on water sample tests.

2.4 Individual Anomaly

Traffic anomalies and unexpected crowds usually have a lot of participators and have a relatively large scale impact. However, there are still some urban anomalies that are caused by abnormal individual activities and have less public influence. For example, taxi fraud [39] and some criminal activities [64].

In this survey, we term such events as individual anomalies and also consider it a critical anomaly category. There are two reasons. On the one hand, the detection of such anomalies can help to discover criminal or illegal activities and protect urban security. On the other hand, a large-scale anomalous event is consist of a lot of individual anomalies. Analyzing the spatial and temporal interactions of a lot of individual anomalies can also help to detect other kinds of anomalies.

3 DATASETS

The datasets are the core part of modern urban anomaly detection and prediction framework. Nowadays, various types of datasets can be obtained from multiple devices. One of the most important sources is smartphones which can function as moving sensors. For instance, when users make phone calls or access the cellular network, their locations can be recorded and used to track the movement of people flows and estimate the density of populations [34]. The text or picture information people post on social media by smartphones is also an ideal semantic description of the urban environment around them, which can help to understand the events happening around. Besides, the distributed sensors located around urban areas are another source of urban data. For example, the surveillance cameras record the surrounding scenes and traffic sensors provide detailed records of traffic status. In this section, we classify the urban datasets into six categories based on the data types and attributes. Four of them are structured data, i.e., trajectory, trip records, urban sensor records, and CDRs. The other two are unstructured data, namely camera surveillance and social media data.

3.1 Trajectory

A trajectory consists of a series of time-location records, which are reported by GPS devices in a sample rate of the minute level. The trajectory datasets provide the most detailed and comprehensive records of vehicle movements.

The primary source of trajectory dataset is the taxis with GPS equipment. As one of the most critical transport modes, taxis are usually widely spread in urban areas and run for almost 24 hours every day. Many works have been done on the urban dynamics understanding with taxi trajectories[15, 19, 125, 131]. In our case, trajectory datasets can be used to detect all the three kinds of anomaly discussed in the above section. For example, gathering events can be detected by predicting trajectories of people in [98], traffic incident is caught based on taxi trajectories [57] and taxi driving fraud is identified by detecting trajectory outliers [39].

We summarize some statistics of trajectory datasets used in studied literature in Table 2. These datasets are all collected from metropolis or countries, especially the areas with a large population such as Beijing, Shanghai, and Singapore. The duration of datasets varies from days to years, while the sample rate is usually at the second or minute level, which is considered dense enough to track the motion of vehicles. The quantity of datasets is given by the number of trajectories(T) or the location points(P) reported by GPS.

Table 2. Trajectory datasets

Dataset	Duration	Sample Rate (s/point)	#T/#P ($\times 10^3$)	#Vehicles
Porto[105]	1.5 years	15	486 T	442
Shanghai[1][105]	10 days	10	757 T	13650
Beijing[1][68]	1 week	-	450 P	8940
Singapore[25]	2 months	-	50 P	-
Shanghai[2][125]	2 years	-	10,000,000 P	10000
Hanzhou[125]	1 year	-	3,000,000 P	5000
Beijing[2][78]	2 months	70.5	19,455 T	13597

3.2 Trip records

Trip records data can be provided by taxi or sharing bike systems. Every trip record usually contains the starting location, ending location, trip distance, trip time and other information about a trip. It can be considered as sparse trajectories, and the primary application of vehicle trip data mining is to understand the human movement in the urban area. For example, the total number of taxi trips from one region to another region reflects the number of people moving from one area to another to some degree. Then people gathering events can be discovered by monitoring the taxi trips.

Table 3. Trip Record Datasets

Dataset	Duration	#Trips ($\times 10^6$)	#Vehicles
NYC-taxi[1] [68]	1 year	3	-
Washington [16]	3 year	8	3296
NYC-bike [130]	1 year	8	6811
NYC-taxi[2] [130]	1 year	165	14144
San Francisco [66]	30 days	0.8	500

Some trip record datasets are listed in Table 3. These datasets are usually published by city taxi or public sharing bike operators and updated every month or season. Thus, the duration of these datasets is up to years, expect the bike trip dataset from San Francisco that is used in a early literature[66]. In taxi trajectory datasets, the starting and ending locations are usually given as the street name or block name, while in sharing bike systems the locations need to be bike stations. Each dataset contains millions of trip records, produced by thousands of individual vehicles in average.

3.3 CDRs

Mobile phone call detail record (CDR) data are the time and location records of phone calls [10]. CDR data use the positions of the cellular towers as the users' location, which is less accurate compared with GPS records. Meanwhile, the time interval between two phone calls made by one user is usually up to hours or even days. That makes the CDRs even sparser than trip records data. However, due to the penetration of mobile phones and the vast number of phone calls made every day, CDRs are more accessible to collect and have a large volume. With this advantage, CDRs are widely used to estimate human mobility and population distributions [83, 97, 120].

The statistics of some CDR datasets are listed in Table 4. These datasets are collected by the internet service provider, ISPs, in major cities or countries, which contain millions of users with the duration from months to one year.

Table 4. CDR Datasets

Dataset	Duration	# of users or divices($\times 10^6$)	# of records ($\times 10^6$)
San Francisco[83]	1 month	1	-
Massachusetts[50]	4 months	1	-
Shen Zhen[120]	1 year	10	435
Haiti[9]	6 weeks	1.9	-
Senegal[34]	5 months	0.05	-
Los Angles[53]	140 days	0.2	321
New York[53]	140 days	0.15	223

3.4 Urban sensors

Apart from smartphones and GPS devices, there are also many sensors distributed around the urban area with a fixed location to collect urban data. The most common sensors are the loop detectors on roads, which are usually installed underneath the pavement at an interval of around half a mile. The loop detectors record the vehicles that pass the locations, and the records can be utilized to evaluate the vehicle speed and road condition [76, 109, 110]. The public transportation card reading machines in bus and subway stations can also act as sensors that record the volume of people flows. These sensors record the urban dynamics of one location from different perspectives. The usage of these data depends on the sensor types. For example, the subway card records reflect the number of passengers who enter and exit the subway station, which can be applied to detect the crowding events surrounding the station [100].

3.5 Social media

Social media data such as Twitter and Weibo are widely utilized for event discovering [69]. However, as we state in the introduction section, the general event detecting from social media is not in our study scope. In urban anomalous event detection, social media datasets are usually not used separately. For example, the topic distribution of social media data is used as a feature together with other types of urban data for multi-view anomaly detection in some works [94, 130]. In some other literature, social media data are also used to get a semantic understanding of detected events [78].

3.6 Camera Surveillance

Camera surveillance plays a vital role in capturing and monitoring scenes of human motions. To better understand the human mobility pattern and detect the abnormal behaviors, several records of camera surveillance are used to train and validate for the abnormal detection algorithms on videos. With an efficient model, then it would be trivial to detect anomaly events automatically.

Here are some datasets which are widely used in research papers. The subway dataset is consist of two videos, 'entrance gate' (1 hour 36 minutes long with 144,249 frames) and 'exit gate' (43 minutes long with 64,900 frames) [4]. The UCSD dataset collects video from walkways in the campus of University of California, San Diego. The data is split into 2 subsets, Peds1 and Peds2. In Peds1, there are groups of people walking towards and away from the camera, and some amount of perspective distortion, which contains 34 training video samples and 36 testing video samples. In Peds2, there are scenes with pedestrian movement parallel to the camera plane. There are 16 training video samples and 12 testing video samples in Peds2 [1]. As for VIRAT dataset [77], it consists of stationary ground camera data with approximately 25 hours long across as well as 16 different scenes with high definition. The CUHK dataset [71] contains 15 sequences with the length of 2 minutes for each. In this dataset, 14 kinds of abnormal events are being documented, which includes running, throwing objects, and loitering. The information about the camera surveillance datasets mentioned above is listed in Table 5.

Table 5. Camera Records Datasets

Dataset	# of Frames	Resolution
Subway-entrance[4]	144,249	512×384
Subway-exit[4]	64,900	512×384
UCSD-Ped1[1]	14,000	158×238
UCSD-Ped2 [1]	5,600	320×240
VIRAT[77]	37,500-45,000	1920×1080
CUHK [71]	35,240	-

3.7 Environment data

The happenings of anomalous events are sometimes affected by environmental factors. Thus, environment data are usually used as the extra information to help anomalies detecting or predicting. For example, the weather condition data are widely used in traffic anomaly detecting since the bad weather is a primary cause of traffic accidents [2, 27, 110, 117]. Concerning environment anomaly detection, environment data are the primary data source. The environment information such as building conditions is used to evaluate fire risk in urban areas. [72, 73, 93]. In the case of water system monitoring, the test results of water samples are used [26, 62, 82].

4 ALGORITHMS

4.1 Detection

In this section, we discuss the state-of-art algorithms on abnormal events detecting from urban data. An urban anomaly detecting algorithm is usually composed of multiple components and combines methods from different fields. We classify these algorithms into five groups by considering only the main thoughts behind them. An overview of our classification is shown in Fig 3. The computer vision based methods involve video processing techniques and are different from the other methods, the further detailed classification of it is separately shown in Fig. 4.

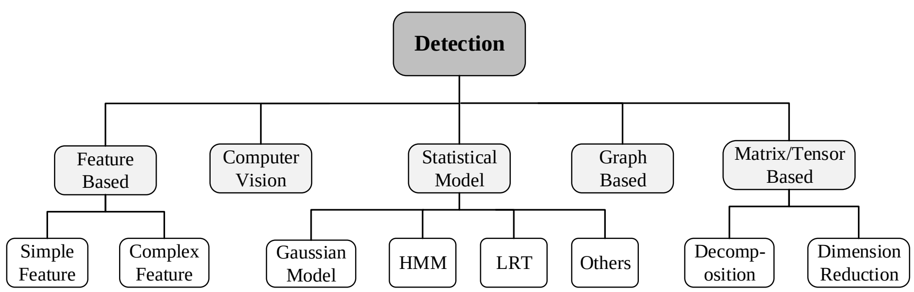

Fig. 3. Classification of urban anomaly detection methods.

4.1.1 Feature Based. The basic idea of feature-based methods is first to extract spatial or temporal features and then apply classical outlier detecting algorithms to identify anomalies. Simple physical features such as speed and distance are frequently used for traffic anomaly detection. Some works also defined complex artificial features based on their application scenarios.

In [25, 101], the average vehicle speed is used to detect traffic accidents or congestion. They divide the urban road network into segments and use trajectories to estimate the speed of traffic flows in a small time slot. [101] computes the change rate of traffic flow speed and identify an anomaly when the change rate exceeds a threshold. [25] designs a non-parametric Kernel Density Estimation function based on historical speed data and then derive a congestion score for the current time slot. Some works use travel distance as a feature. A trip record usually contains a start location, an end location, and the travel distance. Given the start point and destination in a city, the travel distance should also be fixed around a value. Based on this consideration, Zhang et al. [123] proposed a method to detect taxi trips with abnormal travel distance and then further infer traffic congestion and other events. In [39], travel distance is used as evidence for taxi fraud detection. Similarly, Banerjee et al. [7] utilize travel time to discover anomalous trips. For vehicle trajectory outlier detection, the moving direction is another useful feature. In [40], Ge et al. proposed a system that calculated an outlier score for a trajectory based on its moving direction sequence. Besides above simple physical features, higher-level human-designed features are also used for urban anomaly detecting. A widely considered feature is temporal or spatial similarity. Zhang et al. [122] believe different regions in a city could have similar urban dynamics. They combined multiple data sources and define the similarity between each pair of regions as the Pearson Coefficient Correlation of time series in a time window. Then they compute the similarity change rate and identify anomalous regions where the similarity score changes dramatically. Li et al. [66] represents the road network with a directed graph and compute the similarity between edges. They consider every edge has stable neighbors in feature space and detect traffic anomalies by monitoring the change of edge neighbors. Dong et al. considered temporal similarity in [34]. They first identify anomalous users by calculating the similarity between an individual's trajectory with its historical trajectory. Then they detect unexpected crowds by searching gathered abnormal individuals. In case of traffic anomaly detection, routing behavior is modeled as a feature. Pan et al. [78] defined the routing pattern between two location as a vector of traffic volume on each path that connects the two locations. In [39], the authors first represented a trajectory as a sequence of symbols of the locations it passed. Then they assigned a codeword to each symbol. By this meaning, the anomalous trajectories should have unusually high code cost based on coding theory.

After appropriate features are obtained, the urban anomaly detection problem is translated into classical anomaly detection problems. According to a survey [17], classical anomaly detection methods can be grouped into several main categories, including statistical methods, classification based methods, nearest neighborhood-based methods, clustering based methods, information theory methods, and spectral methods. The first two kinds of techniques are most frequently used in the second step of feature-based urban anomaly detection methods. The basic of statistical anomaly detection methods is to estimate the distribution of data. Then instances that have low probabilities is considered as anomalies. There are two different ways to estimate the data distribution: parametric methods and nonparametric methods. The simplest parametric method is a Gaussian-based model. This technique assumes the data is generated by a Gaussian distribution and estimate the mean μ and standard deviation σ with Maximum Likelihood Estimates (MLE). After the parameter is estimated, the anomaly score of an instance can be defined as the degree it deviates from the mean value. A simple example is 3σ test that detects anomalies with a $\mu \pm 3\sigma$ threshold. A gaussian-based parametric method is a dominating method used in urban anomaly detecting application [7, 34, 40, 78, 101, 123]. Another parametric method is based on mixture model, which models the data with the mixture of parametric statistical distributions. For example, Ge et al. [40] modeled the distribution of travel distance between two locations with multiple Gaussian distributions, where each Gaussian distribution describes the distance distribution of one path choice. Compared with parametric methods, nonparametric methods make fewer assumptions about the data distribution. Kernel function based model[32] is a nonparametric method that estimates probability density using kernel functions. It is similar to mixture parametric models, but it does not make prior assumptions about the distribution. This method is used in [22, 25] to model the distribution of traffic speed data and sharing bike renting data. One-class SVM is a classification based anomaly detection algorithm. It learns a region that contains normal points in feature space using kernel function. Instances that fall outside the region is identified as anomalies, and the distance between the instance and region boundary can be used to measure the anomaly degree. In [122] one-class SVM is used to detect anomalous events.

4.1.2 Statistical Model. In the above section, we discussed the use of statistical methods in feature-based urban anomaly detection techniques. However, there are also many works that directly apply statistical methods without feature extracting step. The simplest case is comparing the traffic volume or other urban dynamics in a time slot with historical mean value. Ceapa et al. [16] proposed an algorithm to predict the crowdedness of a subway station by computing the mean number of passengers that enter or exit the station at same hours in historical days. In [125], the authors extract the traffic period by applying Discrete Fourier Transform(DFT). The they evenly split each period into same number of small time intervals and compute a mean traffic volume for each time interval over all periods. In the last, an anomaly score is defined as the degree the traffic volume deviates from the mean value. In [19], Chen et al. label a taxi trip as anomalous trip if the trajectory rarely occurs in dataset. Instead of assuming the dataset follows single Gaussian distribution, Tomaras et al. [95] model the distribution of daily traffic with the combination of multiple Gaussian functions.

Above work make assumptions about the urban dynamics distribution. In some other works, Hidden Markov Model(HMM) is used to model the status transition of urban dynamics. Five basic concepts about HMM are observation sequence, state sequence, initial state, transition probability matrix, emission probability matrix. The states and observations of an HMM are supposed to be discrete and limited. It also assumes that the hidden state at any time only depends on the previous state and the observation at any time only depends on the state at the time. Then the transition probability matrix is used to describe the probability a state transitions to another state. The emission probability matrix defines the probability an observation occurs under a certain state. The three essential components that define an HMM is an initial state, transition probability matrix, and emission probability matrix. Given these three components, the probability that an observation sequence occurs can be computed by the forward-backward algorithm of HMM. Yang et al. [115] define an observation as a vector that represents the number of people located in areas of interest. They adopt the K-means algorithm to group observation vectors into k clusters as the hidden states. Then the initial state and transition probability matrix are estimated based on historical data. To build the emission probability matrix, the authors use the Gaussian Mixture Model (GMM) to model the distribution of observation vectors contained in each cluster. After the model is built, the probability of a new observation sequence can be calculated. When the probability is under a threshold, an anomalous event is considered happening. Witayangkurn et al. [103] proposed a similar algorithm. However, they cluster the observation vectors at first to reduce the number of observations. At the anomaly detection step, they define an anomaly score for each observation rather than the whole observation sequence. HMM is also used to predict urban dynamics in [52]. They build HMM to compute the probability density of the next state and further cooperate with the similarity between different regions to make the final prediction.

Likelihood Ratio Test (LRT) is another statistical technique widely applied in spatial-temporal anomaly detection. LRT is originally used to compare two models in the statistic. Given a dataset X, a model with parameter $\theta \in \Theta_0$, an alternate model with parameter$\theta \in \Theta - \Theta_0$, the likelihood ratio is then defined as,

$$\lambda = \frac{sup_{\Theta_0} L(\theta|X)}{sup_{\Theta} L(\theta|X)},$$

where Θ is the parameter space and Θ_0 is the restricted parameter space. This ratio can be computed by the maximum likelihood estimate (MLE). It is also proved that the asymptotic distribution of $\Lambda = -2log\lambda$ is a chi-square distribution $\chi^2(\Lambda, p-q)$[102], where p and q are respectively the number of free parameters of the null model and the alternative model.

In the application of anomaly detection, the null model is usually associated with the hypothesis that there is no anomaly and the alternate model is associated with the opposite case. A probability density function is first chosen. Then the value Λ can be calculated by MLE. An anomaly is detected at confidence α when $\Lambda > c$, where c is the threshold from which the area under chi-square distribution density function is smaller than α. Wu et al. [106] proposed a framework to detect the spatial anomaly. They first partition the spatial area into $n \times n$ grids and check anomalies from each cell. They make two competing hypotheses on whether the process generating data in a cell is substantially different from the process generating the data outside of that cell. Then based on hypotheses, two models with different parameters are proposed. The parameters of the null model are forced to be identical for every cell while the parameters of the alternate model are customized for each cell. Thus the LRT can be applied to determine which model fits the dataset better. In [79] Pang et al. extended the

LRT framework in [106] to discover traffic anomaly and both persistent and emerging outliers are detected in their work. Khezerlou et al. [55] detect gathering events based on traffic flow. They represent traffic flows in the urban area as a directed graph and propose a likelihood ratio based definition of edge anomaly degree. Finally, they detect gathering events in a location by considering the anomaly degree of in-edges and out-edges. In [130], LRT is also used to calculate an anomaly degree for regions.

There are also some other statistical models applied in urban anomaly detection field. Probabilistic topic model is introduced to describe the underlying states of a region[130] or traffic states on a road[57]. In [105], the authors consider a trajectory as a sequence of routing decision and then apply the maximum entropy inverse reinforcement learning model[133] to compute the probability that a trajectory is produced. Vahedian et al. [98] propose a method to predict gathering events by predicting the destination of uncompleted trajectories. They define the destination prediction problem as a classification problem and adopt a Bayes classifier to compute the probability of a location as the destination conditioned on the uncompleted trajectory. In [51], Hu et al. filter out anomalies by computing the entropy deviaion of the urban data cluster in a region.

4.1.3 Tensor or Matrix Based. Time and location are two fundamental dimensions of urban data. The third dimension of urban data is the feature dimension. In some cases, the feature is a vector. Then the urban data can be formulated as a 3-dimension tensor. In other cases, the feature is a single value such as the indication of a sensor. Then the data can be written as a matrix. By representing the observed urban data as tensors or matrices, some tensor or matrix-based techniques such as decomposition and dimensionality reduction can be applied to detect urban anomalies.

Tensor decomposition is widely used to discover the basic patterns of urban dynamics. Lin et al. [68] formulate the taxi trip data as region-feature-time tensors, where the feature dimension is the traffic volume to and from other regions. Then they use the non-negative CP decomposition[58] method to decompose the tensor into a three-factor matrix, which respectively represents the basic mobility pattern, the temporal and spatial distribution of basic patterns. The mobility pattern matrix is considered shared by urban dynamics in different locations and from the different time period. The upcoming tensor is then decomposed with the mobility pattern matrix fixed. In the last step, the anomalous region that holds a different distribution of basic patterns compared with its history is detected by LOF algorithm[12]. A similar method is proposed in [21]. However, this work decomposes a mobility matrix and a check-in tensor together by forcing them to share two factors.

As mentioned above, in some cases urban data can be formulated as a matrix with columns and rows corresponding to different regions and consecutive times. By this means, dimensionality reduction algorithms can be applied to learning a spatial-temporal feature which combines the spatial and temporal information. Then outliers can be searched in the feature space. Two most commonly used dimensionality reduction algorithms are Principal Component Analysis(PCA) and Locally Linear Embedding(LLE)[87]. The principal idea of PCA is mapping the data to a low dimensional space spanned by a set of new basis, which is called principal components. The principal components are the eigenvectors of the covariance matrix of the original data. By removing eigenvectors with smaller eigenvalues, the dimensionality of original data is reduced. A variant of PCA is Robust PCA[13], which can tolerate non-Gaussian noises. The LLE algorithm constructs a data point with the linear combination of its K nearest neighbors. The weights are derived in the sense of least square construction error. The vector of weights is used as the feature of the original data point. Yang et al. represent the traffic flow[115] monitoring data from distributed sensors in a time window as a matrix. Then they adopt the LLE algorithm to obtain a weight matrix W. Finally, they organize the K eigenvalues of WW^T in ascending order as the LLE-PCA feature of the original matrix. In [18], PCA is used to detect abnormal road traffic from a link-time matrix, which shows the traffic volume on different roads in a time window. Then the root cause of traffic anomaly is located by exploring the linkage relations between roads. Yang et al. [114] introduce the BRPCA[33] algorithm to decompose a matrix into the sum of a low-rank matrix, a sparse matrix, and a noise matrix. The sparse matrix is further written as the element-wise product of a binary matrix and a real matrix that are sampled from specific distributions. Then a binary matrix is used to indicate the occurrences of the traffic event. The authors also extend this framework to combine multiple data matrices from different sensors by making them share the same binary matrix.

4.1.4 Graph Based. Many algorithms have been developed to detect anomalies in graphs [6]. The urban data can also be represented by graphs in different ways. In [94] a dynamic network is constructed from the urban data. The dynamic network contains a series of directed graphs, which represent urban dynamics in consecutive time slots. The nodes in graphs correspond to geographic regions in a city, and the edges are defined as the traffic load between different regions. A spatial feature is assigned to every node, for example, the twitter topic distribution of each region. Then a multi-view hypersphere learning algorithm is used to embed the dynamic network into a latent space shared by both node and edge features. Finally, anomalous nodes are detected in the latent space. Liu et al. [70] use an undirected graph to model the traffic flows in the urban area. Every edge is associated with a feature vector of several traffic properties. Then the edges whose features are significantly different from their spatial and temporal neighbors are identified as anomalies. Rozenshtein et al. [88] model the urban event detection task as the problem detecting active and dense subgraphs, which mean a subset of nodes that are close to each other and have high activity levels. They also propose practical greedy approaches to solve the problem. In [31] heterogeneous graphs are used to represent the urban data. A heterogeneous graph consists of nodes, node attributes and relations between nodes. The three components could be of different types. In this work, the nodes are users, locations, city issue reports and report categories. Then an NPHGS algorithm [20] is applied to detect anomalies from heterogeneous graphs.

4.1.5 Computer Vision. Nowadays, the use of surveillance cameras is increasing rapidly in urban areas, to promote the safety of citizens as well as the stability of society. To detect anomaly behavior from the surveillance videos, the technology of high-level image understanding is needed [81]. However, designing an algorithm to detect the anomaly in road traffic is still challenging as the patterns for the motion of objects are highly complicated [113]. Moreover, there are various kinds of anomaly events with unpredictable variations [56], making the discrimination for genuinely abnormal events more difficult.

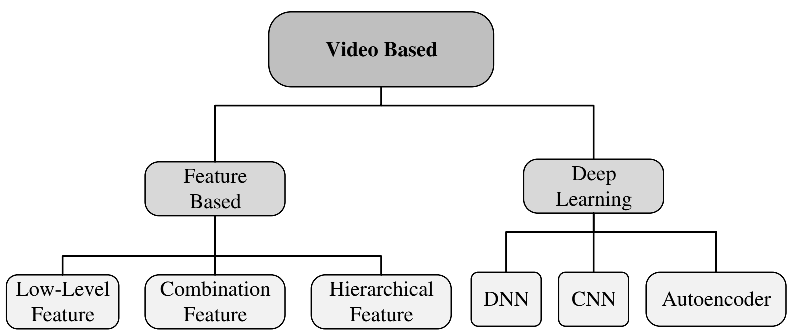

Fig. 4. Classification of video based urban anomaly detection methods.

To capture the feature from videos efficiently and effectively, many methods have been tried. In [80, 124], information about the trajectories of objects in videos are used to describe the motion of objects. By comparing them with the normal trajectory patterns, we can figure out whether they are anomaly event or not. However, when the object is occluded, or the video scenes are crowded, this method would be unable to handle the problem. Under this circumstance, more methods to extract motion patterns are utilized. Benezeth et al. [8] uses the histogram of the pixel change, Kim et al. [56] and Mehran et al. [75] are using the optical flow to measure the dynamic pattern of the object. These approaches mainly emphasize on dynamics but neglect anomalies of object appearance [65].

For better performance, many features need to be considered together. For instance, Li et al. [65] and Mahadevan et al. [74] use the Mixture of Dynamic Textures (MDT) models to detect the spatial abnormality as well as the temporal abnormality. Zhu et al.[132] use the motion and context feature to model the events jointly. Chen et al. [23] calculate the score of anomalous via aggregating the appearance and the mobility features of its nearest neighbor. Besides, to capture the high-level feature, for example, the interactions in videos and the low-level feature, like the motion feature of each video patch, Sabokrou et al. [89] and Cheng et al. [24] use a hierarchical structure to represent for events and interaction.

Recently, with the rapid development of the Deep Neural Network, which proves to be effective in image classification [61], object detection [41] and activity recognition [92] many modern deep architectures are being proposed to replace the hand-crafted features to model activity patterns [84]. In [90], a Binary Fully Convolutional Network (BFCN) is used to capture the feature of video segments. Combining the feature with optical-flow would provide us with the complementary information of both appearance and motion patterns. In [111, 112], appearance and motion features are learned separately by using stacked denoising autoencoders. Then, the learned features are fused for detection. In [46], a fully connected autoencoder is used on optical-flows, then a fully convolutional feed-forward autoencoder is added to learn both the local features and the classifiers to capture the temporal regularity of video sequences. However, as the size of existing datasets with ground truth abnormality samples are small, the Deep Neural Network based methods have the problem that their networks are relatively shallow to prevent the occurrence of overfitting. Moreover, in [111, 112], the feature they learned still have to pass through a classifier for detection, which is not direct enough.

To solve those problems, Generative Adversarial Networks (GANs) [43] are introduced for the task of anomaly event detection. In [85], Ravanbakhsh et al. train GANs with only normal data, thus they are unable to generate abnormal events. By computing the difference between the real video clips with the representations of appearance and motion reconstructed by the GANs, abnormal areas are detected.

4.2 Prediction

Although the precise and reliable prediction of anomalous events is remained as a challenging task, there are some researchers making efforts to forecast the urban anomalies in specific scenarios. There are mainly three study directions among these works. The first one focuses on environment anomalies. Instead of predicting the exact time of anomalous events, these works evaluate if there is a risk of a certain type of anomalies based on observed features. The second kind of works collect real-time urban dynamics to infer whether an anomaly will happen in near future, especially in the case of traffic anomaly prediction. Classification methods are usually used in these two kinds of works. Some other works predict the overall distributions of different anomalies by exploring rules from a large amount of recorded events. Predicting methods such as time series forecasting and deep neural networks are adopted in these works, which we summarize as regression methods in this section.

4.2.1 Classification methods. Classification methods are adopted to judge if an anomalous event will happen in future in studies of environment anomaly and traffic anomaly prediction. The crucial step of making successful classification is to construct and select appropriate features.

In the case of environment anomaly prediction, the features are usually constructed from environment information. Madaio et al. [72] proposed a framework to evaluate the fire risk in Atlanta. They use around 20,000 commercial properties such as fire permits, criminal record, and liquor license to construct the features. In [93] the authors consider the difference between areas of different urban function and respectively select commercial and residential features. Some works predict the water system pollution in cities, in which the residents' family information, health condition, and land information are used as features[3, 26, 82]. After the features are selected, classic classification models are directly used in these works, including Logistic regression(LR)[49], Support Vector Machine (SVM)[30], Random Forest[11] and gradient tree boosting[36].

In the case of traffic anomaly detection, the road condition observed by loop detectors and weather condition are usually utilized the features. Abdel et al.[2] combine the weather information and the statistic features of road condition such as mean speed of vehicles. Xu et al.[110] first adopt Random Forest to select important features[45] and make classification using a Genetic Programming Model[60]. In [109], the authors employ a sequential logistic regression model to predict the severity of traffic accidents. Besides, Yu et al.[117] explore the critical factors of different car crash types and then utilize the hierarchical logistic regression model[104] to predict the traffic crashes.

4.2.2 Regression methods. Instead of predicting the happening of a single anomalous event, some works adopt regression methods to predict the number of anomalies happen in an urban region in a time slot. Wu et al. [107] represent the urban anomalies by a tensor, which has three dimensions of time, region and anomaly categories. Each value in the tensor is the number of different categories of anomalies happened in specific time and region. Then they factorize the tensor into three factors which are the region, category and time matrices. They consider the region and category matrices are constant along time and apply a Vector Autoregression[44] algorithm to estimate the next column of the time matrix. In the last, by reconstructing the tensor with updated time matrix, the number of anomalies in different regions in the next time step is predicted. Recently, deep learning models are also introduced into urban anomaly prediction task. By dividing the urban area into grid regions and representing the urban dynamics in all regions using a matrix or tensor, the deep learning models that make great achievements in image processing domain can be migrated to deal with urban dynamics. Ren et al.[86] propose a new deep neural network structure based on LSTM network[48] to predict the risk of the traffic accident in regions based on historical traffic accident records. Similarly, Yuan et al.[119] utilize the Convolutional LSTM network[108] to predict the number of traffic accidents in different regions, which can capture both spatial and time domain correlations. In [111], the authors further combine human mobility data and abstract the human mobility features by a stacked Denoise Autoencoder[99].

5 OPEN CHALLENGES

Urban big data-based techniques are the future direction of urban anomaly detection. While a considerable number of works have been done in recent years, we believe there are still several challenges need to be conquered. In this section, we discuss the open challenges and problems in urban anomaly detection filed.

5.1 Data Imbalance

The first challenge comes from the imbalance of datasets. While normal urban events happen everyday and everywhere, anomalous events rarely happen and are usually not recorded. That means most of an urban dataset is produced by normal events and human daily activities happen in the urban area and just a tiny part is caused by anomalous events. This extreme imbalance of dataset brings problems from two aspects. First of all, due to the complex types of anomalies, it is hard to capture the patterns of such events and evaluate their effects on urban data with few samples. Secondly, with few recorded anomalous events, it is hard to evaluate proposed methods. Many researchers choose to use some typical events such as important festivals and concerts to test if their methods hit these events. But since these events are just a subset of the anomalies happened in real world, it is far away from a systematical evaluation. Some other researchers use synthetic datasets to evaluate their algorithms. However, the mechanism urban data are produced in real world is extreme complex. It is nearly impossible to simulate the effect of real-world anomalies by simple rules.

5.2 Prediction

Most of the current researches focus on detecting anomalies that already happened from offline urban datasets. However, in practice, the real demand is to detect the early evidence of an anomalous event and alert in advance. For example, in the case of traffic jam detecting, we hope to be aware of the sign that cars flow in one area from many other directions and leave enough time to take actions. To achieve this goal, there are two main difficulties. Firstly, the urban data is full of noises, and the slight sign of anomalous events in their early stage can easily drown in noises. Moreover, the early sign of anomalies sometimes shows up in the different urban data source. For example, the root cause of a traffic jam maybe a protest event. Before the event starts to affect the traffic flow, it might be discussed a lot on the local online community. In this case, a single dataset is not informative and valid enough for anomaly early alert. Effective techniques for data fusion must be developed to capture the interactions between different datasets and evaluate their common effect.

5.3 Explanation

As we state in the introduction section, the final goal of urban anomaly detection is to uncover the underlying anomalous events. However, most of the current works end with detecting abnormal urban dynamics. The gap between current researches and final goal is the explanation of detected anomalies. The explanation is important because it helps to find the root cause and decide what kind of actions should be taken. For example, an unexpected crowd anomaly can be caused by many reasons, such as a concert, a celebration parade or a terrorist attack. These three underlying events are on different emergency levels and should be handled in different ways. Finding the explanation is difficult. The increasing arriving taxi number in one region may be caused by the temporary shutdown of a subway station in another region half an hour ago. To trace back to the root cause, information across time, across regions and across datasets need to be combined.

6 CONCLUSION

The explosion of big data that are generated in urban space has brought new opportunities to solve traditional urban problems. In this paper, we discussed the new rising research area of big urban data based urban anomaly detection. To give a comprehensive introduction and literature review of this topic, We studied a considerable number of relevant works in recent years and answered three questions: What kind of urban data are used in urban anomaly detection? What kind of anomalous events can be detected? And what are the detection methods? In the last, we summarized the shortcomings of current researches and discussed three main challenges in this field that more efforts need to be made on.

REFERENCES

[1] [n. d.]. UCSD Anomaly Detection Dataset. http://www.svcl.ucsd.edu/projects/anomaly/dataset.htm.

[2] Mohamed A Abdel-Aty and Rajashekar Pemmanaboina. 2006. Calibrating a real-time traffic crash-prediction model using archived weather and ITS traffic data. *IEEE Transactions on Intelligent Transportation Systems* 7, 2 (2006), 167–174.

[3] Jacob Abernethy, Alex Chojnacki, Arya Farahi, Eric Schwartz, and Jared Webb. 2018. ActiveRemediation: The Search for Lead Pipes in Flint, Michigan. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 5–14.

[4] 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 & Machine Intelligence* 30, 3 (2008), 555.

[5] Mohamed M Ahmed and Mohamed A Abdel-Aty. 2012. The viability of using automatic vehicle identification data for real-time crash prediction. *IEEE Transactions on Intelligent Transportation Systems* 13, 2 (2012), 459–468.

[6] Leman Akoglu, Hanghang Tong, and Danai Koutra. 2015. Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery* 29, 3 (2015), 626–688.

[7] Prithu Banerjee, Pranali Yawalkar, and Sayan Ranu. 2016. Mantra: a scalable approach to mining temporally anomalous sub-trajectories. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1415–1424.

[8] Y. Benezeth, P. M. Jodoin, V. Saligrama, and C. Rosenberger. 2009. Abnormal events detection based on spatio-temporal co-occurences. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 2458–2465.

[9] Linus Bengtsson, Xin Lu, Anna Thorson, Richard Garfield, and Johan Von Schreeb. 2011. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in Haiti. *PLoS medicine* 8, 8 (2011), e1001083.
[10] Vincent D Blondel, Adeline Decuyper, and Gautier Krings. 2015. A survey of results on mobile phone datasets analysis. *EPJ Data Science* 4, 1 (2015), 10.
[11] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.
[12] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. LOF: identifying density-based local outliers. In *ACM sigmod record*, Vol. 29. ACM, 93–104.
[13] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. 2011. Robust principal component analysis? *Journal of the ACM (JACM)* 58, 3 (2011), 11.
[14] Nan Cao, Chaoguang Lin, Qiuhan Zhu, Yu-Ru Lin, Xian Teng, and Xidao Wen. 2018. Voila: Visual anomaly detection and monitoring with streaming spatiotemporal data. *IEEE transactions on visualization and computer graphics* 24, 1 (2018), 23–33.
[15] Pablo Samuel Castro, Daqing Zhang, Chao Chen, Shijian Li, and Gang Pan. 2013. From taxi GPS traces to social and community dynamics: A survey. *ACM Computing Surveys (CSUR)* 46, 2 (2013), 17.
[16] Irina Ceapa, Chris Smith, and Licia Capra. 2012. Avoiding the crowds: understanding tube station congestion patterns from trip data. In *Proceedings of the ACM SIGKDD international workshop on urban computing*. ACM, 134–141.
[17] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. *ACM computing surveys (CSUR)* 41, 3 (2009), 15.
[18] Sanjay Chawla, Yu Zheng, and Jiafeng Hu. 2012. Inferring the root cause in road traffic anomalies. In *Data Mining (ICDM), 2012 IEEE 12th International Conference on*. IEEE, 141–150.
[19] Chao Chen, Daqing Zhang, Pablo Samuel Castro, Nan Li, Lin Sun, and Shijian Li. 2011. Real-time detection of anomalous taxi trajectories from GPS traces. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Springer, 63–74.
[20] Feng Chen and Daniel B Neill. 2014. Non-parametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1166–1175.
[21] Longbiao Chen, Jérémie Jakubowicz, Dingqi Yang, Daqing Zhang, and Gang Pan. 2017. Fine-Grained Urban Event Detection and Characterization Based on Tensor Cofactorization. *IEEE Trans. Human-Machine Systems* 47, 3 (2017), 380–391.
[22] Longbiao Chen, Dingqi Yang, Jeremie Jakubowicz, Gang Pan, Daqing Zhang, and Shijian Li. 2015. Sensing the pulse of urban activity centers leveraging bike sharing open data. In *Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), 2015 IEEE 12th Intl Conf on*. IEEE, 135–142.
[23] Zhu Chen and V. Saligrama. 2012. Video anomaly detection based on local statistical aggregates. In *Computer Vision and Pattern Recognition*. 2112–2119.
[24] Kai Wen Cheng, Yie Tarng Chen, and Wen Hsien Fang. 2015. Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression. In *Computer Vision and Pattern Recognition*.
[25] Meng-Fen Chiang, Ee-Peng Lim, Wang-Chien Lee, and Agus Trisnajaya Kwee. 2017. BTCI: A new framework for identifying congestion cascades using bus trajectory data. In *Big Data (Big Data), 2017 IEEE International Conference on*. IEEE, 1133–1142.
[26] Alex Chojnacki, Chengyu Dai, Arya Farahi, Guangsha Shi, Jared Webb, Daniel T Zhang, Jacob Abernethy, and Eric Schwartz. 2017. A Data Science Approach to Understanding Residential Water Contamination in Flint. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1407–1416.
[27] Miao M Chong, Ajith Abraham, and Marcin Paprzycki. 2004. Traffic accident analysis using decision trees and neural networks. *arXiv preprint cs/0405050* (2004).
[28] Blerim Cici, Minas Gjoka, Athina Markopoulou, and Carter T Butts. 2015. On the decomposition of cell phone activity patterns and their connection with urban ecology. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, 317–326.
[29] Eddie Copeland. 2015. Big data in the big apple. *Capital City Foundation* (2015).
[30] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297.
[31] Julio De Melo Borges, Till Riedel, and Michael Beigl. 2016. Urban anomaly detection: A use-case for participatory infra-structure monitoring. In *Proceedings of the Second International Conference on IoT in Urban Space*. ACM, 36–38.
[32] MJ Desforges, PJ Jacob, and JE Cooper. 1998. Applications of probability density estimation to the detection of abnormal conditions in engineering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 212, 8 (1998), 687–703.
[33] Xinghao Ding, Lihan He, and Lawrence Carin. 2011. Bayesian robust principal component analysis. *IEEE Transactions on Image Processing* 20, 12 (2011), 3419–3430.
[34] Yuxiao Dong, Fabio Pinelli, Yiannis Gkoufas, Zubair Nabi, Francesco Calabrese, and Nitesh V Chawla. 2015. Inferring unusual crowd events from mobile phone call detail records. In *Joint European conference on machine learning and knowledge discovery in databases*.

Springer, 474–492.
[35] Nivan Ferreira, Jorge Poco, Huy T Vo, Juliana Freire, and Cláudio T Silva. 2013. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2149–2158.
[36] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.
[37] Yue Gao, Sicheng Zhao, Yang Yang, and Tat-Seng Chua. 2015. Multimedia social event detection in microblog. In *International Conference on Multimedia Modeling*. Springer, 269–281.
[38] Victor Garcia-Font, Carles Garrigues, and Helena Rifà-Pous. 2016. A comparative study of anomaly detection techniques for smart city wireless sensor networks. *Sensors* 16, 6 (2016), 868.
[39] Yong Ge, Hui Xiong, Chuanren Liu, and Zhi-Hua Zhou. 2011. A taxi driving fraud detection system. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE, 181–190.
[40] Yong Ge, Hui Xiong, Zhi-hua Zhou, Hasan Ozdemir, Jannite Yu, and Kuo Chu Lee. 2010. Top-eye: Top-k evolving trajectory outlier detection. In *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 1733–1736.
[41] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
[42] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. 2008. Understanding individual human mobility patterns. *nature* 453, 7196 (2008), 779.
[43] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Eds.). Curran Associates, Inc., 2672–2680.
[44] James Douglas Hamilton. 1994. *Time series analysis*. Vol. 2. Princeton university press Princeton, NJ.
[45] Rami Harb, Xuedong Yan, Essam Radwan, and Xiaogang Su. 2009. Exploring precrash maneuvers using classification trees and random forests. *Accident Analysis & Prevention* 41, 1 (2009), 98–107.
[46] Mahmudul Hasan, Jonghyun Choi, Jan Neumann, Amit K. Roy-Chowdhury, and Larry S. Davis. 2016. Learning Temporal Regularity in Video Sequences. In *Computer Vision and Pattern Recognition*. 733–742.
[47] Tianfu He, Jie Bao, Ruiyuan Li, Sijie Ruan, Yanhua Li, Chao Tian, and Yu Zheng. 2018. Detecting Vehicle Illegal Parking Events Using Sharing Bikes' Trajectories. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '18)*. ACM, New York, NY, USA, 340–349. https://doi.org/10.1145/3219819.3219887
[48] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
[49] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. 2013. *Applied logistic regression*. Vol. 398. John Wiley & Sons.
[50] Sahar Hoteit, Stefano Secci, Stanislav Sobolevsky, Carlo Ratti, and Guy Pujolle. 2014. Estimating human trajectories and hotspots through mobile phone data. *Computer Networks* 64 (2014), 296–307.
[51] Jun Hu, Yuxin Wang, and Ping Li. 2017. Online city-scale hyper-local event detection via analysis of social media and human mobility. In *Big Data (Big Data), 2017 IEEE International Conference on*. IEEE, 626–635.
[52] Chao Huang, Xian Wu, and Dong Wang. 2016. Crowdsourcing-based urban anomaly prediction system for smart cities. In *Proceedings of the 25th ACM international on conference on information and knowledge management*. ACM, 1969–1972.
[53] Sibren Isaacman, Richard Becker, Ramón Cáceres, Stephen Kobourov, Margaret Martonosi, James Rowland, and Alexander Varshavsky. 2011. Ranges of human mobility in Los Angeles and New York. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*. IEEE, 88–93.
[54] Shan Jiang, Joseph Ferreira, and Marta C González. 2017. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data* 3, 2 (2017), 208–219.
[55] Amin Vahedian Khezerlou, Xun Zhou, Lufan Li, Zubair Shafiq, Alex X Liu, and Fan Zhang. 2017. A traffic flow approach to early detection of gathering events: Comprehensive results. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, 6 (2017), 74.
[56] Jaechul Kim and K Grauman. 2009. Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 2921–2928.
[57] Akira Kinoshita, Atsuhiro Takasu, and Jun Adachi. 2015. Real-time traffic incident detection using a probabilistic topic model. *Information Systems* 54 (2015), 169–188.
[58] Tamara G Kolda and Brett W Bader. 2009. Tensor decompositions and applications. *SIAM review* 51, 3 (2009), 455–500.
[59] Xiangjie Kong, Ximeng Song, Feng Xia, Haochen Guo, Jinzhong Wang, and Amr Tolba. 2018. LoTAD: long-term traffic anomaly detection based on crowdsourced bus trajectory data. *World Wide Web* 21, 3 (2018), 825–847.
[60] John R Koza. 1994. Genetic programming as a means for programming computers by natural selection. *Statistics and computing* 4, 2 (1994), 87–112.
[61] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems*. 1097–1105.
[62] Avishek Kumar, Syed Ali Asad Rizvi, Benjamin Brooks, R Ali Vanderveld, Kevin H Wilson, Chad Kenney, Sam Edelstein, Adria Finch, Andrew Maxwell, Joe Zuckerbraun, et al. 2018. Using Machine Learning to Assess the Risk of and Prevent Water Main Breaks. *arXiv*

preprint arXiv:1805.03597 (2018).

[63] Jae-Gil Lee, Jiawei Han, and Xiaolei Li. 2008. Trajectory outlier detection: A partition-and-detect framework. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*. IEEE, 140–149.

[64] Wai-Kong Lee, Chun-Farn Leong, Weng-Kin Lai, Lee-Kien Leow, and Thiah-Huat Yap. 2018. ArchCam: Real time expert system for suspicious behaviour detection in ATM site. *Expert Systems with Applications* 109 (2018), 12–24.

[65] W. Li, V Mahadevan, and N Vasconcelos. 2013. Anomaly detection and localization in crowded scenes. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 36, 1 (2013), 18–32.

[66] Xiaolei Li, Zhenhui Li, Jiawei Han, and Jae-Gil Lee. 2009. Temporal outlier detection in vehicle traffic data. In *IEEE International Conference on Data Engineering*. IEEE, 1319–1322.

[67] Zicheng Liao, Yizhou Yu, and Baoquan Chen. 2010. Anomaly detection in GPS data based on visual analytics. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*. IEEE, 51–58.

[68] Chaoguang Lin, Qiuhan Zhu, Shunan Guo, Zhuochen Jin, Yu-Ru Lin, and Nan Cao. 2018. Anomaly detection in spatiotemporal data via regularized non-negative tensor analysis. *Data Mining and Knowledge Discovery* (2018), 1–18.

[69] Xialing Lin, Kenneth A Lachlan, and Patric R Spence. 2016. Exploring extreme events on social media: A comparison of user reposting/retweeting behaviors on twitter and weibo. *Computers in Human Behavior* 65 (2016), 576–581.

[70] Wei Liu, Yu Zheng, Sanjay Chawla, Jing Yuan, and Xie Xing. 2011. Discovering spatio-temporal causal interactions in traffic data streams. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1010–1018.

[71] Cewu Lu, Jianping Shi, and Jiaya Jia. 2014. Abnormal Event Detection at 150 FPS in MATLAB. In *IEEE International Conference on Computer Vision*. 2720–2727.

[72] Michael Madaio, Shang-Tse Chen, Oliver L Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Duen Horng Chau, and Bistra Dilkina. 2016. Firebird: Predicting fire risk and prioritizing fire inspections in atlanta. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 185–194.

[73] Michael Madaio, Oliver L Haimson, Wenwen Zhang, Xiang Cheng, Matthew Hinds-Aldrich, Bistra Dilkina, and Duen Horng Polo Chau. 2015. Identifying and prioritizing fire inspections: a case study of predicting fire risk in Atlanta. In *Bloomberg Data for Good Exchange Conference, New York City, NY, USA*, Vol. 28.

[74] Vijay Mahadevan, Weixin Li, Viral Bhalodia, and Nuno Vasconcelos. 2010. Anomaly detection in crowded scenes. In *Computer Vision and Pattern Recognition*. 1975–1981.

[75] R. Mehran, A. Oyama, and M. Shah. 2009. Abnormal crowd behavior detection using social force model. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 935–942.

[76] Cheol Oh, Jun-Seok Oh, and Stephen G Ritchie. 2005. Real-time hazardous traffic condition warning system: Framework and evaluation. *IEEE Transactions on Intelligent Transportation Systems* 6, 3 (2005), 265–272.

[77] Sangmin Oh, A. Hoogs, A. Perera, N. Cuntoor, Chia Chih Chen, Jong Taek Lee, S. Mukherjee, J. K. Aggarwal, Hyungtae Lee, and L. Davis. 2011. A large-scale benchmark dataset for event recognition in surveillance video. In *IEEE International Conference on Advanced Video and Signal Based Surveillance*. 527–528.

[78] Bei Pan, Yu Zheng, David Wilkie, and Cyrus Shahabi. 2013. Crowd sensing of traffic anomalies based on human mobility and social media. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 344–353.

[79] Linsey Xiaolin Pang, Sanjay Chawla, Wei Liu, and Yu Zheng. 2011. On mining anomalous patterns in road traffic streams. In *International conference on advanced data mining and applications*. Springer, 237–251.

[80] Claudio Piciarelli, Christian Micheloni, and Gian Luca Foresti. 2008. Trajectory-Based Anomalous Event Detection. *IEEE Transactions on Circuits & Systems for Video Technology* 18, 11 (2008), 1544–1554.

[81] Oluwatoyin P. Popoola and Kejun Wang. 2012. Video-Based Abnormal Human Behavior Recognition—A Review. *IEEE Transactions on Systems Man & Cybernetics Part C* 42, 6 (2012), 865–878.

[82] Eric Potash, Joe Brew, Alexander Loewi, Subhabrata Majumdar, Andrew Reece, Joe Walsh, Eric Rozier, Emile Jorgenson, Raed Mansour, and Rayid Ghani. 2015. Predictive modeling for public health: Preventing childhood lead poisoning. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2039–2047.

[83] Gyan Ranjan, Hui Zang, Zhi-Li Zhang, and Jean Bolot. 2012. Are call detail records biased for sampling human mobility? *ACM SIGMOBILE Mobile Computing and Communications Review* 16, 3 (2012), 33–44.

[84] Mahdyar Ravanbakhsh, Moin Nabi, Hossein Mousavi, Enver Sangineto, and Nicu Sebe. 2016. Plug-and-Play CNN for Crowd Motion Analysis: An Application in Abnormal Event Detection. (2016), 1689–1698.

[85] Mahdyar Ravanbakhsh, Moin Nabi, Enver Sangineto, Lucio Marcenaro, Carlo Regazzoni, and Nicu Sebe. 2017. Abnormal Event Detection in Videos using Generative Adversarial Nets. (2017).

[86] Honglei Ren, You Song, Jingwen Wang, Yucheng Hu, and Jinzhi Lei. [n. d.]. A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction. ([n. d.]).

[87] Sam T Roweis and Lawrence K Saul. 2000. Nonlinear dimensionality reduction by locally linear embedding. *science* 290, 5500 (2000), 2323–2326.

[88] Polina Rozenshtein, Aris Anagnostopoulos, Aristides Gionis, and Nikolaj Tatti. 2014. Event detection in activity networks. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1176–1185.

[89] Mohammad Sabokrou, Mahmood Fathy, Mojtaba Hoseini, and Reinhard Klette. 2015. Real-time anomaly detection and localization in crowded scenes. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 56–62.

[90] Mohammad Sabokrou, Mohsen Fayyaz, Mahmood Fathy, and Reinhard Klette. 2016. Fully Convolutional Neural Network for Fast Anomaly Detection in Crowded Scenes. *CoRR* abs/1609.00866 (2016). arXiv:1609.00866 http://arxiv.org/abs/1609.00866

[91] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. 2010. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*. ACM, 851–860.

[92] Karen Simonyan and Andrew Zisserman. 2014. Two-Stream Convolutional Networks for Action Recognition in Videos. 1, 4 (2014), 568–576.

[93] Bhavkaran Singh Walia, Qianyi Hu, Jeffrey Chen, Fangyan Chen, Jessica Lee, Nathan Kuo, Palak Narang, Jason Batts, Geoffrey Arnold, and Michael Madaio. 2018. A Dynamic Pipeline for Spatio-Temporal Fire Risk Prediction. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 764–773.

[94] Xian Teng, Yu-Ru Lin, and Xidao Wen. 2017. Anomaly Detection in Dynamic Networks using Multi-view Time-Series Hypersphere Learning. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 827–836.

[95] Dimitrios Tomaras, Vana Kalogeraki, Nikolas Zygouras, and Dimitrios Gunopulos. 2017. An efficient technique for event location identification using multiple sources of urban data. In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Recommendations for Location-based Services and Social Networks*. ACM, 5.

[96] Jameson L Toole, Michael Ulm, Marta C González, and Dietmar Bauer. 2012. Inferring land use from mobile phone activity. In *Proceedings of the ACM SIGKDD international workshop on urban computing*. ACM, 1–8.

[97] Ionut Trestian, Supranamaya Ranjan, Aleksandar Kuzmanovic, and Antonio Nucci. 2009. Measuring serendipity: connecting people, locations and interests in a mobile 3G network. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement*. Acm, 267–279.

[98] Amin Vahedian, Xun Zhou, Ling Tong, Yanhua Li, and Jun Luo. 2017. Forecasting gathering events through continuous destination prediction on big trajectory data. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 34.

[99] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research* 11, Dec (2010), 3371–3408.

[100] Haiyang Wang, Xiaming Chen, Siwei Qiang, Honglun Zhang, Yongkun Wang, Jianyong Shi, and Yaohui Jin. 2016. Early Warning of City-Scale Unusual Social Event on Public Transportation Smartcard Data. In *Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences*. IEEE, 188–195.

[101] Youcheng Wang, Jian Xu, Ming Xu, Ning Zheng, Jinsheng Jiang, and Kaiwei Kong. 2016. A feature-based method for traffic anomaly detection. In *Proceedings of the 2Nd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*. ACM, 5.

[102] Samuel S Wilks. 1938. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* 9, 1 (1938), 60–62.

[103] Apichon Witayangkurn, Teerayut Horanont, Yoshihide Sekimoto, and Ryosuke Shibasaki. 2013. Anomalous event detection on large-scale gps data from mobile phones using hidden markov model and cloud platform. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. ACM, 1219–1228.

[104] George Y Wong and William M Mason. 1985. The hierarchical logistic regression model for multilevel analysis. *J. Amer. Statist. Assoc.* 80, 391 (1985), 513–524.

[105] Hao Wu, Weiwei Sun, and Baihua Zheng. 2017. A fast trajectory outlier detection approach via driving behavior modeling. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 837–846.

[106] Mingxi Wu, Xiuyao Song, Chris Jermaine, Sanjay Ranka, and John Gums. 2009. A LRT framework for fast spatial anomaly detection. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 887–896.

[107] Xian Wu, Yuxiao Dong, Chao Huang, Jian Xu, Dong Wang, and Nitesh V Chawla. 2017. UAPD: Predicting Urban Anomalies from Spatial-Temporal Data. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 622–638.

[108] SHI Xingjian, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*. 802–810.

[109] Chengcheng Xu, Andrew P Tarko, Wei Wang, and Pan Liu. 2013. Predicting crash likelihood and severity on freeways with real-time loop detector data. *Accident Analysis & Prevention* 57 (2013), 30–39.

[110] Chengcheng Xu, Wei Wang, and Pan Liu. 2013. A Genetic Programming Model for Real-Time Crash Prediction on Freeways. *IEEE Trans. Intelligent Transportation Systems* 14, 2 (2013), 574–586.

[111] Dan Xu, Elisa Ricci, Yan Yan, Jingkuan Song, and Nicu Sebe. 2015. Learning Deep Representations of Appearance and Motion for Anomalous Event Detection. (2015).

[112] Dan Xu, Yan Yan, Elisa Ricci, and Nicu Sebe. 2017. *Detecting anomalous events in videos by learning deep representations of appearance and motion.* Elsevier Science Inc. 117–127 pages.

[113] Yan Xu, Xi Ouyang, Yu Cheng, Shining Yu, Lin Xiong, Sugiri Pranata, Shengmei Shen, and Junliang Xing. 2018. Dual-Mode Vehicle Motion Pattern Learning for High Performance Road Traffic Anomaly Detection. In *CVPR2018 NVIDIA AI CITY CHALLENGE WORKSHOP*.

[114] Shiming Yang, Konstantinos Kalpakis, and Alain Biem. 2014. Detecting road traffic events by coupling multiple timeseries with a nonparametric bayesian method. *IEEE Transactions on Intelligent Transportation Systems* 15, 5 (2014), 1936–1946.

[115] Su Yang and Weihua Liu. 2011. Anomaly detection on collective moving patterns: A hidden markov model based solution. In *2011 IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing*. IEEE, 291–296.

[116] Su Yang and Wenbin Zhou. 2011. Anomaly detection on collective moving patterns: Manifold learning based analysis of traffic streams. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on*. IEEE, 704–707.

[117] Rongjie Yu, Mohamed A Abdel-Aty, Mohamed M Ahmed, and Xuesong Wang. 2014. Utilizing microscopic traffic and weather data to analyze real-time crash patterns in the context of active traffic management. *IEEE Transactions on Intelligent Transportation Systems* 15, 1 (2014), 205–213.

[118] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 186–194.

[119] Zhuoning Yuan, Xun Zhou, and Tianbao Yang. 2018. Hetero-ConvLSTM: A Deep Learning Approach to Traffic Accident Prediction on Heterogeneous Spatio-Temporal Data. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 984–992.

[120] Desheng Zhang, Jun Huang, Ye Li, Fan Zhang, Chengzhong Xu, and Tian He. 2014. Exploring human mobility with multi-source data at extremely large metropolitan scales. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, 201–212.

[121] Daqing Zhang, Nan Li, Zhi-Hua Zhou, Chao Chen, Lin Sun, and Shijian Li. 2011. iBAT: detecting anomalous taxi trajectories from GPS traces. In *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 99–108.

[122] Huichu Zhang, Yu Zheng, and Yong Yu. 2018. Detecting Urban Anomalies Using Multiple Spatio-Temporal Data Sources. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 54.

[123] Jianting Zhang. 2012. Smarter outlier detection and deeper understanding of large-scale taxi trip records: a case study of NYC. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. ACM, 157–162.

[124] Tianzhu Zhang, Hanqing Lu, and Stan Z. Li. 2009. Learning semantic scene models by object classification and trajectory clustering. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. 1940–1947.

[125] Wangsheng Zhang, Guande Qi, Gang Pan, Hua Lu, Shijian Li, and Zhaohui Wu. 2015. City-scale social event detection and evaluation with taxi traces. *ACM Transactions on Intelligent Systems and Technology (TIST)* 6, 3 (2015), 40.

[126] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 3 (2014), 38.

[127] Yu Zheng, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. 2008. Understanding mobility based on GPS data. In *Proceedings of the 10th international conference on Ubiquitous computing*. ACM, 312–321.

[128] Yu Zheng, Furui Liu, and Hsun-Ping Hsieh. 2013. U-Air: When urban air quality inference meets big data. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1436–1444.

[129] Yu Zheng, Tong Liu, Yilun Wang, Yanmin Zhu, Yanchi Liu, and Eric Chang. 2014. Diagnosing New York city's noises with ubiquitous data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 715–725.

[130] Yu Zheng, Huichu Zhang, and Yong Yu. 2015. Detecting collective anomalies from multiple spatio-temporal datasets across different domains. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, 2.

[131] Xi Zhu and Diansheng Guo. 2017. Urban event detection with big data of taxi OD trips: A time series decomposition approach. *Transactions in GIS* 21, 3 (2017), 560–574.

[132] Yingying Zhu, Nandita M. Nayak, and Amit K. Roy-Chowdhury. 2013. Context-Aware Activity Recognition and Anomaly Detection in Video. *Journal of Selected Topics in Signal Processing* 7, 1 (2013), 91–101.

[133] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. 2008. Maximum Entropy Inverse Reinforcement Learning.. In *AAAI*, Vol. 8. Chicago, IL, USA, 1433–1438.