A Review of the Recent Progress on Crowd Anomaly Detection

Sarah Altowairqi, Suhuai Luo, Peter Greer
School of Information and Physical Sciences, The University of Newcastle, Newcastle, Australia

***Abstract*—Surveillance videos are crucial in imparting public security, reducing or avoiding the accidents that occur from anomalies. Crowd anomaly detection is a rapidly growing research field that aims to identify abnormal or suspicious behavior in crowds. This paper provides a comprehensive review of the state-of-the-art in crowd anomaly detection and, different taxonomies, publicly available datasets, challenges, and future research directions. The paper first provides an overview of the field and the importance of crowd anomaly detection in various applications such as public safety, transportation, and surveillance. Secondly, it presents the components of crowd anomaly detection and its different taxonomies based on the availability of labels, and the type of anomalies. Thirdly, it presents the review of the recent progress of crowd anomaly detection. The review also covers publicly available datasets commonly used for evaluating crowd anomaly detection methods. The challenges faced by the field, such as handling variability in crowd behavior, dealing with large and complex data sets, and addressing the imbalance of data, are discussed. Finally, the paper concludes with a discussion of future research directions in crowd anomaly detection, including integrating multiple modalities, addressing privacy concerns, and addressing crowd monitoring systems' ethical and legal implications.**

Keywords—Crowd anomaly detection; advanced computer science; intelligent systems; video surveillance application; machine learning

I. Introduction

The field of crowd anomaly detection is rapidly growing, with increasing interest in identifying abnormal or suspicious behavior in crowds. Crowds can be found in various settings, such as public gatherings, transportation hubs, and shopping centers. The ability to detect anomalies in crowds has many important applications, including public safety, transportation management, and surveillance. The variety and frequency of indoor and outdoor activities that draw big crowds have been fast expanding, which has raised the likelihood of illegal gatherings, disturbance situations, mass stampedes, and other anomalous events greater than before [1]. The size and density of the crowds at huge gatherings have resulted in several public safety crises in recent years, such as the famous stampede on New Year's Eve of 2014 at Shanghai Bund group [2], Mina stampede during Hajj 2015 [3], etc. Surveillance applications are becoming more crucial for efficient crowd-control analysis. Such video surveillance systems must watch for unexpected crowd behavior, like crowd instability or chaos.

Video surveillance should be able to identify violent altercations or traffic accidents quickly and accurately. The effectiveness of traditional methods is significantly constrained by the amount of human effort needed to make judgments manually. However, there is a rising need to express desired irregularities in an automatic and understandable manner as activities become more complicated and there are more possibilities to reason about. It is critical to comprehend the distinction between a crowd and a group. A group is any gathering of people for social interaction that can range in size from a few to many. On the other hand, a crowd is a group of people who have congregated in an uncontrolled or organized way for similar or dissimilar reasons. A crowd might thus be modelled in terms of many aspects, such as size, cohesiveness, structure, period, motive, and closeness. According to [4], the crowd is either structured or unstructured, the structured crowd moves in the same direction, speed, and pattern as in a marathon, a line of people waiting, or people using an escalator, etc. But the unstructured crowd exhibits complete uncertainty in their behaviors. Such crowds can be seen in marketplaces, shopping malls, etc., where the behavior is entirely uncertain.

In summary, this paper provides a comprehensive overview of crowd anomaly detection (CAD), beginning with a brief explanation of CAD systems and their importance in the research field. In Section III, we delve into the typical components of a CAD system, and the different taxonomies of CAD systems. The typical components include object density estimation, object tracking and object behavior analysis. Major taxonomies of crowd anomaly detection systems are presented based on the label availability and anomaly types used to categorize anomalies. Section IV provides an overview of recent advancements in crowd anomaly detection, while Section V presents a detailed analysis of the performance of various existing CAD systems and techniques. In Section VI, we outline publicly available datasets that can be used for crowd anomaly detection research. The challenges faced by existing systems are discussed in Section VII, including accuracy, computational complexity, and handling real-world scenarios. Finally, in Section VIII, we conclude with a summary of the key findings and potential future directions for research in this field. By the end of this paper, readers will have a comprehensive understanding of the current state of the art in crowd anomaly detection and will be equipped with the necessary knowledge to overcome existing challenges and push the boundaries of this research field even further.

II. Crowd Anomaly Detection (CAD)

Crowd anomaly detection (CAD) is the process of understanding the overall characteristics of a crowd in a video, such as density, flow, and demographic information. The crowd’s density is the number of people per unit area and is used to measure the congestion level in a crowd. The flow of a crowd is the direction and speed of movement of people and is used to measure the level of mobility in a crowd. Several techniques have been developed to detect crowd anomaly in recent years. These techniques include image processing, computer vision, and machine learning. Image processing techniques, such as background subtraction and blob detection, are used to extract information about the density and flow of a crowd from video footage. Computer vision techniques, such as object detection and tracking, are used to extract information about the demographic characteristics of a crowd. Machine learning techniques, such as neural networks and, recently, deep learning models, are used to analyze the extracted information and make predictions about the crowd.

Anomaly detection in crowds can further be defined as identifying specific individuals or groups of people behaving abnormally, such as loitering, running, or moving against the flow of the crowd. This method involves tracking the movement of individuals in a crowd using computer vision techniques such as object detection and object tracking. Once an individual is tracked, their behavior can be analyzed to determine if it is normal or abnormal. Recently, anomaly detection and its analysis in social crowds have become a significant area of research. Due to the variety of anomalous events, crowd anomaly detection is a practical and challenging topic for computer vision. Automatic security analysis of crowd behavior is now possible when there are odd crowds or anomalous congestion. Because of activities like terrorist activities, fights, strange and suspicious movements, etc. automated detection of abnormal behavior in the crowd is of utmost relevance. In traditional systems, becomes the operators’ responsibility to supervise the security surveillance to ensure safety closely. This is a significant challenge, resulting in costly and inaccurate decision-making. Therefore, creating a system that is free from errors and without any fatigue, providing real-time functionality, will have sufficient effects on managing crowd behavior.

The emergence of several sophisticated algorithms and the availability of high computational powers heightened the quantity and quality of the research in crowd anomaly detection. Computer vision algorithms that utilise image processing, machine learning and pattern recognition depict the challenges in crowd behavior analysis [5-7]. Some of its most crucial applications are crowd control, video surveillance, and the design of intelligent public spaces [8-14]. The intelligent environment, which is essential for public safety, can help to redirect the crowd and help the planner to design the public area with the most available space [4].

The increased number of research publications in the top publications related to crowd anomaly detection indicates the growing interest and demand in this field [5]. Fig. 1 presents the number of recent papers published on crowd anomaly detection. Crowd anomaly detection (CAD) is identifying unusual or abnormal behavior in a crowd using data analysis techniques, typically with the help of video cameras, sensors, or other monitoring devices. The goal of crowd anomaly detection is to identify situations that may pose a risk to public safety or security, such as the presence of suspicious individuals or activities, overcrowding, or potential hazards. In order to achieve this goal, crowd anomaly detection systems employ machine learning algorithms and computer vision techniques to analyze data sources in real-time. By recognizing typical patterns of behavior, such as standing, sitting, or walking in specific areas, these algorithms can be trained to detect any deviations from the norm. In such cases, the system sends an alert to relevant authorities or security personnel, who can take the necessary steps to investigate the matter.

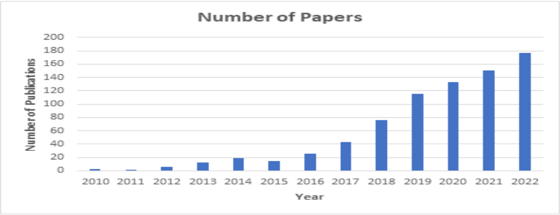

Fig. 1. Number of papers published on crowd anomaly detection.

III. Component and Taxonomy of Crowd Anomaly Detection

The general structure and flow of a crowd anomaly detection system is shown in Fig. 2. The anomaly detection system starts with the raw video data collected by the CCTV cameras. The sensor data are then pre-processed using various methods to lower signal noise and go through a feature extraction process. These features might include things like color, texture, motion, or shape. The goal is to identify patterns in the video that can help distinguish normal behavior from anomalous behavior.

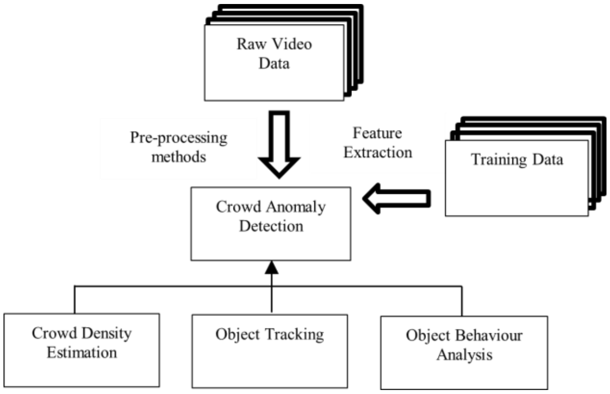

Fig. 2. Flow and structure of crowd anomaly detection.

Crowd anomaly detection system typically consists of three main components: crowd density estimation, object tracking, and object behavior analysis. Here is an overview of how these components work together. Crowd density estimation: The first step in crowd anomaly detection is to estimate the density of people in the scene. This can be done using various computer vision techniques such as background subtraction, foreground detection, or optical flow. The goal is to identify the regions of the scene where people are present and estimate the number of people in each region. Object tracking: Once the crowd density has been estimated, the next step is to track individual objects (i.e., people) in the scene over time. This is typically done using a tracking algorithm that assigns a unique ID to each object and updates its position as it moves through the scene. Many different tracking algorithms are available, but some common techniques include Kalman filtering, particle filtering, and graph-based tracking. Object behavior analysis: Once individual objects have been tracked, the next step is to analyze their behavior to detect anomalies. This typically involves comparing the behavior of each object to some predefined normal behavior model. For example, the system might look for objects that are moving in an unusual direction, moving faster or slower than expected, or loitering in one area for an extended period. There are many different techniques for object behavior analysis, including rule-based methods, machine learning, and anomaly detection algorithms. The overall flow of the system is typically iterative, with the crowd density estimation and object tracking components running continuously in real time. The object behavior analysis component typically runs periodically (e.g. every few seconds) to detect any anomalies in the behavior of the tracked objects. When an anomaly is detected, the system can generate an alert or trigger some other action (e.g. turning on lights, sounding an alarm, or notifying security personnel).

Crowd density estimation is an important area of research with practical applications in managing and monitoring crowds in density populated locations such as subway stations, sports stadiums, and convention centers. With increasing population and urbanization, it is common for a large crowd to gather quickly. Precisely predicting the emergence of crowds and gauging their density is crucial for effective event planning and crowd management. The recent COVID-19 pandemic further highlights the importance of crowd density estimation, as social distancing policies were implemented to prevent the spread of the virus [15]. There are two primary methods for crowd density estimation: counting objects and estimating the density map [16]. CNN-based algorithms are preferred due to their better image and video sequence performance. A generic deep learning model for the automatic feature extraction from crowd scenes for crowd anomaly detection has been shown in Fig. 3. Techniques based on CNN, such as Scale Pyramid Module Network [17] and Attention Networks [18], are being used for crowd density estimation and counting. Attention Networks are capable of counting individuals in photos while considering scales by selecting appropriate global and local scales using the attention mechanism. Tracking the crowd is crucial in CAD systems, as it involves analyzing image sequences to determine the motion and trajectory of objects, specifically pedestrians. The process begins with detecting objects in a video and filtering them for tracking. The monitoring of pedestrian movement is an essential aspect of understanding crowd behavior. Object tracking can be challenging, as it requires following one or more objects over time. Automated tracking systems are needed to keep up with the movement of pedestrians in a crowd. Identifying and defining regions of interest (ROIs) is the first and most important step in this process. This can be difficult due to various factors, such as the camera's view, orientation, and resolution. Once the ROI features have been extracted, the tracker can then follow the object of interest.

These models receive either supervised or unsupervised training. In order to take the necessary action, such as dispersing the crowd, when the crowd density exceeds a predetermined threshold, the degree of congestion can be calculated. The objects are tracked, and the anomaly is analyzed by utilizing the objects under tracking and their behavior. A final decision can be made in real-time whether the crowd state is normal or abnormal.

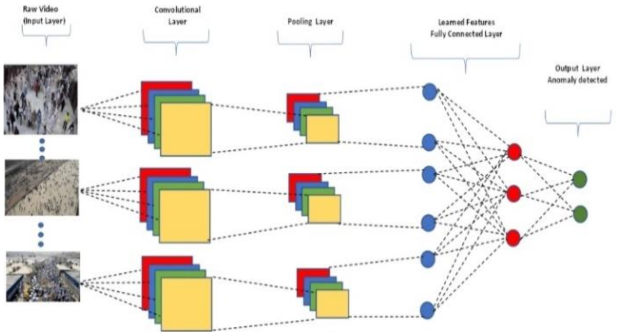
Raw Video (Input Layer)
Convolutional Layer
Pooling Layer
Learned Features Fully Connected Layer
Output Layer Anomaly detected

Fig. 3. A deep learning model for the automatic feature extraction from crowd scenes for crowd anomaly detection.

The research on crowd anomaly detection can be categorized into two types based on the availability of labels and the type of anomalies. The different types of crowd anomaly detection methods and related works on them are given in the following subsections.

A. CAD based on the Availability of Labels

Based on the availability of labels, the models used for crowd anomaly detection are divided into supervised/semi-supervised, unsupervised, hybrid, and one-class neural networks, as shown in Fig. 4. Supervised learning: This type of anomaly detection relies on labelled data, where the anomalies and normal behavior are defined beforehand. The model is trained on this labelled data, which can then be used to detect anomalies in new, unseen data [15], [16]. Semi-supervised learning: This type of anomaly detection also relies on labelled data, but it also uses unlabeled data to enhance the model's performance. The model is trained on labelled and unlabeled data, which can then be used to detect anomalies in new, unseen data. Unsupervised learning: This type of anomaly detection does not rely on labelled data. Instead, it uses techniques like clustering and dimensionality reduction to identify patterns in the data. Any deviation from these patterns can then be considered an anomaly. Given that they use labelled data, supervised anomaly detection approaches outperform unsupervised ones in terms of performance. From a series of annotated data instances (training), supervised anomaly detection learns the separation border. Using the learned model, it then divides a test instance into normal and anomalous classes (testing).

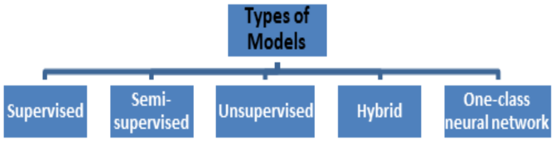

Fig. 4. Crowd anomaly detection based on the availability of labels.

Overviews of deep learning-based semi-supervised algorithms for anomaly identification were done in [15] and [6]. Unsupervised crowd anomaly detection is a significant study area, both for fundamental machine learning research and practical applications. One of the deep core architectures for unsupervised anomaly detection is autoencoders, as described in [17]. Discriminative boundary surrounding the majority class is learned by different anomaly detection techniques, as stated by Perera and Patel in their works [18], [19]. Any test instance that falls outside of this boundary is considered anomalous. To discover robust features, deep learning models are frequently utilized as feature extractors [20]. The hybrid models use two-step learning and are demonstrated to provide cutting-edge outcomes [21]. Robust characteristics are retrieved from the deep neural network's hidden layers to help separate them from the irrelevant features that can hide the existence of abnormalities. In deep hybrid models, one class Radial Basis Function (RBF) and Support Vector Machine (SVM) classifiers are used as inputs instead of more conventional techniques. A robust anomaly detection model needs both an anomaly detector and a feature extractor to be built on complicated, high-dimensional domains. One-class neural networks (OC-NN) combine deep network capabilities to extract more rich representations of data with a one-class objective, such as a hyperplane [22] or hypersphere [23], to distinguish all the typical data points from the outliers. Data representation in the hidden layer is learned by optimizing the objective function designed for anomaly detection. The experimental findings in [22] show that OC-NN may achieve equivalent or higher performance than current state-of-the-art approaches for complex datasets while having reasonable training and testing times in comparison to the existing methods.

B. CAD based on the Type of Anomaly

Crowd anomaly detection techniques can also be categorized based on the type of anomaly they handle. Mainly three types of anomalies are handled: point, contextual or conditional, and collective or group, as shown in Fig. 5. Most of the literature is devoted to pointing out anomalies. Point anomalies frequently signify an irregularity or deviation that occurs at random and may not have a specific meaning. A data instance that might be regarded as anomalous in a certain context is known as a contextual anomaly, also known as a conditional anomaly. By considering contextual and behavioral variables, a contextual abnormality is found. Time and space are two contextual elements that are frequently used. In contrast, the behavioral characteristics could be a pattern of financial spending, the occurrence of system log events, or any characteristic that characterizes typical behavior. Collective or group anomalies are abnormal groups of individual data points where each individual point, when viewed separately, appears to be a typical instance of data but, when observed collectively, exhibit an unexpected characteristic.

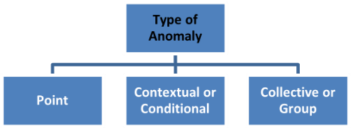

Fig. 5. Crowd anomaly detection based on the type of anomaly.

IV. Recent Progress on Crowd Anomaly Detection

In recent years, there has been a growing body of literature on crowd anomaly detection. Studies focus on various techniques and applications, including computer vision-based approaches, machine learning algorithms, and real-time anomaly detection. In this section, we review the most recent works on crowd anomaly detection and highlight their key findings and contributions to the field. The field of crowd anomaly detection and analysis has seen significant growth in recent years as the importance of understanding and managing crowd behavior has become increasingly apparent. With the proliferation of surveillance cameras and social media, there is a vast amount of data available for analyzing crowd behavior. As a result, numerous studies have been conducted on various aspects of crowd anomaly detection. However, with so much literature available, it can be difficult to gain a comprehensive understanding of the current state of research in this field. This literature review aims to address this issue by providing a comprehensive overview of the current state of research in crowd anomaly detection and analysis. By combining all available works on this topic, this literature review will provide a holistic view of the field, highlighting the most important contributions and identifying gaps in existing literature. A comprehensive examination of the various techniques used in the visual analysis of crowd behavior for surveillance purposes was conducted in [6]. In this study, recent research on crowd anomaly detection was classified based on the level of analysis and the types of anomalies observed. Several machine learning and deep learning approaches have been proposed to analyze crowd anomaly detection. A review of the different classical methods, such as the Spatial-Temporal Technique (STT), optical flow, Gaussian mixture model (GMM) and Hidden Markov model (HMM), has been done for the identification of the crowd abnormal behavior in [13]. Many recent studies in crowd anomaly analysis utilise deep learning methods. The different attributes of Convolution Neural Network (CNN) and the various optimization methods used in the context of crowd behavior analysis have been presented in [24]. Another literature review on intelligent video surveillance using various deep learning techniques for crowd detection was conducted in [17]. This work examined the use of Long-Short Term Memory (LSTM) networks, VGG16, and YOLO specifically. Additionally, [18] also conducted a detailed study of the various deep learning methods used for crowd counting and analysis, which are key components of crowd anomaly detection. Different methodologies utilized for counting the crowd have been reviewed and compared, presenting the various trends in CNN and traditional machine learning methods [25, 26].

Aggregation of Ensembles (AOE), a combination of four classifiers over sub-ensembles of three tuned Convolutional Neural Networks (CNNs) on crowd datasets, is proposed by [27] as a method to detect abnormalities in movies of crowded scenes via majority voting. A different classifier is used to process each sub-ensemble independently. A sub-ensemble of CNNs is composed of the CIFAR-10 AlexNet [28], GoogleNet [29], and VGGNet [30] networks. In this situation, CNNs acted as feature extractors, feeding Linear SVM, Quadratic SVM, Cubic SVM, and a SoftMax classifier and other classifiers. Several video frames are selected and analyzed to extract the features. A video is deemed abnormal if more than 10% of the batch's frames fall into this category. The publicly available datasets such as the Avenue dataset, UCSD Ped 1 and UCSD Ped 2, and AOE training and evaluation, were employed for this analysis.

A dual branch network was proposed in [35] as a framework for social multiple-instance learning. This approach uses a two- stream neural network consisting of an interactive dynamic stream and a spatiotemporal stream. The spatiotemporal stream processes RGB video clips. A 3D ConvNets (C3D) model that has already been trained on Kinetics and UCF-101 is used to extract features from the video after it has been divided into smaller pieces using a video segmentation method. The output is fed into a fully linked network after the features have been fed into a one-dimensional dependency attention capture module. The second channel receives force maps of social interactions as input. Maps of social interactions in a scene are created using the social force model [14]. They evaluate the effectiveness of their strategy on the UCF Crime dataset using the receiver operating characteristic (ROC) curve and area under the curve (AUC) metrics [31]. Comparisons with four more techniques were made throughout the evaluation. The dynamic network allows for a compact representation of moving video frames, which reduces false-positive anomaly alarms due to spatial limitations. This is the second advantage of their use. They employ the perturbation visual interpretation technique for identifying anomalies in order to give the results more credibility. The results presented were competitive with many of the similar works.

A general adversarial network-based abnormal behavior analysis in the massive crowd was proposed in [53], where a case study of the Hajj pilgrimage is considered. In this work, the dynamic features were extracted using optical flow. It uses a transfer learning strategy and U-Net and Flownet to distinguish large crowd scenes' normal and abnormal behaviors. This system has shown a very high accuracy in smaller video scenes such as UMN and UCSD, with 99.4% and 97.1%, respectively, whereas a considerably low accuracy on large datasets such as the Hajj dataset. The authors describe the need for improvement in anomaly behavior accuracy by collecting more annotated Hajj datasets and extracting complex features that utilise deeper models.

Behavior understanding becomes a difficult task with the fact that anomalies are not well defined and would occur very less frequently. Researchers have been trying to address these issues to make the learning algorithms robust in detecting the anomaly in the video. In [32], a new approach based on Generative Cooperative Learning (GCL) has been employed for addressing the low frequency of anomalies and contributing towards the avoidance of manual annotations. The generator and discriminator networks in the model get trained in a cooperative style, thus enabling unsupervised learning. This approach has been found to be showing consistent improvement in UCF crime and ShanghaiTech, which are considered to be the well-cited large-scale video datasets.

A deep convolutional neural network (DCNN) architecture was proposed in [21] for detecting crowd anomalies. The architecture utilized the VGG16 model, which included ten convolutional layers and three max pooling layers. For crowd counting, they employed six convolutional layers with a dilation rate of 2 and a kernel size of 3x3. Another approach was proposed by Sagar [22], who presented a network architecture for crowd counting using a feature extractor based on ResNet. It extracts the details of an object at different scales by down-sampling the block with dilated convolutions and further up-samples the block using transposed convolution. In the analysis of crowd dynamics, Recurrent Neural Networks (RNNs) based models have been utilized [23]. In this work, the Bhattacharya distance was applied to detect a given frame's emotional state order to select the optimal keyframe for the video. To describe the scene, the Space-Time Interest Points (STIP) descriptor was used, with features being extracted from the keyframes. The RNN model was trained using an improved version of the Butterfly optimization algorithm, which enables it to distinguish between normal crowd behavior and behaviors associated with fighting, fleeing, walking, anger, happiness, and violence. The problem of vanishing and exploding gradients is addressed by Long-Short Term Memory (LSTM) networks [24], which are an extension of Recurrent Neural Networks (RNNs). LSTM has a longer memory capacity and can retain information for an extended period.

Most crowd anomaly detection robustness is analyzed based on the temporal consistency among the frames. The temporal features were considered for anomaly prediction based on the motion information using optical flow analysis [33, 34]. But there should be some system to distinguish between fake and real sequences for temporal consistency that lead to an anomaly or normal behavior. The anomaly behavior can be well detected by the optical flow methods, which are good for analyzing the short-term temporal relationship between adjacent frames. But eventually, these methods fail, especially in videos where events based on a long-term temporal relationship occur. To overcome such issues, a novel method based on a bi-directional architecture that introduces the inconsistencies on three different levels such as temporal-sequence, cross-modal, and pixel, has been proposed [35]. The bidirectional predictive network introduced in this work regularizes the predictive consistency. The long-term temporal relationship in the video sequences is identified by the discriminator developed in work. This method outperformed all other state-of-the-art learning methods on several datasets such as UCSD Pred2, ShanghaiTech and CUHK Avenue. Recent literature reviews have been conducted on the topic of crowd anomaly detection, providing in-depth coverage of the various frameworks, taxonomies, methods, and techniques used. These reviews also included information on various datasets, such as the Hajj dataset used for video surveillance during the Hajj pilgrimage in Saudi Arabia. The Hajj pilgrimage is known to be the largest human gathering in the world, with an estimated of 2.5 to 3 million participants from various regions globally [9].

A summary of the different types of crowd anomaly detection is given in Table I.

TABLE I. SUMMARY OF THE CROWD ANOMALY DETECTION METHODS

Approach	Anomaly	Dataset	Performance
SSD-VGG16 [15]	Bullet train, pedestrian	PASCAL VOC, Railway	Overall accuracy= 98.01% Detection accuracy=99.55%
SSD-VGG16 [16]	Small object	ILSVRC CLS-LOC, Railway	Accuracy =96.6%

3D-CNN LSTM [36]	Panics, fighting, protest	UMN,
		CAVIA,
		Web
			Accuracy=0.995%
			Accuracy= 0.974%
			Accuracy= 0.926%

CNN RNN [37]	Use mobile in class, fighting, fainting	KTH, CAVIAR	Accuracy=87.15%

CNN Residual LSTM [38]	Fighting, explosion, accidents, shooting, robbery, shoplifting,	UCF-Crime,
		UMN,
		CUHK Avenue
			Accuracy=78.43 %
			Accuracy=98.20 %
			Accuracy=98.80%

GAN [39]	Biking, fighting, vehicle, running	CUHK Avenue
		UCSD,
		ShanghaiTech Campus.
			AUC=86.6%
			AUC=96.9%
			AUC=82.5%
			AUC=73.8%

Optical Flow GAN [40]	Standing, sitting, sleeping, running, moving in opposite, non-pedestrian	UMNScence1
		UMNScence2
		UMNScence3
		UCSD,
		HAJJ datasets
			Accuracy=99.4%
			Accuracy=97.1%,
			Accuracy=97.6%,
			Accuracy=89.26%
			Accuracy= 79.63%

Convolutional Neural Networks (CNNs) and Random Forests (RFs) [41]	standing, running, moving in opposite or different crowd directions, and non-pedestrian entities	UMN,
		UCSD,
		HAJJv2 dataset
			Accuracy= 99.77%
			Accuracy= 93.71%.
			Accuracy=76.08%.

Convolutional Neural Network (CNN) [42]	Wheelchairs, skateboarers, motor vehicles, bicycles and crossing pedestrian tracks.	Violent Flows,
		UCSD,
		CUHK Avenue
			Accuracy =90%
			Accuracy= 99.98%
			Accuracy=95%

Convolutional Long–Short-Term Memory (ConvLSTM) network and a Convolutional AutoEncoder. [43]	cyclists, skaters, cars	UCSD ped2,
		Shanghai Tech Campus.
			AUC=95.6%
			AUC=73.1%

3DConv, Convolution Long Short-Term Memory (ConvLSTM) [44]	Vehicle and bicycle movement, throwing objects, running, Arrest, Abuse, Accident, Burglary, Explosion.	UCSD Ped1,
		UCSD Ped2,
		Avenue
		UCF-crime dataset
			AUC=80.7%
			AUC=85.3%
			AUC=81.0%
			AUC=75.82%

CNN, RNN KNN, Optical Flow [45]	Bicycles, skateboards, wheelchairs	CUHK Avenue
		UCSD,
		UR fall
		Shanghai Tech Campus,
			AUC=80.68%
			AUC=96.01%
			AUC=91.28% when k=10
			AUC= 0.703 Optical flow module

Conv-LSTM [46]	Violence	Standard crowd anomaly	Accuracy =95.16%
Vgg-16 and LSTM [47]	Non-pedestrian	UCSD Ped2 CUHK Avenue	frame level: 95.0%, pixel level: 72.5% frame level:87.3%, pixel level: 93.8%

Cascaded attention model, Two Convolutional layers, Adam [48]	Fighting, Running, Robbery, lying down, crossing and car accident.	UCSD Ped2
		CUHK Avenue,
		Shanghai Tech Campus
			AUC =0.974,
			AUC=0.867,
			AUC= 0.736

3DCN, Transformer Adam [49]	Fighting, Running, Burglary, Fire and Assault.	Shanghai Tech Campus,
		UCF-Crime
			AUC = 0.976
			AUC = 0.832

GCN technique [50]	Fighting, Running, Burglary, Fire, Abnormal walking, lying down, group gathering and Assault.	UCSD, UCF-Crime,
		Shanghai Tech campus.
			AUC =0.93,
			AUC=0.82
			AUC=0.84.

motion attention, location attention. [51]	Fighting, Running, Burglary, Fire, Abnormal walking, lying down, group gathering, Assault, theft and Explosion.	UCSD Ped1,
		UCSD Ped2,
		CUHK Avenue,
		Shanghai Tech campus, UMN,
		Street Scene.
			AUC=0.942,
			AUC=0.929,
			AUC=0.805,
			AUC=0.803,
			AUC=0.988
			AUC=0.730.

V. PERFORMANCE COMPARISON

There are advantages and disadvantages to the crowd anomaly detection models that have been put forth so far. Most models just provide the output and estimate the findings in terms of accuracy, sensitivity, and specificity, failing to address the issue of output uncertainty. One of the issues that various anomaly detection techniques in a video frameset have in common is that they don't look at various situations including computational cost, pixel occlusion, noise and efficiency. To evaluate the models' performance and contrast it with other approaches, the following factors were considered:

1) The first is an analysis based on the amount of time required to run the algorithm for model estimation and a cost analysis based on an estimate of the overall expenses associated with evaluation and error analysis.

2) Analysis of the uncertainty based on the dispersion of mean squares of error in different iterations and the average weight estimate of precision and recall on evaluating the performance.

3) Investigating the sensitivity to noise based on the classification of crowd behaviour, particularly in the presence of artifacts such as noise, a changing temperature, pixel occlusion, and low received frame quality.

4) Generalizations of the approaches for identifying and classifying person and group behavior in unseen frames.

Performance comparison of the most common datasets used in the CAD with different methods was shown in this section. The datasets used for the comparison are UMN [14], UCSD (University of California, San Diego) and UCSD Ped [53]. Performance comparison of UCSD on six different methods has been shown in the Fig. 6. CNN has the highest accuracy, with a score of 99.98%. GAN and CNN+KNN+Optical flow methods also have high accuracy, with scores of 96.90% and 96.01%, respectively. CNN+RF has a moderate accuracy score of 93.71%, while Optical Flow GAN has the lowest accuracy score of 89.25%.

In Table II, a performance comparison of UMN and UCSD Ped2 datasets are given. It can be observed that on the UMN dataset, the combination of CNN and random forest method achieved the highest performance, while on the UCSD Ped2 dataset; the Cascaded Attention CNN achieved the highest performance. It is also interesting to note that the UMN-Method achieved a higher performance than any of the methods applied to the UCSD Ped2 dataset.

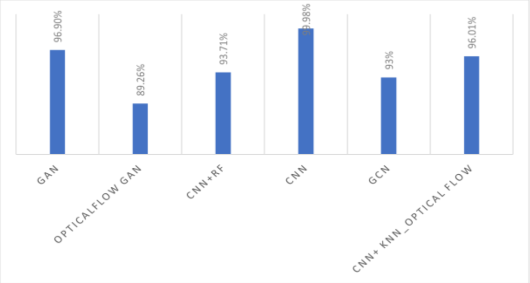

Fig. 6. Performance comparison of different methods on UCSD.

TABLE II. PERFORMANCE COMPARISON ON UCSD AND UMN DATASETS

UMN-Method	**3D-CNN-LSTSM**	**CNN- Residual LSTM**	**Optical Flow GAN**	**CNN +RF**
Performance	99.50%	98.20%	98.03%	99.77%
UCSD- Ped2 Method	Conv LSTM	Conv LSTM+ Conv Encoder	Vgg-16 LSTM	Cascaded Attention CNN
Performance	83.00%	95.20%	95%	97.40%

IV. PUBLICLY AVAILABLE DATASETS

Over the past few years, there has been a surge in the number of datasets devoted to crowd anomaly detection. These data sets can be used to analyze, compare and improve the performance of crowd anomaly detection systems. Applications related to crowds, including counting, density estimates, categorization, activity recognition, and anomaly detection, use real crowd datasets. The majority of visual real crowd datasets have focused on counting tasks, including UCSD, PETS2009, UCF-CC-50, Mall, Shanghai Tech, etc. In a recent study, SIMulated Crowd Datasets (SIMCD) were presented for use in creating models for predicting and detecting crowd anomalies [52]. The details of these datasets along with many other datasets available for crowd anomaly detection have been given in Table III. The name of the dataset, the size of the dataset, a small description and a short scenario are mentioned in the table.

TABLE III. DETAILS OF THE PUBLICLY AVAILABLE DATASETS FOR CROWD ANOMALY DETECTION

Name	**Scale**	**Description**	**Scenario**
UMN [14]	Small	The collection consists of films from 11 different escape event scenarios shot in 3 various indoor and outdoor settings.	Crowd behavior anomaly detection
UCSD Peds 1 UCSD Peds 2 [53]	Small	Videos by a stationary camera gazing down on pedestrian walkways. peds1: perspective distortion and groups of people walking. Peds2: Camera-parallel pedestrian movement.	Abnormal crowd behavior detection
CVCS [54]	Medium	Cross-view, cross-scene, multi-view counting using a synthetic dataset. Each scenario in the dataset has roughly 100 camera views and consists of 31 scenes. For each scene, 100 crowd multi-view photos were taken.	Multi-view crowd counting
Grand Central [55]	Medium	It is collected from the New York Grand Central Station.	Crowd train station dataset
HAJJv1 [40]	Large	It is collected from pilgrims passing through hall passages in Haram masjid, Mecca, Saudi Arabia	Human abnormal behavior in Hajj
Shanghai Tech Part A Part B [56]	Large	It has 13 scenes with complex light conditions and different camera angles. It contains 130 abnormal events and over 270, 000 training frames.	Crowd counting and density estimation
Violent flows [57]	Large	Video footage of crowd fighting together with industry-recognized benchmark standards for testing the accuracy of both the classification of violent and non-violent crowd behavior and the identification of violent outbreaks. 246 videos are included in the data collection. All of the videos	Classify and detect violent and non-violent behavior

		were downloaded from YouTube.	
WWW Crowd [58]	Large	A rich dataset for crowd understanding is provided by 10,000 videos with more than 8 million frames from 8,257 different scenes.	Crowd understanding
UCF-CC-50 [59]	Large	Extremely dense crowd dataset for crowd counting	for crowd counting
Multi-Task Crowd [60]	Large	A recent 100 image dataset that is completely annotated for crowd recognition, violent behavior detection, and categorization of density level.	Crowd counting, violence detection, and density level classification
UCF-Crime [31]	Large	It is made up of 1900 uncut, lengthy real-world surveillance movies that include 13 actual oddities like fighting, car accidents, robberies, and other crimes in addition to everyday occurrences.	Crowd behavior anomaly recognition, crowd behavior
Mall [61]	Medium	Webcams with public access are used to gather data. The video has 2000 frames, and each frame's annotations note the head position of each pedestrian.	Crowd counting
Street Scene [62]	Large	It is made up of video clips shot from a stationary USB camera looking down on a two-lane street with bike lanes and sidewalks, with another 35 clips used for testing.	Video anomaly detection.
CUHK-Avenue [63]	Small	It is a collection of short clips recorded by a single outdoor surveillance camera aimed at the side of a building facing a sidewalk. It has 15 sequences. Each one lasts approximately 2 minutes. There are 14 unusual events, including running, throwing objects, and loitering.	Abnormal event detection.

VI. Challenges and Limitations

In crowd scenarios with a variety of conditions, a reliable crowd anomaly detection algorithm tries to evaluate both local and global density and reliably anticipate crowd behaviour. The model's performance is significantly impacted by the changing circumstances. Therefore, it's crucial to first comprehend these difficulties and how they could affect the performance of the model. The development of more reliable models is aided by a thorough grasp of these difficulties.

a) Difficult to monitor crowd behaviour in various settings. For instance, it is significantly simpler to count and track the individuals in images captured by a single CCTV camera (such as those in the Mall dataset [61]) than to count people in images captured by numerous security cameras (such as those in the WorldExpo'10 dataset [64]). Drone surveillance often involves changing the scene, which makes it more difficult when combined with other variations such as scale variations.

b) When items belonging to the same class (in this case, humans) appear at various sizes in both a single photograph and across various images, it's referred to as scale variation. The distance (between the camera and the objects) and the perspective impact in the same image are the two factors that affect scale. Scale changes are additionally seen in photos with various resolutions. In crowd anomaly detection research, scale variation is one of the most prevalent issues that significantly affect model performance.

c) In different images, there are different numbers of people or other subjects of interest. Typically, low density visuals are simpler to understand than high density images. Similarly, another difficulty is when the same image shows various densities of people in multiple areas.

d) The distribution of objects in crowd photos may vary in different scenarios. For example, seats in a sports arena are evenly distributed among people, with constant spacing between objects, whereas in crowded streets, things might be distributed at random. In the absence of other features that would alter the accuracy, uniformly dispersed crowds can be estimated more precisely than non-uniform crowds.

e) Occlusion has been a major challenge in video analysis, and it is even difficult for crowd anomaly detection. The term occlusion describes how objects overlap. Intra-class occlusion refers to the overlapping of similar items (like humans), but inter-class occlusion refers to the overlapping of distinct objects (like automobiles, walls, and other people). Dealing with occlusion is frequently difficult. In the presence of occlusion, it is problematic for both the object detectors and the annotators to accurately annotate the objects as well as forecast them. Occlusion makes it challenging to distinguish between object borders in frames by interweaving semantic elements. It can also be challenging to learn when the object's pixel values are comparable to those of the background. Occlusion can limit the effectiveness of crowd anomaly detection systems as it makes it difficult to identify individuals in a crowd, leading to false positives or false negatives. Techniques like using multiple cameras or machine learning algorithms can help mitigate the effects of occlusion, but it can still be a significant challenge in dense and complex crowds.

f) Due to the different lighting conditions, the illumination in an image can change during the day and in various areas of the same image. This makes learning difficult because the same object (like people) in the same image will have varying pixel values. Other conditions which make CAD difficult include changes in the weather, noise and pixelation in images, rotation of objects, etc.

Some of the limitations of current crowd anomaly detection methods are given below:

- High computational cost: Some methods, such as deep learning-based methods, require a large number of computational resources to train and test, which can be a limitation in real-world applications where computational resources are limited.

- Need for annotated data: Many methods, such as deep learning-based methods, require a large amount of annotated data to train the models, which can be a limitation in real-world applications where data is limited or difficult to annotate.

- Ethical and privacy concerns: some methods, such as facial recognition-based methods, may raise privacy concerns, and it is important to ensure that the method is used in compliance with relevant laws and regulations.

- Limited ability to detect novel anomalies: Many methods are designed to detect known anomalies and may not be able to detect novel or unknown anomalies. Many methods focus on detecting anomalies but don't provide any insight into the root cause of the anomaly.

- Limited scalability: Many methods are not designed to handle large crowds and may not be able to scale to handle large amounts of data. Limited ability to handle multiple anomalies: Many methods focus on detecting a single type of anomaly and may not be able to handle multiple types of anomalies simultaneously.

VI. Summary

This review paper provides an in-depth analysis of crowd anomaly detection and its importance in real-world surveillance and security. The latest research on crowd analysis is reviewed and summarized, with a focus on the major components of crowd anomaly detection, such as crowd density estimation, object tracking, and object behavior analysis. The paper also provides a summary of research based on different taxonomies, including the type of anomaly, and the type of dataset labels of the anomaly. Publicly available datasets used for crowd analysis were also reviewed, including the types of anomalies they address. Considerations for evaluating model performance and current challenges in the field were also discussed. In light of the review, the paper provides directions for future research, including the need for model generalization for different anomalies in different scenarios, designing application-specific crowd anomaly detection s, and effectively selecting the most appropriate models for analysis to reduce unnecessary resource usage and carbon emissions. The paper also provides directions for future research, including the incorporation of generative models, graph-based methods, reinforcement learning, transfer learning, online learning, ensemble methods, multi-task learning, domain adaptation, active anomaly detection, and meta-learning which have the potential to significantly improve the performance of crowd anomaly detection systems and address current limitations in the field. Overall, this review paper offers valuable insights and a comprehensive understanding of the field of crowd anomaly detection, which will aid researchers in developing robust solutions to address the current limitations of the system.

VII. Future Directions

Some of the research directions based on the challenges identified by the various researches are mentioned next. The same benchmark datasets are used to train and evaluate the majority of crowd counting methods. So, model generalization hasn't been studied much, or studies are limited to the models that were fine-tuned on one dataset after being pretrained on another. However, the results that are normally published come from the refined model. This causes a significant disparity in the generalization of crowd counting models across different scenes, which warrants more research. It would be fascinating to observe model generalization in a variety of unusual situations, such as interior and outdoor videos, CCTV photos, drone photographs, etc.

An effective model must be able to run on a variety of hardware platforms with diverse computational capabilities, including servers, drones, cameras, mobile phones, etc., in order to support the potential applications of crowd anomaly detection. Applications (such as real-time or non-real-time), types of surveillance (such as CCTV-based surveillance or drone-based surveillance), and scenarios (shopping malls, metro stations, stadiums, etc.) all have different performance requirements. Therefore, it is ineffective to create a single optimal model with the highest accuracy for all applications, surveillance techniques, and circumstances. In actuality, such a model will be big, need a lot of computing power for fine-tuning, and have lengthier inference delays given the current trends in crowd anomaly detection model design. Applications requiring real-time inference, limited on-chip memory, and battery-powered devices will not be compatible with such a strategy. As a result, we anticipate and have also seen some recent attempts to have application-specific model designing, such as lightweight models for real-time applications on resource-constrained devices, and dense models for maximum accuracy over dense crowds in server-based systems.

The selection of deep or shallow neural networks is to be studied in detail. For sparse datasets with low crowd-density images, such as those from UCSD [53], Mall [61], and ShanghaiTech Part B [56], shallow models provide reasonably sufficient accuracy, and deeper models may not be necessary for the circumstances depicted in these datasets. Deeper models are typically used to achieve higher accuracy over large datasets, but these efforts often result in deeper and more complicated architectures. Unsurprisingly, single-scene crowd analysis and sparse multi-scene crowd detection are the two tasks with the fewest requirements. Even relatively tiny accuracy gains are the focus of most research efforts, and the ensuing model complexity is frequently disregarded. This leads to an increase in model complexity for a minor and frequently insignificant gain in accuracy. We think it's important to look into benchmarking for model training and inference times for crowd models. It’s high time that researchers think about green computing and help to conduct low carbon emission systems for their research.

Incorporating domain knowledge, temporal information and multiple modalities into anomaly detection methods can improve their performance and make them more robust to different scenarios. Incorporating generative models, such as generative adversarial networks (GANs) and variational autoencoders (VAEs) can improve the ability to detect novel or unknown anomalies and can also be used to generate synthetic data for training models. Incorporating graph-based methods, such as graph convolutional networks (GCNs) and graph recurrent networks (GRNs), can improve the ability to model the relationships and interactions between individuals in the crowd. Ensemble methods, such as an ensemble of classifiers and an ensemble of experts, can improve the robustness and generalizability of the models by combining the predictions of multiple models. The concept of meta-learning can be used, which allows the model to learn how to learn, can improve the ability to adapt to new data and tasks, and can also improve the interpretability of the models. Developing interpretable models, such as decision trees and rule-based models, that can provide insight into why a certain behavior is considered abnormal can make the methods more understandable and trustworthy. Incorporating explainable AI(XAI) can reveal the reason for a certain decision and can improve the interpretability of the models and make the methods more trustworthy.

There are further many possible future directions in terms of techniques and methods used for crowd anomaly detection, such as incorporating generative models, graph-based methods, online learning, ensemble methods, multi-task learning, domain adaptation, active anomaly detection, and meta-learning. These techniques and methods can help to improve the performance and robustness of crowd anomaly detection models and make them more suitable for real-world applications.

REFERENCES

[1] Y. Zhou, M. Qin, X. Wang, and C. Zhang, "Regional Crowd Status Analysis based on GeoVideo and Multimedia Data Collaboration," in 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), 2021, vol. 4: IEEE, pp. 1278-1282.

[2] J. Duan, W. Zhai, and C. Cheng, “Crowd Detection in Mass Gatherings Based on Social Media Data: A Case Study of the 2014 Shanghai New Year’s Eve Stampede,” Int. J. Environ. Res. Public Heal. 2020, Vol. 17, Page 8640, vol. 17, no. 22, p. 8640, Nov. 2020, doi: 10.3390/IJERPH17228640.

[3] M. Yamin, “Managing crowds with technology: cases of Hajj and Kumbh Mela,” Int. J. Inf. Technol., vol. 11, no. 2, pp. 229–237, Jun. 2019, https://doi.org/10.1007/s41870-018-0266-1.

[4] J. Ma, Y. Dai, and K. Hirota, "A Survey of Video-Based Crowd Anomaly Detection in Dense Scenes," Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 21, no. 2, pp. 235-246, 2017, doi: 10.20965/jaciii.2017.p0235.

[5] M. S. Zitouni, A. Sluzek, and H. Bhaskar, "Visual analysis of socio-cognitive crowd behaviors for surveillance: A survey and categorization of trends and methods," Engineering Applications of Artificial Intelligence, vol. 82, pp. 294-312, 2019.

[6] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L.-Q. Xu, "Crowd analysis: a survey," Machine Vision and Applications, vol. 19, no. 5, pp. 345-357, 2008.

[7] J. C. S. J. Junior, S. R. Musse, and C. R. Jung, "Crowd analysis using computer vision techniques," IEEE Signal Processing Magazine, vol. 27, no. 5, pp. 66-77, 2010.

[8] S. D. Bansod and A. V. Nandedkar, "Crowd anomaly detection and localization using histogram of magnitude and momentum," The Visual Computer, vol. 36, no. 3, pp. 609-620, 2020.

[9] V. J. Kok, M. K. Lim, and C. S. Chan, "Crowd behavior analysis: A review where physics meets biology," Neurocomputing, vol. 177, pp. 342-362, 2016.

[10] B. Yogameena and C. Nagananthini, "Computer vision based crowd disaster avoidance system: A survey," International journal of disaster risk reduction, vol. 22, pp. 95-129, 2017.

[11] X. Zhang, Q. Yu, and H. Yu, "Physics inspired methods for crowd video surveillance and analysis: a survey," IEEE Access, vol. 6, pp. 66816-66830, 2018.

[12] N. Nayan, S. S. Sahu, and S. Kumar, "Detecting anomalous crowd behavior using correlation analysis of optical flow," Signal, Image and Video Processing, vol. 13, no. 6, pp. 1233-1241, 2019.

[13] A. Afiq et al., "A review on classifying abnormal behavior in crowd scene," Journal of Visual Communication and Image Representation, vol. 58, pp. 285-303, 2019.

[14] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in 2009 IEEE conference on computer vision and pattern recognition, 2009: IEEE, pp. 935-942.

[15] B. Guo, J. Shi, L. Zhu, and Z. Yu, "High-speed railway clearance intrusion detection with improved SSD network," Applied Sciences, vol. 9, no. 15, p. 2981, 2019.

[16] L. Yundong, D. Han, L. Hongguang, X. Zhang, B. Zhang, and X. Zhifeng, "Multi-block SSD based on small object detection for UAV railway scene surveillance," Chinese Journal of Aeronautics, vol. 33, no. 6, pp. 1747-1755, 2020.

[17] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in Proceedings of ICML workshop on unsupervised and transfer learning, 2012: JMLR Workshop and Conference Proceedings, pp. 37-49.

[18] P. Perera and V. M. Patel, "Learning deep features for one-class classification," IEEE Transactions on Image Processing, vol. 28, no. 11, pp. 5450-5463, 2019.

[19] G. Blanchard, G. Lee, and C. Scott, "Semi-supervised novelty detection," The Journal of Machine Learning Research, vol. 11, pp. 2973-3009, 2010.

[20] J. Andrews, T. Tanay, E. J. Morton, and L. D. Griffin, "Transfer representation-learning for anomaly detection," 2016: JMLR.

[21] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," Pattern Recognition, vol. 58, pp. 121-134, 2016.

[22] R. Chalapathy, A. K. Menon, and S. Chawla, "Anomaly detection using one-class neural networks," arXiv preprint arXiv:1802.06360, 2018.

[23] L. Ruff et al., "Deep one-class classification," in International conference on machine learning, 2018: PMLR, pp. 4393-4402.

[24] G. Tripathi, K. Singh, and D. K. Vishwakarma, "Convolutional neural networks for crowd behaviour analysis: a survey," The Visual Computer, vol. 35, no. 5, pp. 753-776, 2019.

[25] Y. Luo, J. Lu, and B. Zhang, "Crowd counting for static images: a survey of methodology," in 2020 39th Chinese control conference (CCC), 2020: IEEE, pp. 6602-6607.

[26] M. Bendali-Braham, J. Weber, G. Forestier, L. Idoumghar, and P.-A. Muller, "Recent trends in crowd analysis: A review," Machine Learning with Applications, vol. 4, p. 100023, 2021.

[27] K. Singh, S. Rajora, D. K. Vishwakarma, G. Tripathi, S. Kumar, and G. S. Walia, "Crowd anomaly detection using Aggregation of Ensembles of fine-tuned ConvNets," Neurocomputing, vol. 371, pp. 188-198, 2020, doi: 10.1016/j.neucom.2019.08.059.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[29] X. Hu, Y. Huang, X. Gao, L. Luo, and Q. Duan, "Squirrel-cage local binary pattern and its application in video anomaly detection," IEEE Transactions on Information Forensics and Security, vol. 14, no. 4, pp. 1007-1022, 2018.

[30] K. Simonyan and A. Zisserman, "VGGNet," in 3rd Int. Conf. Learn. Represent. ICLR, 2015.

[31] W. Sultani, C. Chen, and M. Shah, "Real-world anomaly detection in surveillance videos," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6479-6488.

[32] M. Z. Zaheer, A. Mahmood, M. H. Khan, M. Segu, F. Yu, and S.-I. Lee, "Generative Cooperative Learning for Unsupervised Video Anomaly Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 14744-14754.

[33] R. Cai, H. Zhang, W. Liu, S. Gao, and Z. Hao, "Appearance-motion memory consistency network for video anomaly detection," in Proceedings of the AAAI Conference on Artificial Intelligence, 2021, vol. 35, no. 2, pp. 938-946.

[34] G. Yu et al., "Cloze test helps: Effective video anomaly detection via learning to complete video events," in Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 583-591.

[35] C. Chen et al., "Comprehensive Regularization in a Bi-directional Predictive Network for Video Anomaly Detection," in Proceedings of the American association for artificial intelligence, 2022, pp. 1-9.

[36] Y. Guan, W. Hu, and X. Hu, "Abnormal behavior recognition using 3D-CNN combined with LSTM," Multimedia Tools and Applications, vol. 80, no. 12, pp. 18787-18801, 2021.

[37] C. Amrutha, C. Jyotsna, and J. Amudha, "Deep learning approach for suspicious activity detection from surveillance video," in 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020: IEEE, pp. 335-339.

[38] W. Ullah, A. Ullah, T. Hussain, Z. A. Khan, and S. W. Baik, "An efficient anomaly recognition framework using an attention residual LSTM in surveillance videos," Sensors, vol. 21, no. 8, p. 2811, 2021.

[39] Y. Hao, J. Li, N. Wang, X. Wang, and X. Gao, "Spatiotemporal consistency-enhanced network for video anomaly detection," Pattern Recognition, vol. 121, p. 108232, 2022.

[40] T. Alafif, B. Alzahrani, Y. Cao, R. Alotaibi, A. Barnawi, and M. Chen, "Generative adversarial network based abnormal behavior detection in massive crowd videos: a Hajj case study," Journal of Ambient Intelligence and Humanized Computing, vol. 13, no. 8, pp. 4077-4088, 2021, doi: 10.1007/s12652-021-03323-5.

[41] T. Alafif et al., "Hybrid classifiers for spatio-temporal real-time abnormal behaviors detection, tracking, and recognition in massive hajj crowds," arXiv preprint arXiv:2207.11931, 2022.

[42] R. Lalit, R. K. Purwar, S. Verma, and A. Jain, "Crowd abnormality detection in video sequences using supervised convolutional neural network," Multimedia Tools and Applications, vol. 81, no. 4, pp. 5259-5277, 2021, doi: 10.1007/s11042-021-11781-4.

[43] B. Wang and C. Yang, "Video Anomaly Detection Based on Convolutional Recurrent AutoEncoder," Sensors, vol. 22, no. 12, p. 4647, 2022.

[44] X. Hu, J. Lian, D. Zhang, X. Gao, L. Jiang, and W. Chen, "Video anomaly detection based on 3D convolutional auto-encoder," Signal, Image and Video Processing, pp. 1-9, 2022.

[45] K. Doshi and Y. Yilmaz, "A modular and unified framework for detecting and localizing video anomalies," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 3982-3991.

[46] T. Saba, "Real time anomalies detection in crowd using convolutional long short-term memory network," Journal of Information Science, p. 01655515211022665, 2021.

[47] L. Xia and Z. Li, "A new method of abnormal behavior detection using LSTM network with temporal attention mechanism," The Journal of Supercomputing, vol. 77, no. 4, pp. 3223-3241, 2021.

[48] V.-T. Le and Y.-G. Kim, "Attention-based residual autoencoder for video anomaly detection," Applied Intelligence, vol. 53, no. 3, pp. 3240-3254, 2023.

[49] D. Zhang, C. Huang, C. Liu, and Y. Xu, "Weakly supervised video anomaly detection via transformer-enabled temporal relation learning," IEEE Signal Processing Letters, vol. 29, pp. 1197-1201, 2022.

[50] N. Li, J.-X. Zhong, X. Shu, and H. Guo, "Weakly-supervised anomaly detection in video surveillance via graph convolutional label noise cleaning," Neurocomputing, vol. 481, pp. 154-167, 2022.

[51] S. Zhang et al., "Influence-aware attention networks for anomaly detection in surveillance videos," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 8, pp. 5427-5437, 2022.

[52] A. Bamaqa, M. Sedky, T. Bosakowski, B. Bakhtiari Bastaki, and N. O. Alshammari, "SIMCD: SIMulated crowd data for anomaly detection and prediction," Expert Systems with Applications, vol. 203, 2022, doi: 10.1016/j.eswa.2022.117475.

[53] A. B. Chan, Z.-S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in 2008 IEEE conference on computer vision and pattern recognition, 2008: IEEE, pp. 1-7.

[54] Q. Zhang, W. Lin, and A. B. Chan, "Cross-view cross-scene multi-view crowd counting," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 557-567.

[55] B. Zhou, X. Wang, and X. Tang, "Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012: IEEE, pp. 2871-2878.

[56] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 589-597.

[57] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2012: IEEE, pp. 1-6.

[58] J. Shao, K. Kang, C. Change Loy, and X. Wang, "Deeply learned attributes for crowded scene understanding," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 4657-4666.

[59] H. Idrees, I. Saleemi, C. Seibert, and M. Shah, "Multi-source multi-scale counting in extremely dense crowd images," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 2547-2554.

[60] M. Marsden, K. McGuinness, S. Little, and N. E. O'Connor, "Resnetcrowd: A residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification," in 2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS), 2017: IEEE, pp. 1-7.

[61] K. Chen, C. C. Loy, S. Gong, and T. Xiang, "Feature mining for localised crowd counting," in Bmvc, 2012, vol. 1, no. 2, p. 3.

[62] B. Ramachandra and M. Jones, "Street scene: A new dataset and evaluation protocol for video anomaly detection," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 2569-2578.

[63] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 fps in matlab," in Proceedings of the IEEE international conference on computer vision, 2013, pp. 2720-2727.

[64] C. Zhang, H. Li, X. Wang, and X. Yang, "Cross-scene crowd counting via deep convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 833-841.