Personal and Ubiquitous Computing (2024) 28:135–151
https://doi.org/10.1007/s00779-021-01586-5

ORIGINAL ARTICLE

A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance

Khosro Rezaee[1] · Sara Mohammad Rezakhani[2] · Mohammad R. Khosravi[3] · Mohammad Kazem Moghimi[4]

Received: 16 February 2021 / Accepted: 8 June 2021 / Published online: 25 June 2021

Abstract

Fast and automated recognizing of abnormal behaviors in crowded scenes is significantly effective in increasing public security. The traditional procedure of recognizing abnormalities in the Web of Thing (WoT) platform comprises monitoring the activities and describing the crowd properties such as density, trajectory, and motion pattern from the visual frames. Accordingly, incorporating real-time security monitoring based on the WoT platform and machine learning algorithms would significantly enhance the influential detection of abnormal behaviors in the crowds. This paper addresses various automatic and real-time surveillance methods for abnormal event detection to recognize the dynamic crowd behavior in security applications. The critical aspect of security and protection of public places is that we cannot manually monitor the unpredictable and complex crowded environments. The abnormal behavior algorithms have attempted to improve efficiency, robustness against pixel occlusion, generalizability, computational complexity, and execution time. Similar to the state-of-the-art abnormal behavior detection of crowded scenes, we broadly classified methods into different categories such as tracking, classification based on handcrafted extracted features, classification based on deep learning, and hybrid approaches. Hybrid and deep learning methods have been found to have more satisfactory results in the classification stage. A set of video frames called Motion Emotion Dataset (MED) is employed in this study to examine the various conditions governing these methods. Incorporating an appropriate real-time approach with considering WoT platform can facilitate the analysis of crowd and individuals' behavior for security screening of abnormal events.

Keywords Crowd analysis · Security surveillance · Abnormal behavior · Deep learning · Tracking · Hybrid models

1 Introduction

Security surveillance systems are employed to prevent violations in private and public areas. Analysis of public environments described by the phenomenon of overcrowding in the form of the population can be considered one of the challenging issues in machine vision and image processing [1]. The high volume of the crowd needs the surveillance and participation of many individuals such as personnel and operators to visually monitor and control abnormal events [2]. Human error is one of the challenges that can make the routine of crowd surveillance a difficult and complex procedure [3]. By monitoring human activities in sensitive crowded situations via real-time manner, we can detect abnormal and unconventional behaviors [4]. This real-time process will improve the security condition and prevent abnormal and unconventional behaviors in crowded public environments. Abnormal behaviors are actions that are unexpected and often assessed negatively because they differ from conventional or expected behavior.

✉ Khosro Rezaee
kh.rezaee@meybod.ac.ir

Sara Mohammad Rezakhani
sararezakhani73@yahoo.com

Mohammad R. Khosravi
m.khosravi@sutech.ac.ir

Mohammad Kazem Moghimi
k.moghimies@pgs.usb.ac.ir

[1] Department of Biomedical Engineering, Meybod University, Meybod, Iran

[2] Department of Biomedical Engineering, Semnan University, Semnan, Iran

[3] Department of Computer Engineering, Persian Gulf University, Bushehr, Iran

[4] Department of Communication Engineering, University of Sistan and Baluchestan, Zahedan, Iran

When the behavior of a person or individuals seems abnormal, this phenomenon is called an abnormal action [5–7].

Abnormal behavior is strongly dependent on the norms defined in the considered environment and cannot be precisely defined. For instance, moving clockwise around the Kaaba is an abnormal behavior, although this behavior may be perfectly normal in other situations. Congestion is generally considered an abnormal category. Sometimes it is a security challenge, meaning that an abnormal behavior has occurred to create an action outside the framework.

The evolution of the WoT is associated with machine learning and computer vision Web-based technologies for organizing a fast hybrid decision-making system. Besides, the WoT is proceeding to more control over our living conditions, allowing more facilitation in obtaining things done. WoT represents a collection of criteria by the W3C for determining the interoperability problems of various Internet of Things (IoT) application fields and principles [8]. Moreover, the Thing Description of WoT is the essential part of the WoT building blocks. In this definition, a Thing Description helps realize WoT as a physical or virtual thing. Therefore, a Thing based on semantic vocabulary and a serialization based on JavaScript Object Notation (JSON) are considered the model's information.

We required robust machine learning algorithms with various capabilities such as automatic behavior detection in real-time conditions [9–11]. One of the new and automated methods that have recently been the main focus of researchers in the field of machine learning is deep learning-based methods [12]. Even in less crowded environments, monitoring the abnormal behavior of humans is necessary as a real-time procedure. Some events also occur unintentionally as a result of inherent challenges in the population itself. In 2010, a tragic event occurred during a music festival in Duisburg, Germany, which led to the death of 20 people and the injury of nearly 500 others [13]. In 1989, over 96 people died due to overcrowding during a football match, and 766 people lost their lives at Hillsborough Stadium, Liverpool, England [14]. The incident started when people were about to leave the stadium, while it was possible to control the issue and prevent the event [14]. Similarly, at least 2431 people lost their lives in congestion in Mecca, Saudi Arabia, in 2015 [15]. At the beginning of 2020 in Kerman, Iran, about 58 people died due to overcrowding during a mourning ceremony [16]. In all these cases, by observing the crowd's behavior, it was possible to prevent the occurrence of unfortunate accidents. Furthermore, it has been observed that terrorist attacks can occur with the sudden entry of a person or runaway vehicles into the crowd, and the entry of people with suspicious tools, equipment, and even bag. If these incidents have been predicted or observed beforehand, they may have been prevented. When the operator controls crowded environments, there is a possibility of a sudden loss of data due to a lack of attention and accuracy. The purpose of automated security surveillance systems based on IoT or WoT platforms is to minimize false errors and control crowd behavior in unconventional forms of congestions. Hence, the issues involve crowds of people and abrupt changes in their behavior.

2 Security surveillance and real-time processing

Automated security surveillance approaches are necessary to protect a country's crucial infrastructure and public environments (e.g., metro, airports, city centers, mall shops, and stadiums) against the warning of criminal activity, civil unrest, cyber-attacks, and terrorism. Thus, increased security surveillance is further expected for any occurrence that represents high-density crowds [17]. Urban security monitoring and real-time detection of crowd behavior rely heavily on the condition of CCTV cameras. Urban security monitoring includes dynamic, evolving crowd scenes that place more significant requirements on visual search processes. Some studies have shown that although the development of security surveillance has various aspects of progress, it cannot make full decisions on behalf of the operator [17].

Real-time video processing is one of the most cardinal topics in big data analysis. Accordingly, it is required for uninterrupted security surveillance of various events, messages, and processing and analysis in network infrastructure [18]. Huge amounts of data that continuously reach pipelines can be generated in any format, such as structured, unstructured, and semi-structured. Therefore, the information exchanged for video processing include messages and events. Processing events such as real-time activity recognition of abnormal crowd behavior will improve the correlation. This possibility creates pattern recognition procedures at the scale of observing a large number of events and transmitting information at microsecond speeds. Hence, real-time detection of abnormal events in practical video processing applications has rarely been considered in state-of-the-art abnormal behavior algorithms. The processing of abnormal behavior algorithms is associated with transferring high volumes of information, requiring a powerful hardware platform and software designs. However, it is not always possible to access powerful hardware, so comparisons between abnormal crowd behavior detection methods, like other video analysis methods, are based on comparisons between the algorithms used.

In offline systems, one can yield to employ the time to obtain optimal or near-optimal approaches [19]. Instead, working methods operate on online situations and need effective solutions. Online processing indicates some interaction; however, it does not impose a delay limit. Besides, a real-time manner means limited latency, and we can define online

processing as consecutive logging of transactions for real-time computer methods [20].

In real-time abnormal event detection, due to computational flexibility, statistical techniques are often used in video frame processing designs as well as fast algorithms with low computational complexity [21]. With these assumptions, some abnormal event detection methods' time performance and computational complexity increase significantly and cannot generate real-time outputs. Some ways reduce the size of video frames to resolve the computational complexity, which can disrupt pattern recognition and even affect the body shape of people in the crowd, resulting in inadequate tracking or high error in classification [22, 23].

Evolutionary and meta-heuristic algorithms in many applications cannot work as real-time models. These algorithms need a lot of time to process information due to various parameters such as population, parameters initializing, different loops such as different population generations, and other similar challenges. Furthermore, the need to converge to the optimal value requires several calling of the cost function. However, due to the heavy processing of video information and population analysis, some studies have stated that in the future, it can be hoped that optimization algorithms will be used extensively in the network space and the challenge of computational complexity in conventional computers [24]. It is solved for applications such as video processing.

Other abnormal event detection methods utilize structures based on deep learning, but these structures can provide low-error responses even for low-quality images. Nonetheless, the deep learning structure is lazy in data processing and requires a considerable amount of time, especially during the training phase. Putting all these together, one can consider trade-offs of methods, based on which we may discard some optimal outcomes [24, 25]. For this reason, in the strategies designed to identify abnormal events for crowd behavior, no attention has been paid to the real-time aspect of the technique, and if the method is real-time, it is associated with some other challenges such as reduced detection accuracy.

3 Overview and motivation

When humans monitor crowded environments due to fatigue or lack of operator focus, it is always possible to miss a critical event with unpleasant consequences. In this regard, the purpose of security surveillance systems is to minimize the risk of false alarm rate (FAR).

In recent years, security surveillance automation of these places has attracted many researchers in the field of real-time image and video processing [26–31]. The designed system must detect abnormal events to make security surveillance systems more intelligent and automated. In addition, other problems such as noise and pixel occlusion, the interaction of objects and people, the simultaneous existence of several unusual events, computational complexity, and unstructured events can affect security surveillance. These challenges are common in all environments, and abnormal behavior detection algorithms must deal with them.

We also need suitable techniques to solve the existing challenges and to analyze them properly. There are many methods in the field of machine vision that have been used for crowd behavior analysis. Most of them, like heavy deep learning structures, are computationally overwhelming. Therefore, they are not suitable for real-time processing. Moreover, [32] has shown that the use of optical flow in the analysis of noise-impregnated frames is also effective. There are other methods such as speeded up robust features (SURF) [33] and scale-invariant feature transform (SIFT) [34] that analyze crowd behavior based on features. Considering all the problems mentioned, solving them will be a difficult and complicated process. Consequently, using an algorithm robust against these challenges would be a good solution. Various algorithms have been proposed in this field, which may detect the behavior as an anomaly; hence, the methods mentioned in the detection of abnormal behavior can have their classification. These divisions are broadly categorized into supervised and unsupervised methods.

The present study aims to analyze abnormal behavior detection and classification methods based on algorithms that have been proposed as state-of-the-art real-time or near-real-time approaches in security surveillance applications. Automatic detection is considered in environments where there is a high probability of people walking and commuting. The occurrence of abnormal conditions varies, but the presence of bicycles, cars, skates, throwing objects, and the like, which are faster than pedestrians, can usually be identified as abnormal behavior. Even fleeing, fighting, gathering, and moving out of the ordinary are in some ways considered abnormal crowd situations.

The remainder of the research describes some related methods. Then, some efficient and similar algorithms will be introduced, and finally, their results and interpretation will be presented. Above all, there will be reliance on analysis using novel methods and algorithms such as deep learning. In Section 4, similar algorithms for identifying abnormal behavior in the crowd scenes will be discussed. Section 5 provides a general comparison in terms of functional ability, and eventually, Section 6 gives a summary of the conclusions based on the performed research.

4 Crowd anomaly behavior detection

Motion detection can also be defined from a visual point of view, which is a process during the combination of modeling algorithms and machine vision [35, 36]. The main purpose of

human motion video analysis algorithms is to develop and improve detection systems in the field of human motion detection and analysis. In the meantime, comprehensive datasets that contain the main human movement patterns will cause the systems that are invented and proposed in this field every day to operate based on harmonious and common principles.

This section presents the abnormal crowd detection algorithms applicable for automated security surveillance platforms. We scrutinize the recent algorithms such as tracking, classification based on handcrafted extracted features, classification based on deep learning, and hybrid methods.

4.1 Tracking

One of the traditional methods in analyzing crowd behavior is optical flow, an apparent pattern movement of objects, surfaces, and edges in a visual frame, which arises from the relative movement between the observer and a scene [37, 38]. Some events are subject to the study of a specific behavior in specific situations, and such a study can be found in [39]. So that, the Kalmen filter has been used to segment the background and foreground of video frames. However, applying backgrounds such as optical flow can be used to distribute the apparent velocities of the motion of a light pattern in an image [40].

Figure 1 shows an outline of the real-time classification plan used to detect normal and abnormal crowd behaviors in the UCSD ped2 and UMN databases, which employ a Gaussian distribution model [41]. This schematic includes four sections: input video, global and local descriptors, abnormal behavior classification, and fusion scheme. In the first section, each frame scene is split into different non-overlapping cubes. The second section of this construction involves global and local descriptors.

The local descriptor uses the structural similarity index method (SSIM) method to calculate the similarity between patches, a type of local patch descriptor. Thus, two types of local descriptors are carried out based on the space-time neighborhood approach and inner temporal approach (TIA). Concerning the first local description, the space-time neighborhood sections of each patch include a section of the spatial neighborhood, including itself in the center, and a section of the temporal neighborhood following the patch. The SSIM values result from the first local descriptor [d_0, …, d_9]. Concerning the TIA, the SSIM value for all frames in the patch is computed as [D_0…, D_{t-1}]. Finally, the SSIM values are combined from both methods to obtain a local descriptor [d_0, …, d_9, D_0, …,D_{t-1}].

In the classification process, the abnormal behaviors of two Gaussian classifiers, including C_1 and C_2, are estimated through two sets of features from the global and local models. Classifiers [42] use the Gaussian distribution procedure to design the regular activities in each patch of the video, and the Mahalanobis distance technique is applied to identify outlier data. The outcomes in their work depict that the real-time technique is comparable to a state-of-the-art approach on UCSD ped2 and UMN benchmarks, but with even more time to analyze the frames [41]. They compared their work with Li et al. [43] based on run-time network situation. The processing times of their work and Li et al.'s study were about 0.04 s per frame and 1.38 s per frame, respectively. They also measure the effect of anomaly based on frame level, pixel level, and dual pixel-level evaluations, as shown in Fig. 2.

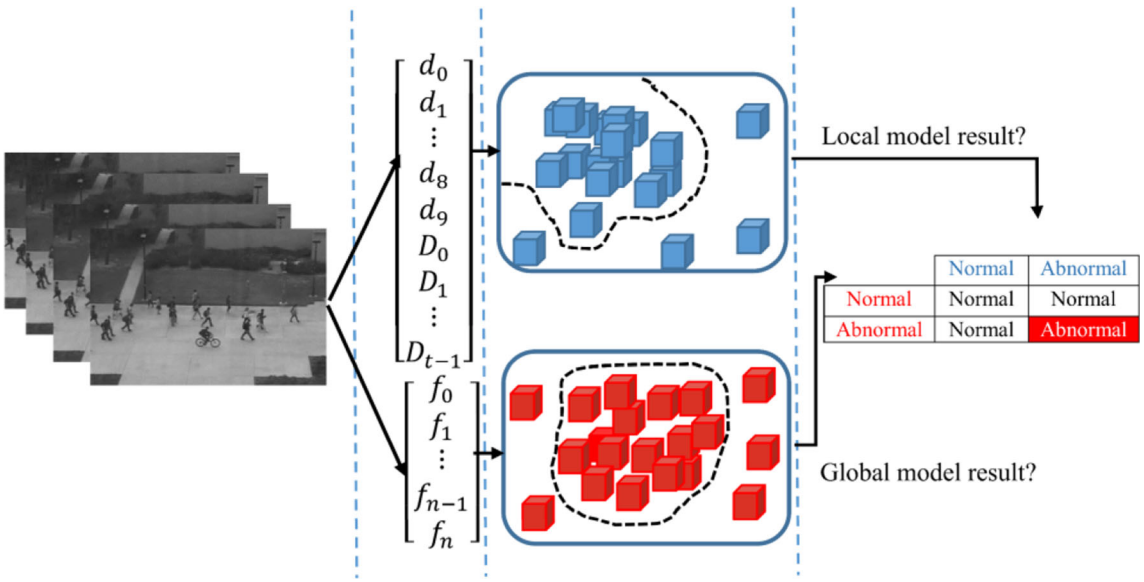

Fig. 1 The structure of real-time method in [34] (*left to right*): input visual image, global and local views of patches, modeling the information employing Gaussian distributions, and the final decision

Yu et al. [44] proposed a method for detecting abnormal behaviors using the Gaussian-Poisson mixture model (GPMM).

Inspired by the Gaussian mixed model, this algorithm generates information about the movement pattern of crowd behavior and the number of events with normal and abnormal behaviors. In principle, the expectation-maximization (EM) procedure is used to teach the GPMM [45]. A predetermined threshold is also utilized to detect abnormal behavior. Because the value obtained is considered to be less than the threshold, the event is considered abnormal. They also used a classification scheme that analysis the abnormal behaviors based on the behavior's temporal and spatial frequency. Region 1 in Fig. 3 shows a high temporal frequency and low spatial frequency of behavioral patterns. Consequently, it is ranked as a global abnormal behavior. In region 2, the behavior pattern shows a high temporal frequency and a high spatial frequency. Thus, this behavior is supposed a local abnormal behavior.

Other similar studies such as Sabokrou et al. [41], Leyva et al. [46], Lu et al. [47], Marsden et al. [48] used tracking-based methods. The latest one employs a real-time method claimed to have a short response time in the analysis of crowd abnormal behavior. Other optical flow-based tracking methods have been proposed that combine optical flow and HOG or use it in association with techniques such as GMM and EM [49–51]. The method proposed by Pennisi et al. [50] is a real-time and online crowd abnormal behavior detection method. It is a combination of visual feature extraction and image segmentation that operates without requiring a training phase. Other tracking methods are also of concern, including the Markov hidden model [52–57], which is used as a dynamic probabilistic method in areas where crowd and security surveillance is possible. These include security surveillance [52, 53], path analysis [54, 55], and action recognition [56, 57]. Other research has used a combination of similar tracking methods, including GM-HMM modeling [58]. Such method consists of applying a combined tracking method and utilizing feature extraction by principal component analysis (PCA) and histogram of gradient (HOG), which is based on the K-means++ clustering method and tracking using a Gaussian mixed model.

The application of other similar methods using Gaussian mixed models in tracking as well as extracting features from abnormal behavior can be found in [59–61], in some of which the analysis of the key components plays a key role in feature extraction and size reduction of the features. Some research has also used multi-objective tracking methods considering the possibility of people overlapping in the crowd or non-static and dynamics movement conditions and trajectories [62]. This method is based on an analysis of the trajectory of people and detecting several objectives, which is known as the extended K-shortest path (E-KSP) and is a type of search for optical flow with the minimum cost in the greedy search for paths.

Another paramount tracking method that is highly effective in detecting abnormal behaviors of individuals and crowds is the Kanade-Lucas-Tomasi (KLT) method, which extracts trajectory information using clustering in motion patterns among the crowd [63, 64]. The application of this method [65] is provided in Fig. 4. The method is based on the input frames obtained from the crowd and the extraction of features derived from the accelerator sections called FAST, along with the optimized KLT method used in detecting motion trajectory.

Other similar studies related to crowd abnormal behavior tracking include the method presented by Biswas and Babu [66] and Luo et al. [67].

4.2 Learning models

In learning-based anomaly detection methods, the aim is to use methods working on non-automatic feature extraction as well as automatic feature extraction. Non-automatic feature extraction methods obtain the features through conventional feature extraction methods, and automated methods are considered among deep learning models. The following describes some of the proposed methods in this field that have been published in recent years for abnormal behavior detection.

4.2.1 Handcrafted features

The real-time crowd anomaly detection algorithm for security surveillance has been proposed by Wang et al. [9]. Their

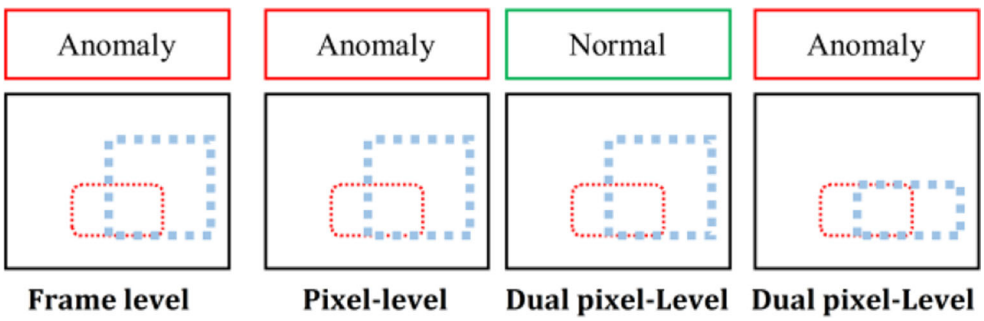

Fig. 2 Measure of anomaly assessment. The blue and red rectangles represent the output of the method and anomaly ground truth, respectively [41]

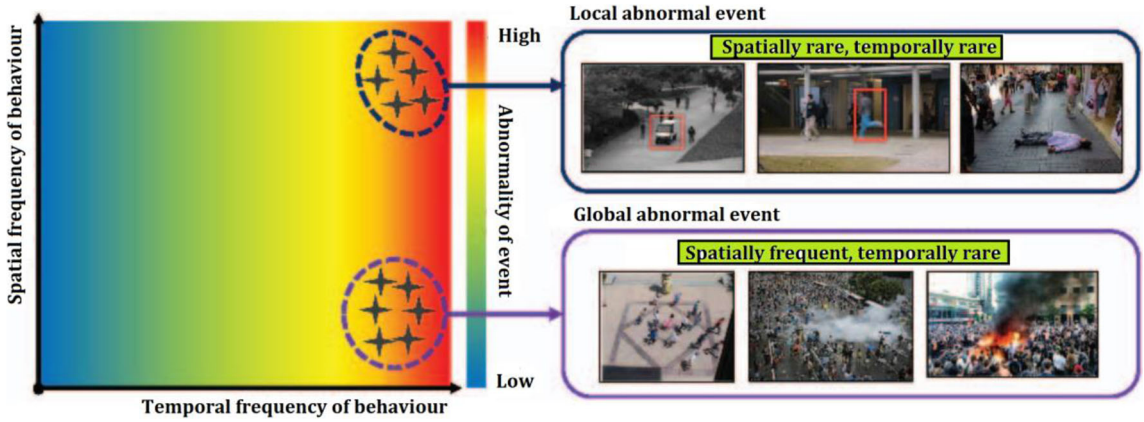

Fig. 3 Classification of abnormal behavior based on the spatiotemporal frequencies of behaviors that show the behavioral pattern [44]

research has developed a spatiotemporal texture model for feature extraction. They established a redundant texture feature space using the wavelet transform. The detection algorithm is fast and robust, and the system has shown improved accuracy and performance compared with similar methods.

The proposed method in [68] used features based on motion information instead of detecting actions or events in order to detect the abnormality. The EM algorithm is used to cluster seven-dimensional sample vectors with a predetermined number for clusters. Events that are not related to any of these predetermined clusters are considered unusual events. The method introduced in [69] represents the individual movement label categorizing events using the two-state Markov chain model. In [70], a statistical framework for modeling activity and discovering anomalies has been provided. They generally described a family of unsupervised methods for video anomaly recognition based on handcrafted extracted features and statistical activity analysis of video sequences.

In [71], events are modeled using spatiotemporal cubes, and the decision tree classification technique is used to identify the type of event. At the core of most of these techniques, the probabilistic modeling is realized based on location and the data obtained from the tracking part. The conventional method used in transport monitoring is to cluster the trajectory of identified moving targets and track their movements. The resulting clusters are used as normal models to detect and estimate abnormal and abnormal behavior. In [72], a classifier-based approach to recognize dynamic events in security surveillance sequences has been presented.

They have also suggested handcraft local patterns of features, and an ensemble of randomized trees has a spatiotemporal organization of patterns. In most researches, attempts

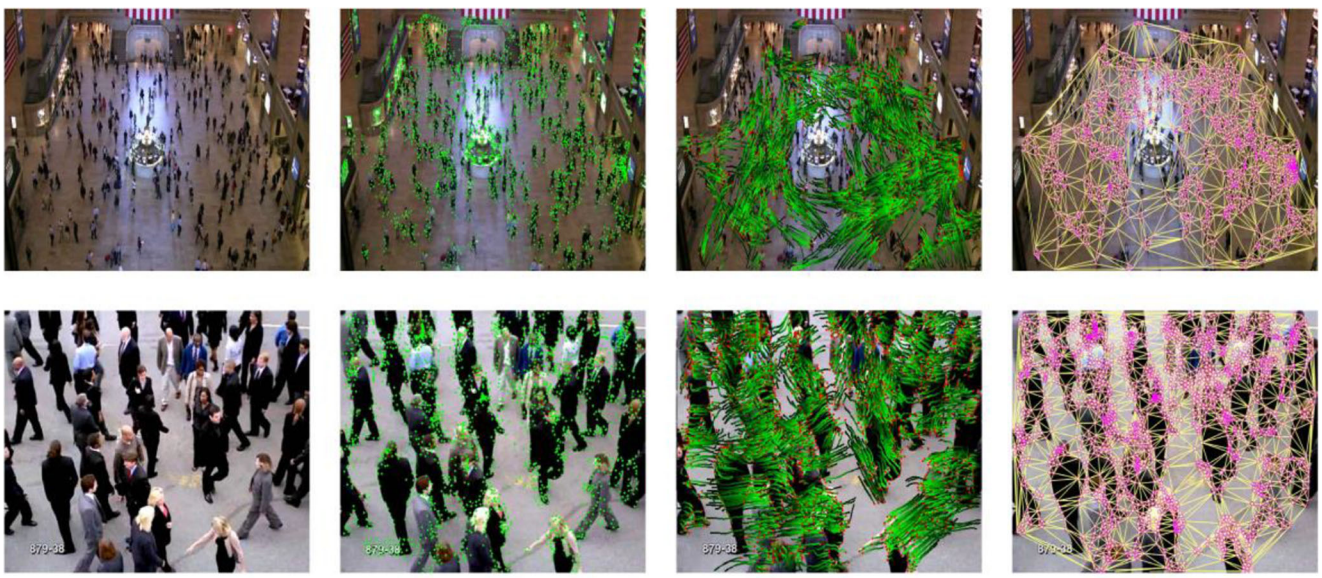

Fig. 4 The performance of the KLT method in detecting abnormal behavior: **a** input frames, **b** the FAST features, **c** feature tracks over time, and **d** spatial proximity using Delaunay triangulation [65]

have been made to use the trajectory of people as appropriate features to analyze behavior type. Similar methods are effective in helping create a suitable context for object tracking and trajectories.

These techniques are used for abnormal security events in the traffic category, as various tracking techniques have been proposed that we can use to greatly assess the ultimate goal of behavior recognition [73]. One of the common methods in this field is to extract the path taken by the vehicles in a normal and completely common way and search for their deviation from a specific path in the received traffic videos or car traffic [74–78]. The vehicle that is being tracked is taken into account in the evaluation step. Therefore, its trajectory is compared with normal or conventional features. Too much deviation from all the features is considered to represent an abnormal path.

Since the advent of video processing, machine vision, and behavioral pattern detection, especially abnormal movements, analysis methods of moving object trajectories have played a crucial role. Through learning from the analysis of individuals' trajectories or moving objects, some of the proposed methods operate in real-time [79], which has resulted in tracking targets, especially humans [80]. In this case, countless people or objects will be tracked during the training and learning phases over some time. In the next step, the resulting paths will be converted into a set of paths and generally displayed as an overview of the activities of the people in the background of the frames. The detection and testing phase of the trajectories obtained from the video is compared with the trajectories examined in the learning phase. Other similar methods can be found in Zou et al. [81], Chaker et al. [82], and Singh et al. [83]. In these methods, SVM and similarity level measurement have been used as classifiers of abnormal movements of people, respectively. In some other studies, the definition of tracklets has been utilized for optimizing the classification process [4, 64]. While many methods for crowd anomaly detection suggest offline solutions, few studies have considered real-time analysis of crowd behavior. However, the reason for this concerns the dynamics of the crowd, which sometimes requires cumbersome calculations.

One of the crowd anomalies is panic, which Aldissi et al. [21] analyzed and proposed a real-time distinguishing method that examined the crowd's movements according to a simple and efficient algorithm. The main idea of their method is to study the interactions between moving edges along with the video in the frequency domain.

4.2.2 Deep learned features

Deep learning is a more specialized form of machine learning proposed based on the definition of depth for simple neural networks [84, 85]. Since the layers of the deep neural network (see Fig. 5 for details) are completely interconnected, they have complexity in processing high-dimensional inputs. Hence, they may encounter problems such as over-fitting if they lack access to the appropriate dataset or processor hardware, thus failing to create a model with real-time capability in decision-making. Generally, we convolved the n-frame sequence with 3D filters in Fig. 5. After that, $m*c*k$ part filters are applied to convolve k feature maps for the classification category such as abnormal event detection to recognize the crowd behavior.

Various patterns and configurations of deep convolutional neural networks have been considered in the study [86].

In this study, the appropriate function to separate the frames obtained from security surveillance and natural frames recorded in case of abnormal states of individuals have been analyzed by bringing suitcases or items suspected of being at risk of terrorist explosions. Learning in the network is realized by using data differences between consecutive frames. Most of their activities concern various procedures of crowd surveillance and automatic analysis of an uncontrolled environment, which is considered target detection or individual effect of moving in public places.

Creating a surveillance environment and visual perception of the frames received from the crowd is a challenging and important issue in various categories of computer vision. In [87], a new configuration called Deep-Crowd was inspired by the deep learning method of residual neural network (ResNet) to extract and separate spatial traits fully. A unique dataset of nearly 6000 image frames has been generated to learn and evaluate the proposed system. The different evaluation criteria of their proposed system help achieve the appropriate accuracy of 83.11%, which can be compared with other efficient methods.

Also, Kotapalle et al. [88] used deep convolutional neural networks to detect and recognize the individuals' traffic in which video frames were analyzed for security surveillance. In this method, initially, some pre-processing methods were performed on images and video frames to enable high-precision detection.

A new model for detecting abnormal movements within the crowd is presented [89]. This model somehow improves user behavioral patterns and behaviors, and adopts a method with new similarity criteria. Experiments based on image data show that the detection model presented in this study provides satisfactory detection performance. Some methods detect abnormal movements of individuals by combining deep learning with patterns, such as spatial-temporal volume (STV) [90].

Hu et al. [92] also used convolutional neural networks called ConvNet to extract features from the crowd. The ConvNet structure consists of three convolutional layers with three max-pooling layers and one full-connected layer, as shown in Fig. 6. After each convolutional layer, the rectified linear units (ReLU) process continues with the max-pooling and non-linear processes. Aggregation and integration

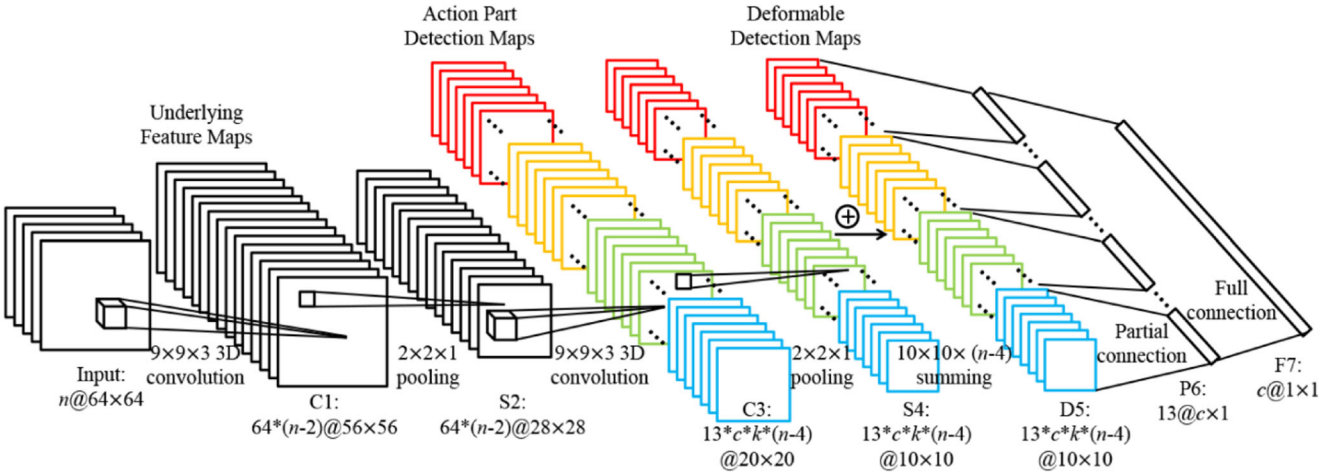

Fig. 5 The conventional 3D-CNN is used for classification application [91]

processes to extract visual features hierarchically from local low-level features to global high-level features have been considered in their method. Slicing convolutional neural network (S-CNN) has been proposed to classify crowd anomaly events [93]. Some popular strategies for optimizing hyper-parameters are applied in related studies of abnormal behavior detection. Manual hyper-parameter tuning, grid search, random search, Bayesian optimization, gradient-based optimization, and evolutionary optimization are the traditional techniques widely used in machine learning methods.

Moreover, learning rate, batch size, momentum, and weight decay are hyper-parameter tuning techniques in deep learning. Adjusting and tuning hyper-parameters related to deep learning is considerably complex, and yet, if done, it can have a positive effect on improving the classification process. The S-CNN method is based on 3D feature mapping to display spatial and temporal sections in 2D format. Similar to the methods presented in [94, 95], the convolutional neural network method has been employed to detect crowd anomaly behavior.

Fan et al. [11] proposed an algorithm to detect abnormal behavior in a set of video frames that used the spatiotemporal auto-encoder to solve the less negative samples challenge to extract features and improve learning. They designed a spatiotemporal convolutional neural network (sCNN) with a simple structure and low computational complexity. Their experiments were applied to the UCSD and UMN datasets. The algorithm was able to operate in real-time only by using a CPU.

4.3 Real-time design

Notwithstanding the fact that tracking models have been employed extensively in computer vision applications, they may not always be accurate and noise-robustness. On the other hand, one of the obvious problems with this method is that

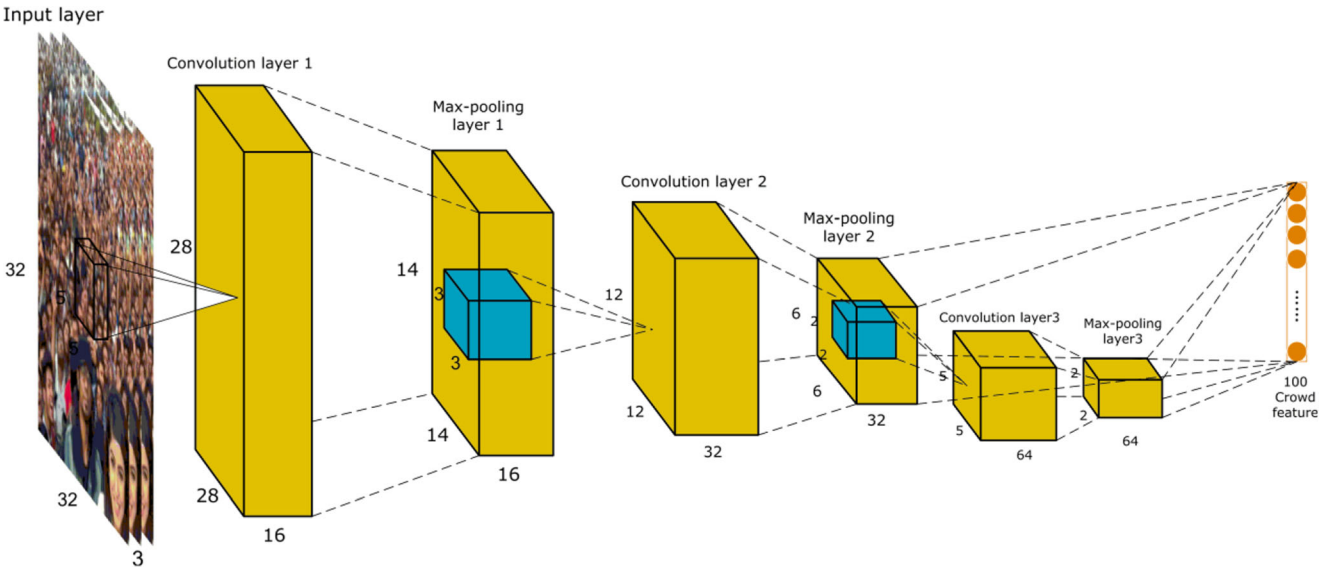

Fig. 6 The ConvNet structure for automated feature extraction from crowd scenes [92]

the object stays static. If the people in the scenes stop in their place, it will be difficult for them to be tracked. Although suggestions have been proposed to address this problem, in general, the appropriate method that can currently be effective in estimating people's location and tracking is the tracking strategy. However, when the number of objects in the frame increases, tracking is slightly delayed. Hence, to solve this difficulty, we implemented a scheme that reduces the complexity of processing time. Due to the deep learning network's need for the frames received from the tracking models, we can reduce the frames' dimensions and resolution to some extent. Figure 7 illustrates the flow plan for real-time processing of received frames from the social activities. If the number of frames contains moving people with high and low congestion be F, the processing time to refine the algorithm's responses is on average constant and in the range of milliseconds (t_R). However, the time required for tracking using the tracking model (t_K) varies due to changes in the number of people present in the crowded scenes. The proposed design can operate as the real-time model to detect the abnormal behavior based on people's movement when the inequality holds as (1).

$$(t_{k_i} + t_{R_i}) < t_p \tag{1}$$

4.4 Deep transfer learning

In many video processing applications, transfer learning models were employed to re-train deep learning approaches [96], and the behavior classification tasks in-crowd are assessed in terms of efficiency and accuracy of the model [97]. Deep learning approaches include layered architecture with various layers to learn multiple features, and eventually, all mentioned layers are joined to a fully connected layer to generate the concluding outcomes [98, 99]. In transfer learning, the layered structure can handle the pre-trained models such as VGG, ResNet, and AlexNet without its terminative classification layer as an accommodated feature extractor to obtain more reliable classification performance with less training time. The primary AlexNet includes five convolutional layers, three max-pooling layers, and three fully connected layers [100]. The frame input layer needs the frame of size 227 × 227 × 3. ReLU is implemented following each convolution and the fully connected layer, which extends the nonlinear attributes of the network design. Consequently, cross-channel normalization is used, and a dropout ratio of 0.5 is assumed.

A more direct path throughout the network results in the proper performance with very deep architectures for propagating information in the deep residual network (ResNet) [101]. The accuracy begins to saturate due to degradation difficulty. The degradation challenge happens due to the increment of network layers. The backpropagation procedure prevents from vanishing gradient problem; and therefore, ResNet uses backpropagation, which has shortcut connections parallel to the regular convolutional layers. This network helps to extract global features.

GoogleNet [102] is a deep convolutional neural network design that obtained proper classification outcomes with enhanced computational efficiency in various applications applying transfer learning [103, 104]. GoogleNet or Inception design includes 22 layers in deep consisting two convolution layers. Other layers are four max-pooling layers, one average pooling used at the end of the last inception module, and nine inception modules linearly stacked.

The depth of design is extended to 19 and 16 layers in the Visual Geometry Group (VGG) net [105]. The number of parameters is decreased by using very small convolution filters (3 × 3), which are VGG-19 and VGG-16. VGG design includes convolution layers with several continuous 3 × 3 convolutions. Two fully connected layers then support the 2 × 2 max-pooling layer with the ultimate layer as the Softmax output.

Sánchez et al. [106] proposed a taxonomic organization of current achievements following a pipeline. They discussed crowd behavior interpretation through deep learning and considered crowd emotions, datasets, anomaly detection, and other perspectives in crowd analysis. In [107], violent scene detection has been discussed considering action anomaly recognition in the crowd based on deep transfer learning.

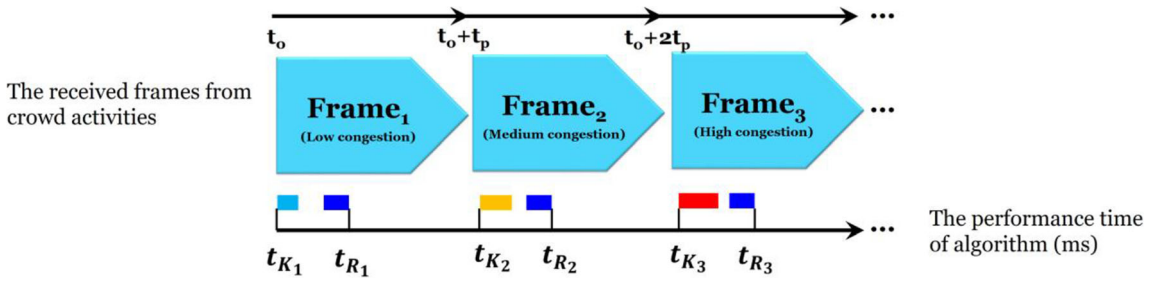

Fig. 7 The flow plan for real-time processing of received frames from the crowd activities

5 Performance comparison

The models that have been proposed so far for crowd anomaly detection have pros and cons.

Most models do not solve the problem of uncertainty of outputs and only report the output and estimate the results in accuracy, sensitivity, and specificity. One of the common problems among different condition analysis algorithms in a video frameset is not examining different conditions such as computational complexity, noise, pixel occlusion, efficiency, and cost. Visual image is received from a collection of library video frames [17]. In these frames, in addition to the various crowd scenarios that have been used to implement different behavioral classes, abnormal objects that can be a threat to a crowd have also been used. For example, a motorcycle that suddenly speeds through a crowd of pedestrians with a suspicious backpack dropped by a person in the crowd. In the proposed database, five different behavioral classes are defined. For each behavioral class, at least two different scenarios are extracted in terms of scenario structure, camera angle, and crowd density in the scene. The defined behavioral classes include (1) panic, (2) fighting, (3) congestion, (4) obstacle or abnormal object, (5) natural, and (6) nothing conditions in the crowd. To classify crowd behavior, we categorized natural states versus abnormal states. In total, about 29,220 video frames related to normal forms and 14,910 video frames related to abnormal conditions were recorded. In Table 1, the attributes related to the video data used are summarized. Figure 8 illustrates a frame Motion Emotion Dataset (MED) [108] that includes normal and abnormal crowd situations.

General methods based on tracking from the perspective of machine vision, such as optical flow and GMM to analyze different scenarios and several similar systems, are compared with learning-based methods such as handcrafted extracted features and automated feature extraction (i.e., deep learning) and hybrid approaches. In general, we have measured normal and abnormal behaviors, and the problem here is a binary classification. Evaluation criteria for the general assessment of the categorized methods include criteria such as inverse computational complexity (P_1), inefficiency (P_2), time (P_3), uncertainty (P_4), noise sensitivity and artifacts (P_5), and the loss of generalization (P_6) in achieving the solution. Figure 9 shows the relative estimates of the performance for the methods separately by computing the general evaluation methods with two iterations.

To assess the performance of the abnormal behavior detection model in-crowd, accuracy, sensitivity, and specificity are evaluated. Furthermore, the outcomes have been compared with similar methods in Table 2.

In Table 2, the values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are computed each time the algorithm is executed. Besides, the accuracy (Acc), sensitivity (Se), and specificity (Sp) outcomes are shown in Table 2 for simple, adjusted, and hybrid models. We have determined the accuracy of abnormal behavior detection in the crowd for various versions of combined models and the datasets employed for low- and high-congestion volumes in scenes. The following items were considered to analyze the performance of the models and compare them with similar methods:

1. Cost analysis according to the estimation of total costs resulting from model estimation and error analysis
2. Performance evaluation according to the average weight estimate of precision and recall indices
3. Analysis of the time according to the estimated time spent to run the algorithm in one model estimation
4. Analysis of uncertainty based on calculating the difference and dispersion of mean squares of error in different iterations
5. Investigation of noise sensitivity based on the calculation of the class related to crowd behavior and in the presence of artifacts such as noise application, climate change, pixel occlusion, and poor quality of received frames
6. Generalization analysis following the application of unseen frames on the classification and detection methods of individual and crowd behavior

Similar methods either use only video frames containing crowd anomaly conditions or exclusively analyze general behavior and do not analyze the influential features that cause a correlation in the response. Figure 9 also compares the abnormal behavior detection methods, including tracking, classification handcrafted, classification deep, and hybrid model methods according to criteria P_1 to P_6. It is observed that the smaller the area (i.e., which is scored from 1 to 5), the better

Table 1 Details of MED dataset

Behavior type	No. of frames	Details
Neutral	29,713	Moving with fixed velocity
Congestion	2364	Demonstration
Panic	2002	Backpack, hoodlum attack, terrorist firework
Obstacle	5120	Backpack, motorcycle crossing, bag theft
Fight	4423	Bad physical contact
Total	43,626	-

Fig. 8 Frames from the MED database that contain crowd anomaly behavior: **a** natural, **b** obstacle or abnormal object, **c** panic, **d** nothing, **e** fighting, and **f** congestion [108]

the performance of the method. Hybrid methods have yielded better results. Despite providing the desired accuracy, deep learning methods sometimes cannot develop a real-time method for detecting crowd behavior. The reason is that the application of the image type and the lack of special features in these methods require the adjustment of multiple parameters and high computational complexity. Moreover, the need for large volumes of data for training, the definition of the input data, the need for defining broad parameters of the deep classifier, processing time, ambiguity in feature processing, and the strong dependence on the definition of different combined layers must be included in the calculations.

Although the accuracy of previous studies for a limited number of frames was 68.2% to a maximum of 73% for identifying the accuracy of the crowd anomaly behavior classifier, the maximum accuracy despite high standard deviation is another drawback [88, 109]. Nonetheless, the lack of a need to define specific feature descriptors is one of their notable merits.

In methods such as those given in [110–112], the outputs are calculated with a high standard deviation level. The methods are focused only on examining the status of the crowd in limited classes. In these methods, the SVM classifier is used, which requires defining many parameters and cannot create a proper hyper-plane if it is not precisely adjusted. Compared to the methods mentioned above, approaches such as [41] employ neural networks separately, sometimes associated with over-fitting issues. However, due to the dynamic nature of the neural network, the dispersion of responses is high; yet, if the parameters are adjusted, it can create more optimal responses.

Some methods for tracking or classification are less accurate, and no solution has been adopted regarding the possibility of using them in real-time. However, it is not possible to

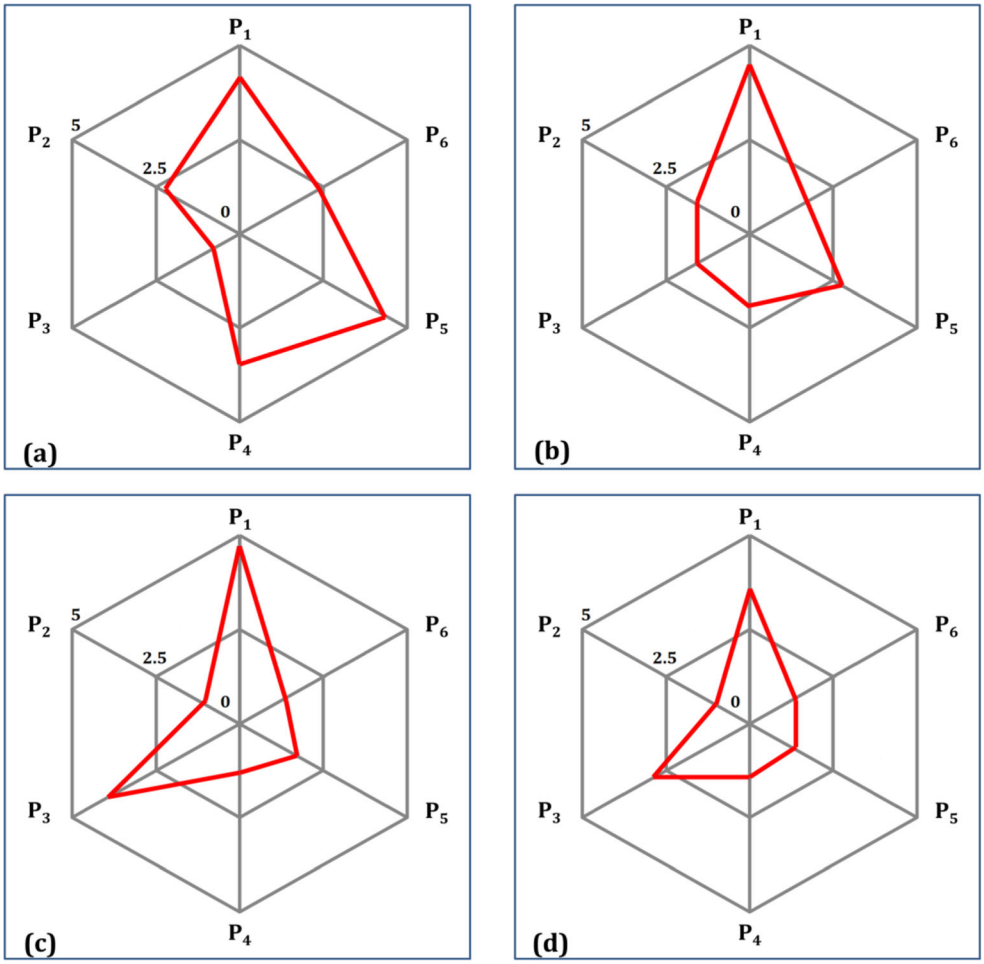

Fig. 9 Calculation of the overall performance of algorithms in detecting crowd anomaly behavior. **a** Tracking methods, **b** classification handcrafted methods, **c** classification deep methods, and **d** hybrid model according to evaluating the calculation of the introduced criteria P_1 to P_6

Table 2 Assessment of the performance of the automated and hybrid abnormal behavior detection model

Method	Accuracy	Sensitivity	Specificity
Tracking (simple)	0.6813	0.6920	0.6721
Handcrafted features (simple)	0.7641	0.7783	0.7554
Deep learning features (simple)	0.8538	0.8891	0.8316
Tracking (adjusted)	0.7849	0.7938	0.7667
Handcrafted features (adjusted)	0.8113	0.8228	0.7944
Deep learning features (adjusted)	0.8813	0.8934	0.8607
Tracking (hybrid)	0.8254	0.8416	0.8011
Handcrafted features (hybrid)	0.8862	0.9033	0.8749
Deep learning features (hybrid)	0.9413	0.9672	0.9381

draw a direct line between different methods in the analysis of crowd anomaly behavior. This issue is related to the type of data and the purpose of using the mentioned method. Some ways are different from other similar techniques in video processing and have a clear function for specific topics. We have compared the automated extracted feature-hybrid technique with similar models in Table 3. The deep transfer learning based on AlexNet structure and Kalman filter as tracking model have been used to conduct abnormal behavior detection.

Some abnormal event data have been collected according to users' needs and to assess the situation and the lack of efficient methods [64, 86, 94, 113]. With this approach, an experiment is performed between similar scenarios, where the MED and UCSD datasets are used in general to analyze the obtained frames. In crowd behavior analysis, criteria such as computational complexity and frame dimensions are compared between different methods in Fig. 10. We have considered three different sizes of frames in this experiment for two datasets. Each method is divided into three other ways. For

Table 3 Comparison of the criteria between the automated extracted features-hybrid model and other similar methods

Ref.	Features type	Dataset	Model type	Accuracy (%)	Computational complexity
[108]	Histogram of optical flow (HOF)	MED	SVM	37.69	Medium
[108]	Tracklet	MED	SVM and k-NN	38.17	Low
[108]	Motion boundary histogram (MBH)	MED	SVM	38.80	Medium
[114]	Automated	MED	3DCNN	34.05	High
[114]	Automated	MED	V_3G	36.99	High
[114]	Automated	MED	C_3D	51.22	High
[115]	Automated	MED	Dense Trajectories	43.64	High
[116]	Automated	MED	CNN	71.70	High
[117]	Automated	MED	3DCNN	90.91	High
[118]	Automated	MED	Cognitive deep model	93.82	High
Deep learning and hybrid	Automated	MED	Tracking and AlexNet	94.13	Medium

example, tracking methods are divided into simple, adjusted, and hybrid procedures, and then the benchmarks mentioned for them will be estimated.

Moreover, Fig. 10 depicts experiments for classification based on handcrafted extracted features and classification based on deep learning methods. This comparison is only a relative estimation and is not taken into account as an absolute estimation. However, most tracking-based models and handcrafted extracted features can be implemented in real-time or near-real-time. This can also involve a small part of deep learning systems and hybrid methods.

6 Conclusion

In the case of crowd anomalies or unusual congestions, automatic security analysis of the crowd behavior becomes possible. Automated detection of abnormal behavior in the crowd is

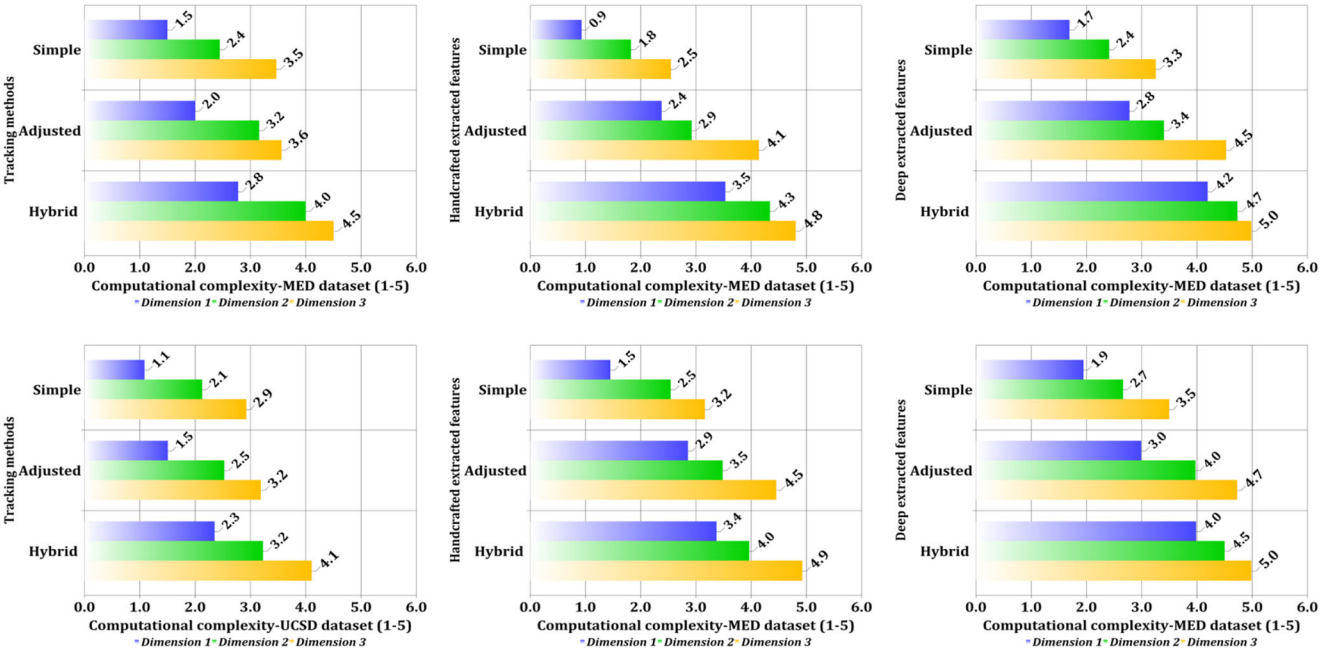

Fig. 10 Computational complexity evaluation for tracking, classification based on handcrafted extracted features, and classification based on deep learning methods of analyzing abnormal behavior in the crowds. Dimensions 1, 2, and 3 are frames with sizes equal to the half, equal to same, and equal to twice the sizes of original frames, respectively

of high importance because activities such as terrorist activities, fights, unusual and suspicious movements, etc., all require the supervision of operators and vigilant personnel to participate in security surveillance. However, this is a considerable challenge and leads to high cost and low accuracy in decision making. Therefore, designing a fatigue-free and error-free system that simultaneously offers real-time capability based on WoT platform will provide satisfactory impacts on controlling the crowd behavior. In this paper, different crowd anomaly detection methods are studied. Various aspects such as individual tracking, classification based on handcrafted extracted features, classification based on deep learning, and hybrid models are examined. It is found that deep learning methods and hybrid models have more satisfactory performance characteristics and can identify and predict crowd anomaly behavior. Nevertheless, in crowd behavior analysis, computational complexity has been considered in a few methods, where the reaction time to abnormal behavior can be reduced by reducing the processing time. The authors are looking to implement inferred patterns based on hybrid models and WoT platforms and reduce computational time and complexity by improving accuracy in future studies.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Varghese EB, Thampi SM (2020) Towards the cognitive and psychological perspectives of crowd behaviour: a vision-based analysis. Connect Sci 3:1–26
2. Yuan Y, Fang J, Wang Q (2014) Online anomaly detection in crowd scenes via structure analysis. IEEE transactions on cybernetics 45(3):548–561
3. Singh K, Rajora S, Vishwakarma DK, Tripathi G, Kumar S, Walia GS (2020) Crowd anomaly detection using aggregation of ensembles of fine-tuned ConvNets. Neurocomputing. 371:188–198
4. Mousavi H, Mohammadi S, Perina A, Chellali R, Murino V (2015) Analyzing tracklets for the detection of abnormal crowd behavior. In 2015 IEEE Winter Conference on Applications of Computer Vision (pp. 148-155)
5. Hatirnaz E, Sah M, Direkoglu C (2020) A novel framework and concept-based semantic search interface for abnormal crowd behaviour analysis in surveillance videos. Multimed Tools Appl 20: 1–39
6. Tripathi G, Singh K, Vishwakarma DK (2019) Convolutional neural networks for crowd behaviour analysis: a survey. Vis Comput 35(5):753–776
7. Tewell J, O'Sullivan D, Maiden N, Lockerbie J, Stumpf S (2019) Monitoring meaningful activities using small low-cost devices in a smart home. Pers Ubiquit Comput 23(2):339–357
8. Aguzzi C, Gigli L, Sciullo L, Trotta A, Di Felice M (2020) From cloud to edge: seamless software migration at the era of the web of things. IEEE Access 8:228118–228135
9. Wang J, Xu Z (2016) Spatio-temporal texture modelling for real-time crowd anomaly detection. Comput Vis Image Underst 144: 177–187
10. Zhang X, Ma D, Yu H, Huang Y, Howell P, Stevens B (2020) Scene perception guided crowd anomaly detection. Neurocomputing. 414:291–302
11. Fan Z, Yin J, Song Y, Liu Z (2020) Real-time and accurate abnormal behavior detection in videos. Mach Vis Appl 31(7):1–3
12. Fagette A, Courty N, Racoceanu D, Dufour JY (2014) Unsupervised dense crowd detection by multiscale texture analysis. Pattern Recogn Lett 44:126–133
13. Lamba S, Nain N (2017) Crowd monitoring and classification: a survey. In Advances in computer and computational sciences (pp. 21-31). Springer, Singapore
14. Marana AN, Costa LD, Lotufo RA, Velastin SA (1999) Estimating crowd density with Minkowski fractal dimension. In1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258) (Vol. 6, pp. 3521-3524). IEEE
15. Alabdulkarim L, Alrajhi W, Aloboud E (2016) Urban analytics in crowd management in the context of Hajj. In International Conference on Social Computing and Social Media (pp. 249-257). Springer, Cham
16. Ibrion M (2020) Iran: The impact of the beliefscape on the risk culture, resilience and disaster risk governance. Forensic Science and Humanitarian Action: Interacting with the Dead and the Living 10:117–134
17. Hodgetts HM, Vachon F, Chamberland C, Tremblay S (2017) See no evil: cognitive challenges of security surveillance and monitoring. Journal of applied research in memory and cognition 6(3): 230–243
18. Hao H, Li X, Li M A Detection method of abnormal event in crowds based on image entropy. InProceedings of the 2019 4th International Conference on Intelligent Information Processing 2019 Nov 16 (pp. 362-367)
19. Steiger C, Walder H, Platzner M (2004) Operating systems for reconfigurable embedded platforms: online scheduling of real-time tasks. IEEE Trans Comput 53(11):1393–1407
20. Rezaee K, Alavi SR, Madanian M, Ghezelbash MR, Khavari H, Haddadnia J (2013) Real-time intelligent alarm system of driver fatigue based on video sequences. In2013 First RSI/ISM International Conference on Robotics and Mechatronics (ICRoM)Feb 13 (pp. 378-383).
21. Aldissi B, Ammar H (2020) Real-time frequency-based detection of a panic behavior in human crowds. Multimed Tools Appl 79(33):24851–24871
22. Kh R, Ghezelbash MR, Haddadnia J, Delbari A (2012) An intelligent surveillance system for falling elderly detection based on video sequences. In19th Iranian Conference of Biomedical Engineering (ICBME), Tehran, Iran Dec (pp. 20-21)
23. Qasim T, Bhatti N (2019) A low dimensional descriptor for detection of anomalies in crowd videos. Math Comput Simul 166: 245–252
24. Indrusiak LS, Davis RI, Dziurzanski P (2019) Evolutionary optimisation of real-time systems and networks. arXiv preprint arXiv 1905:01888
25. Hu Y (2020) Design and implementation of abnormal behavior detection based on deep intelligent analysis algorithms in massive video surveillance. Journal of Grid Computing 1:1–1
26. Leyva R, Sanchez V Li CT. The LV dataset: a realistic surveillance video dataset for abnormal event detection. In2017 5th International Workshop on Biometrics and Forensics (IWBF) 2017 Apr 4 (pp. 1-6). IEEE
27. Popoola OP, Wang K (2012) Video-based abnormal human behavior recognition—a review. IEEE Transactions on Systems,

Man, and Cybernetics. Part C (Applications and Reviews) 42(6): 865–878
28. Mabrouk AB, Zagrouba E (2018 Jan 1) Abnormal behavior recognition for intelligent video surveillance systems: a review. Expert Syst Appl 91:480–491
29. Wang L, Dong M (2012) Real-time detection of abnormal crowd behavior using a matrix approximation-based approach. In2012 19th IEEE International Conference on Image Processing Sep (pp. 2701-2704). IEEE.
30. Ryan D, Denman S, Fookes C, Sridharan S (2011) Textures of optical flow for real-time anomaly detection in crowds. In2011 8th IEEE international conference on advanced video and signal based surveillance (AVSS) Aug 30 (pp. 230-235). IEEE
31. Ihaddadene N, Djeraba C (2008) Real-time crowd motion analysis. In2008 19th International Conference on Pattern Recognition Dec 8 (pp. 1-4). IEEE
32. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In2009 IEEE Conference on Computer Vision and Pattern Recognition Jun 20 (pp. 935-942). IEEE
33. Bay H, Tuytelaars T, Van Gool L (2006) SURF: speeded up robust features. InEuropean conference on computer vision (pp. 404-417). Springer, Berlin, Heidelberg
34. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110
35. Boghossian BA, Velastin SA (1999) Motion-based machine vision techniques for the management of large crowds. InICECS'99. Proceedings of ICECS'99. 6th IEEE International Conference on Electronics, Circuits and Systems (Cat. No. 99EX357) (Vol. 2, pp. 961-964). IEEE
36. Wang B, Ye M, Li X, Zhao F, Ding J (2012) Abnormal crowd behavior detection using high-frequency and spatio-temporal features. Mach Vis Appl 23(3):501–511
37. Horn BK, Schunck BG (1981) Determining optical flow. InTechniques and Applications of Image Understanding (Vol. 281, pp. 319-331). International Society for Optics and Photonics
38. Beauchemin SS, Barron JL (1995) The computation of optical flow. ACM computing surveys (CSUR) 27(3):433–466
39. Rezaee K, Haddadnia J, Delbari A (2015) Modeling abnormal walking of the elderly to predict risk of the falls using Kalman filter and motion estimation approach. Comput Electr Eng 46: 471–486
40. Ravanbakhsh M, Nabi M, Mousavi H, Sangineto E, Sebe N (2018) Plug-and-play CNN for crowd motion analysis: an application in abnormal event detection. In2018 IEEE Winter Conference on Applications of Computer Vision (WACV) pp. 1689-1698). IEEE
41. Sabokrou M, Fathy M, Hoseini M, Klette R (2015) Real-time anomaly detection and localization in crowded scenes. InProceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 56-62).
42. Chen YC, Su CT (2016) Distance-based margin support vector machine for classification. Appl Math Comput 283:141–152
43. Li W, Mahadevan V, Vasconcelos N (2013) Anomaly detection and localization in crowded scenes. IEEE Trans Pattern Anal Mach Intell 36(1):18–32
44. Yu J, Gwak J, Jeon M (2016) Gaussian-Poisson mixture model for anomaly detection of crowd behaviour. In2016 International Conference on Control, Automation and Information Sciences (ICCAIS) (pp. 106-111). IEEE
45. Lim KL, Wang H, Mou X (2016) Learning Gaussian mixture model with a maximization-maximization algorithm for image classification. In2016 12th IEEE International Conference on Control and Automation (ICCA) pp. 887-891). IEEE
46. Leyva R, Sanchez V, Li CT (2017) Video anomaly detection with compact feature sets for online performance. IEEE Trans Image Process 26(7):3463–3478
47. Lu C, Shi J, Wang W, Jia J (2019) Fast abnormal event detection. Int J Comput Vis 127(8):993–1011
48. Marsden M, McGuinness K, Little S, O'Connor NE (2016) Holistic features for real-time crowd behaviour anomaly detection. IEEE International Conference on Image Processing (ICIP):918–922 IEEE
49. Kaltsa V, Briassouli A, Kompatsiaris I, Hadjileontiadis LJ, Strintzis MG (2015) Swarm intelligence for detecting interesting events in crowded environments. IEEE Trans Image Process 24(7):2153–2166
50. Pennisi A, Bloisi DD, Iocchi L (2016) Online real-time crowd behavior detection in video sequences. Comput Vis Image Underst 144:166–176
51. Wang Q, Ma Q, Luo CH, Liu HY, Zhang CL (2016) Hybrid histogram of oriented optical flow for abnormal behavior detection in crowd scenes. Int J Pattern Recognit Artif Intell 30(02): 1655007
52. Cai Y, Wang H, Chen X, Jiang H (2015) Trajectory-based anomalous behaviour detection for intelligent traffic surveillance. IET Intell Transp Syst 9(8):810–816
53. Zhu G, Song K, Zhang P, Wang L (2016) A traffic flow state transition model for urban road network based on Hidden Markov Model. Neurocomputing. 214:567–574
54. Kwon Y, Kang K, Jin J, Moon J, Park J (2017) Hierarchically linked infinite hidden Markov model based trajectory analysis and semantic region retrieval in a trajectory dataset. Expert Syst Appl 78:386–395
55. Sun S, Zhao J, Gao Q (2015) Modeling and recognizing human trajectories with beta process hidden Markov models. Pattern Recogn 48(8):2407–2417
56. Ding W, Liu K, Fu X, Cheng F (2016) Profile HMMs for skeleton-based human action recognition. Signal Process Image Commun 42:109–119
57. Zhou L, Li W, Ogunbona P, Zhang Z (2017) Semantic action recognition by learning a pose lexicon. Pattern Recogn 72:548–562
58. Wang Y, Zhang X, Li M, Jiang P, Wang F (2015) A GM-HMM based abnormal pedestrian behavior detection method. IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC):1, IEEE–6
59. Zheng CH, Pei WJ, Yan Q, Chong YW (2017 Mar 8) Pedestrian detection based on gradient and texture feature integration. Neurocomputing. 228:71–78
60. Güngör E, Özmen A (2017 Mar 1) Distance and density based clustering algorithm using Gaussian kernel. Expert Syst Appl 69: 10–20
61. Zang X, Li G, Li Z, Li N, Wang W (2016) An object-aware anomaly detection and localization in surveillance videos. IEEE Second International Conference on Multimedia Big Data (BigMM):113–116 IEEE
62. Wang X, Fan B, Chang S, Wang Z, Liu X, Tao D, Huang TS (2017) Greedy batch-based minimum-cost flows for tracking multiple objects. IEEE Trans Image Process 26(10):4765–4776
63. Zhou S, Shen W, Zeng D, Zhang Z (2015) Unusual event detection in crowded scenes by trajectory analysis. In2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 1300-1304). IEEE
64. Afiq AA, Zakariya MA, Saad MN, Nurfarzana AA, Khir MH, Fadzil AF, Jale A, Gunawan W, Izuddin ZA, Faizari M (2019) A review on classifying abnormal behavior in crowd scene. J Vis Commun Image Represent 58:285–303

65. Fradi H, Luvison B, Pham QC (2016) Crowd behavior analysis using local mid-level visual descriptors. IEEE Transactions on Circuits and Systems for Video Technology 27(3):589–602
66. Biswas S, Babu RV (2017) Anomaly detection via short local trajectories. Neurocomputing. 242:63–72
67. Luo X, Tan H, Guan Q, Liu T, Zhuo HH, Shen B (2016) Abnormal activity detection using pyroelectric infrared sensors. Sensors. 16(6):822
68. Zweng A, Kampel M (2010) Unexpected human behavior recognition in image sequences using multiple features. In2010 20th International Conference on Pattern Recognition (pp. 368-371). IEEE.
69. Xiang T, Gong S (2008) Video behavior profiling for anomaly detection. IEEE Trans Pattern Anal Mach Intell 30(5):893–908
70. Saligrama V, Konrad J, Jodoin PM (2010) Video anomaly identification. IEEE Signal Process Mag 27(5):18–33
71. Wang X, Ma X, Grimson WE (2008) Unsupervised activity perception in crowded and complicated scenes using hierarchical Bayesian models. IEEE Trans Pattern Anal Mach Intell 31(3): 539–555
72. Simon C, Meessen J, De Vleeschouwer C (2010) Visual event recognition using decision trees. Multimed Tools Appl 50(1): 95–121
73. Johnson N, Hogg D (1996) Learning the distribution of object trajectories for event recognition. Image Vis Comput 14(8):609–615
74. Yilmaz A, Javed O, Shah M (2004) Object tracking: a survey. Acm computing surveys (CSUR). 2006 Dec 25;38(4):13-es
75. Junejo IN, Javed O, Shah M (2004) Multi feature path modeling for video surveillance. InProceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. (Vol. 2, pp. 716-719). IEEE
76. Hu W, Tan T, Wang L, Maybank S (2004) A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews).;34(3):334-52
77. Fu Z, Hu W, Tan T (2005) Similarity based vehicle trajectory clustering and anomaly detection. InIEEE International Conference on Image Processing 2005 Sep 14 (Vol. 2, pp. II-602). Ieee
78. Piciarelli C, Micheloni C, Foresti GL (2008) Trajectory-based anomalous event detection. IEEE transactions on circuits and systems for video technology 18(11):1544–1554
79. Kumar P, Ranganath S, Weimin H, Sengupta K (2005) Framework for real-time behavior interpretation from traffic video. IEEE Trans Intell Transp Syst 6(1):43–53
80. Vaswani N, Roy-Chowdhury AK, Chellappa R (2005) " Shape activity": a continuous-state HMM for moving/deforming shapes with application to abnormal activity detection. IEEE Trans Image Process 14(10):1603–1616
81. Zou J, Ye Q, Cui Y, Wan F, Fu K, Jiao J (2016) Collective motion pattern inference via locally consistent latent Dirichlet allocation. Neurocomputing. 184:221–231
82. Chaker R, Al Aghbari Z, Junejo IN (2017) Social network model for crowd anomaly detection and localization. Pattern Recogn 61: 266–281
83. Singh D, Mohan CK (2017) Graph formulation of video activities for abnormal activity recognition. Pattern Recogn 65:265–272
84. Riveiro M, Lebram M, Elmer M (2017) Anomaly detection for road traffic: a visual analytics framework. IEEE Trans Intell Transp Syst 18(8):2260–2270
85. Yan W, Zou Z, Xie J, Liu T, Li P (2018) The detecting of abnormal crowd activities based on motion vector. Optik. 166:248–256
86. Swathi HY, Shivakumar G, Mohana HS (2017) Crowd behavior analysis: a survey. In2017 international conference on recent advances in electronics and communication technology (ICRAECT) (pp. 169-178). IEEE.
87. Contractor U, Dixit C, Mahajan D (2018) CNNs for surveillance footage scene classification. arXiv preprint arXiv 1809:02766
88. Kotapalle GR, Kotni S (2018) Security using image processing and deep convolutional neural networks. In2018 IEEE International Conference on Innovative Research and Development (ICIRD) pp. 1-6). IEEE
89. Xie S, Zhang X, Cai J (2019) Video crowd detection and abnormal behavior model detection based on machine learning method. Neural Comput & Applic 31(1):175–184
90. Zhou S, Shen W, Zeng D, Fang M, Wei Y, Zhang Z (2016) Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes. Signal Process Image Commun 47:358–368
91. Liu M, Li S, Shan S, Wang R, Chen X (2014) Deeply learning deformable facial action parts model for dynamic expression analysis. InAsian conference on computer vision (pp. 143-157). Springer, Cham
92. Hu Y, Chang H, Nian F, Wang Y, Li T (2016) Dense crowd counting from still images with convolutional neural networks. J Vis Commun Image Represent 38:530–539
93. Shao J, Loy CC, Kang K, Wang X (2016) Slicing convolutional neural network for crowd video understanding. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 5620-5628)
94. Zitouni MS, Sluzek A, Bhaskar H (2019) Visual analysis of socio-cognitive crowd behaviors for surveillance: a survey and categorization of trends and methods. Eng Appl Artif Intell 82:294–312
95. Yi S, Li H, Wang X (2016) Pedestrian behavior understanding and prediction with deep neural networks. InEuropean Conference on Computer Vision Oct 8 (pp. 263-279). Springer, Cham
96. Rezaee K, Badiei A, Meshgini S (2020) A hybrid deep transfer learning based approach for COVID-19 classification in chest X-ray images. In2020 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME) (pp. 234-241)
97. Chaturvedi I, Ong YS, Arumugam RV (2015) Deep transfer learning for classification of time-delayed Gaussian networks. Signal Process 110:250–262
98. Bendali-Braham M, Weber J, Forestier G, Idoumghar L, Muller PA (2019) Transfer learning for the classification of video-recorded crowd movements. In2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA) (pp. 271-276)
99. Da Silva FL, Costa AH (2019) A survey on transfer learning for multiagent reinforcement learning systems. J Artif Intell Res 64: 645–703
100. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Adv Neural Inf Proces Syst 25:1097–1105
101. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. InProceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778)
102. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. InProceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9)
103. Canziani A, Paszke A, Culurciello E (2016) An analysis of deep neural network models for practical applications. arXiv preprint arXiv 1605:07678
104. Ballester P, Araujo R(2016) On the performance of GoogLeNet and AlexNet applied to sketches. InProceedings of the AAAI Conference on Artificial Intelligence (Vol. 30, No. 1)
105. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv 1409:1556

106. Sánchez FL, Hupont I, Tabik S, Herrera F (2020) Revisiting crowd behaviour analysis through deep learning: taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects. Information Fusion 29
107. Keçeli AS, Kaya AY (2017 Jun 20) Violent activity detection with transfer learning method. Electron Lett 53(15):1047–1048
108. Rabiee H, Haddadnia J, Mousavi H, Kalantarzadeh M, Nabi M, Murino V (2016) Novel dataset for fine-grained abnormal behavior understanding in crowd. In2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) Aug 23 (pp. 95-101). IEEE
109. Khan G, Farooq MA, Hussain J, Tariq Z, Khan MU (2019) Categorization of crowd varieties using deep concurrent convolution neural network. In2019 2nd International Conference on Advancements in Computational Sciences (ICACS) Feb 18 (pp. 1-6). IEEE
110. Yogameena B, Komagal E, Archana M, Abhaikumar SR (2010) Support vector machine-based human behavior classification in crowd through projection and star skeletonization. J Comput Sci 6(9):1008–1013
111. Wang T, Snoussi H (2014) Detection of abnormal visual events via global optical flow orientation histogram. IEEE Transactions on Information Forensics and Security 9(6):988–998
112. Wang T, Snoussi H (2012) Histograms of optical flow orientation for visual abnormal events detection. In2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance Sep 18 (pp. 13-18). IEEE
113. Yogameena B, Nagananthini C (2017) Computer vision based crowd disaster avoidance system: a survey. International journal of disaster risk reduction 22:95–129
114. Dupont C, Tobias L, Luvison B (2017) Crowd-11: a dataset for fine grained crowd behaviour analysis. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (pp. 9-16)
115. Rabiee H, Haddadnia J, Mousavi H (2016) Crowd behavior representation: an attribute-based approach. SpringerPlus. 5(1):1–7
116. Lazaridis L, Dimou A, Daras P (2018) Abnormal behavior detection in crowded scenes using density heatmaps and optical flow. In2018 26th European Signal Processing Conference (EUSIPCO) Sep 3 (pp. 2060-2064). IEEE
117. Varghese EB, Thampi SM (2018) A deep learning approach to predict crowd behavior based on emotion. InInternational Conference on Smart Multimedia Aug 24 (pp. 296-307). Springer, Cham
118. Varghese E, Thampi SM, Berretti S (2020) A psychologically inspired fuzzy cognitive deep learning framework to predict crowd behavior. IEEE Trans Affect Comput 13