Multimedia Tools and Applications (2023) 82:15673–15694
https://doi.org/10.1007/s11042-022-13967-w

Anomalous event detection and localization in dense crowd scenes

Areej Alhothali[1] · **Amal Balabid**[1] · **Reem Alharthi**[1] · **Bander Alzahrani**[1] · **Reem Alotaibi**[1] · **Ahmed Barnawi**[1]

Received: 13 May 2022 / Revised: 18 August 2022 / Accepted: 12 September 2022 /
Published online: 15 October 2022

Abstract
Recognizing and localizing anomalous events in crowd scenes is a challenging problem that has attracted the attention of researchers in computer vision. Surveillance cameras record scenes that require an automated examination to identify anomalous events. Existing approaches in the field have utilized different feature descriptors, modeling methods, and recognition strategies to accurately and efficiently detect anomalies in the scene. Existing techniques in the field have focused mainly on performing global frame-level identification of abnormal events. Only a small number of studies have considered locating abnormal action in the frame. Proposed methods are also often evaluated on scenes that contain a sparse number of individuals performing abnormal and normal staged acts. This research aims to detect and locate anomalies in a structured and unstructured dense crowd scene. The proposed model first detects moving objects and individuals in the scene using a deep convolutional neural network and tracks objects and individuals using spatial and temporal features. Then, spatial-temporal features are extracted from consecutive frames of interest points. The extracted features include the histogram of optical flow, velocity and direction of moving objects, and other features that can indicate sudden motion change. A support vector machine model is then used to classify abnormal events into one of seven classes. The proposed methodology is evaluated on Hajj2 dataset that has 18 videos and 7 different types of abnormal events.

Keywords Crowd anomaly detection · Hajj Scene · Magnitude of optical flow · Spatial-temporal features

✉ Areej Alhothali
aalhothali@kau.edu.sa

✉ Bander Alzahrani
baalzahrani@kau.edu.sa

[1] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

1 Introduction

Growing attention has been paid to building intelligent monitoring and surveillance systems that detect and recognize human activities in real time. This interest has led researchers to investigate techniques that detect and locate abnormal events in video scenes to decrease the dependency on the timely and costly operation of manual inspection. Anomalous events can be generally defined as events that deviate from the prevalent behaviors in the scene. The definition of abnormal events varies according to the context of the event. For example, the presence of a car on the road might be considered normal in some contexts and abnormal in others.

Researchers have followed one of two methodologies to detect abnormal behavior in crowd scenes. The first methodology comprises feature descriptors to extract motion or/and appearance features from pixels, patches, or volumes of frames, along with a modeling approach to characterize normal behavior and detect anomalies. The second methodology uses deep learning models that jointly extract indicative features from consecutive frames and learn to detect abnormalities. To extract motions or appearance features, researchers have focused on trajectory-based feature descriptors, dense feature descriptors, and spatial-temporal or volume-based feature descriptors. The trajectory-based [14, 35, 38, 49, 81, 85, 86] or tracklets-based approaches [54, 55] aim to track interesting points in the scene either by using deep learning objects detection approaches or by segmenting frames into motion blocks, patches, or superpixels and then tracking objects or frame segments throughout the entire or some part of the scene. The trajectory features are then used to model normal behavior using unsupervised methods such as k-means clustering or Gaussian mixture model (GMM). Abnormal events deviating from normal actions were identified using a judging formula and an experimental threshold.

To eliminate the problem and complications of tracking moving objects in crowd scenes, researchers use dense-based descriptors that extract features from the entire frame pixels or patches. Several dense feature descriptors have been used, such as the optical flow feature [11, 13, 69] and histogram of optical flow (HOF) descriptor [73] representing pixel displacement in two consecutive frames. Researchers have also examined the histogram of oriented gradient (HOG) [31, 52], the mixture of dynamics texture [18, 47], and kinetic energy [36, 83]. Several modeling approaches have been used to detect anomalies, such as social force [29, 56], dictionary and sparse representation [24, 34], Gaussian mixture model (GMM) [78, 85], and once class support vector machine [25, 59]. Recent studies have developed models that extract motion features from 3D volume or cuboids to incorporate spatial features [22, 27]. Spatial-temporal approaches have shown promising results in comparison with dense features descriptors. Other studies have developed deep learning models such as convolutional neural networks, convolutional auto-encoder [37, 53, 80], and convolutional long short-term memory [60] for feature extraction and anomaly detection in crowd scenes.

One of the observed limitations in crowd anomaly detection is that most of the datasets used in the field are small and have limited types of anomalies that focus on objects' appearance, motion, or location [48, 51]. This could entail, for example, the presence of a rare object, panic escape behavior, or the presence of non-human objects on the sidewalk. Less focus has been given to anomalous behaviors that involve interaction among individuals and those that appear in crowd scenes. In addition, most of the proposed methods were formulated as binary classification or outlier detection, and few studies have looked into the multi-class classification of abnormal behavior.

To summarize, the main contribution of this work is to propose a deep learning method that detects, localizes, and recognizes anomalous behavior in high-density crowds. The proposed model first detects and tracks objects and then extracts features from the detected objects, including the velocity and histogram of optical flow. The extracted features are then classified using a support vector machine. The model is evaluated on a real-world high-dense anomaly dataset. The dataset depicts a typical surveillance recording of Hajj scenes that shows different types of anomalies that each requires specific kinds of spatial-temporal features to be identified. Thus, the proposed model classifies abnormal behaviors into seven anomaly categories rather than the traditional binary classification used in the field.

The proposed framework outperforms a previous study that addressed the same problem by 12.88 AUC [3]. The rest of the paper is organized as follows: we provide a brief summary of some of the related works in Section 2. The details of the used dataset are described in Section 3. The suggested method is presented in Section 4. Experimental evaluations are detailed in Section 5. Finally, Section 6 summarizes this work.

2 Related work

Anomaly detection has gained extensive attention in recent years due to the availability of surveillance data and the need for automatic inspection of surveillance videos. To identify abnormal behavior in surveillance videos, researchers have investigated a wide range of feature descriptors to describe and characterize abnormal and normal events in some contexts. They also studied a large number of supervised [25, 59], unsupervised [32, 61, 85], and weakly supervised [43, 68] machine learning models to identify, locate, and recognize abnormal events. Deep learning techniques have also been investigated for jointly extracting features and recognizing irregularities. Identifying abnormal actions can be achieved by modeling normal actions and identifying actions that deviate from the norm or by classifying abnormal behavior into normal and abnormal actions. Models such as convolutional autoencoders [26], unsupervised double stream variational autoencoder and long-short term memory (LSTM) [76], weakly supervised convolutional graph neural networks [39], generative adversarial networks (GAN) with spatial and motion attention [84], and graph adversarial convolutional neural networks [15] have also been recently investigated for anomaly detection. The anomaly detection methodologies are evaluated against other approaches on benchmark datasets such as UCSD and UMN.

Detecting abnormal events in crowd scenes is also widely approached using features descriptor and modeling algorithms to extract features, model normal behavior, and identify abnormal actions that deviate from normal behavior. Recent studies have tackled the problem using end-to-end deep learning methods to extract features and identify irregularities. Feature descriptors are often categorized into three major categories: trajectory-based, dense-based, and deep learning-based approaches. Trajectory-based approaches aim at tracking moving objects through the entire or part of the scene to detect spatial and temporal anomalies in crowd scenes. Bera et al. [8] modeled pedestrians' local and global behavior based on trajectory features and Bayesian inference techniques to detect abnormal events. Zhao et al. [85] proposed a two-phase approach to extract point trajectory-based histogram of optical flow features and used GMM and k-mean clustering to model normal motion and identify abnormal events. Li et al. [35] utilized trajectory features as a post-processing stage to track anomaly candidates and obtain their global motion pattern.

Several other approaches analyzed and grouped trajectories into various clusters based on their motion characteristics, such as direction, distance, and speed. Then they estimated the normal group trajectory to identify anomalous cases deviating from other groups [14, 46]. Researchers also used tracklets or short local trajectories to provide a more robust motion representation than long trajectories. Moustafa and Gomaa [54] developed a model that utilizes tracklets features and long short-term memory to detect abnormal events. Marsden et al. [49] modeled crowd motions based on low-level holistic tracklets features. Biswas and Venkatesh Babu [10] developed a crowd anomaly detection approach using a short history of local motions and a hidden Markov model. Despite the importance of tracking pedestrians throughout the entire scene and analyzing their temporal motion history to detect and recognize abnormal behavior, real-world pedestrian tracking and prediction remain difficult, particularly as crowd density rises as a result of intra-pedestrian occlusion.

To overcome the challenges associated with tracking individuals in crowd scenes, researchers have focused more on extracting dense features, which are often computed on the basis of the differences between the displacement of pixels, patches or superpixels in consecutive frames. Tomé and Salgado [69] developed an approach based on an optical flow feature descriptor that extracts the magnitude and textures of optical flow from cuboid volumes. A Gaussian mixture model was used to model normal behavior and detect abnormalities. Chen and Shao [11] developed an optical flow feature descriptor that extracts magnitude, location, direction, and weighted velocity to describe escape behavior in a crowd scene and employed diverging centers to identify anomalous behavior in crowd scenarios. Guo et al. [20] proposed an approach that uses the optical flow and sparse linear models for feature extractions, estimates Gaussian prior distributions over the extracted features, and employs the Infinite Hidden Markov model to identify abnormal events. Khan et al. [32] proposed a method that obtains optical flow features from super-pixels to represent motion spatially over consecutive frames and uses k-means clustering algorithm and univariate Gaussian discriminant analysis for anomalous behavior detection. Bansod and Nandedkar [7] developed a model that combines appearance and motion attributes with the momentum and histogram of the magnitude of foreground objects. The normal crowd behavior is learned using an unsupervised clustering approach, while abnormalities are located using positional characteristics. Guo et al. [21] proposed a model that uses an optical flow velocity field to represent crowd motions with an enhanced k-means algorithm to detect abnormal events in the crowd scene. Histograms of optical flow were investigated in various studies to detect abnormal events [12, 33, 34, 57, 74]. Chen and Wang [12] used a weighted multi-histogram of oriented optical flow (WMOF) feature descriptor and sparse representation learning method to detect abnormal events. Li et al. [34] utilized histogram of maximal optical flow feature descriptor and online dictionary learning with sparse reconstruction method to identify abnormal behavior. Patil and Biswas [57] presented a model that utilizes a spatial-temporal feature descriptor to extract HOF and magnitude of optical flow from different sizes of frame's blocks and uses a one-class SVM model for abnormal event detection. Wang et al. [74] proposed a model that utilizes optical flow descriptor for spatial-temporal feature extraction from foreground objects. They also used principal component analysis foreground texture selection and SVM for classification. Li et al. [33] proposed a model that extracts histogram of maximal optical flow (HMOFP) features based on the saliency map of the optical flow field and uses online dictionary learning trained on normal samples. To identify abnormal behavior, sparse reconstruction coefficients (SRC) are calculated for testing samples. Lin et al. [41] developed a model that extracts HOF from spatial-temporal patches and temporal patches, then employs an enhanced one-class SVM for anomalous events detection.

Recent studies in the field have used deep learning models to learn representative features and identify anomalies in crowd scenes. Several researchers developed a deep learning method with optical flow. Almazroey and Jarraya [4] proposed an approach that computes 2D optical flow features of scenes' keyframes, extracts high-level features using pre-trained convolution neural networks (CNN), and identifies anomalies using an SVM classifier. Feng et al. [17] proposed a deep GMM model that trained on appearance and motion features extracted using a PCANet model from 3D gradients. Bansod and Nandedkar [6] proposed a model that detects and localizes anomalies using optical flow and stack autoencoder. Mondal and Chanda [53] proposed model that computes and compares the magnitude of optical flow against the mean flow magnitude of normal motion. An autoencoder model is then trained to reconstruct the mean optical flow patch given the corresponding flow patch from each frame. In the testing phase, high reconstruction errors indicate anomalous events accrued. Sabokrou et al. [63] proposed an auto-encoders feature descriptor that extracts from spatial-temporal cubic patches. Abnormal events that deviate from normal behavior were identified using classifiers trained to model global and local normal events.

Other studies used both handcrafted features and deep learning features. For instance, Ilyas et al. [28] developed a hybrid deep network approach that combines handcrafted features with deep learning features to detect anomalous events. Hu et al. [25] developed a model that identifies moving objects using Fast R-CNN model and extracts motion information from identified regions using the Histogram of Large Scale Optical Flow (HLSOF) to construct a magnitude and direction map. The extracted features were then down-sampled and used to train a multi-label SVM. The use of CNNs with motion features was investigated by several researchers to identify anomalous events. Direkoglu [16] introduced an approach that combines motion information images (MIIs) and CNN models to detect anomalies. Generative adversarial network (GAN) and motion features were also used in the field [22, 72]. Sabih and Vishwakarma [62] used CNN and bidirectional LSTM to learn motion features of optical flow. Zhang et al [82] developed an approach that reconstructs frames using HOF and HOG features or autoencoder features, and the reconstruction error used to determine anomalous events.

Researchers have also investigated deep learning descriptors to extract spatial and temporal characteristics of crowd behavior and identify abnormal behavior. Mehmood [50] developed pre-trained 2D-CNN models to detect and localize anomalous events. Joshi and Patel [30] proposed a CNN-based approach for global anomaly detection. Autoencoders and convolutional autoencoders (CAEs) has been widely used for anomalous event detection. Ramchandran and Sangaiah [60] proposed a convolutional autoencoder and convolutional LSTM model to reconstruct frame and frame edges. Anomaly events are identified using the reconstruction error. Aqeel et al. [5] introduced a method that uses a convolutional autoencoder and GAN with different classification models to extract features and detect anomalous events. Sabokrou et al. [64] developed a cascade deep learning model using 3D auto-encoders that examine small cubic patches to detect normal patches and use a 3D CNN model to further evaluate the region of interest. Also, Li et al. [37] proposed a 3D spatial-temporal cascade autoencoder for local and global anomaly detection. Xu et al. [79] presented a convolutional variational auto-encoder to learn appearance and motion features, with Gaussian models to model normal behavior and detect anomalies. Wang et al. [75] developed a model that consists of two stages: a stacked fully connected variational autoencoder and a convolutional variational auto-encoder. The first stage is a shallow network that filters some visible normal samples, and the second stage learns hierarchical and local relationship between features from the sampled input. Gnouma et al. [19] developed a model that first identifies the region of interest using a binary quantization map (BQM) and then uses a stacked autoencoder to extract and detect anomalies. Graph-based representation with deep learning features was employed for anomaly detection [42, 43].

3 Dataset

The Hajj2 dataset published by Alafif et al [3] was used in this research, including abnormal behaviors in a high-density crowd. This dataset was manually collected and labeled by a research group. The dataset presents scenes from Hajj and Umrah and includes 18 videos captured at four different locations in the Hajj. These locations include Tawaf, Mas'a, Jamarat, and Arafat. Video locations vary in terms of the degree of crowding, the conditions of acquisition, and the types of abnormal behavior present. Figure 1 shows an example of each location in the dataset.

The abnormal behavior in this dataset is categorized into seven abnormal classes that may present a risk to large-scale crowd movements. These classes include standing, sitting, sleeping, running, moving in the opposite direction, and moving in a different direction to the crowd, in addition to non-human objects such as vehicles and wheelchairs. Figure 2 shows an example of each of these classes. The dataset was divided into two sets, training and testing. The training set contains nine videos; each is about 25 seconds and includes 700 frames. While the testing set had seven videos, each video is about 20 seconds and consists of 500 frames. More specifically, 170,772 subjects showing abnormal behavior were labeled and identified in the training set. In contrast, the test set consists of 129,769 samples with abnormal behavior.

4 Methodology

The suggested framework is divided into two parts: the first detects and locates individuals at the frame level, and then tracks them in successive frames. The second detects abnormal behavior in the crowd by extracting features for each individual and then classifying their behavior. Figure 3 shows these two parts. The following subsections discuss the proposed techniques accordingly.

4.1 Part1: multi-objects detection and tracking

The Hajj dataset includes scenes of a high-density crowd with heavy occlusions. Individual detection at these scenes is a tricky problem that significantly impacts the framework's performance. The proposed solution to tackle this problem and detect and track pilgrims is

(a) Masa'a (b) Jamarat (c) Tawaf (d) Arafat

Fig. 1 Images of the test dataset at the four different locations

Fig. 2 Abnormal behaviors examples in the HAJJ dataset

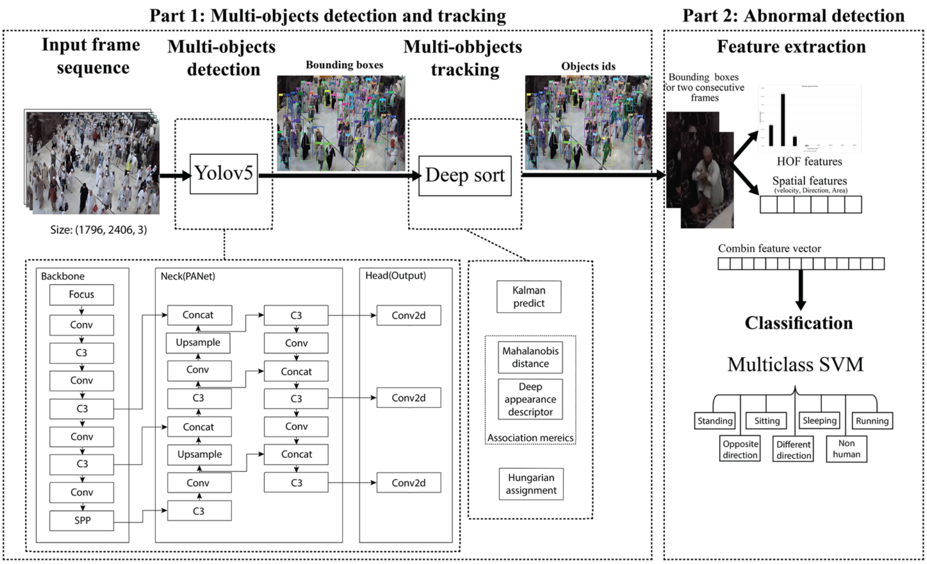


Fig. 3 The Architecture of the proposed framework

covered in the following subsections. Algorithm 1 also provides an overview of the applied procedures in this part.

Input: v, video stream

Output: Bounding boxes $b_{f_i,n}$ with track-ids list where f_i is the ith frame and n is the bounding box number within f_i frame

1: **procedure** DETECTION(v)
2: $boxes \leftarrow [][]$
3: **for each** frame f_i in v **do**
4: $boxes_{fi} \leftarrow finetuned\_YOLO(f_i)$
5: **procedure** TRACKING($boxes$)
6: $track\_ids \leftarrow [][]$
7: **for each** box $b_{f,n}$ in boxes **do**
8: $track\_ids_{f_i,n} \leftarrow DeepSort(b_{f_i,n})$

Algorithm 1 Part 1: Detection and tracking.

4.1.1 Localization and object detection

This study utilizes a deep CNN model (Yolov5) [58] that was fine-tuned on the CrowdHuman dataset [66] to detect individual people in crowd scenes. The CrowdHuman is a benchmark dataset that includes annotated images of crowds. This dataset has an average of 23 persons per image, with various kinds of occlusions. The Hajj dataset, in comparison, includes an average of 119 persons per image. In this case, an object detector that has been trained or fine-tuned on a crowd dataset is strongly recommended.

Yolov5 is an object detection model trained on the COCO dataset [40] that is believed to be much faster and lighter than previous YOLO releases. At the same time, the accuracy of Yolov5 is on par with the Yolov4 benchmark. The three main parts of the Yolov5 are the backbone, neck, and the head. The backbone is the Cross Stage Partial Networks (CSP), which is used to extract important features from an input image. The Yolov5 neck, on the other hand, generates the feature pyramids using PANet [44].

Figure 3 depicts the Yolov5 structure. The structure begins with the focus layer that divides the image of size (1796, 2406, 3)into several segments, and these segments are subsequently divided into layers. Conv runs Conv2d, calculates the value of batchNorm2d and LeakyRELU, and outputs the result. C3 transforms the input data, computes it at the bottleneck layer, adds the value of Conv of the initial input to the value of the computation at the bottleneck layer in Concat, then converges and outputs it. The bottleneck continues with Conv(1, 1) on the input value and outputs the calculated value of Conv(3, 1). Following the Conv operation, SPP exports three MaxPooling values (5×5, 9×9, and 13×13), merges them with the Conv value from the current input value in Concat, and then sends them out. Then the number of each array of feature maps in the structure values is doubled by upsampling. Concat is responsible for combining input layers. Consequently, combining the three Conv2d values detects and outputs them.

The initial phase of the proposed framework involves using the fine-tuned version of Yolov5 to recognize individual persons in each frame. The output of this phase is a list of the bounding boxes of all the individuals as shown in Fig. 4. However, numerous misdetections

Fig. 4 Yolov5 output on a single frame

occur as a result of the high density of the crowd and occlusions. Table 1 shows the true positive and the false negative with the recall result of the detection model on multiple videos.

4.1.2 Tracking assignment

The tracking assignment is the process of assigning a unique ID for each detected object in successive frames. This work utilizes the DeepSORT tracking algorithm [77] for this step. DeepSORT is a simple online real-time algorithm for multiple object tracking (MOT) tasks. DeepSORT employs the motion and appearance information to improve the performance of SORT algorithm [9]. In such way, a convolutional neural network (CNN) trained to discriminate pedestrians is used to overcome the low performance of SORT algorithm in the case of the occlusions. Figure 3 illustrates the DeepSORT. When given a new bounding box tracked using a Kalman filter utilizing the assignment problem, DeepSORT connects a new detection with a new prediction. The Hungarian Algorithm connects the separately produced findings after quantifying the association with the Mahalanobis distance. The Hungarian method is applied by taking into account the added value of the Kalman filter and the deep learning feature. Figure 5 shows the result of this step.

Table 1 The true positive, the false negative and recall results of the detection model

	True positive	False negative	Recall
Video 2	2155	1999	0.51%
Video 3	8737	6590	0.57%
Video 5	5104	7046	0.42%
Video 7	6420	8663	0.42%
Video 8	3760	4345	0.46%

Fig. 5 Deep sort output on a single frame

4.2 Part 2: abnormal detection

The second part of the proposed framework is detecting and recognizing the abnormal behaviors in the crowd. This phase works in two stages: feature extraction and classification. Another overview of the procedures used in this part is given by Algorithm 2.

Input: v, video stream; Detected bounding boxes $b_{f_i,n}$ in each frame; $true\_labels$ list where f_i is the ith frame and n is the bounding box number within f_i frame

Output: Classification results

procedure FEATURE EXTRACTION(v, $boxes$)

  $feature\_vector \leftarrow [\ ]$

  for each frame f_i in v **do**

    $magnitude_{f_i}, direction_{f_i} \leftarrow calculate\_optical\_flow(f_i, f_{i+1})$

    for each box $b_{f_i,n}$ in frame f_i **do**

      $m_{b_{f_i,n}} \leftarrow extract\_box\_magnitude(magnitude_{f_i})$

      $d_{b_{f_i,n}} \leftarrow extract\_box\_direction(direction_{f_i})$

      $hof_{b_{f_i,n}} \leftarrow calculate\_hof\_features(m_{b_{f_i,n}}, d_{b_{f_i,n}})$

      $spatial_{b_{f_i,n}} \leftarrow calculate\_spatial\_features(m_{b_{f_i,n}}, d_{b_{f_i,n}})$

      $vector \leftarrow concat(hof_{b_{f_i,n}}, spatial_{b_{f_i,n}})$

      $feature\_vector.append(vector)$

procedure CLASSIFICATION($feature\_vector$, $true\_labels$)

  $prediction \leftarrow SVM(feature\_vector, true\_labels)$

Algorithm 2 Part 2: Abnormal detection.

4.2.1 Feature extraction

Several global and local features are proposed to identify abnormal behaviors. The global features are extracted from the entire frame, such as the main direction of the crowd, while local features are obtained from each bounding box in the frame, which resulted from the first part. For each video, 27 frames per second (fps) are extracted. Then, 14 features are extracted for each bounding box in every two consecutive frames. The time complexity is calculated for this stage, as it ranges from 3 to 15 fps with an average of 4 fps. It varies based on the number of samples needed to extract its features in each frame. According to recent studies, a system is considered to operate in real time if it can process at least 25 frames per second [71]. In fact, the frame rate is increased by decreasing the computational cost per frame, so the computations must be mitigated in order to use the system in real time. The following describes the two types of extracted features: HOF features and spatial features.

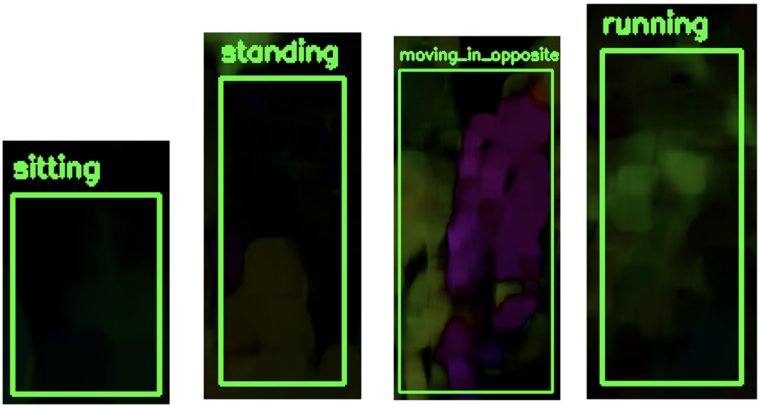
Histogram of optical flow features Optical flow is the visual motion of an object in a scene, and the apparent flow of pixels in relation to its surroundings [70]. In typical crowded settings such as in the Hajj dataset, people density is high and individual motion is confined by other people's movement, thus individual movement is typically sluggish. Individuals' speed and direction do not alter dramatically in a short period of time. However, both the direction and magnitude of optical flow become important characteristics in describing crowd motions and give an indication of abnormal movements. The dense optical flow looks at all of the points and recognizes pixel intensity changes between the two frames, resulting in a picture with highlighted pixels. It takes an array of flow vectors, i.e., $(\frac{dx}{dt}, \frac{dy}{dt})$ and calculates the magnitude and direction of optical flow. The Hue value of the image visualizes the optical flow direction, and the Value plane visualizes the optical flow magnitude, as shown in Fig. 6. In Fig. 6, the result of the optical flow of four different classes is depicted, where in the case of sitting (Fig. 6a), the Hue (direction) is zero, and the Value (the magnitude of movement) is also zero, so the image appears completely black. Similarly, these results are observed for the standing class (Fig. 6b). Since these two classes have the same Hue and Value, spatial features are employed to distinguish them, as will be discussed later. Also, the exact figure shows that moving in opposite direction class (Fig. 6c) and running class (Fig. 6d) both have an amount of Hue and Value which correspond to their direction and magnitude, respectively. After calculating the magnitude and direction vectors, Hu et al. [25] approach is followed to calculate the histogram of optical flow for each bounding box. The direction angle range [0°, 360°] was quantized into 8 bins. Then, two flow thresholds were set (γ maximum flow threshold and δ minimum flow threshold) to calculate the h vector, as shown in (1). The h vector represents the sum of the optical flow in the k-th direction. Equations (2, 3) show the method of calculating the h vector, where m_{ij} and θ_{ij} represent the magnitude and direction at pixels (i, j).

$$h = [h_0 + h_1 + \cdots + h_k],\ s.t\ 0 \leqslant k \leqslant 8 \tag{1}$$

$$h_k = \sum sign\ .\ m_{ij},\ s.t.\ round(\theta_{ij}/2\pi) = k,\ 1 \leqslant k \leqslant 8 \tag{2}$$

$$sign = \begin{cases} 0 & m_{ij} \leqslant \delta \\ m_{ij} * 2 & m_{ij} \geqslant \gamma \\ m_{ij} & otherwise \end{cases} \tag{3}$$

Figure 7 shows examples of the difference in the histogram bins resulting from the optical flow of different classes. The x-axis represents the quantized direction angles, while the



a) Sitting b) Standing c) Moving in the opposite direction d) Running

Fig. 6 Examples of optical flow results of different classes

y-axis represents the summation of the magnitude values for these angles. Histogram bins for the sitting class in the Fig. 7a are zeros, which obviously represent the magnitude and direction associated with the sitting behavior due to immobility. However, there are different levels of movement associated with other classes. These levels are owing to the presence of the magnitude and direction of the movement in these classes.

Spatial features As discussed earlier, optical flow features alone are not sufficient to distinguish between classes. For this reason, eight spatial features were used. These spatial features are extracted directly from the bounding boxes of individuals as described below.

- **Difference in horizontal and vertical axes:**
  According to the position information (p_x, p_y) of each bounding box b, provided by the Yolo model, the difference in x and y axes for each bounding box is calculated as shown in (4). These differences give the shift in movement context and serve as useful indicators for identifying the various abnormal classes.

$$\begin{aligned} x_{diff} &= x - x_0 \\ y_{diff} &= y - y_0 \end{aligned} \tag{4}$$

  Figure 8 shows the difference in the y-axis in two consecutive frames for a person moving in the opposite direction, which results in a negative value representing that change.
- **Velocity in horizontal and vertical axes:**
  The velocity vector is computed in the x and y axes. This assumes that the speed s is constant uses the initial (x0, y0) and final (x, y) positions of the bounding box in two consecutive frames as shown in Equation The velocity vector is computed in the x and y axes. Assuming that the speed s is constant, the speed is estimated on the basis of the initial (x_0, y_0) and final (x, y) positions of the bounding box in two consecutive

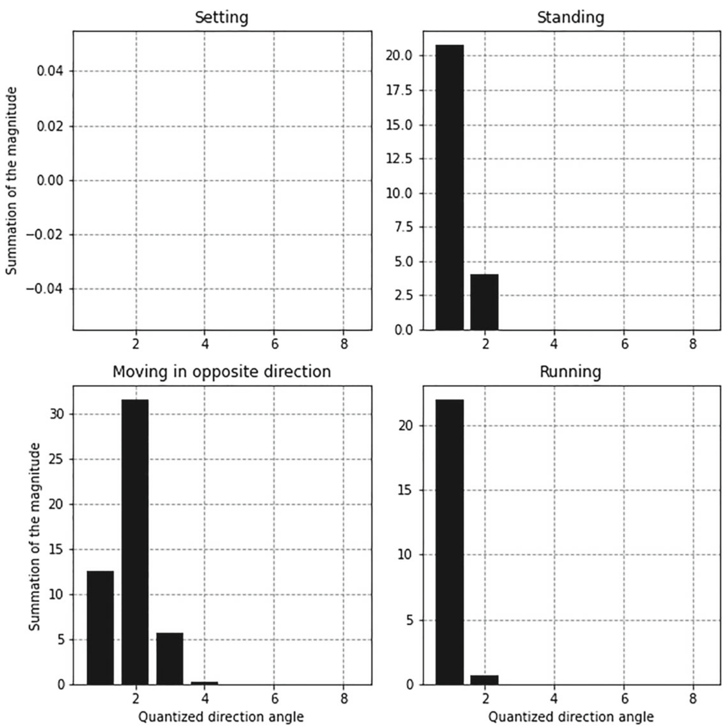


Fig. 7 Examples of histogram of optical flow for different classes

frames, as shown in (5). Figure 8 shows an example of velocity in the x-axis for a person moving in the opposite crowd direction.

$$velocity = \begin{cases} v_x = \frac{s}{d}(x - x_0) \\ v_y = \frac{s}{d}(y - y_0) \end{cases} \tag{5}$$

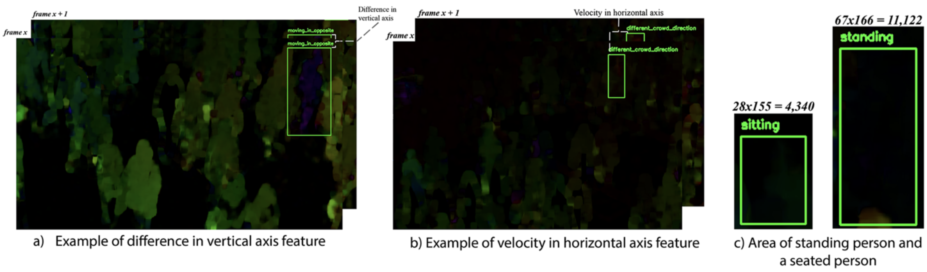


Fig. 8 Examples of extracted features

- **Area:**
  The area of the bounding box is calculated based on the position information on x and y axes (p_w, p_h). Figure 8 shows the difference in the area for a standing person and a seated person.
- **Overall direction of the bounding box:**
  This value is estimated based on the obtained optical flow as follow: for each bounding box, the direction is filtered where the motion has a magnitude higher than a threshold in order to removes some motion noises. The mode of these movements is then computed to determine which is the most common in that box.

4.2.2 Classification

The Support Vector Machine (SVM), a standard machine learning algorithm, was utilized to categorize anomalous behavior into seven different abnormal classifications (standing, sitting, sleeping, running, moving in the opposite direction, moving in different crowd direction, and non-pedestrian entities). It is used in binary classification tasks to identify the maximum margin hyperplane between the two classes (true or false), and it can also handle multiclass classification utilizing ECOC, OvO, and OvR approaches.

In order to find the optimal classifier, a grid search is used, which is an iterative search through the predetermined values for each parameter in the classifier. The grids are chosen with parameter space for c $\in \{2^x | x \in \{0, 1, 2, \cdots, 10\}\}$, $\gamma \in \{0.1^x | x \in \{0, 1, 2, \cdots, 5\}\}$, kernel $\in \{linear, poly, rbf, sigmoid\}$, and decision function $\in \{ovo, ovr, ecoc\}$.

After performing the grid search, the optimal parameters of the SVM were C = 1, gamma = 0.01, kenerl= Gaussian Radial Basis Function (RBF). The kernel is defined as:

$$k(x_i, x_j) = e^{\left(-\frac{d(x_i, x_j)^2}{2l^2}\right)} \tag{6}$$

where l in (6) denotes the kernel's length scale and d denotes the Euclidean distance.

The error correcting output codes (ECOC) method is used with the SVM model. It decomposes multi-class classification into many binary classification tasks [1]. The method involves two stages: (a) encoding to construct coding matrix ($2^{(n-1)} - 1$, where n is the number of classes) and assigning each class a unique codeword, and (b) decoding to assign the data points to the class with the closest codeword [45, 65].

The SVM model with chosen parameters is trained on the extracted features mentioned in Section 4.2.1. After the training, the classifier is tested with the test dataset, and the performance is measured in terms of the performance metrics.

4.3 Implementation platform

In order to implement the proposed system, PyCharm was used as an IDE with the Python programming language using NVIDIA Tesla $V100S$ GPU server with 32 GB of RAM.

5 Experiments and results

To verify the performance and effectiveness of the proposed solution for detecting abnormal behaviors, metric calculations were performed on the results of the used models.

Table 2 SVM performance on the testing set using the ground truth

Video's location	Accuracy	Precision	Recall	F1	AUC
Masa'a	75.30%	74.93%	75.32%	73.65%	91.01%
Jamarat	97.40%	96.37%	97.38%	96.12%	68.91%
Arafat	75.60%	76.99%	75.64%	75.62%	82.70%
Tawaf	70.23%	69.17%	70.23%	66.38%	68.71%
All	75.08%	73.73%	75.08%	71.81%	89.02%

In the testing phase of this study, the SVM classifier was tested in two ways: (i) tested separately on the extracted features from the ground truth, with errors resulting from the detection and tracking phase ignored and (ii) tested on the detection and tracking results resulting from the first stage described in Sections 4.1.2 and 4.1.1.

Since the dataset contains four different locations in Hajj as described in Section 3, the model was trained and tested on each location separately and then on all locations jointly. Tables 2 and 3 present the SVM performance on the testing set using the ground truth and the detection and tracking results, respectively. In this context, recall determines the percentage of true anomalies that are identified while precision indicates the proportion of identified anomalies that are true anomalies. Additionally, the F1 score is reported, which determines the anomaly detection model's overall performance by combining Recall and Precision, using harmonic mean (Fig. 9).

It is clear from the results that the different imaging locations and pilgrim density had a significant impact on the performance of detection, tracking, and classification, as the models achieved higher results in the data captured in the Masa'a with an F1 of 73.65% on ground truth and 79.94% on our previous stage results. In contrast, the Tawaf was the hardest for detection and classification as it achieved an F1 of 66.38% on the ground truth. Moreover, the detection and tracking models did not recognize pilgrims due to the capturing distance and the extreme density of the crowd, as shown in Fig. 10. Also, Fig. 9 shows examples of abnormal behaviors detected at different locations. A few detected behaviors are indicated for clarification purposes in the figure. Figure 9a and b shows the detection of sitting despite the overlap with other pedestrians in Masa'a, also the detection of the non-human moving object (the wheelchair). While in Fig. 9c shows the model's ability to detect the standing behavior in different lighting and imaging conditions. Figure 9d also presents the detection of standing and moving in opposite direction behaviors in other capture conditions.

Furthermore, the results for each class separately are reported in Table 4. As seen from the table, the dataset is imbalanced, which can lead the model to perform poorly due to the

Table 3 SVM performance on the testing set using the detection and tracking results

Video's location	Accuracy	Precision	Recall	F1	AUC
Masa'a	80.48%	81.30%	80.48%	79.94%	95.30%
Jamarat	96.21%	94.97%	96.21%	94.41%	64.60%
Arafat	75.06%	76.16%	75.06%	74.81%	92.11%
Tawaf	–%	–%	–%	–%	–%
All	78.90%	78.16%	78.90%	77.22%	88.96%

Fig. 9 Model's results under different conditions

bias toward the majority class. The Synthetic Minority Oversampling Technique (SMOTE) was used [23], to augment the minority class samples by synthesizing new samples from the

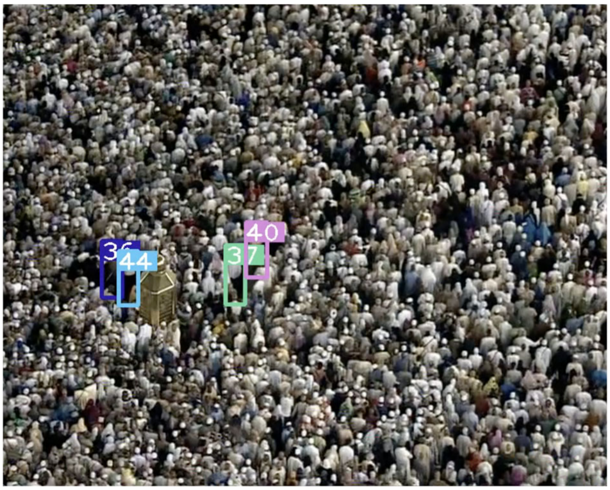


Fig. 10 The extreme density of the crowd in Tawaf

Table 4 Results for each class separately without using SMOTE

Class	Precision	Recall	F1	No. of samples
Different crowd direction	80%	30%	43%	2143
Moving in opposite	78%	63%	69%	10972
Moving non-human object	63%	29%	40%	1252
Running	0%	0%	0%	15
Sitting	81%	96%	88%	30130
Sleeping	62%	35%	45%	720
Standing	70%	57%	63%	6071
Overall	78%	78%	77%	51303

existing samples. Table 5 shows the results after applying the SMOTE technique. The results improved significantly for the minority classes; however, the overall results are relatively low compared to those shown in Table 4.

As an integrated system from the detection and tracking stage to the classi-fication stage, the SVM model achieved AUC of 95.30% for the Masa'a, 65.60% for Jamarat, 92.11% for Arafat, and an overall AUC of 88.96%.

Compared to existing studies focusing on the same problem, our proposed system has achieved promising results. Our model achieved an overall AUC of 88.96% while the other study achieved an AUC of 76.08%. According to the proportions test, there is a statistical difference between these two results (p ¡ 0.0001). In addition, our study was distinguished by the fact that it approached anomalous behaviors as a multi-class problem, unlike other studies that dealt with it as a binary classification problem such as [2] in Hajj.

6 Conclusion

In this paper, a solution is developed to detect abnormal behaviors in Hajj. This problem is considered essential and needs extensive studies because the Hajj in Mecca, Saudi Arabia, is the greatest human gathering globally, with about 2.5 million pilgrims in 2019 [67]. In these gatherings, many violations appear that need to be mitigated. Abnormal behaviors in this study were divided into seven classes (standing, sitting, sleeping, running, moving in the opposite direction, moving in the different crowd direction, and non-human objects such

Table 5 Results for each class separately using SMOTE

Class	Precision	Recall	F1	No. of samples
Different crowd direction	69%	84%	76%	30130
Moving in opposite	74%	54%	62%	30130
Moving non-human object	60%	46%	52%	30130
Running	98%	90%	94%	30130
Sitting	45%	82%	58%	30130
Sleeping	89%	63%	74%	30130
Standing	75%	65%	70%	30130
Overall	72%	69%	69%	210910

as vehicles and wheelchairs). These abnormal behaviors were detected in two steps: the first step was the detection and tracking of pilgrims. The Yolov5 model was used for detection, while the DeepSORT model was used for tracking. The second step was to classify the abnormal behavior by extracting the features for each detected bounding box using optical flow and other spatial features; then, it was classified based on these extracted features using SVM. The SVM model achieved an average of 88.96% AUC.

To sum up, our proposed solution differs from existing approaches in the field regarding the nature of abnormal behaviors. In other studies, abnormal behaviors are represented by seeing vehicles like bicycles and cars between pedestrians, as in the UCSD dataset [48], or by observing a group of people starting to run when they receive a signal, as in the UMN dataset [51]. The anomalous behaviors recognized in this proposed remedy differ in that they are more challenging, intricate, and densely crowded.

Our proposed solution is also distinguished from other studies in the field that detect abnormal behaviors in the same context (Hajj) in that our proposed solution works on the problem of abnormal behaviors as a multi-class problem, unlike other studies [2] that work on them as a binary classification problem. Our system overcame the difficulty of multi-class classifications and achieved high results compared to the binary classification [2].

All these indicate that our proposed framework greatly impacts anomaly detection in dense crowds such as those seen during Hajj. At the same time, our proposed framework indicates promising results compared to the study by Alafif et al.'s [3] study which, classified abnormal behaviors in Hajj into seven different categories. Our proposed solution achieved a result that outperformed the previous results by 12.88% AUC.

The proposed approach can effectively identify anomalous behavior in a huge dense crowd, although it must be faster to operate in real time. In subsequent work, we will extract more advanced classification features with greater value and rapid time to maximize the system's effectiveness in real time. In order to boost the speed of detection, we will also aim to combine the detection and tracking phases into a single phase.

Acknowledgments The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number (227).

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interests The authors declare that they have no known conflicts of interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Al-Shargie F, Tang TB, Badruddin N et al (2018) Towards multilevel mental stress assessment using svm with ecoc: an eeg approach. Med Bio Eng Comput 56(1):125–136
2. Alafif T, Alzahrani B, Cao Y et al (2022) Generative adversarial network based abnormal behavior detection in massive crowd videos: A hajj case study. J Ambient Intell Humaniz Comput 13(8):4077–4088
3. Alafif T, Hadi A, Allahyani M et al (2022) Hybrid classifiers for spatio-temporal real-time abnormal behaviors detection, tracking, and recognition in massive hajj crowds. https://doi.org/10.48550/ARXIV.2207.11931

4. Almazroey AA, Jarraya SK (2020) Abnormal events and behavior detection in crowd scenes based on deep learning and neighborhood component analysis feature selection. In: Joint European-US workshop on applications of invariance in computer vision. Springer, pp 258–267
5. Aqeel M, Khan KB, Azam MA et al (2020) Detection of anomaly in videos using convolutional autoencoder and generative adversarial network model. In: 2020 IEEE 23rd International Multitopic Conference (INMIC). IEEE, pp 1–6
6. Bansod SD, Nandedkar AV (2019) Anomalous event detection and localization using stacked autoencoder. In: International conference on computer vision and image processing. Springer, pp 117–129
7. Bansod SD, Nandedkar AV (2020) Crowd anomaly detection and localization using histogram of magnitude and momentum. Vis Comput 36(3):609–620
8. Bera A, Kim S, Manocha D (2016) Interactive crowd-behavior learning for surveillance and training. IEEE Comput Graph Appl 36(6):37–45. https://doi.org/10.1109/MCG.2016.113
9. Bewley A, Ge Z, Ott L et al (2016) Simple online and realtime tracking. In: IEEE International Conference on Image Processing (ICIP), vol 2016. IEEE. https://doi.org/10.1109/icip.2016.7533003
10. Biswas S, Venkatesh Babu R (2017) Anomaly detection via short local trajectories. Neurocomputing 242:63–72
11. Chen CY, Shao Y (2015) Crowd escape behavior detection and localization based on divergent centers. IEEE Sensors J 15(4):2431–2439. https://doi.org/10.1109/JSEN.2014.2381260
12. Chen Y, Wang S (2017) A weighted mhof and sparse representation based crowd anomaly detection algorithm. In: IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, pp 760–765
13. Chibloun A, El Fkihi S, Mliki H et al (2018) Abnormal crowd behavior detection using speed and direction models. In: 2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC). IEEE, pp 197–202
14. Das D, Mishra D (2018) Unsupervised anomalous trajectory detection for crowded scenes. In: 2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS) pp 27–31
15. Deng L, Lian D, Huang Z et al (2022) Graph convolutional adversarial networks for spatiotemporal anomaly detection. IEEE Trans Neur Netw Learn Syst 33(6):2416–2428. https://doi.org/10.1109/TNNLS.2021.3136171
16. Direkoglu C (2020) Abnormal crowd behavior detection using motion information images and convolutional neural networks. IEEE Access 8:80,408–80,416
17. Feng Y, Yuan Y, Lu X (2017) Learning deep event models for crowd anomaly detection. Neurocomputing 219:548–556
18. Gnouma M, Ejbali R, Zaied M (2018) Abnormal events' detection in crowded scenes. Multimed Tools Applic 77(19):24,843–24,864
19. Gnouma M, Ejbali R, Zaied M (2019) Video anomaly detection and localization in crowded scenes. In: International Joint Conference: 12th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2019) and 10th International Conference on EUropean Transnational Education (ICEUTE 2019). Springer, pp 87–96
20. Guo C, Lin H, He Z et al (2019) Crowd abnormal event detection based on sparse coding. Int J Humanoid Robot 16(04):1941,005
21. Guo S, Bai Q, Gao S et al (2019) An analysis method of crowd abnormal behavior for video service robot. IEEE Access 7:169,577–169,585. https://doi.org/10.1109/ACCESS.2019.2954544
22. Han Q, Wang H, Yang L et al (2020) Real-time adversarial gan-based abnormal crowd behavior detection. J Real-Time Image Proc 17(6):2153–2162
23. He H, Ma Y (2013) Imbalanced learning: foundations, algorithms, and applications. Wiley-IEEE Press
24. Hu L, Hu F (2017) Anomaly detection in crowded scenes via sa-mhof and sparse combination. In: 2017 10th International Symposium on Computational Intelligence and Design (ISCID), pp 421–424. https://doi.org/10.1109/ISCID.2017.130
25. Hu X, Dai J, Huang Y et al (2020) A weakly supervised framework for abnormal behavior detection and localization in crowded scenes. Neurocomputing 383:270–281
26. Hu X, Lian J, Zhang D et al (2022) Video anomaly detection based on 3d convolutional auto-encoder. SIViP, 1–9
27. ping Hu Z, Zhang L, fang Li S et al (2020) Parallel spatial-temporal convolutional neural networks for anomaly detection and location in crowded scenes. J Vis Commun Image Represent 67:102,765
28. Ilyas Z, Aziz Z, Qasim T et al (2021) A hybrid deep network based approach for crowd anomaly detection. Multimed Tools Appl, 1–15
29. Ji QG, Chi R, Lu ZM (2018) Anomaly detection and localisation in the crowd scenes using a block-based social force model. IET Image Process 12(1):133–137

30. Joshi KV, Patel NM (2021) A cnn based approach for crowd anomaly detection. Int J Next-Gen Comput, 12(1)
31. Kaltsa V, Briassouli A, Kompatsiaris I et al (2015) Swarm intelligence for detecting interesting events in crowded environments. IEEE Trans Image Process 24(7):2153–2166. https://doi.org/10.1109/TIP.2015.2409559
32. Khan MUK, Park HS, Kyung CM (2018) Rejecting motion outliers for efficient crowd anomaly detection. IEEE Trans Inform Forens Secur 14(2):541–556
33. Li A, Miao Z, Cen Y (2016) Global anomaly detection in crowded scenes based on optical flow saliency, IEEE
34. Li A, Miao Z, Cen Y et al (2017) Anomaly detection using sparse reconstruction in crowded scenes. Multimed Tools Applic 76(24):26,249–26,271
35. Li X, Li W, Liu B et al (2018) Object-oriented anomaly detection in surveillance videos. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 1907–1911. https://doi.org/10.1109/ICASSP.2018.8461422
36. Li X, Li W, Liu B et al (2019) Object and patch based anomaly detection and localization in crowded scenes. Multimed Tools Appl, 1–16
37. Li N, Chang F, Liu C (2020) Spatial-temporal cascade autoencoder for video anomaly detection in crowded scenes. IEEE Trans Multimed 23:203–215
38. Li X, Yang Y, Xu Y et al (2021) Rapid detection of crowd abnormal behavior based on the hierarchical thinking. In: Barolli L, Li KF, Miwa H (eds) Advances in intelligent networking and collaborative systems. Springer International Publishing, Cham, pp 361–371
39. Li N, Zhong JX, Shu X et al (2022) Weakly-supervised anomaly detection in video surveillance via graph convolutional label noise cleaning. Neurocomputing 481:154–167
40. Lin TY, Maire M, Belongie S et al (2014) Microsoft coco: common objects in context. https://doi.org/10.48550/ARXIV.1405.0312
41. Lin H, Deng JD, Woodford BJ (2015) Anomaly detection in crowd scenes via online adaptive one-class support vector machines. In: 2015 IEEE International Conference on Image Processing (ICIP), pp 2434–2438. https://doi.org/10.1109/ICIP.2015.7351239
42. Lin S, Yang H, Tang X et al (2019) Social mil: interaction-aware for crowd anomaly detection. In: 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp 1–8. https://doi.org/10.1109/AVSS.2019.8909882
43. Lin S, Yang H (2021) Dual-mode iterative denoiser: tackling the weak label for anomaly detection. In: 2020 25th International Conference on Pattern Recognition (ICPR), pp 6742–6749. https://doi.org/10.1109/ICPR48806.2021.9412673
44. Liu S, Qi L, Qin H et al (2018) Path aggregation network for instance segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 8759–8768
45. Liu S, Xu L, Li Q et al (2018) Fault diagnosis of water quality monitoring devices based on multiclass support vector machines and rule-based decision trees. IEEE Access 6:22,184–22,195
46. Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: Proceedings of the 7th international joint conference on artificial intelligence, vol 2. Morgan Kaufmann Publishers Inc., San Francisco, pp 674–679. IJCAI'81
47. Luo Z, He W, Liwang M et al (2017) Real-time detection algorithm of abnormal behavior in crowds based on gaussian mixture model. In: 2017 12th International Conference on Computer Science and Education (ICCSE), pp 183–187. https://doi.org/10.1109/ICCSE.2017.8085486
48. Mahadevan V, Li W, Bhalodia V et al (2010) Anomaly detection in crowded scenes. In: 2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, pp 1975–1981
49. Marsden M, McGuinness K, Little S et al (2016) Holistic features for real-time crowd behaviour anomaly detection. In: 2016 IEEE International Conference on Image Processing (ICIP), pp 918–922. https://doi.org/10.1109/ICIP.2016.7532491
50. Mehmood A (2021) Efficient anomaly detection in crowd videos using pre-trained 2d convolutional neural networks. IEEE Access 9:138,283–138,295. https://doi.org/10.1109/ACCESS.2021.3118009
51. Mehran R, Oyama A, Shah M (2009) Abnormal crowd behavior detection using social force model. In: IEEE conference on computer vision and pattern recognition. IEEE, pp 935–942
52. Mishra SR, Mishra TK, Sarkar A et al (2020) Detection of anomalies in human action using optical flow and gradient tensor. In: Smart intelligent computing and applications. Springer, pp 561–570
53. Mondal R, Chanda B (2018) Anomaly detection using context dependent optical flow. In: Proceedings of the 11th Indian conference on computer vision, graphics and image processing, vol 2018. Association for Computing Machinery, New York, pp 1–8. ICVGIP
54. Moustafa AN, Gomaa W (2020) Gate and common pathway detection in crowd scenes and anomaly detection using motion units and lstm predictive models. Multimed Tools Applic 79(29):20,689–20,728

55. Nain N, Lamba S (2018) Oriented tracklets approach for anomalous scene detection in high density crowd. In: 2018 14th International conference on signal-image technology internet-based systems (SITIS), pp 435–441. https://doi.org/10.1109/SITIS.2018.00073
56. Pan J, Liang D (2017) Holistic crowd interaction modelling for anomaly detection. In: Chinese conference on biometric recognition. Springer, pp 642–649
57. Patil N, Biswas PK (2016) Global abnormal events detection in surveillance video — a hierarchical approach. In: 2016 Sixth International Symposium on Embedded Computing and System Design (ISED), pp 217–222. https://doi.org/10.1109/ISED.2016.7977085
58. Jocher G, Chaurasia A, Stoken A, Borovec J, NanoCode012, Kwon Y, TaoXie, Fang J, imyhxy, Michael K, Lorna, Abhiram V, Montes D, Nadar J, Laughing, tkianai, yxNONG, Skalski P, Wang Z, Hogan A, Fati C, Mammana L, AlexWang1900, Patel D, Yiwei D, You F, Hajek J, Diaconu L, Thanh Minh M (2022) ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference. Zenodo, v6.1. https://doi.org/10.5281/zenodo.6222936
59. Qasim T, Bhatti N (2019) A low dimensional descriptor for detection of anomalies in crowd videos. Math Comput Simul 166:245–252
60. Ramchandran A, Sangaiah AK (2020) Unsupervised deep learning system for local anomaly event detection in crowded scenes. Multimed Tools Applic 79(47):35,275–35,295
61. Ramos J, Nedjah N, de Macedo Mourelle L et al (2018) Visual data mining for crowd anomaly detection using artificial bacteria colony. Multimed Tools Applic 77(14):35,755–17,777
62. Sabih M, Vishwakarma DK (2021) Crowd anomaly detection with lstms using optical features and domain knowledge for improved inferring. Vis Comput, 1–12
63. Sabokrou M, Fathy M, Hoseini M et al (2015) Real-time anomaly detection and localization in crowded scenes. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 56–62. https://doi.org/10.1109/CVPRW.2015.7301284
64. Sabokrou M, Fayyaz M, Fathy M et al (2017) Deep-cascade: Cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes. IEEE Trans Image Process 26(4):1992–2004. https://doi.org/10.1109/TIP.2017.2670780
65. Samat A, Yokoya N, Du P et al (2019) Direct, ecoc, nd and end frameworks—which one is the best? An empirical study of sentinel-2a msil1c image classification for arid-land vegetation mapping in the ili river delta, kazakhstan. Remote Sens 11(16):1953
66. Shao S, Zhao Z, Li B et al (2018) Crowdhuman: a benchmark for detecting human in a crowd. https://doi.org/10.48550/ARXIV.1805.00123
67. statista (2022) Saudi Arabia: Hajj population. Annual number of Hajj pilgrims to Saudi Arabia from 1999 to 2019. https://www.statista.com/statistics/617696/saudi-arabia-total-hajj-pilgrims/#statisticContainer. Accessed 2022-02-06
68. Tian Y, Pang G, Chen Y et al (2021) Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp 4975–4986
69. Tomé A, Salgado L (2017) Detection of anomalies in surveillance scenarios using mixture models. In: 2017 International Carnahan Conference on Security Technology (ICCST), pp 1–4. https://doi.org/10.1109/CCST.2017.8167830
70. Turaga P, Chellappa R, Veeraraghavan A (2010) Advances in video-based human activity analysis: challenges and approaches. In: Zelkowitz MV (ed) Advances in computers, advances in computers, vol 80. Elsevier, pp 237–290. https://doi.org/10.1016/S0065-2458(10)80007-5
71. Vaquero L, Brea VM, Mucientes M (2022) Tracking more than 100 arbitrary objects at 25 fps through deep learning. Pattern Recogn 121:108,205. https://doi.org/10.1016/j.patcog.2021.108205, https://www.sciencedirect.com/science/article/pii/S0031320321003861
72. Wang D, Wang S (2021) Abnormal event detection algorithm based on dual attention future frame prediction and gap fusion discrimination. J Electron Imag 30:023,009–023,009
73. Wang Q, Wang QWKZC, Ma M et al (2016) Hybrid histogram of oriented optical flow for abnormal behavior detection in crowd scenes. Int J Pattern Recognit Artif Intell 30:1655,007:1–1655,007:14
74. Wang S, Zhu E, Yin J et al (2016) Anomaly detection in crowded scenes by sl-hof descriptor and foreground classification. In: 2016 23rd International Conference on Pattern Recognition (ICPR), pp 3398–3403. https://doi.org/10.1109/ICPR.2016.7900159
75. Wang T, Qiao M, Lin Z et al (2019) Generative neural networks for anomaly detection in crowded scenes. IEEE Trans Inform Forens Secur 14(5):1390–1399. https://doi.org/10.1109/TIFS.2018.2878538
76. Wang L, Tan H, Zhou F et al (2022) Unsupervised anomaly video detection via a double-flow convlstm variational autoencoder. IEEE Access 10:44,278–44,289. https://doi.org/10.1109/ACCESS.2022.3165977

77. Wojke N, Bewley A, Paulus D (2017) Simple online and realtime tracking with a deep association metric. https://doi.org/10.48550/ARXIV.1703.07402
78. Xu Y, Lu L, Xu Z et al (2018) Towards intelligent crowd behavior understanding through the stfd descriptor exploration. Sensing and Imaging 19(1):1–22
79. Xu M, Yu X, Chen D et al (2019) An efficient anomaly detection system for crowded scenes using variational autoencoders. Appl Sci, 9(16). https://doi.org/10.3390/app9163337, https://www.mdpi.com/2076-3417/9/16/3337
80. Yang B, Cao J, Ni R et al (2018) Anomaly detection in moving crowds through spatiotemporal autoencoding and additional attention. Adv Multim 2018:2087,574:1–2087,574:8
81. Yang M, Rashidi L, Rajasegarar S et al (2018) Crowd activity change point detection in videos via graph stream mining. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp 328–3288. https://doi.org/10.1109/CVPRW.2018.00059
82. Zhang Y, Lu H, Zhang L et al (2016) Combining motion and appearance cues for anomaly detection. Pattern Recogn 51:443–452
83. Zhang X, Zhang Q, Hu S et al (2018) Energy level-based abnormal crowd behavior detection. Sensors 18(2). https://doi.org/10.3390/s18020423, https://www.mdpi.com/1424-8220/18/2/423
84. Zhang S, Gong M, Xie Y et al (2022) Influence-aware attention networks for anomaly detection in surveillance videos. IEEE Trans Circuits Syst Video Technol 32(8):5427–5437. https://doi.org/10.1109/https://doi.org/10.1109/
85. Zhao K, Liu B, Li W et al (2018) Anomaly detection and localization: a novel two-phase framework based on trajectory-level characteristics. In: 2018 IEEE International Conference on Multimedia Expo Workshops (ICMEW), pp 1–6. https://doi.org/10.1109/ICMEW.2018.8551517
86. Zhou S, Shen W, Zeng D et al (2015) Unusual event detection in crowded scenes by trajectory analysis. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp 1300–1304. https://doi.org/10.1109/ICASSP.2015.7178180