Bhuiyan *et al. Journal of Big Data* (2023) 10:86
https://doi.org/10.1186/s40537-023-00779-4

Journal of Big Data

RESEARCH Open Access

Hajj pilgrimage abnormal crowd movement monitoring using optical flow and FCNN

Md Roman Bhuiyan[1*], Junaidi Abdullah[1], Noramiza Hashim[1], Fahmid Al Farid[1] and Jia Uddin[2]

*Correspondence: romanbhuiyanpv@gmail.com

[1] Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia
[2] Technology Studies Department, Endicott College, Woosong University, Daejeon 34000, South Korea

Abstract

This article discusses an effective technique for detecting abnormalities in Hajj crowd videos. In order to guarantee the identification of anomalies in scenes, a trained and supervised FCNN is turned into an FCNN using FCNNs and temporal data. By minimizing computational complexity, incorrect movement detection is utilized to achieve high performance in terms of speed and precision. This FCNN-based architecture is designed to handle two primary tasks: feature representation and the detection of incorrect movement outliers. Additionally, to overcome the aforementioned issues, this research will generate a new crowd anomaly video dataset based on the Hajj pilgrimage scenario. On the proposed dataset, the UCSD Ped2, Subway Entry, and Subway Exit datasets, the proposed FCNN-based technique obtained ultimate accuracy of 100%, 90%, 95%, and 89%, respectively. Additionally, the ResNet50-based technique achieved ultimate accuracy of 96%, 89%, 94%, and 92%, respectively, for the proposed dataset, the UCSD Ped2, Subway Entry, and Subway Exit datasets.

Keywords: Crowd anomaly classification, FCNN, Optical Flow and Hajj crowd video dataset

Introduction

The deployment of security cameras involves the processing of very huge amounts of video data using computer vision technologies. A common application in this discipline is the detection of anomalies in recorded scenes. Due to the ambiguous, subjective, or situational character of the term "anomaly," detecting and localising it in video analysis is a difficult task. In general, an event refers to an incident that happens seldom or unexpectedly [1].

Unlike the previously stated deep-cascade approach, this paper suggests and assesses a novel method for anomaly identification [1]. The purpose of this study is to describe and assess a pre-trained convolutional neural network (CNN) that has been adjusted for the detection and localization of abnormalities. In comparison to [1], the proposed CNN is not entirely "ne-tuned" [1]. Suggested a method for processing a video frame that separated the frame into patches and then arranged the anomaly detection according to patch levels. In contrast, the CNN approach suggested in this research requires the input of a whole video frame. To summarise, the new procedure is more methodologically straightforward but also quicker in terms of training and testing, with an accuracy

equivalent to that stated in [1]. Anomalies in crowd scene recordings are caused by distinctive shapes or movements. Due to the time requirements associated with looking for unknown shapes or motions, cutting-edge approaches use sections or patches of normal frames as reference models. Indeed, these reference models include conventional motions or shapes for each segment of the training data. Areas that differ from the conventional model are categorised as abnormal throughout the testing phase. Classification of these regions as normal or abnormal requires a large number of training samples in order to characterise the characteristics of each region properly [1].

Numerous descriptors are applicable for characterising the features of the location. The behaviour of objects has been described using trajectory-based methodologies. Recent work has used low-level characteristics such as histograms of gradients (HoG) and histograms of optic flow (HoF) to describe the spatiotemporal dynamics of video data. These trajectory-based approaches have two major drawbacks. They are incapable of addressing occlusion issues and also display a high degree of complexity, which is particularly noticeable in crowded situations. CNNs have lately been shown to be effective in the development of efficient data processing algorithms for a variety of applications.

In a range of areas, including image classification [2], object identification [3], and activity recognition [4], CNN-based strategies outperformed state-of-the-art methods. Handcrafted characteristics, it is believed, are incapable of replicating natural video properly [5–7]. Despite these advantages, CNNs are inefficient computationally, particularly more so when block-wise techniques are applied [3]. Thus, after segmenting a video into patches and modelling them using CNNs, it is necessary to investigate other methods for speeding up the process.

The following are the primary concerns that occur when CNNs are used to detect anomalies:

1. While FCNN is considered a time-consuming technique, it is inefficient for patch-based solutions.
2. Because CNNs are totally supervised in their training, identifying anomalies in real-world videos is essentially unachievable owing to the difficulties of training large quantities of data on non-existent classes of anomalies.

As a consequence of these constraints, there has been a recent trend toward improving CNN-based algorithms in order to make them more practical. Faster-RCNN [8] finds objects by creating a feature map for each region in the input data using convolutional layers. To extract regional features for semantic segmentation, methods such as [9] use fully convolutional networks (FCNs) in place of conventional convolutional neural networks (CNNs).

Computing expenses are reduced by using a regional feature extractor and turning typical classification CNNs to fully convolutional networks. By and large, since CNNs and FCNs are supervised approaches, they are unable to solve issues involving anomaly detection.

To solve the aforementioned problems, we provide a novel FCN-based approach for extracting video region-specific information. This novel method employs several convolutional layers inside a pre-trained CNN based on an AlexNet model [2], as well as

an extra convolutional layer. AlexNet, like [10], is a pre-trained image classification network that was suggested utilising ImageNet [11, 12] and the MIT Places dataset [13]. The developed criteria are sufficiently discriminative for detecting problems in this kind of video data.

Generally, the suggested FCNN is given whole frames. As a consequence, all regions' features are efficiently retrieved. The output analysis is used to extract and localise video abnormalities. Convolution and pooling operations occur simultaneously in all CNN layers. A typical NVIDIA TITAN GPU processes 370 frames per second (fps) when examining 1920 x 1080 (low-resolution) frames. This is regarded as very rapid.

CNNs utilise convolution and pooling techniques to extract areas from input data based on a stride and size specification. These patch-based techniques return information about the extracted regions. The output characteristics and their accompanying descriptions are used to estimate the position of an area within a sequence of video frames. The operations of convolution and pooling are both invertible. On the other hand, a roll-back operation develops a receptive field (a area inside a frame) from the network's deeper layers to its more shallow levels. This receptive field is what generates feature vectors. We provide a strategy for finding and localising aberrant regions in a frame by analysing the output of deep layers in an FCN. Localization of a receptive field was influenced by the faster-RCNN method described in [8] and the Over Feat technique described in [4, 14].

As with [15], we use a transfer learning strategy to improve the description of each area. We demonstrate our technique for determining the optimal intermediate convolutional layer for a CNN. Then, after the best-performing CNN layer, a new convolutional layer is created. Our FCN modifies and uses the kernels of a pre-trained CNN as constants; the final new convolutional layer's parameters are taught using our training frames.

Throughout the testing process, confident anomalies are locations that deviate considerably from the initial Gaussian model. The term "normal zones" refers to zones that are totally consistent with the first model. The remaining areas, which are separated by a very little difference, are represented by a sparse-auto-encoder and assessed more precisely using the second Gaussian model. As described in further depth in the next sections, this strategy is comparable to a two-stage cascade classifier.

The following are the paper's significant contributions:

1. This is the first time, to our knowledge, that an FCNN and optical flow have been used to discover abnormalities.
2. We provide an unique FCNN architecture for rapidly identifying erroneous movement abnormalities.
3. The proposed technique exceeds state-of-the-art approaches in terms of performance, it outperforms them in terms of time since most applications are real-time.
4. On a typical GPU, we reached a processing speed of 370 frames per second, more than five times faster than the previous quickest approach.

The remaining paper is arranged as follows: "Related works" Section. presents the related work; "Proposed method" Section expounds on the proposed method; "Experimental

design and result analysis" Section gives an experiment; 5 results and Discussion, and Section 6 presents the conclusion. Table 1 shows the list of abbreviations.

Related works

Estimation of object trajectory is critical in a variety of anomaly detection applications [16–27]. Anomalies occur when an object deviates from previously taught conventional pathways. This technique often has a number of drawbacks, including an inability to handle occlusions effectively and being too complicated for processing dense images. To circumvent these two drawbacks, it is recommended to use spatiotemporal low-level characteristics such as optical ow or gradients. Zhang et al. [28] replicate the normal patterns in a film using a Markov random field (MRF) and a range of parameters. According to Boiman and Irani [29], an event is anomalous if it cannot be recreated using just historical data. Adam et al. [30] depict the histograms of optical ow in tiny regions using an exponential distribution.

Mahadevan et al. [31] suggest describing video using a combination of dynamic textures. This technique entails fitting the represented features to a Gaussian mixture model using a Gaussian mixture model. [32] delves further into the explores it in more depth. Kim and Grauman [33] depict local optical wave patterns using a combination of probabilistic (PCA) models. Additionally, they use an MRF to identify recurring trends.

Table 1 List of Abbreviations

Short form	Full form
FCNN	Fully convulation neural network
UCSD Ped2	University of California San Diego
CNN	convolutional neural network
HoG	Histograms of gradients
GPU	Graphics processing unit
MRF	Markov random field
PCA	Patterns using a combination of probabilistic
GMMs	Gaussian mixture model
HMM	Hidden Markov model
SF	Social force
BOV	bag of videos
HOT	Histogram of oriented tracklets
SCD	Structural context descriptor
HVS	Human visual system to determine spatial
SSMF	Sparse semi-nonnegative matrix factorization
CL	Convulation layer
DL	Dropout layer
FCL	Fully convulation layer
PL	Pooling layer
BPS	Backward propagation stage
DCNN	Deep convulation neural network
LK	Lucas-Kandae
ROC	Receiver operating characteristic
AUC	Area under the ROC curve

Benezeth et al. [34] suggested a technique for modelling behaviour using the motion characteristics of individual pixels. They described the video by constructing a matrix of co-occurrences for frequently occurring events across space and time. [35] fits a Gaussian model to the spatiotemporal gradient characteristics and identifies anomalous occurrences using a hidden Markov model (HMM). Mehran et al. [36] suggest the use of social force (SF) to model crowd actions that are unconventional. As mentioned in [37], aberrant behaviours are identified using an approach based on spatial-temporal directed energy filtering. Cong et al. [38] use normal data to construct an enlarged normal basis set. If it is difficult to rebuild a patch using this basis set, it is termed abnormal.

Antic et al. describe a scene parsing approach in [?]. Normal training explains all object hypotheses for the foreground of a frame. Anomalies are irrational beliefs based on conventional wisdom. Saligrama et al. [39] provide a technique for classifying test data according to optical-ow properties. Ullah et al. [40] proposed segmenting crowd movements using a cut/max-ow technique. When a flow deviates from the established model of motion, anomalies develop. Lu et al. [41] suggest a sparse representation-based technique for fast (140–150 frames per second) anomaly detection.

In [42], Roshtkhari et al. enhance the bag of video words approach (BOV). The authors of [43] provide a technique for identifying anomalies in videos that is context-aware. They define the movie based on its movements and context. In [44], they provide an approach for describing both motion and form in terms of a descriptor (dubbed the "motion context"), and they treat anomaly detection as a matching issue. Roshkhari et al. [45] propose a method for learning about the events in a film by creating a hierarchical codebook for the film's most key events. Ullah et al. [46] use learned particles and an MLP neural network to extract video activity. Using the recovered characteristics, Gaussian mixture models (GMMs) are used to learn the behaviour of particles. Additionally, [47] suggests using an MLP neural network to extract corner properties from standard training samples; the authors also utilise the MLP to identify test samples. The authors of [48] extract and evaluate corner characteristics for anomalous sample identification using an enthalpy model, a random forest with corner features. Xu et al. [49] propose a unified anomaly energy function based on the finding of hierarchical activity patterns for the aim of identifying anomalies.

Sabokrou et al. [5, 6] highlights research that use auto-encoders to imitate common events [50]. They recognise outliers from the target (normal) class using a one-class classifier. Additionally, Section 1 summarises the work presented in [1]; this article presents a cascaded classifier for anomaly detection that utilises two deep neural networks. Here, incorrect patches are initially recognised using a tiny deep network and then classified further using another deep network. [51] uses a histogram of oriented tracklets to characterise video and identify anomalies (HOT). Furthermore, this work proposes a novel technique for enhancing HOT. Yuan et al. [52] suggest an effective structural context descriptor (SCD) for identifying distinct populations. The present approach analyses a population's (spatiotemporal) SCD variation in order to identify the anomalous location.

Feng et al. [17] employs an unsupervised deep learning algorithm to extract abnormalities from cluttered photos. This approach extracts shapes and attributes from 3D gradients using a PCANet [18]. A deep Gaussian mixture model is then used to characterise the event patterns (GMM). [19] also makes use of a PCANet. The authors of this

work use the human visual system to determine spatial qualities (HVS). On the other hand, the motion of the film is represented by a histogram of optical owls at many scales (MHOF). PCANet is used to discriminate between normal and abnormal events by using these spatiotemporal properties.

Cheng et al. [53] establishes a hierarchical architecture for detecting locally and globally distributed anomalies. The authors find common geometric correlations between sparse interest areas and then use Gaussian process regression to construct a template for normal interactions. Xiao et al. [54] derive the local pattern of pixels using sparse semi-nonnegative matrix factorization (SSMF). Their approach includes generating a probability model from localized pixel patterns that account for spatial and temporal context. Their technique is very surprising. The trained model is used to identify anomalies.

Sabokrou et al. [55] introduces the notion of the "motion impact map," a very effective tool for describing human actions in video data in terms of their motion characteristics. Aberrant blocks are ones that appear rarely in a frame. To explain anomalies in densely populated settings, we built a spatio-temporal CNN model capable of recognising traits in both spatial and temporal dimensions using spatio-temporal convolutions. Li et al. [56] provide an unsupervised strategy for finding anomalies that leverages clustering and sparse coding to discover global activity patterns and locally significant behaviour patterns.

Bhuiyan et al. [57] the study provides in-depth analyses of current crowd analysis methods and approaches, with a focus on deep learning techniques for identifying anomalous behaviour in crowd videos. For this reason, we're launching an exhausting but rewarding adventure into crowd analysis, categorization, and the spotting of any irregular movements among Hajj pilgrims. It also drives us to conduct a large-scale critical analysis of the crowd, given that the Hajj pilgrimage is the most heavily populated area for video-related substantial research activities. Bhuiyan et al. [58] In this research area intends to solve the technical constraints of video analysis in a situation where the movement of large numbers of pilgrims with densities ranging from 7 to 8 per square metre is taking place A novel dataset based on the Hajj pilgrimage scenario will be developed in this project to solve this difficulty.

Alafif et al. [59] Two key issues are raised in this research. First, author provide the collection of large-scale aberrant Hajj crowd behavior that has been identified and categorized (HAJJv2). Secondly, the author propose two approaches to detect and characterize spatial and temporal anomalous behaviors in small- and large-scale crowd data using mixed convolutional neural networks (CNNs) and random forests (RFs). A ResNet-50 CNN model that has previously been trained is modified to determine whether or not each frame in small-scale crowd footage is normal. [60] The goal of this study is to find anomalies in dense crowd scenes that are both organized and unstructured. The suggested model uses a deep convolutional neural network to first identify moving objects and people in the scene before utilizing spatial and temporal data to follow them.

With a particular emphasis on the visual surveillance in the Hajj, this study tries to highlight the research studies pertinent to the larger area of video analytics employing deep learning. The paper highlights the difficulties and cutting-edge approaches to visual surveillance in general that may be skilfully used to the purposes of Hajj and Umrah.

The study provides in-depth analyses of methods currently in use for crowd analysis from crowd footage, particularly those that apply deep learning to identify anomalous behaviours.

Proposed method

Details FCNN technique

The suggested model is as follows, as indicated: The input layer is the first of the proposed model's 10 layers, which end with two convolutional layers (CLs). Two pooling layers (PLs), two dropout layers (DLs), two fully connected layers (FCLs), and one output layer make up the architecture. SGI specifies the dimensions of the input layer (256 x 256 x 1). To reduce the amount of parameters and increase training effectiveness, a 5 x 5 kernel size was used. Each CL1 and CL2 has 64 and 32 filters, respectively. Compared to CL1, the PL2 sample size is lower. The FCL1 converts the feature maps from the CL2 into a one-dimensional representation. The FCL2 enables the last layer to categorize the input data according to its intended use. This neural architecture's valid convolution method makes sure that the size of the feature maps stays constant. Additionally, by allowing the network to generalize the input, the two dropout layers reduce over teaching [61, 62]. The neural network (BPS) is trained using the backward propagation step, as was already explained. By using the BPS update of weights and biases, network training's primary purpose is to lower objective function error. A deep learning rate is taken into account to choose the DCNN structure during the training phase. This deep learning technique boosts the neural network's effectiveness and prevents it from settling at a local minimum. Additionally, it is advised to update the DCNN's weights using the adaptive moment estimation approach (Adam) [63]. When dealing with sparse gradients, Adam combines the benefits of deep gradient algorithms (AdaGrad) and a non-stationary root-mean-square propagation (RMSProp) approach. Adam monitors the gradient's square and its exponential moving average (EMA), which are connected as follows:

$$w = w - \alpha \frac{B_{mt_1}}{\sqrt{B_{mt_2} + \varepsilon}} \tag{1}$$

$$B_{mt_1} = \beta_1 B_{mt_1 - 1} + (1 - \beta_1) \frac{\partial}{\partial w} \cos t(w) \quad \text{here } \beta_1 \approx 1 \tag{2}$$

$$B_{mt_2} = \beta_1 B_{mt_2 - 1} + (1 - \beta_2) \frac{\partial^2}{\partial w^2} \cos t(w) \quad \text{here } \beta_2 \approx 1 \tag{3}$$

where is the step size in the positive scalar? w is the weight measure, and α is the weight. The first and second moment bias fixes are B_{mt_1} and B_{mt_2}, respectively. The decay rates are β_1, β_2. Eqs. (2 and 3) show that both the step size α and the decline rates β_1, β_2 are small. As a consequence, the weight update approach in Eq. (1) yields a nearly optimal-201learning rate choice [63]. As a result, the proposed model is used to refer to the final structure, which is made up of the CNN, the deep learning rate, and Adam in this study. Finally, the hyperparameters (i.e., dropout rate, learning rate, momentum, number of epochs, and batch size) of the recommended architectures are optimized using a grid search-based 5-fold Cross Validation (5-CV). Figure 1 shows the proposed model

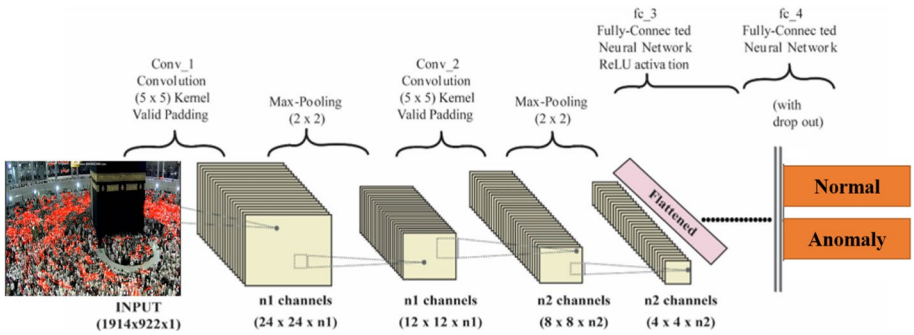

Fig. 1 The proposed FCNN model

```
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 118, 118, 32)      896

 max_pooling2d (MaxPooling2D  (None, 59, 59, 32)       0
 )

 conv2d_1 (Conv2D)           (None, 57, 57, 64)        18496

 max_pooling2d_1 (MaxPooling  (None, 28, 28, 64)       0
 2D)

 conv2d_2 (Conv2D)           (None, 26, 26, 64)        36928

 max_pooling2d_2 (MaxPooling  (None, 13, 13, 64)       0
 2D)

 flatten (Flatten)           (None, 10816)             0

 dense (Dense)               (None, 128)               1384576

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 2)                 258

=================================================================
Total params: 1,441,154
Trainable params: 1,441,154
Non-trainable params: 0
_________________________________________________________________
```

Fig. 2 Parameter's details in proposed FCNN model

in detail, along with layer specifications. The model is constructed by drawing hyperparameters from each of the specified distributions. These settings govern the activation function, dropout layer probability, learning rate, optimizer, and learning decay for the optimization technique. This paradigm's primary objective is classification. Figure 2 shows the parameters details in the proposed model.

Datasets

Hajj-Crowd-2021 anomaly dataset

This section will go through the new Hajj-Crowd dataset for anomalies. Its objective is to establish a standard for optical flow, with a particular emphasis on applications in crowd

analysis. The primary objective of optical flow algorithms in this sector is to estimate pedestrian errors in movements, which is very difficult to do in heavily packed regions. This motion estimate accuracy is critical for following algorithms such as crowd flow analysis, segmentation, and tracking. Between 2015 and 2019, data on the HAJJ audience was acquired through YouTube live telecasts in Mecca and Hajj. Several towns around the Kaaba (Tawaf region) shot video sequences depicting typical crowd circumstances, such as touching the black stone in the Kaaba area and flinging it into the mina area. We limited our investigation to films taken in the vicinity of the Kaaba (Tawaf area). We gathered films from two different classes, and each ten-second video has 300 frame sequences. The dataset comprises 100 videos that are considered normal and 100 videos that are considered abnormal. Each video, whether normal or abnormal, has 60,000 frames. The whole scene was shot in high definition at a frame rate of 25Hz frames per second. Figure 3 shows the example of the Hajj-Crowd anomaly dataset.

Justification

Created a novel Hajj crowd video anomaly dataset. Since there is no predefined dataset for crowd anomalies based on Hajj, we created a new one as part of this study. There have been a few public datasets in the previous several years from various institutions across the globe (such Sub way, UCSD, and UMN), but our datasets are completely unique. The reason why our generated dataset is superior is because it is entirely based on the Hajj domain, while state-of-the-art public datasets only partially cover this topic.

UCSD ped2 [57]

In this dataset, walkers are the most prevalent dynamic objects, with crowd density ranging from low to high. A phenomenon is defined as the appearance of an item, such as a vehicle, skateboarder, wheelchair, or bicycle. Each training frame in this

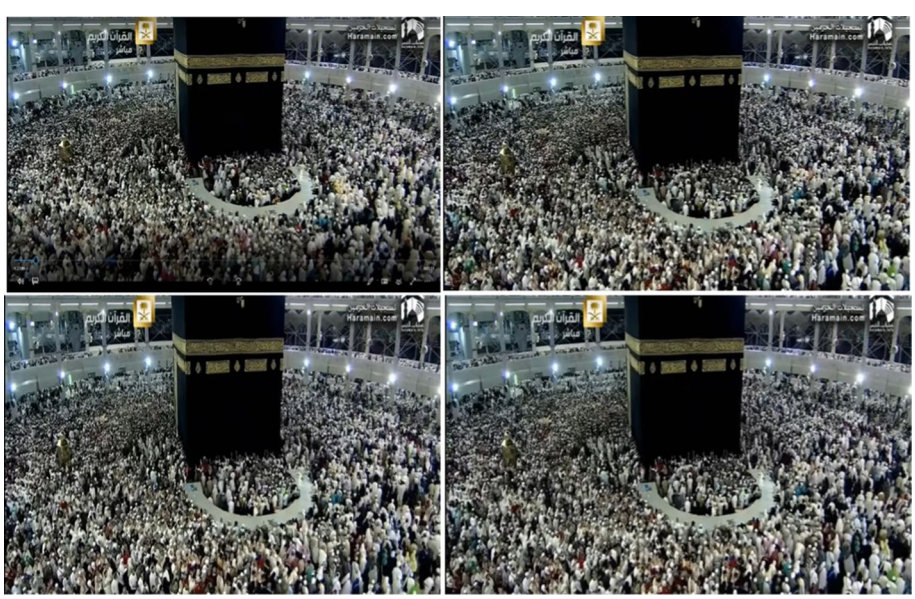

Fig. 3 Example of the Hajj-Crowd anomaly datasetl

dataset is normal and consists entirely of pedestrians. At 320 x 240 resolution, this dataset includes 12 video tests and 16 video tests for training. All test frames' ground truth may be accessed to evaluate the localization. 2,384 abnormal and 2,566 normal frames make up the 2,384 abnormal frames.

Subway [30]

This dataset contains two sequences gathered during the entrance (1 h and 36 min, 144,249 frames) and exit of a subway station (43 min, 64,900 frames). The majority of individuals that enter and exit the station behave nicely. Individuals moving in the wrong direction (i.e., leaving from or entering the entry) or escaping payment are classified as atypical instances. This dataset is subject to the following constraints: There are few abnormalities, and their geographical localizations are predictable (at entrance or exit regions).

Abnormal detection process (Labeling)

The data labelling process shown in Fig. 4. For this process, our dataset is 10 s video. We have developed three rules for the anomaly detection labelling process.

A) Abnormality happens when the crowd movement does not follow the normal Tawaf movement.

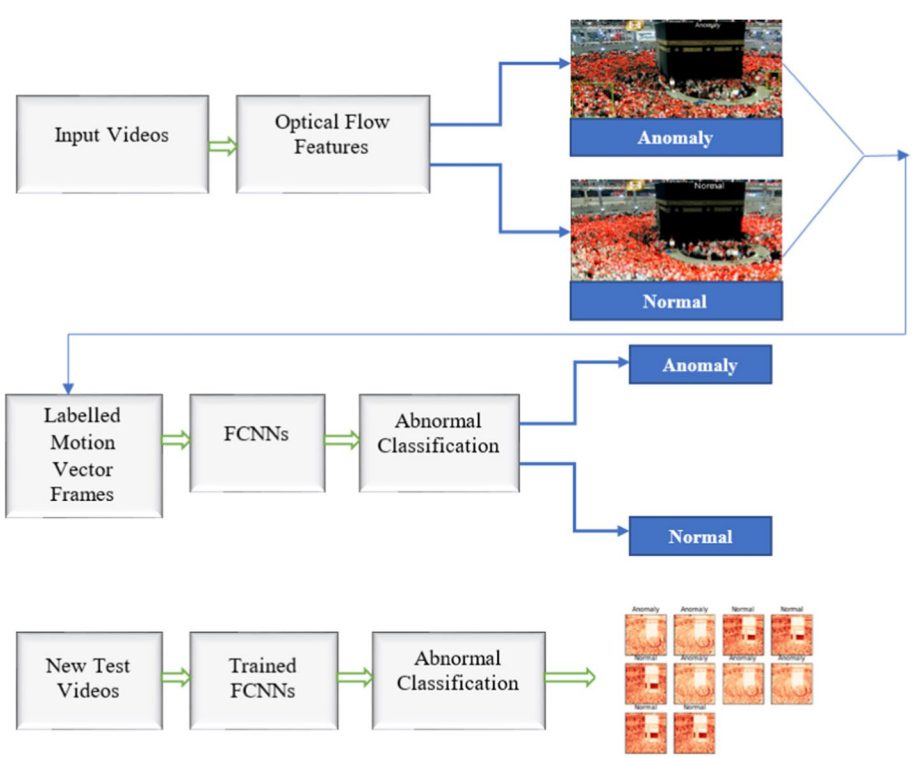

Fig. 4 Anomaly detection model

B) From a 10 s video, we extract the image frames. Select every frame and check the abnormal motion vector for each image. If the image has abnormal motion vectors of more than around 20% across 10 s, we label it abnormal.
C) We have checked manually if the labelling has not been corrected.

Annotation

We need labeled videos for training purposes in order for our anomaly detection technology to perform. While this is true, we must know the start and end frames of each abnormal testing video in order to analyze the abnormal event's influence on testing images. Numerous annotators document the chronological duration of each abnormality by assigning them the same movie. Final temporal annotations are generated by averaging the annotations of many annotators. After several months of work, the dataset was ultimately finalized. Figure 5 shows the normal and anomaly labelled frames.

Optical flow technique

The optical flow method would be ideal since the goal is to build a low-cost system with little implementation complexity while quick enough to process and efficient with motion-only outcomes.

When compared to other methods, the optical flow has the advantage of requiring fewer data storage, reducing complexity due to the feature vectors provided by optical flow being sufficient to identify motion and objects of interest, and decreasing processing costs due to the need for less bandwidth to transmit only the flow vectors rather than monitoring the entire video. Lucas-Kandae (LK) has also been used in an optical flow, which is another benefit. Due to its method of solving the video frame in tiny parts, LK outperformed

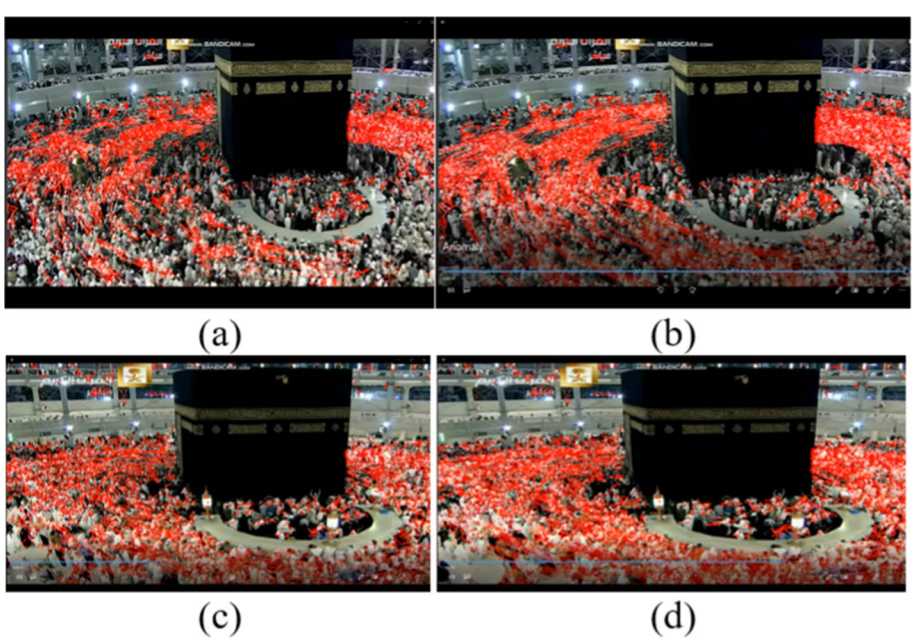

Fig. 5 **a** and **b** shows the anomaly labelled frames,**c** and **d** shows the normal labelled frames

Horn-Schunck in this application, which assumes optical flow is smooth across the whole image. The optical flow may be calculated using the following equation:

$$I_x u + I_y v + I_t = 0 \tag{4}$$

The spatial-temporal image brightness derivatives are *Ix*, *Iy* and *It*, while *u* represents the horizontal part of the optical flow and *v* represents the vertical part of the optical flow. Following the division of the image into small sections, a weighted least-square fit of (1) is applied to a constant model *[U V]T* in each segment, by minimising:

$$\sum_{x \in \Omega} W^2 \left[I_x u + I_y v + I_t \right]^2 \tag{5}$$

W is the window function in the above equation, which emphasises the constraints at the top. The following is the solution to the minimization equation:

$$\begin{bmatrix} \sum w^2 I_x^2 & \sum w^2 I_x I_y \\ \sum w^2 I_y I_x & \sum w^2 I_y^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum^2 I_x I_t \\ \sum W^2 I_y I_t \end{bmatrix} \tag{6}$$

Anomaly detection

The video is represented in this article using a collection of regional characteristics. Feature vectors in the output of the kth convolutional layer equation 7 are extensively used to extract and define these characteristics: All normal regional features obtained by the FCN are fitted with a Gaussian classifier $G1(.)$. Regional characteristics whose distance from $G1(.)$ s greater than the criterion *alpha* are regarded abnormal. When the distance between $G1$ two points is less than or equal to the threshold value *beta*, the point is considered normal.

$$T_{k,n} = \left\{ T_k^t(i,j,n) \right\}_{(i,j)=(1,1)}^{(w'_k, h'_k)}, \text{ for } n = 1, 2, \ldots, h \tag{7}$$

where h is the size of the auto-encoder-generated feature vectors, which is equal to the size of the hidden layers.

This stage processes just the anomaly locations. Thus, certain locations (i, j) in grid (w_k, h_k) are disregarded and do not undergo analysis in the grid $\left(w'_k, h'_k\right)$. As with G_1, we generate a Gaussian classifier G_2 on all of the standard training regional characteristics that our auto-encoder represents. Anomalies are defined as areas that are not satisfactorily suited to G_2.

Equation (8) outline the identification of anomalies via the use of Gaussian classifiers. We possess that,

$$G_1\left(f_k^t\left(i,j,1:m_k\right)\right) = \begin{cases} \text{Normal} & \text{if } d\left(G_1, f_k^t\left(i,j,1:m_k\right)\right) \\ \leq \beta & \\ \text{Abnormal} & \text{if} d\left(G_1, f_k^t\left(i,j,1:m_k\right)\right) \\ \geq \alpha & \end{cases} \tag{8}$$

Experimental design and result analysis

Experimental setup

The processing of high-resolution images in a fully connected network (e.g., 1914 x 922 pixels) presents a range of challenges and constraints, particularly with the use of GPU memories. Only specific kernels and layer scans have our FCNN convolutional (i.e., model capacity). We, therefore, aim to create the best possible FCNN architecture to process images like those in a UCSD dataset with the highest possible resolution. We used an NVidia GTX 1660Ti 6GB RAM 16GB card for our testing. In the end, we utilized python3 in conjunction with deep learning programs such as open-cv2, NumPy, SciPy, matplotlib, Tensor Flow GPU, CUDA, Keras, and other similar tools.

Matrix evaluation

The efficacy of the proposed Hajj-Crowd framework can be evaluated using the following performance criteria: 1. Precision 2. Recall, 3. F1 score, 4. Accuracy final, 5. Obtain graph illustrating the separation of classes. Precision, Recall, and F1 scores can be calculated using the following formula ([64]). For measuring detection accuracy at the frame level, the Receiver Operating Characteristic (ROC) curve is utilized. The relationship between the true positive rate (TPR) and false positive rate (FPR) is known as the ROC curve.

$$\text{Accuracy} = \frac{\text{Correct Prediction}}{\text{Correct Prediction} + \text{Incorrect Prediction}} \tag{9}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{10}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{11}$$

$$\text{Flscore} = 2 * \frac{\text{Recall} * \text{Precision}}{\text{Recall} + \text{Precision}} \tag{12}$$

$$TPR = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{13}$$

$$FPR = \frac{False\ Positive}{True\ Negative + False\ Positive} \tag{14}$$

In Equations (9) through (14), TP, TN, FN, and FP stand for true positive, true negative, false negative, and false positive, respectively. The perplexity matrix illustrates the performance's clarity while evaluating the suggested Hajj-Crowd output. Experiment 1 and Experiment 2 both result in the addition of all metrics.

Training and testing set

The dataset is split into training and testing groups of 100 normal and 100 abnormal videos each. The two anomalies occur in both the training and testing sets at various

temporal positions. Further, several of the videos exhibit multiple irregularities. The duration of the training films, in 10 s, is determined by their distribution.

Abnormal detection process (training)

The abnormal detection training procedure is shown in Fig. 4. For completing this process, at first, we used labelled motion vector images. Secondly, we collected all the labelled motion vector images for training using fully convolutional neural networks (FCNN). We used FCNN for abnormal classification and found out the Abnormal and Normal. Finally, we have done the classification based on the three rules.

Abnormal detection process (testing)

The method of crowd abnormal detection testing using FCNN is shown in Fig. 4. Firstly, we have prepared the new test video dataset. Then we have passed the labelled motion vector images for testing. Secondly, we have completed testing two classes according to three rules using FCNN. Finally, we have gotten the two classes classification results abnormal and normal based on three rules.

Comparison with the state-of-the-art

Reconstruction errors were used by Lu et al. [10] to learn typical behaviour and identify anomalies. Each typical training video yields 7000 cuboids, which the researchers use to calculate gradient-based characteristics in each volume. After PCA reduces the function dimension, the dictionary learns using sparse representation. To comprehend local features and classifiers, a fully convolutional feed forward deep autoencoder-based method was presented by Hasan et al. [4]. To get the most out of the network, use their implementation on movies with a 40-frame time window. Similar to [10], reconstruction error is used to estimate anomaly. The model training setup for this technique is the same as for the strategy we propose, with 32 video segments in each bag and C3D features calculated.

Furthermore, as a starting point, they use a binary SVM classifier. They sort all videos based on whether they include anomalies or not and then separate those videos again. Each movie has its own set of C3D characteristics, and a linear kernel is used to train a binary classifier. This classifier determines the probability that a video clip will be considered abnormal for testing. Table 2 shown the comparison of the anomaly dataset.

Results and discussion

Experiment using FCNN

There are 60,000 frames in the Hajj-Crowd anomaly dataset, with 30,000 frames in each class. The UCSD Ped2 dataset comprises 4,950 frames total, with 23,84 aberrant and 2,566 normal frames for each class. The subway dataset has divided into two. (1) Subway entry and Subway exit. The entire number of datasets for subway entries is 144,249, while the total number of datasets for subway exits is 64,900. For each of the two datasets, we utilized 80% for training and 20% for testing. We separated each dataset into two folds.

Each of these comparison investigations used the same experimental dataset. For experiment 1, the suggested dataset, UCSD Ped2, the subway entry dataset, and the

Table 2 Comparison of dataset with anomaly

Dataset name	Number of videos	Number of frame	Dataset length	Type of anomaly
UCSD Ped1 [57]	70	201	5 min	Bikers, small carts, walking across walkways
UCSD Ped2 [57]	28	163	5 min	Bikers, small carts, walking across walkways
Subway entrance [30]	1	121,749	1.5 h	Wrong direction, No payment
Subwa exit [30]	1	64,901	1.5 h	Wrong direction, No payment
UMN [65]	5	1290	5 min	Run
Abnormal crowd [66]	31	1408	24 min	Panic, fight, congestion, obstacle, neutral
Proposed (Hajj-Crowd-2021)	200	60,000	2 h	Normal and Wrong Movement

Table 3 Result for the two classes proposed Hajj-Crowd dataset using FCNN model

Class	Precision	Recall	f1-Score	Support
Anomaly	1.00	1.00	1.00	6000
Normal	1.00	1.00	1.00	6000
Micro avg	1.00	1.00	1.00	12000
Macro avg	1.00	1.00	1.00	12000

Table 4 Result for the two classes for UCSD Ped2 using FCNN model

Class	Precision	Recall	f1-Score	Support
Anomaly	1.00	0.90	1.00	495
Normal	1.00	1.00	0.89	495
Micro avg	0.92	0.92	0.92	990

Table 5 Result for the two classes for subway Entry dataset using FCNN model

Class	Precision	Recall	f1-Score	Support
Anomaly	1.00	1.00	0.75	14424
Normal	1.00	0.90	1.00	14424
micro avg	0.96	0.96	0.96	28849

subway exit dataset, respectively, the FCNN technique obtained final accuracy of 100%, 90%, 95%, and 89 %. The average microprecision, microrecall, and microF1 scores for the recommended approach are shown in Tables 3, 4, 5, 6. The equations from [67] were used to build each of these evaluation matrices. The average microaccuracy, microrecall, and microF1 score for the proposed dataset are 100%, 100%, 100%, respectively, compared to 92.0%, 92.0%, 96.0%, 96.0%, and 90.0% for the UCSD Ped2, Subway Entry and Subway Exit dataset. Fig. 5 shows the ROC AUC graph for FCNN model. In Fig. 5a, we got the AUROC 96 % using UCSD PED1 dataset and AUROC 99% using Proposed Hajj-Crowd dataset, respectively. In Fig. 5b, we achieved AUROC 85% using Subway Entry dataset and Subway Exit dataset we got the AUROC 94%. Using a state-of-the-art

Table 6 Result for the two classes for Subway Exit dataset using FCNN model

Class	Precision	Recall	f1-Score	Support
Anomaly	0.80	1.00	1.00	6490
Normal	0.95	1.00	0.90	6490
micro avg	0.90	0.90	0.90	12980

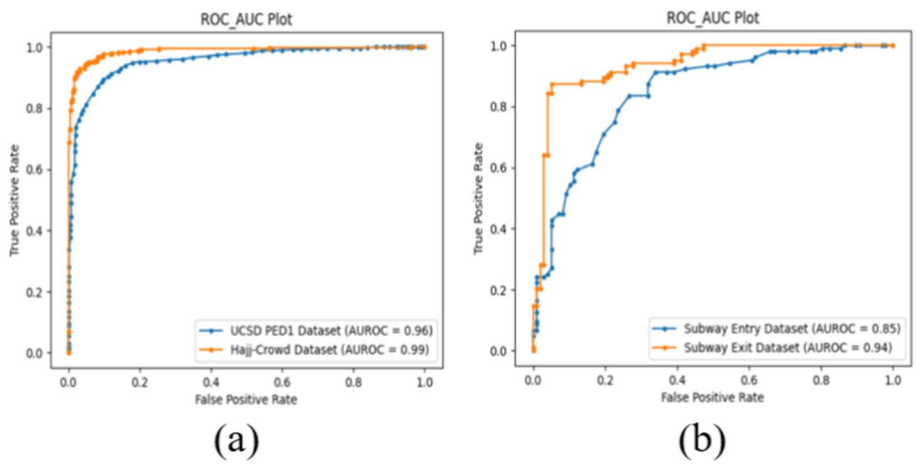

Fig. 6 ROC AUC graph for FCNN model

Table 7 Result for the two classes proposed Hajj-Crowd dataset using ResNet50 model

Class	Precision	Recall	f1-Score	Support
Anomaly	1.00	0.95	1.00	6000
Normal	1.00	1.00	0.92	6000
Macro avg	0.95	0.95	0.95	12000

Table 8 Result for the two classes for UCSD Ped2 using ResNet50 model

Class	Precision	Recall	f1-Score	Support
Anomaly	0.99	1.00	0.89	495
Normal	1.00	1.00	0.92	495
Micro avg	94	94	94	990

dataset including JHU-CROWD and UCSD, the proposed FCNN model outperforms in terms of overall precision. In addition, as far as we are aware, our proposed dataset is the only one of its kind in this discipline. Fig. 6 shows the ROC AUC graph for proposed FCNN Model.

Experiment 2 using ResNet50

In the second trial, ResNet50 achieved accuracies of 96%, 89%, 94%, and 92% on the suggested dataset, UCSD Ped2, and the Subway Entry and Exit datasets, respectively. Tables 7, 8, 910 show the average microprecision, microrecall, and microF1 score

Table 9 Result for the two classes for Subway Entry dataset using ResNet50 Model

Class	Precision	Recall	f1-Score	Support
Anomaly	0.80	1.00	0.90	14424
Normal	1.00	0.99	1.00	14424
Micro avg	0.91	0.91	0.91	28849

Table 10 Result for the two classes for Subway Exit dataset using ResNet50 Model

Class	Precision	Recall	f1-Score	Support
Anomaly	0.75	1.00	0.98	6490
Normal	1.00	0.83	0.85	6490
Micro avg	0.89	0.89	0.89	12980

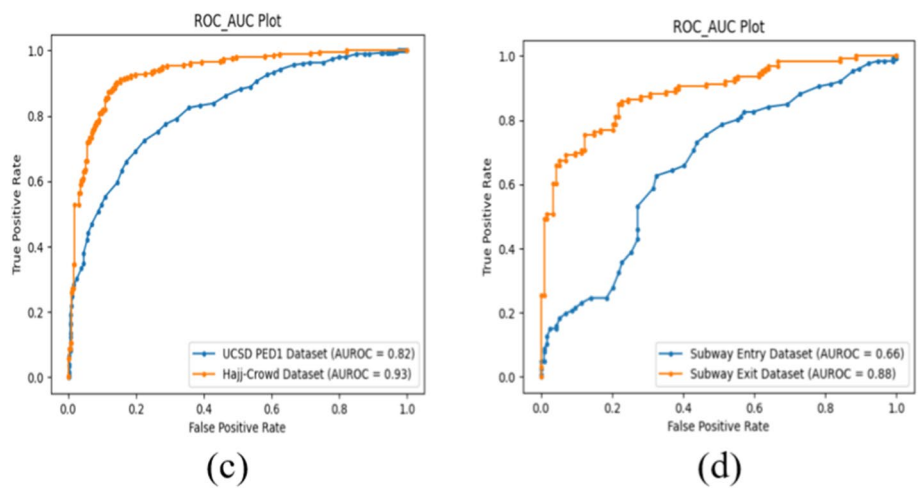

Fig. 7 ROC AUC graph for ResNet50 model

achieved by the proposed method. The evaluative matrices shown here were computed using the formulas from [67]. The suggested dataset has an average microaccuracy, microrecall, and microF1 score of 95%, 94%, 94%, and 95%, respectively, whereas the UCSD Ped2, Subway Entry, and Subway Exit datasets have scores of 94%, 94%, and 94%, and 89%, 89%, and 89%, respectively. Fig 6 shows the ROC AUC graph for ResNet50 model. In Fig. 6c, we got the AUROC 82 % using UCSD PED1 dataset and AUROC 93% using Proposed Hajj-Crowd dataset, respectively. In Fig. 6d, we achieved AUROC 66% using Subway Entry dataset and Subway Exit dataset we got the AUROC 88%. The proposed FCNN model exhibits impressive overall accuracy when compared to cutting-edge datasets like the UCSD PED1 dataset and the Subway Entry and Exit dataset. The data collecting method we suggest could be the first of its kind in this industry. The ROC AUC graph for the ResNet50 Model is shown in Fig. 7.

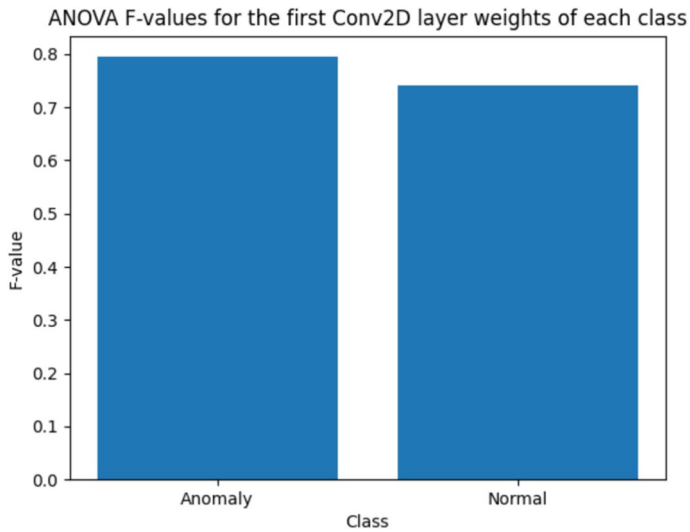

Fig. 8 ANOVA test for F-values

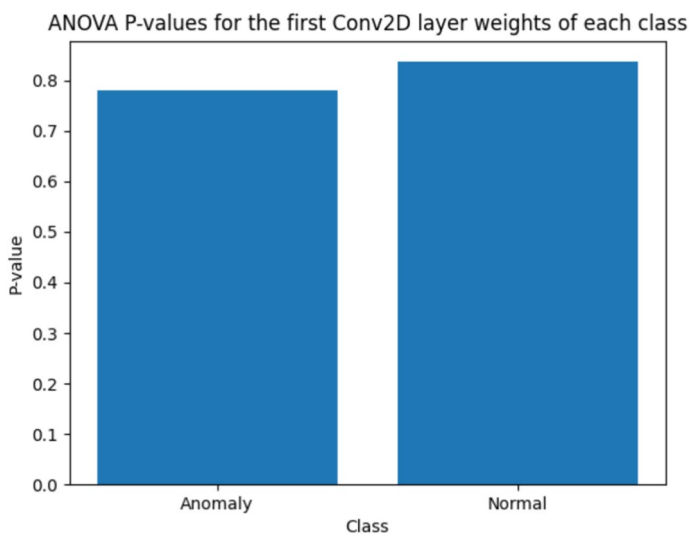

Fig. 9 ANOVA test for P-values

Anova test result

To ascertain if there were statistically significant variations in the means of the weights for each class in the first Conv2D layer of a neural network, we used an analysis of variance (ANOVA) F-test. The norms of the other classes' classes were compared to the mean weights of each class to see whether there was a statistically significant difference. P values: 0.6626972046077602 was established using the ANOVA test, which had an F-value (F values: 0.8760922406451678) of X and degrees of freedom for Y and Z in the numerator and denominator, respectively. As a result, we draw the conclusion that there are statistically significant differences between the means of the weights for each class in the first Conv2D layer and reject the null hypothesis. Figs. 8 and 9 shows the ANOVA test F-values and P-values graphical results.

Conclusion

This article describes a novel FCNN architecture with optical flow for creating and identifying anomalous areas in videos. Context free regional features are produced by combining the capabilities of FCNN with optical flow architecture for patch-wise actions on input data. A new convolutional layer based on kernels learnt from the training video is also included in the FCNN. The suggested FCNN's final convolutional layer must be taught. In terms of processing speed, the suggested methodology beats previous approaches. Additionally, it is a method for circumventing constraints in the training samples needed to learn a full CNN.We can run a deep learning-based algorithm at a frame rate of roughly 370 frames per second using this method. Video abnormalities can be quickly and accurately detected using the above-mentioned method. The difficulties of incorporating input from mobile security cameras in crowd video analysis were not addressed. There must be a human verification step in the crowd density tagging procedure.

Acknowledgements
Multimedia University, Cyberjaya, Malaysia fully supported this research. This research supported from the grant of GRA (Grant No. MMUI/190031.) and FRDGS (Grant No. MMUE/210,030).

Author contributions
RB collected the both datasets, wrote the full paper and review the paper. JA helped to fix the writing, checked the paper formate as well as reviewed the full paper. NH gave some important feedback on this paper. FF helped with the structured full paper revision. JU helped format the full paper. All authors read and approved the final manuscript.

Funding
This research is supported by the Multimedia University MMU GRA grant (Grant No.MMUI/190031).

Availability of data and materials
The 2015–2019 Mecca Hajji crowd density dataset has a total of 30,000 images, while the crowd anomaly dataset contains a total of 200 videos. Both datasets were taken during the Hajj in Mecca. The information includes three categories of crowd density in the Tawaf region. The Methods section provides comprehensive details that will allow the study to be repeated. If you have any queries about the method, please email the author relating to your inquiry.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that there is no competing interest in this paper.

Received: 21 September 2022 Accepted: 17 May 2023
Published online: 28 May 2023

References

1. Sabokrou M, Fayyaz M, Fathy M, Klette R. Deep-cascade: cascading 3d deep neural networks for fast anomaly detection and localization in crowded scenes. IEEE Trans Image Process. 2017;26(4):1992–2004.
2. Krizhevsky A, Hinton Sutskever I, GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012. https://doi.org/10.1145/3065386.
3. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation, in proceedings of the IEEE conference on computer vision and pattern recognition, 2014;580–587.
4. Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos, Advances in neural information processing systems, 2014;27.
5. Sabokrou M, Fathy M, Hoseini M, Klette R, Real-time anomaly detection and localization in crowded scenes, in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015;56–62.
6. Xu D, Ricci E, Yan Y, Song J, Sebe N. Learning deep representations of appearance and motion for anomalous event detection. arXiv Prepr. 2015. https://doi.org/10.48550/arXiv.1510.01553.
7. Sabokrou M, Fathy M, Hoseini M. Video anomaly detection and localisation based on the sparsity and reconstruction error of auto- encoder. Electron Lett. 2016;52(13):1122–4.

8. Ren S, He K, Girshick R, Sun J. Faster r-cnn: towards real-time object detection with region proposal networks, advances in neural information processing systems, 2015;28.
9. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation, in proceedings of the IEEE conference on computer vision and pattern recognition, 2015;3431–3440.
10. Zhou B, Lapedriza A, Xiao J, Torralba A. Oliva A. Learning deep features for scene recognition using places database: advances in neural information processing systems, 2014; 27.
11. Fie-Fie L, Li K. Imagenet, image-net. org, 2016.
12. J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.
13. places M. database, places.csail.mit.edu, in IEEE conference on computer vision and pattern recognition. Ieee. 2009;2017: 248–55.
14. Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, Le- Cun Y. Overfeat: integrated recognition, localization and detection using convolutional networks. arXiv Preprint. 2013. https://doi.org/10.48550/arXiv.1312.6229.
15. Oquab M, Bottou L, Laptev I, Sivic J. Learning and transferring mid-level image representations using convolutional neural networks, in proceedings of the IEEE conference on computer vision and pattern recognition, 2014;1717–1724.
16. Jiang F, Yuan J, Tsaftaris SA, Katsaggelos AK. Anomalous video event detection using spatiotemporal context. Comput Vis Image Underst. 2011;115(3):323–33.
17. Feng Y, Yuan Y, Lu X. Learning deep event models for crowd anomaly detection. Neurocomputing. 2017;219:548–56.
18. Chan T-H, Jia K, Gao S, Lu J, Zeng Z, Ma Y. Pcanet: a simple deep learning baseline for image classification? IEEE Trans Image Process. 2015;24(12):5017–32.
19. Fang Z, Fei F, Fang Y, Lee C, Xiong N, Shu L, Chen S. Abnormal event detection in crowded scenes based on deep learning. Multimed Tools Appl. 2016;75(22):14617–39.
20. Wu S, Moore BE, Shah M, Chaotic invariants of lagrangian particle trajectories for anomaly detection in crowded scenes, in IEEE computer society conference on computer vision and pattern recognition. IEEE. 2010;2010:2054–60.
21. Piciarelli C, Foresti GL. On-line trajectory clustering for anomalous events detection. Pattern Recognit Lett. 2006;27(15):1835–42.
22. Piciarelli C, Micheloni C, Foresti GL. Trajectory-based anomalous event detection. IEEE Trans Circuits Syst video Technol. 2008;18(11):1544–54.
23. Antonakaki P, Kosmopoulos D, Perantonis SJ. Detecting abnormal human behaviour using multiple cameras. Signal Process. 2009;89(9):1723–38.
24. Calderara S, Heinemann U, Prati A, Cucchiara R, Tishby N. Detecting anomalies in people's trajectories using spectral graph analysis. Comput Vision Image Underst. 2011;115(8):1099–111.
25. Morris BT, Trivedi MM. Trajectory learning for activity under- standing: unsupervised, multilevel, and long-term adaptive approach. IEEE Trans Pattern Anal Mach Intell. 2011;33(11):2287–301.
26. Hu W, Xiao X, Fu Z, Xie D, Tan T, Maybank S. A system for learning statistical motion patterns. IEEE Trans Pattern Anal Mach Intell. 2006;28(9):1450–64.
27. Tung F, Zelek JS, Clausi DA. Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance. Image Vision Comput. 2011;29(4):230–40.
28. Zhang D, Gatica-Perez D, Bengio S, McCowan I. Semi- supervised adapted hmms for unusual event detection, in 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1. IEEE, 2005;611–618.
29. Boiman O, Irani M. Detecting irregularities in images and in video. Int J Comput Vision. 2007;74(1):17–31.
30. Adam A, Rivlin E, Shimshoni I, Reinitz D. Robust real-time unusual event detection using multiple fixed-location monitors. IEEE Trans Pattern Anal Mach Intell. 2008;30(3):555–60.
31. Mahadevan V, Li W, Bhalodia V, Vasconcelos N, Anomaly detection in crowded scenes, in IEEE computer society conference on computer vision and pattern recognition. IEEE. 2010;2010:1975–81.
32. Li W, Mahadevan V, Vasconcelos N. Anomaly detection and localization in crowded scenes. IEEE Trans Pattern Anal Mach Intell. 2013;36(1):18–32.
33. Kim J, Grauman K. Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates, in IEEE conference on computer vision and pattern recognition. IEEE. 2009;2009:2921–8.
34. Benezeth Y, Jodoin P-M, Saligrama V, Rosenberger C, Abnormal events detection based on spatio-temporal co-occurences, in IEEE conference on computer vision and pattern recognition. IEEE. 2009;2009:2458–65.
35. Kratz L, Nishino K. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models, in IEEE conference on computer vision and pattern recognition. IEEE. 2009;2009:1446–53.
36. Mehran R, Oyama A, Shah M. Abnormal crowd behavior detection using social force model, in IEEE conference on computer vision and pattern recognition. IEEE. 2009;2009:935–42.
37. Zaharescu A, Wildes R. Anomalous behaviour detection using spatiotemporal oriented energies, subset inclusion histogram comparison and event-driven processing, in European conference on computer vision. Springer, 2010;563–576.
38. Cong Y, Yuan J, Liu J. Sparse reconstruction cost for abnormal event detection, in CVPR. IEEE. 2011;2011:3449–56.
39. Saligrama V, Chen Z, Video anomaly detection based on local statistical aggregates, in IEEE conference on computer vision and pattern recognition. IEEE. 2012;2012:2112–9.
40. Ullah H, Conci N. Crowd motion segmentation and anomaly detection via multi-label optimization, in ICPR workshop on pattern recognition and crowd analysis, 2012; 75.
41. Lu C, Shi J, Jia J. Abnormal event detection at 150 fps in matlab, in proceedings of the IEEE international conference on computer vision, 2013; 2720–2727.
42. Roshtkhari MJ, Levine MD. An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal com- positions. Comput Vision Image Underst. 2013;117(10):1436–52.

43. Zhu Y, Nayak NM, Roy-Chowdhury AK. Context-aware modeling and recognition of activities in video, in proceedings of the IEEE conference on computer vision and pattern recognition, 2013;2491–2498.
44. Cong Y, Yuan J, Tang Y. Video anomaly search in crowded scenes via spatio-temporal motion context. IEEE Trans Inform Forensics Secur. 2013;8(10):1590–9.
45. Roshtkhari M Javan, Levine MD. Online dominant and anomalous behavior detection in videos, in proceedings of the IEEE conference on computer vision and pattern recognition, 2013;2611–2618.
46. Ullah H, Tenuti L, Conci N. Gaussian mixtures for anomaly detection in crowded scenes, in video surveillance and transportation imaging applications, vol. 8663. International society for optics and photonics, 2013;866303.
47. Ullah H, Ullah M, Conci N. Real-time anomaly detection in dense crowded scenes, in video surveillance and transportation imaging applications 2014;9026. SPIE, 2014, pp. 51–57.
48. Ullah H, Ullah M, Conci N. Dominant motion analysis in regular and irregular crowd scenes, in international workshop on human behavior understanding. Springer, 2014; 62–72.
49. Xu D, Song R, Wu X, Li N, Feng W, Qian H. Video anomaly detection based on a hierarchical activity discovery within spatiotemporal contexts. Neurocomputing. 2014;143:144–52.
50. Vincent P, Larochelle H, Bengio Y, Manzagol P-A. Extracting and composing robust features with denoising autoencoders, in proceedings of the 25th international conference on machine learning, 2008;1096–1103.
51. Mousavi H, Nabi M, Galoogahi HK, Perina A, Murino V. Abnormality detection with improved histogram of oriented tracklets, in international conference on image analysis and processing. Springer, 2015;722–732.
52. Yuan Y, Fang J, Wang Q. Online anomaly detection in crowd scenes via structure analysis. IEEE Trans cybern. 2014;45(3):548–61.
53. Cheng K-W, Chen Y-T, Fang W-H. Video anomaly detection and localization using hierarchical feature representation and gaussian process regression, in proceedings of the IEEE conference on computer vision and pattern recognition, 2015;2909–2917.
54. Xiao T, Zhang C, Zha H. Learning to detect anomalies in surveillance video. IEEE Signal Process Lett. 2015;22(9):1477–81.
55. Sabokrou M, Fayyaz M, Fathy M, Moayed Z, Klette R. Deepanomaly: fully convolutional neural network for fast anomaly detection in crowded scenes. Comput Vision Image Underst. 2018;172:88–97.
56. Li N, Wu X, Xu D, Guo H, Feng W. Spatiotemporal context analysis within video volumes for anomalous-event detection and localization. Neurocomputing. 2015;155:309–19.
57. Bhuiyan MR, Abdullah J, Hashim N, Al Farid F. Video analytics using deep learning for crowd analysis: a review. Multimed Tools Appl. 2022;81:1–28.
58. Rabiee H, Haddadnia J, Mousavi H, Kalantarzadeh M, Nabi M, Murino V. Novel dataset for fine-grained abnormal behavior understanding in crowd, in 2016 13th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE, 2016;95–101.
59. Alafif T, et al. Hybrid classifiers for spatio-temporal abnormal behavior detection. Tracking, and recognition in massive hajj crowds. Electronics. 2023;12(5):1165. https://doi.org/10.3390/electronics12051165.
60. Alhothali A, Balabid A, Alharthi R, et al. Anomalous event detection and localization in dense crowd scenes. Multimed Tools Appl. 2023;82:15673–94. https://doi.org/10.1007/s11042-022-13967-w.
61. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 2014.
62. Vijay Mahadevan, Weixin Li, Viral Bhalodia and Nuno Vasconcelos. Anomaly Detection in Crowded Scenes, In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, 2010.
63. Ma K, Doescher M, Bodden C. Anomaly detection in crowded scenes using dense trajectories. University of Wisconsin-Madison. 2015.
64. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In Advances in Information Retrieval: 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21–23, 2005. Proceedings 27 2005 (pp. 345–359). Springer Berlin Heidelberg.
65. Bhuiyan MR, Abdullah J, Hashim N, Al Farid F, Haque MA, Uddin J, Isa WNM, Husen MN, Abdullah N. A deep crowd density classification model for hajj pilgrimage using fully convolutional neural network. PeerJ Comput Sci. 2022;8: e895.
66. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res. 2014;15(1):1929–58.
67. Dahl GE, Sainath TN, Hinton GE, Improving deep neural networks for lvcsr using rectified linear units and dropout, in IEEE international conference on acoustics, speech and signal processing. IEEE. 2013;2013:8609–13.

Publisher's Note