



# Motion-shape-based deep learning approach for divergence behavior detection in high-density crowd

Muhammad Umer Farooq<sup>1</sup> · Mohamad Naufal M. Saad<sup>1</sup> · Sultan Daud Khan<sup>2</sup>

Accepted: 8 February 2021 / Published online: 26 February 2021  
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

## Abstract

We propose a novel method of abnormal crowd behavior detection in surveillance videos. Mainly, our work focuses on detecting crowd divergence behavior that can lead to serious disasters like a stampede. We introduce a notion of physically capturing motion in the form of images and classify crowd behavior using a convolution neural network (CNN) trained on motion-shape images (MSIs). First, the optical flow (OPF) is computed, and finite-time Lyapunov exponent (FTLE) field is obtained by integrating OPF. Lagrangian coherent structure (LCS) in the FTLE field represents crowd-dominant motion. A ridge extraction scheme is proposed for the conversion of LCS-to-grayscale MSIs. Lastly, a supervised training approach is utilized with CNN to predict normal or divergence behavior for any unknown image. We test our method on six real-world low- as well as high-density crowd datasets and compare performance with state-of-the-art methods. Experimental results show that our method is not only robust for any type of scene but also outperform existing state-of-the-art methods in terms of accuracy. We also propose a divergence localization method that not only identifies divergence starting (source) points but also comes with a new feature of generating a ‘localization mask’ around the diverging crowd showing the size of divergence. Finally, we also introduce two new datasets containing videos of crowd normal and divergence behaviors at the high density.

**Keywords** FTLE · Divergence · Anomaly · Motion · Blob · Shape · Localization

## 1 Introduction

Human lives are always at stake during mass crowd gathering if any undesired event happens, e.g., stampede congestion or bottleneck [1–3]. Studies [4, 5] show that dangerous events at high-density crowds do not occur at once; instead, there is a progression of smaller abnormal pre-events that lead to larger disasters. A few examples of abnormal pre-event include abnormal crowd density increase, panicking or life-threatening overcrowding, jamming (person falling creates a hurdle for others to move), distress due to reduced inter-person distance, people escaping from normal crowd motion path, etc. The most dominant crowd motion observe

in the said abnormal situations is ‘escape’ or ‘divergence’ of the crowd [6].

There have been several attempts in the past for crowd divergence or escape behavior detection [7, 8]. However, in the literature, the description of divergence for a high-density crowd is unclear. In this work, we provide a precise definition of divergence, i.e., ‘the escape behavior of the crowd from a predetermined motion/walking route/path in which high-density crowd is approaching a common destination.’ In the above definition of divergence anomaly for high-/very-high-density crowd, two terms are important and require further elaboration. The first term is ‘predetermined motion/walking path/route’ and second is ‘common destination.’ The reason why we call here motion/walking path/route as ‘predetermined’ is that in the majority of real-life scenarios, there is a certain event/happening at some ‘common destination’ that the crowd wants to see or participate in. Therefore, to reach that event/happening at a common destination, the crowd starts walking with the same goal of reaching a common destination. This results in having the same walking goal for the whole crowd and thus determines in advance where the crowd is heading

✉ Muhammad Umer Farooq  
m.umer.farooq82@gmail.com

<sup>1</sup> Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, 32610 Bandar Seri Iskandar, Perak Darul Ridzuan, Malaysia

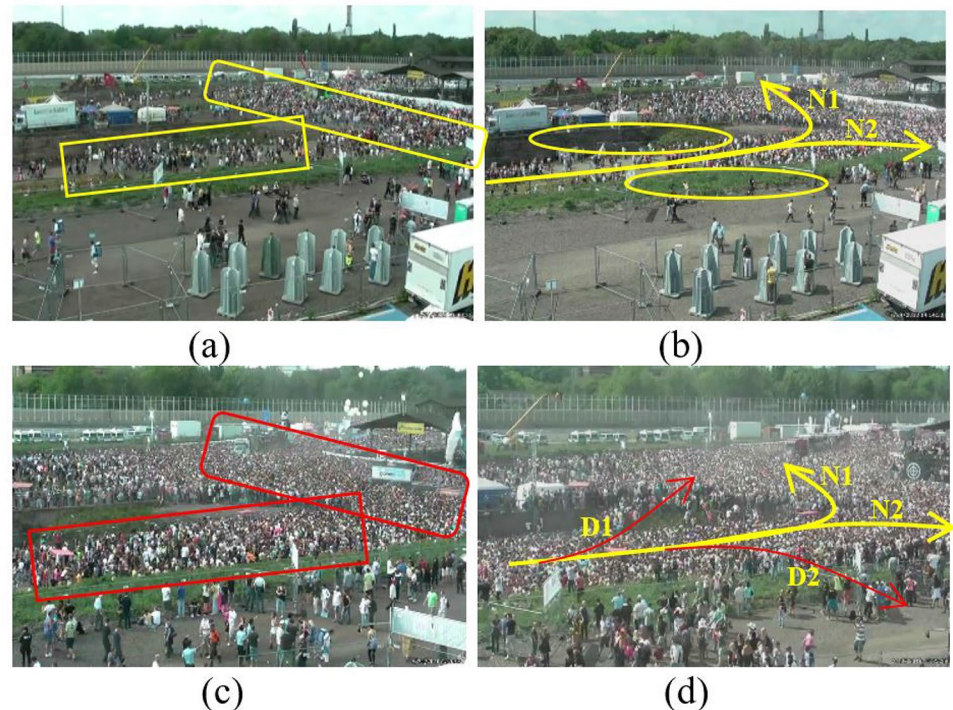
<sup>2</sup> Department of Computer Science, National University of Technology, Islamabad, Pakistan

to. To elaborating more on the predetermined path or the normal motion path, there can be different types of crowd motion paths in real life. Existing art does not mention the type of normal motion patterns from which divergence can occur. We introduce two types of normal motion patterns for divergence: straight and circular motions. If the crowd moves away from straight or circular walking paths, it is considered as crowd divergence. We illustrate the above definition with the help of two examples below. The first example of crowd divergence from a straight walking path can be found in the videos of the Loveparade dataset [4] shown in Fig. 1.

Under normal conditions, the crowd is walking ‘straight’ from the tunnel exit (left side) toward the musical event stage (right side) shown in Fig. 1a. Crowd normal ‘walking paths’ to reach the musical stage are indicated by N1 and N2 (Fig. 1b). Circular regions in Fig. 1b are the green grassy areas that should not be used for walking. However, as the crowd density increases tremendously between the ramp and the musical stage (indicated with a red box in Fig. 1c), crowd starts diverging left and right through green areas, hence leaving the normal straight path. Paths D1 and D2 in Fig. 1d show crowd divergence from normal walking paths N1 and N2.

Similarly, another example of crowd divergence from the circular path is shown in Fig. 2. Under the normal conditions, the crowd follows semicircular path N. During the stampede, the crowd was forced to leave path N due to congestion and diverge through D1 and D2 paths.

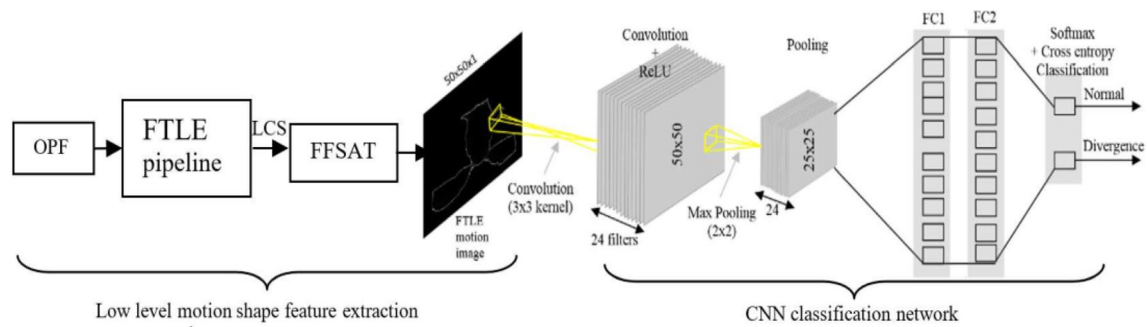
**Fig. 1** Demonstration of crowd divergence from the straight walking pattern at Loveparade 2010: **a** crowd walking under normal conditions with low density, **b** crowd walking paths N1, N2 under normal conditions, **c** high-density crowd within the same region, **d** crowd diverging through paths D1 and D2



**Fig. 2** Example of high-density crowd circular motion (N) with divergence paths (D1 and D2) at the stampede event of the Hindu festival in south India 2015 [9]

In this work, we propose a novel concept of capturing motion shape information in the form of an image and use the images with a CNN classifier for divergence behavior detection. Figure 3 provides a complete pipeline for divergence behavior detection.

There are two key stages of the proposed pipeline: low-level MSI extraction and the CNN-based classification. Crowd motion is estimated by employing the FTLE method. One of the key advantages of FTLE is that it provides clean ridges at the crowd boundaries (also known as Lagrangian coherent structures). LCS ridges are extracted from the FTLE field using a shape extraction scheme known



**Fig. 3** Pipeline of divergence detection using MSI obtained from FTLE and trained with CNN classification network

as FTLE field-strength adaptive thresholding (FFSAT), and grayscale MSI is produced at each FTLE integration time instant. A supervised learning approach is used to train CNN with MSIs to predict the normal or divergent class for an unknown incoming image.

We also propose a divergence localization method that not only indicates divergence starting points but also provides a ‘growing’ divergence mask around the crowd at each video frame. Divergence mask provides divergence size information, and a large size divergence mask can indicate the severity of the crowd situation.

Lastly, we produce two new datasets containing divergence behavior from the straight and circular path at low and high density. The two datasets are synthetic (SYN) and real (PILGRIM and CONCERT) datasets. The SYN dataset is generated using MassMotion crowd simulation software [10] and is a collection of a large number of videos of long durations. The second dataset contains real videos of the high-density crowd normal and divergence behaviors recorded at a CONCERT in Milan, Italy, and PILGRIMS walking in Makkah, Saudi Arabia. To the best of our knowledge, currently, there are no real videos of high-density crowd available with both normal and divergent scenes.

The main contributions of this paper are summarized below: (1) we introduce a notion of physically capturing crowd ‘motion shape’ in the form of grayscale images. Unlike existing arts producing numerical values for low-level features, our method produces MSIs. One of the key advantages of this hybrid approach is that it can be used as plug and play with CNN-based classification, i.e., in the future, any new motion estimation method can be replaced with FTLE, generates MSIs and train CNN for behavior classification. (2) A divergence localization scheme is introduced that not only identifies divergence starting (source) points but also comes with a new feature of generating a ‘localization mask’ around the diverging crowd that varies size each frame according to crowd size. (3) Two new high-density datasets are generated containing scenes of both normal and divergent crowd behaviors. A big SYN dataset

contains high-quality long-duration videos and two real high-density crowd videos capture at different locations. The remainder of this paper is organized as follows: In Sect. 2, relevant works on crowd ME and abnormal behavior detection are provided. In Sect. 3, we discuss our methodology of crowd divergence behavior detection with MSI’s generation from FTLE- and CNN-based classification for divergence behavior. In Sect. 4, experimentation results of divergence detection and localization are provided, and comparisons are made with state-of-the-art existing divergence detection and localization methods on real benchmark datasets. Finally, Sect. 5 concludes our work with a discussion on the limitations of our proposed method with future improvements and research opportunities in this area.

## 2 Related work

Anomaly detection is a blend of two main phases: ME and feature classification. In this work, we mainly concentrate on anomaly detection at the high-density crowd and reviewed the literature in this context. Mainly two types of ME methods are covered in this section for anomaly detection, i.e., OPF and fluid dynamics-based ME methods.

OPF is considered to be one of the most fundamental motion flow models [11–14] that has been widely employed for ME [15–17], crowd flow segmentation [18], behavior understanding [19–21], and tracking in the crowd [22]. Kratz and Nishino [23, 24] model motion in the high-density crowded scene through 3D Gaussian distribution of spatiotemporal gradients (applied on each pixel’s intensity function) that obtained local spatiotemporal motion patterns. Temporal motion is obtained through distribution-based HMM and spatial motion through coupled HMM. Steady-state motion patterns are captured through spatiotemporal relationships of obtained temporal/spatial motion patterns and anomaly in the crowd (pedestrian moving in an irregular direction, individuals obstructing others in indoor subway scene, etc.) is detected as a large statistical deviation

from steady state. This method relies on appearance-based events in the scene for abnormality detection that acts as an outlier for normal scene motion patterns. Unfortunately, results shown are for an indoor crowd, and appearance-based anomalous patterns are not easy to detect in open location high-density crowd scenes, e.g., the high-density crowd at Hajj, etc. Many researchers cluster OPF to obtain motion patterns and anomaly detection. Min Hu et al. [19] group motion vector obtained from OPF neighborhood graphs, and typical motion patterns in crowded scenes are detected by employing a hierarchical agglomerative clustering algorithm. Unfortunately, no results are reported for abnormal behavior detection. Chen and Huang [21] apply adjacency-matrix-based clustering (AMC) extracting orientation and position features and detect anomaly based on the new orientation of crowd appearance. Cong et al. [6, 25] obtain a motion and appearance descriptor from OPF for each patch of the image. For motion descriptors, they employ a multilayer histogram of optical flow (MHOF), and for appearance descriptor, edge orientation histogram (EOH) is used. To obtain motion patterns, they propose a method called dynamic patch grouping (DPG) that adaptively clusters similar image patches into one group based on spatiotemporal information obtained from MHOF and EOH descriptors. Abnormal event detection is performed by measuring the spatiotemporal similarity between a query image and a training dataset using a compact projection method. Wu et al. [26] perform density-based clustering on OPF to obtain local and global coherent motions having arbitrary shapes and varying densities. The approach is named collective density clustering (CDC). Collective density is obtained through the estimation of position and orientation in the coherent motions. These motion features are extracted using the KLT tracker. For local coherent motion detection, collective clustering is performed on collective density estimates. And, for global coherent motion detections, collective merging process is used. However, the method is not tested for anomaly detection in human crowded scenes and is available for traffic flow only (where density cannot exceed certain limits). Majority of OPF ME methods discussed above are efficient for abnormal behavior detection at low- to medium-density crowded scenes, whereas for real-world high-/very-high-density crowd scenes, OPF suffers from various problems like motion discontinuities, lack of spatial and temporal motion representation, variations in illumination conditions, severe clutter and occlusion, etc.

To overcome problems of OPF ME, researchers employ particle advection concepts from fluid dynamics into the computer vision domain [27] and obtain long-term ‘motion trajectories’ under the influence of the OPF field. Particle advection ME methods are termed as Lagrangian or particle flow methods. Lagrangian motion trajectories are proven to be valuable in determining the ‘global’

dynamic structure of the crowd at different temporal scales [28] ignoring pedestrian-level details in the image. Wu et al. [29] employ chaotic invariants on Lagrangian trajectories to characterize crowd motion and extract two chaotic invariant features, maximum Lyapunov exponent and correlation dimension, which measure the extent of particle separation over time and attractor size, respectively. A Gaussian mixture model was used to model chaotic invariant distributions for normal crowd scene, and based on trajectories likelihood, it was determined that either behavior of the crowd is normal or not. They also perform localization of anomaly by determining the source and size of the anomaly. Unfortunately, no results were reported for the high-density crowd. Similarly, Ali et al. [27] obtain Lagrangian coherent structures (LCS) from particle trajectories by integrating trajectories over a finite interval of time termed as finite-time Lyapunov exponent (FTLE). LCS appears as ridges and valleys in the FTLE field at the locations where different segments of the crowd behave differently. Authors perform crowd segmentation and instability detection in the high-density crowd using LCS in FTLE; however, actual anomalies of the high-density crowd like crowd divergence, escape behavior detection, etc., are not performed. Similarly, authors in [30, 31] obtain particle trajectories using high-accuracy variational model for crowd flow and perform crowd segmentation only. Mehran et al. [32] obtain streak flow by spatial integration of streaklines that are extracted from particle trajectories. For anomaly detection, they decompose streak flow field into curl-free and divergence-free components using Helmholtz decomposition theorem and observe variations in potential and streak functions used with SVM to detect anomalies like crowd divergence/convergence, escape behavior, etc. However, results are reported for anomaly detection and segmentation at low-density crowd, and efficacy is still questionable for anomalies at the high-density crowd. Pereira et al. [33] obtain long-range motion trajectories by using the farthest point seeding method called streamline diffusion on streamlines instead of spatial integration. Behavior analysis is performed by linking short streamlines using Markov random field (MRF). However, only normal behavior detection and crowd segmentation results are reported. Although particle flow methods discussed are considered good candidates for ME of the high-density crowd, they are rarely employed for abnormal behavior detection at high-density crowded scenes. Other approaches perform patch grouping-based image denoising to obtain better semantic information [34–37]. However, these methods apply to the low-density crowd scene where background covers a larger area of image and object(s) have less area, and background noise effects are more prominent. At high crowd density, the majority

of the image area is covered by the crowd and the background is almost invisible. Motion estimation methods discussed above obtain motion through OPF. The OPF applies a smoothing process that will filter out noise at high-density crowd and object(s) motion information is easily captured among two frames.

In this work, we propose to estimate motion directly from the crowd motion pattern using the FTLE method and save the crowd motion in the form of MSIs. Figure 4 shows the motion shape obtained by various state-of-the-art ME methods and our approach to motion shape. ME methods used for motion shape at high density include multiple object tracking (MOT) using Histogram of Oriented Gradient (HOG) for object detection and Kalman filtering for tracking [38]; background subtraction algorithm Vibe [39]; OPF method from [11] Social Force Model (SFM) [40]; Streakflow representation of flow (STF) [32] and finite-time Lyapunov exponent (FTLE) [27]. Motion patterns obtained from the above methods are plotted by using respective features and overlaying on the image in Fig. 4: MOT: bounding box is shown as object detected and tracked; Vibe, object mask detected after background subtraction; SFM: the jet colormap is used to overlay detected interaction force ( $F_{int}$ ) over the image; STF: use a grid of color where Hue indicates the streak flow direction, and saturation indicates streak flow magnitude; FTLE: use jet map to display LCS in (FTLE) field. It is clear from Fig. 4 that MOT, Vibe, STF, and SFM features are unable to produce clear motion shapes, whereas LCS ridges in the FTLE field provide a clean motion shape at crowd boundaries (both at low- and high-density scenarios columns left to right). Therefore, in this work, we utilize the FTLE method to obtain crowd motion shape and translate motion shape into a single-channel grayscale image (Fig. 4 second row last image).

### 3 Motion-shape-based divergence detection using CNN

#### 3.1 Low-level FTLE features

Motion shape is extracted from the FTLE field obtained at the crowd boundaries. FTLE computation (performed in the forward and backward time directions) is a three-step process: (1) obtaining flow maps from OPF; (2) calculating Cauchy green tensor (CGT) or spatial gradients of flow maps; (3) performing eigenvalue analysis on CGT to obtain LCS in the FTLE field [27, 41]. The FTLE pipeline is shown in Fig. 5.

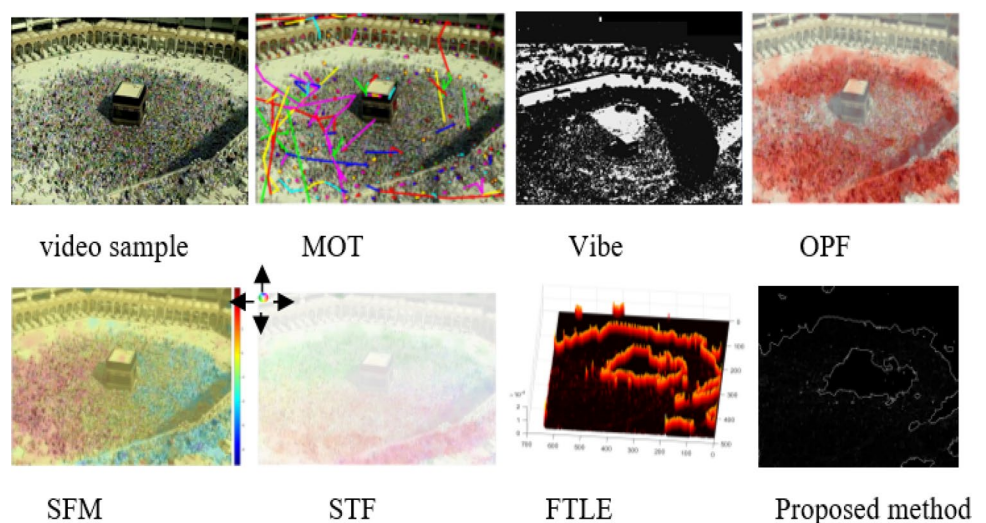
Initially, flow maps (also known as Lagrangian trajectories) are obtained by placing a rectangular grid of particles and performing advection by integrating a system of a differential Eq. (1) subject to initial conditions  $(x_0, y_0)$  with integration interval  $T$ .

$$\frac{dx}{dt} = u(x, y, t), \quad \frac{dy}{dt} = v(x, y, t) \tag{1}$$

where  $u(x, y, t)$  represents optical flow horizontal velocity computed at grid locations  $(x, y)$  and at time instant  $t$ .  $v(x, y, t)$  represents optical flow vertical velocity computed at grid locations  $(x, y)$  and at time instant  $t$ .  $\frac{dx}{dt}$  represents the change in particle position in the x-direction obtained at time instant  $t$ .  $\frac{dy}{dt}$  represents the change in particle position in the y-direction obtained at time instant  $t$ .

The position of particles obtained over time interval  $T$  after the integration Eq. (1) is called a flow map represented by  $(\phi_x, \phi_y)$ . For a given integration interval  $T$ , the flow map is computed in the forward or backward directions. The forward flow map represented by  $(\phi_x, \phi_y)_f$  and backward flow map  $(\phi_x, \phi_y)_b$  is obtained by integrating (1) in the forward direction

Fig. 4 Motion shape at high-density crowds by state-of-the-art ME methods and proposed approach



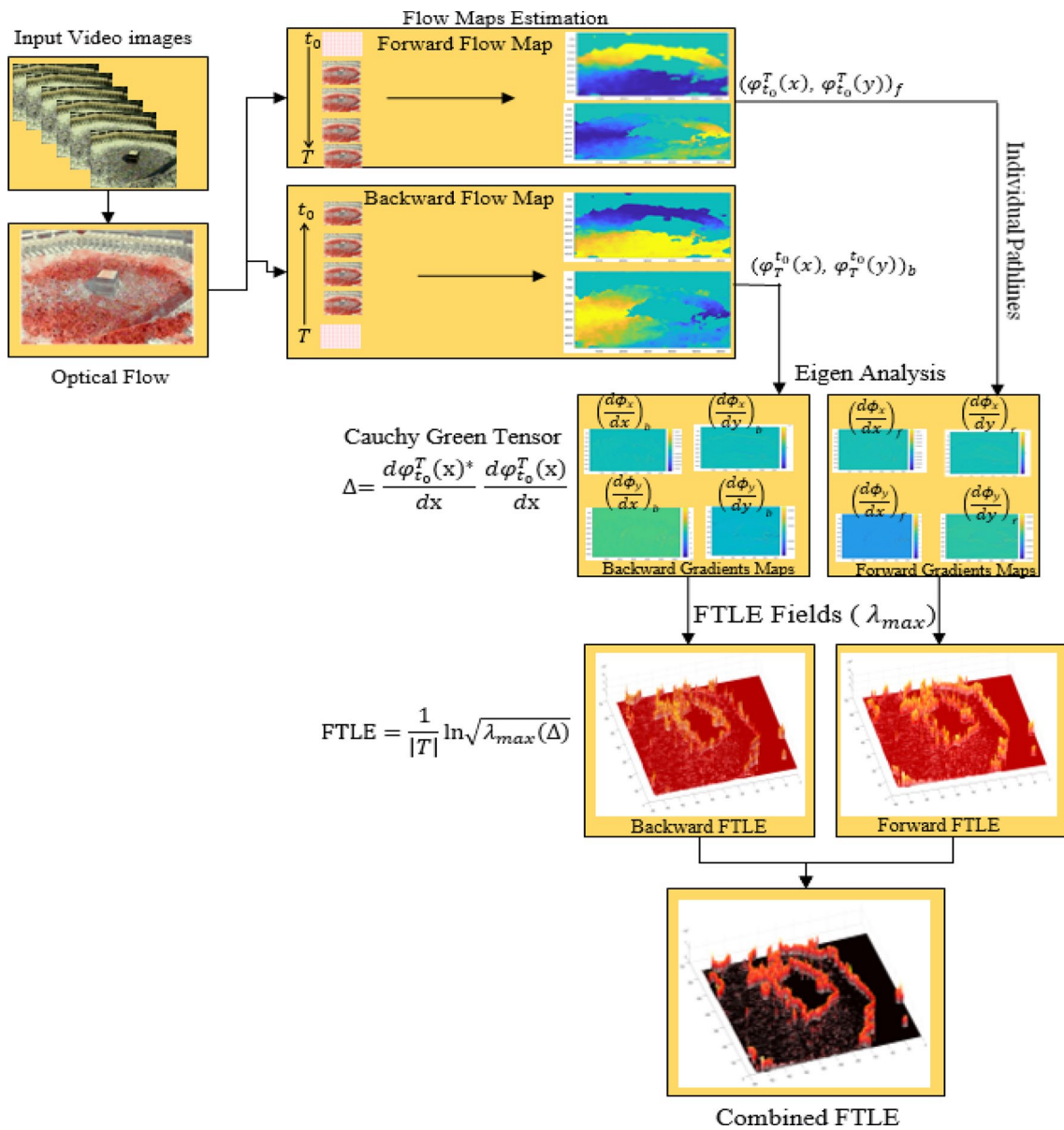


Fig. 5 FTLE pipeline

from  $t_0$  to  $T$  and in the backward direction from  $T$  to  $t_0$ , respectively. In this work, the integration interval  $T=25$ .

The next step is to compute CGT obtained by computing spatial gradients of flow maps, i.e.,  $\frac{d\phi_x}{dx}, \frac{d\phi_x}{dy}, \frac{d\phi_y}{dx}, \frac{d\phi_y}{dy}$  in both forward and backward directions. Finally, eigenvalues analysis is performed on spatial gradients and maximum eigenvalues  $\lambda_{max}$  of gradients are plugged into Eq. (2) to obtain the FTLE field (both in the forward and in the backward directions).

$$FTLE = \frac{1}{|T|} \ln \sqrt{\lambda_{max}(\Delta)} \tag{2}$$

where  $\Delta$  represents spatial gradients of flow maps  $(\frac{d\phi_x}{dx}, \frac{d\phi_x}{dy}, \frac{d\phi_y}{dx}, \frac{d\phi_y}{dy})$ .  $\lambda_{max}$  represents the maximum value in spatial gradients  $\Delta$ .

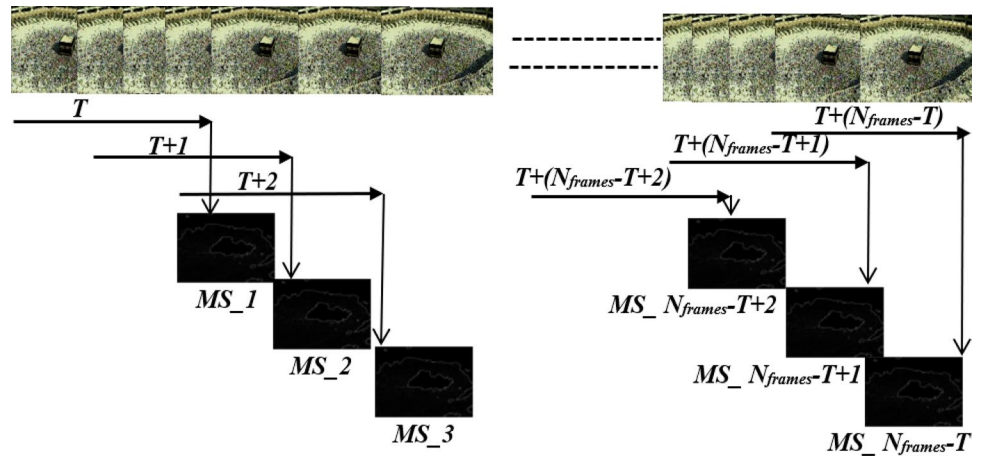
The FTLE field is obtained by adding forward FTLE field ( $FTLE_f$ ) and backward FTLE field ( $FTLE_b$ ). LCS ridges appear in the FTLE field in the regions where two neighboring particles behave differently over the length of integration time  $T$ .

### 3.2 Motion-shape extraction

Motion shape is produced at every integration step of the differential Eq. (1) as shown in Fig. 6. Crowd motion shape is obtained by extracting LCS ridges from the FTLE field and converted into a grayscale image. Various methods exist in literature for LCS extraction [42, 43]. In this work, a simple FTLE field-strength adaptive thresholding (FFSAT) scheme is developed for LCS ridge extract. At every integration step,

maximum Eulerian distance ( $d_{max}$ ) is calculated between LCS absolute peak value and average FTLE field strength, and a threshold ( $ffsat\_thr$ ) is set for  $d_{max}$ . LCS values crossing  $ffsat\_thr$  are extracted and converted into a single-channel grayscale image. FFSAT algorithm ensures only strong magnitude LCS values from the FTLE field are extracted and noise is filtered out. Figure 7 shows two examples of crowd motion shapes for synthetic Loveparade and Kabbah scenarios.

Fig. 6 MSI generation at every integration step of Eq. (1)



Scenario	Video Sample	LCS in FTLE field	Extracted FTLE ridges	Gray-scale Motion image
Loveparade	Normal			
	Divergent			
Kabbah	Normal			
	Divergent			

Fig. 7 Examples of MSIs obtained for synthetic datasets of Loveparade and Tawaf around Kabbah

### 3.3 Divergence detection using CNN

It is observed that motion shapes of normal and divergent crowd behaviors are mostly identical and differ mainly at crowd divergence regions. Hence, the relationship among normal and divergent motion shapes can be best described through the ‘convolution process,’ i.e., positive peaks of convolution field indicate the regions where both normal and divergent MSIs overlap or are the same, whereas the negative peak of convolution field will indicate the regions where only divergence shape exists. Analysis of motion shape through convolution is achieved by implementing a convolution neural network (CNN), shown in Fig. 8, using a single convolution layer. The MSI is rescaled to  $50 \times 50$  pixels at network input. Convolution layer uses convolution filters (24 filters) with rectified linear unit (ReLU) activations.

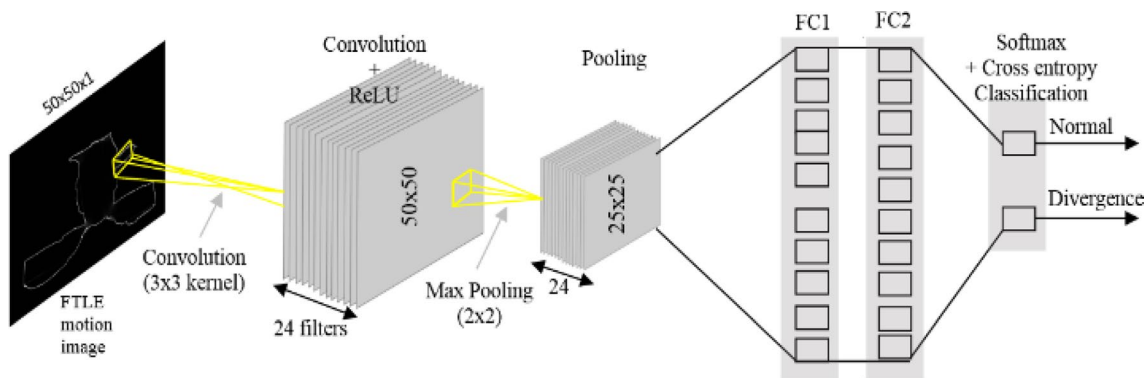
The purpose of using many convolution filters is to ensure all important receptive fields of CNN are excited about a given motion shape. ReLU is adopted as the activation function because of its good performance for CNNs [44], and

Max pooling is used for each  $2 \times 2$  region. Complete configuration of the proposed network with activation shape and learnable parameters is provided in Table 1. Representative layers’ activation plots are shown in Fig. 9, where the top row of Fig. 9 shows a feature map of the convolution layer (24 filters outputs), the middle layer shows 24 ReLU activations output, and the last row is a feature map obtained from max pool outputs.

### 3.4 Divergence localization

The block diagram of the divergence localization framework is shown in Fig. 10. Our divergence localization scheme is simple; it takes the difference of incoming MSI from a reference motion shape. The reference motion shape is obtained by saving the incoming motion shape at the time instant the divergence detection is indicated by the CNN network, as shown in Fig. 10.

However, motion shapes obtained at every integration step are not similar and exhibit both local and global



**Fig. 8** CNN architecture for normal and divergence detection from motion images

**Table 1** Divergence behavior detection of CNN configurations

No.	Name	Activations	Learnable	Total learnable parameters
1	Image input $50 \times 50 \times 1$ images with ‘zero-center’ normalization	$50 \times 50 \times 1$	–	0
2	Conv 24 $3 \times 3 \times 1$ convolutions with stride [1 1] and pooling [1 1 1 1]	$50 \times 50 \times 24$	Weights $3 \times 3 \times 1 \times 24$ Bias $1 \times 1 \times 24$	240
3	Batchnorm batch normalization with 24 channels	$50 \times 50 \times 24$	Offset $1 \times 1 \times 24$ Scale $1 \times 1 \times 24$	48
4	ReLU	$50 \times 50 \times 24$	–	0
5	Maxpool $2 \times 2$ maxpool with stride [2 2] and padding [0 0 0 0]	$25 \times 25 \times 24$	–	0
6	Fc 2 fully connected layers	$1 \times 1 \times 2$	Weights $2 \times 15,000$ Bias $2 \times 1$	30,002
7	Softmax	$1 \times 1 \times 2$	–	0
8	Classout Crossentropyex with classes ‘normal’ and ‘divergence’	–	–	0



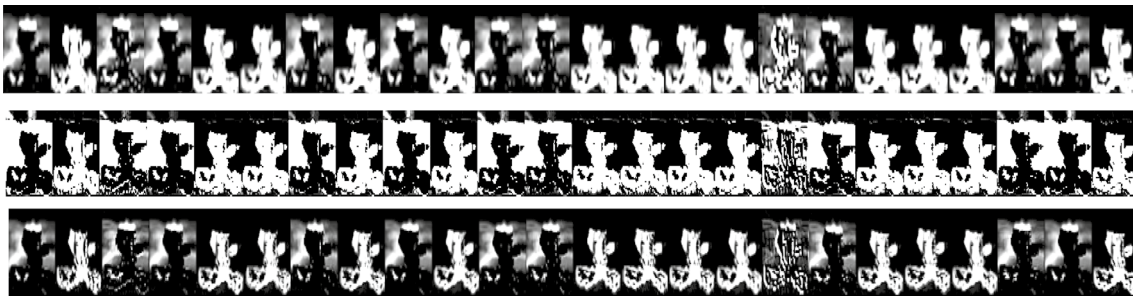


Fig. 9 First row: convolution layer activation map, second row: ReLU activation map, third row: max pool activation map

Fig. 10 Framework for divergence localization

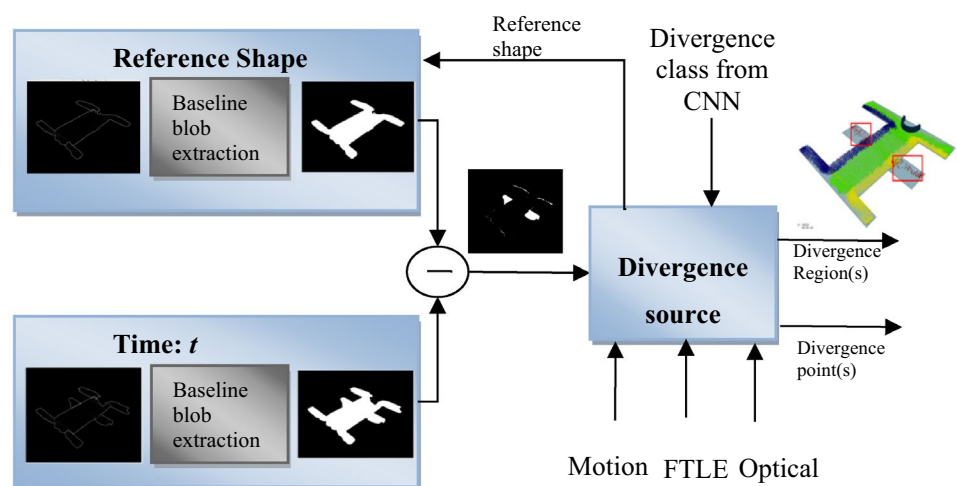
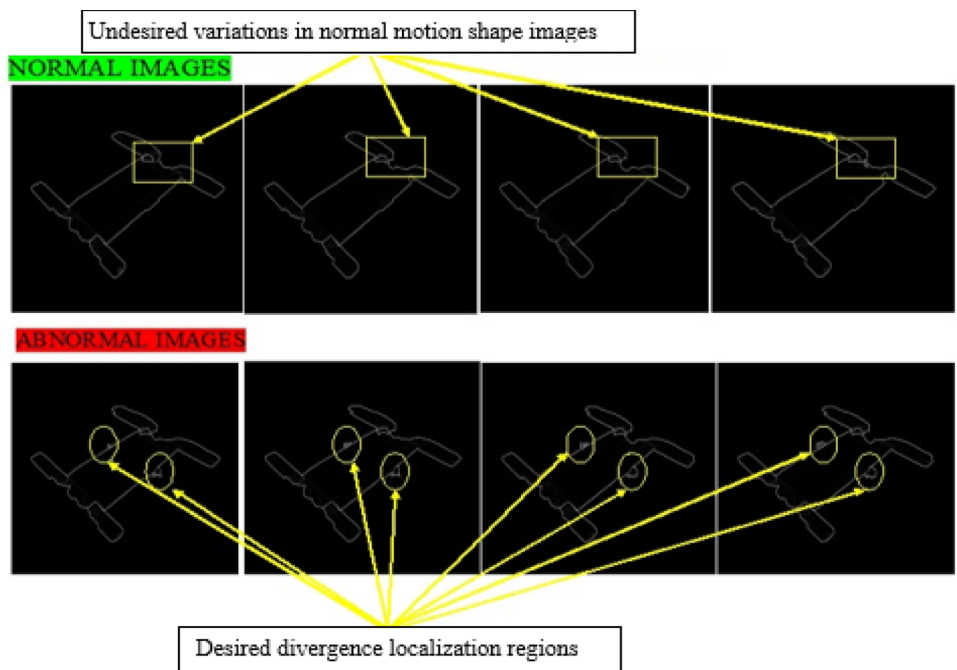


Fig. 11 Top row: undesired motion-shape variations due to crowd oscillatory motion; bottom row: real-shape change due to divergence



shape variations. Figure 11 showing samples of shape variations for normal (top row) and divergence (bottom row) behaviors for the crowd at Loveparade scenario (crowd on the ramp entering from left and right tunnels and heading toward ramp top for exiting). Regions marked with squares in Fig. 11 (upper row) are motion-shape variations experienced during normal behavior frames. Simply differencing normal and divergence motion shapes (in the presence of shape variations) produces many undesired blobs that are not actual divergence sources and can lead to false divergence-source detections. As mentioned above, shape variations occur locally and globally at the high-density crowd. Local motion shape variations are due to to-and-fro motion [45], and global crowd motion-shape changes are due to the segment(s) of the crowd that become ‘stationary’ and result in no FTLE field. For the former problem, the crowd naturally starts oscillating left and right at their central axis, generating crowd waves. As the waves reach boundaries, the crowd naturally expands and shrinks. As a result, the FTLE field at crowd boundary also expands/shrinks causing local (minor) motion-shape variations.

Global motion-shape variation occurs when a certain crowd segment stops motion and crowd velocities reduce to zero; results are zero OPF/ FTLE field in that region. Thus, LCS ridges in the FTLE field disappear from stationary segments and appear at other moving segments, causing global motion shape variation. We address the shape variation problems discussed above by implementing a blob preprocessing pipeline (Fig. 12) before differencing motion shapes.

Pipeline in Fig. 12 extracts ‘baseline blob’ for normal and divergent behaviors. Subtraction of divergent baseline blob from reference (normal) baseline blob not only generates blob(s) representing actual divergence region(s) but also smaller residual blocks act as noise and require filtering to avoid false divergence source detections. Residuals are filtered out by applying temporal and spatial filtering on OPF and FTLE fields on current and past  $N_{\text{filter}}$  frames. Blob(s) obtained after the filtering process is(are) marked as divergence source region(s).

Divergence source points are also extracted from the divergence region detected. The contour of the divergence region is obtained, and pixels distance from contour to baseline motion shape blob is computed. The distance of pixels less than a threshold is marked as divergence source points. Hence, our algorithm not only identifies divergence source

points but also divergence region masks representing the size of divergence. It is important to mention here that existing state-of-the-art divergence localization algorithms [7, 8, 46] only identify divergence source points or starting points, and later in the video there is no update on localization information, i.e., where divergence leads to over the time and how big is the divergence size, whereas our divergence localization algorithm not only identifies divergence source points but also masks region of divergence that evolves temporarily in the video. Divergence mask provides useful information about the severity of divergence, i.e., the smaller mask would mean less or low-density crowd divergence, whereas a larger divergence mask would represent high-density crowd diverging. Another difference with the state-of-the-art divergence method is that there is no divergence direction information in the existing art as they only provide divergence starting point information, whereas our method shows the direction divergence is progressing. Divergence direction information can help to deploy rescue efforts at the right locations of disaster.

### 3.5 Crowd datasets

#### 3.5.1 Previous datasets

In this section, a review of benchmark crowd anomaly datasets is provided containing crowd divergence sequences. **UCF** dataset [47] contains a zebra crossing video where the crowd going through zebra crossing with the expected normal behavior of the crowd remains over the zebra crossing area. However, near the ending of the video, a group of people start walking off the zebra crossing creating a divergence pattern. Overall crowd density in the video is low. **UMN** dataset [48] consists of three sequences (two recorded outdoor and one indoor) with a total of 11 different activities. Initially, the crowd moves randomly within a certain region and then suddenly disperses in all directions. In this case, crowd dispersion is taken as divergence behavior. These scenes also consist of low crowd density. In **PETS2009** [49] dataset, sequence 2 contains crowd divergence scenes where a group starts walking left off the normal straight walking path on the road. Videos are recorded with four different camera views. Crowd density in the PETS dataset ranges from low to medium. **NGSIM** dataset [50] consists of real CCTV surveillance videos monitoring traffic mounted at a

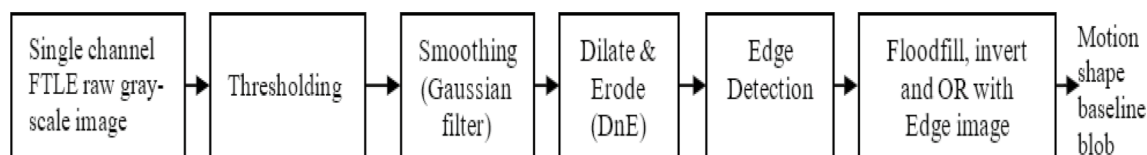


Fig. 12 Baseline blob extraction pipeline for normal and divergence behaviors

pole with top view. Various car divergences are available in different directions; however, the number of cars on road constitutes medium density. Overall, the existing video dataset contains crowd divergence at low- to medium-density crowd (objects) and the duration of video clips is relatively short. Existing video datasets limitations compel us to generate a new rich dataset containing divergence scenes at the (very) high-density crowd. Details of our new dataset are discussed in the next subsection. Table 2 provides a comparison and properties of existing video datasets and the last column showing crowd normal walking pattern, i.e., *straight* or *circular*, as our focus is on divergence detection from these two types of crowd motions.

### 3.5.2 Our dataset

A new large-scale dataset is developed to evaluate our method under realistic high-density crowd divergence scenarios. Mainly we construct three datasets: (1) synthetic (SYN) dataset, (2) MELAN CONCERT dataset and (3) PILGRIMS dataset. SYN dataset is generated using MassMotion crowd simulation software [10], and videos of the low- and high-density crowd are produced with normal and divergent behaviors. Low-density synthetic videos contain the same behaviors of real datasets such as PETS2009,

UMN, UCF, and NGSIM. The length of each rendered video is 20 min. For high density, we design locations of Holly mosque of Masjid Al-Haram (Kabbah) and Loveparade in Massmotion. In the Kabbah scenario, agents performing Tawaf around Kabbah are considered normal behavior, and the agents leaving Tawaf after completing seven laps are considered diverging from the normal walking crowd. Since Tawaf is circular walking behavior and theoretically divergence can happen at any angle in 360 degrees, we generate 25 different videos with crowd diverging every 14 degrees on the circular path. Each video length is 20 min, and the number of agents varies from 10 K (starting) to 50 K (during divergence). Samples of the Kabbah dataset and FTLE images with divergence from different locations are shown in Fig. 13, where a rectangular area appearing at the circle edge presents a crowd divergence region.

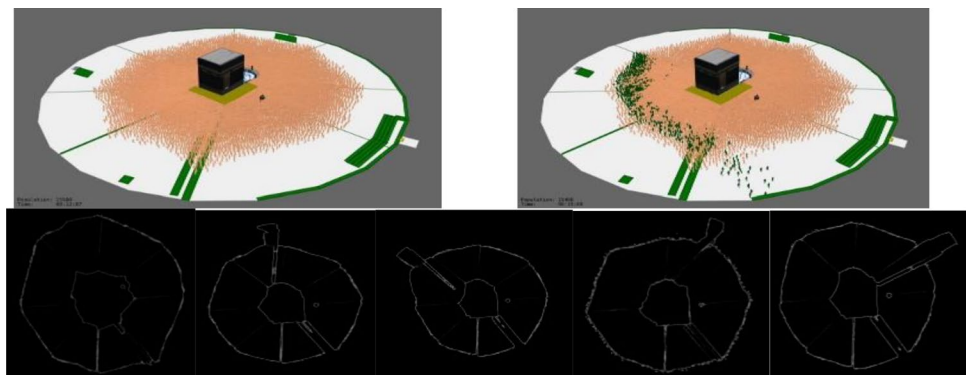
Synthetic Loveparade scenario videos consist of real stampede disaster events. Just before the stampede occurred, crowd conditions at ramp top were very critical, crowd motion was stopped due to congestion and there was no more space for the new incoming crowd at the musical band's performance area. To avoid congestion, the new entering crowd starts diverging left and right off the ramp toward safe areas. We model similar conditions and crowd divergence off the ramp in Massmotion. In the real scenario, various

**Table 2** Video datasets comparison (existing benchmark and ours)

Dataset	No. of frames		No. of scenes	Resolution	Frame rate	Density level	Type (Real/synthetic)	Normal motion type
	Normal	Abnormal						
UCF	110	140	1	480×360	25	Low	Real-ped	Straight
UMN	500*	100*	11	320×240	30	Low	Real-ped	Circular
PETS2009	52*	50*	8	768×576	7	Low/medium	Real-ped	Straight
NGSIM	250*	241*	8	640×480	10	Low	Real-traffic	Straight
OUR-SYN	10,000*	10,000*	100	1280×720	24	High	Syn-ped	Straight/circular
OUR—MELAN CONCERT	553*	553*	4	1920×1080	50	High	Real-ped	Straight
OUR—PILGRIM	1500*	1500*	3	1280×720	25	High	Real-ped	Straight

\*Number of frames averaged for the respective number of scenes (normal or abnormal)

**Fig. 13** Synthetic Kabbah dataset: top row—normal (circular) and divergent behavior sample scenes; bottom row—samples of FTLE images with crowd diverging from different locations of the circular motion

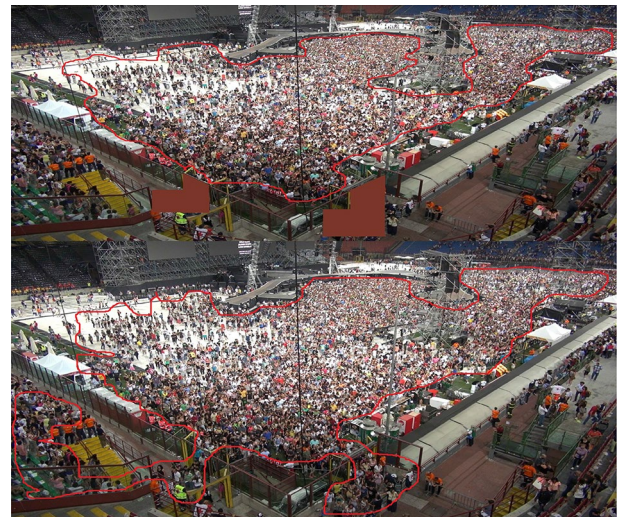
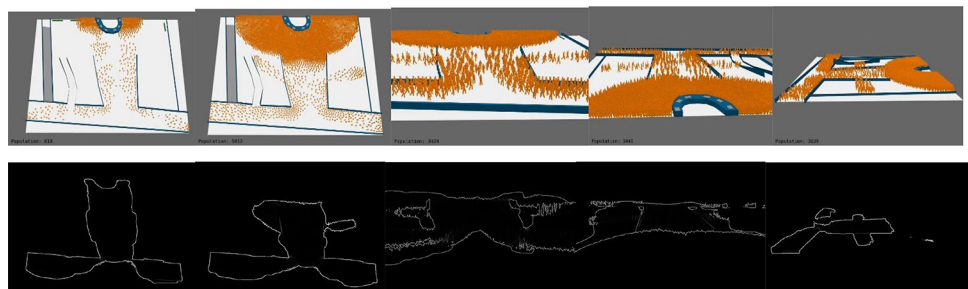


CCTV cameras were installed at different locations covering the same crowd motions with different camera angles. We mimic similar camera arrangements in MassMotion and generate crowd videos views and angles. Examples of different camera view videos rendered in MassMotion are shown in Fig. 14 together with corresponding FTLE shape images. Also, for dataset diversity, agents in the software are made to diverge from different ramp locations (15 divergence locations) and 15 divergence videos are generated. Agents' density is varied from  $\sim 15$  K (under normal conditions) to  $\sim 65$  K (during divergence). The average length of each video is 20 min.

**MELAN CONCERT** dataset is captured with a fixed camera mounted at a height (with a top view). The high-density crowd coming from the concert area is exiting through two gates. Here, crowd exiting through gates is considered divergence behavior. Video length is one minute. We generate three divergence videos out of the single video. Normal crowd behavior video is generated by blurring exit gate areas. For crowd right divergence, the left exit gate is blurred, and for left crowd divergence, the right-side exit gate is blurred. Finally, the crowd exiting from both gates is considered a dual divergence simultaneously. Samples of the MELAN dataset with FTLE shapes superimposed on actual images are shown in Fig. 15.

**PILGRIMS** dataset shows pilgrims walking in a Y-shape path (Fig. 16). Video is recorded from the Makkah TV channel live broadcast for one minute and then the camera moves to another scene. We generate three different behavior videos from a single video. The crowd toward the left direction is blurred and the only crowd walking right side is considered as normal crowd behavior. Similarly, a crowd walking right is blurred and the crowd flowing left side crowd is taken as normal behavior. For both normal behaviors, the actual video serves as a divergence scenario. The summary of our large-scale crowd dataset is provided in Table 2.

**Fig. 14** Loveparade synthetic video dataset: 1st column—top view normal behavior (straight motion); 2nd col. Top view divergence left (L) and Right (R); 3rd col. K13 view LR divergence; 4th col. K5 view LR divergence; 5th col. K12 view LR divergence from straight crowd motions



**Fig. 15** MELAN Concert dataset samples: in the first image, both gates are blurred and crowd normal behavior is straight motion pattern with crowd exiting from the stadium and approaching exit gates (crowd normal motion shape is superimposed on the image); in the second image, crowd diverge from both gates, crowd divergence motion shape is superimposed on the image

## 4 Experimentation

### 4.1 Divergence detection

#### 4.1.1 Experimentation setup

We compute the dense OPF using the Brox method [12]. Integration time  $T$  in the forward /backward FTLE computation is set to 25.  $ffsat\_thr$  in the FFSAT scheme is set to 65% of  $dmax$  for motion shape extraction. The extracted motion shape is resized to  $50 \times 50$  pixels and saved for model training. Our model training is not performed on a raw image dataset, instead of on small size ( $50 \times 50$ ) MSIs (hence reduced 4 GB raw image dataset to  $\sim 500$  MB). Data is divided into 70–10–20 configurations where 70% of random data is used for training, 10% for validation, and 20% for testing. We train the CNN model on a core i5 processor with 8 GB system memory. With a small yet diverse



**Fig. 16** PILGRIM dataset: the first image crowd left normal (straight walking pattern); the second image right normal (also straight walking pattern); the third image is the divergence for both left and right normal behaviors. Corresponding crowd motion shapes superimposed on images

image dataset of motion shapes, we were able to complete model training in less than 1.5-h on a normal CPU without GPU support. Other training settings include no. of epochs = 150 and 100 batches in each epoch. We apply stochastic gradient descent with momentum (SGDM) optimizer with a learning rate of 0.01. Model output scores for normal and divergent classes with values lie between 0 and 1.

We compare our divergence detection method with state-of-the-art divergence detection methods including a Bayesian Model (BM) for escape behavior detection [7], the method based on chaotic invariants (CI) [46], the method using curl and divergence of motion trajectories (CDT) [51] and streak flow (STF) method [32]. Experiments are run at five datasets discussed in Sect. 3.5.2. We perform both qualitative and quantitative evaluations in this section following the protocol shown in Table 3.

**Table 3** Divergence behavior detection experimentation protocol

Method	Features (quantitative evaluation)	Classifier (qualitative evaluation)
Bayesian model (BM)	Mean optical flow velocity magnitude	Bayesian classifier
Chaotic invariants (CI)	Representative trajectories	Gaussian mixture model (GMM)
Curl and divergence of motion trajectories (CDT)	Divergence descriptor	Support vector machine (SVM)
Streak flow	Velocity potential function	Support vector machine (SVM)
Ours	Motion shape	CNN

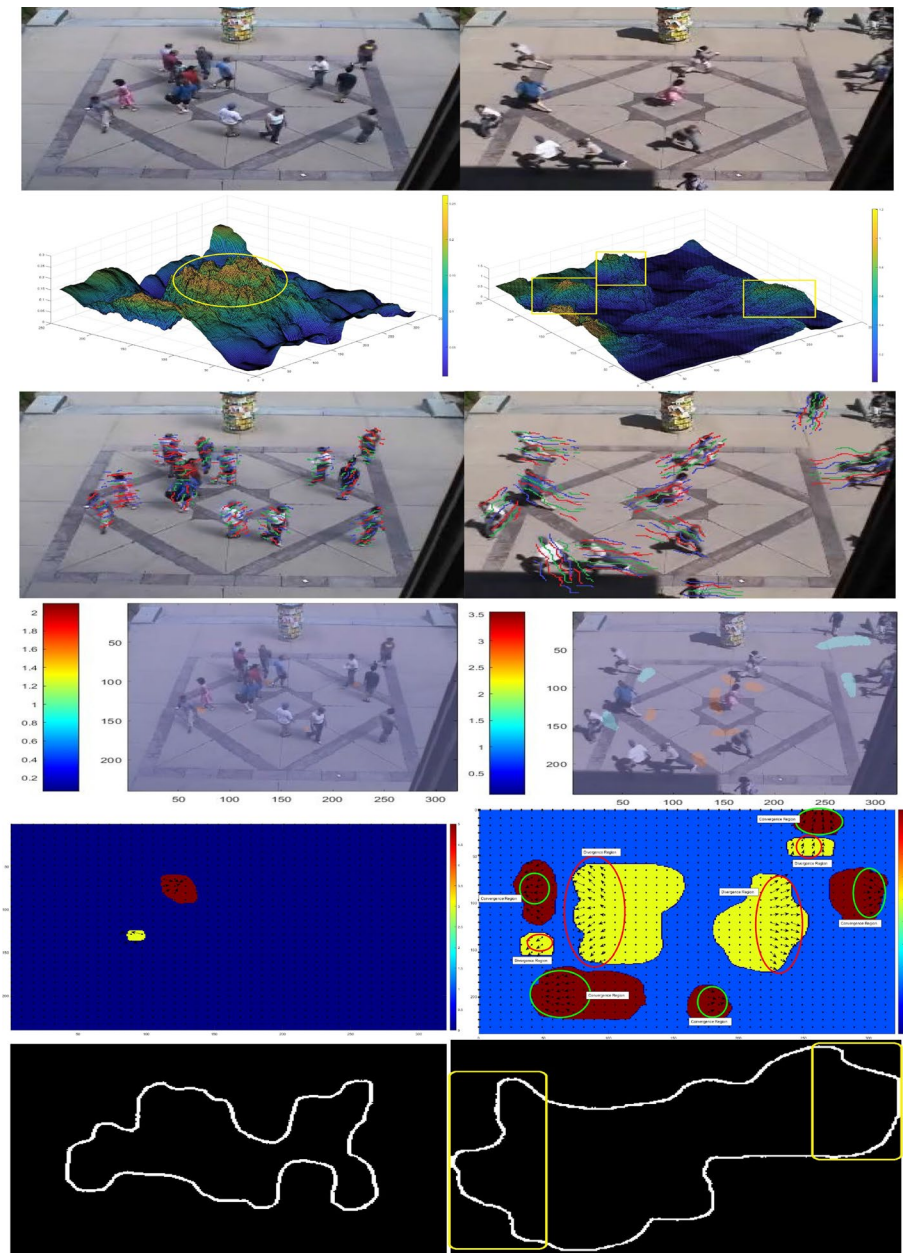
#### 4.1.2 Qualitative evaluation

In qualitative evaluation, we determine how discriminating a method's feature is for normal and divergent crowd behaviors. It is well known that a classifier performance heavily reliant on the input features and robust features (that are strongly discriminative) ensures efficient classification results.

We extract low-level features mentioned in Table 3 at two crowd density levels (medium and high) for both normal and divergent crowd behaviors. Figure 17 shows the features of each method plotted for normal and divergent behaviors at a medium-density crowd scene taken from the PETS2009 dataset. It is obvious from Fig. 17 that existing art and our method perform well on a medium-density divergence scenario. Velocity magnitude features of the BM method are different in this scenario for normal and divergent behaviors. The concentration of velocity magnitude is more at the center in the normal scenario, whereas velocity magnitude is scattered in divergence behavior and concentration is more at outer locations of the scene (Fig. 17 second row). The representative trajectories of the CI method are quite apparent in the divergent scenario and can be easily differentiated from normal behavior where representative trajectories are not significant. The CDT method can efficiently represent no/less divergent regions at the crowd normal behavior and capture most possible divergent regions at the abnormal behavior. Similarly, STF methods' velocity potential can also efficiently capture divergent regions of abnormal behavior with minor detections of the divergent region at normal behavior. Finally, our motion shape for normal and divergent behaviors is significantly different and can be easily interpreted by the classifier as a normal or divergent scene.

Low-level features' comparison for high-density crowd scenario is shown in Fig. 18. It is important to notice that existing methods rely on inter-person distance for their motion features to be strongly discriminative. For example, the distribution of velocity magnitude in the BM method creates the difference between normal and divergent behaviors. This distribution requires inter-person distance to increase so that velocity magnitude is uniformly distributed over the scene. But at the high-density scenario, where crowd density remains the same for normal

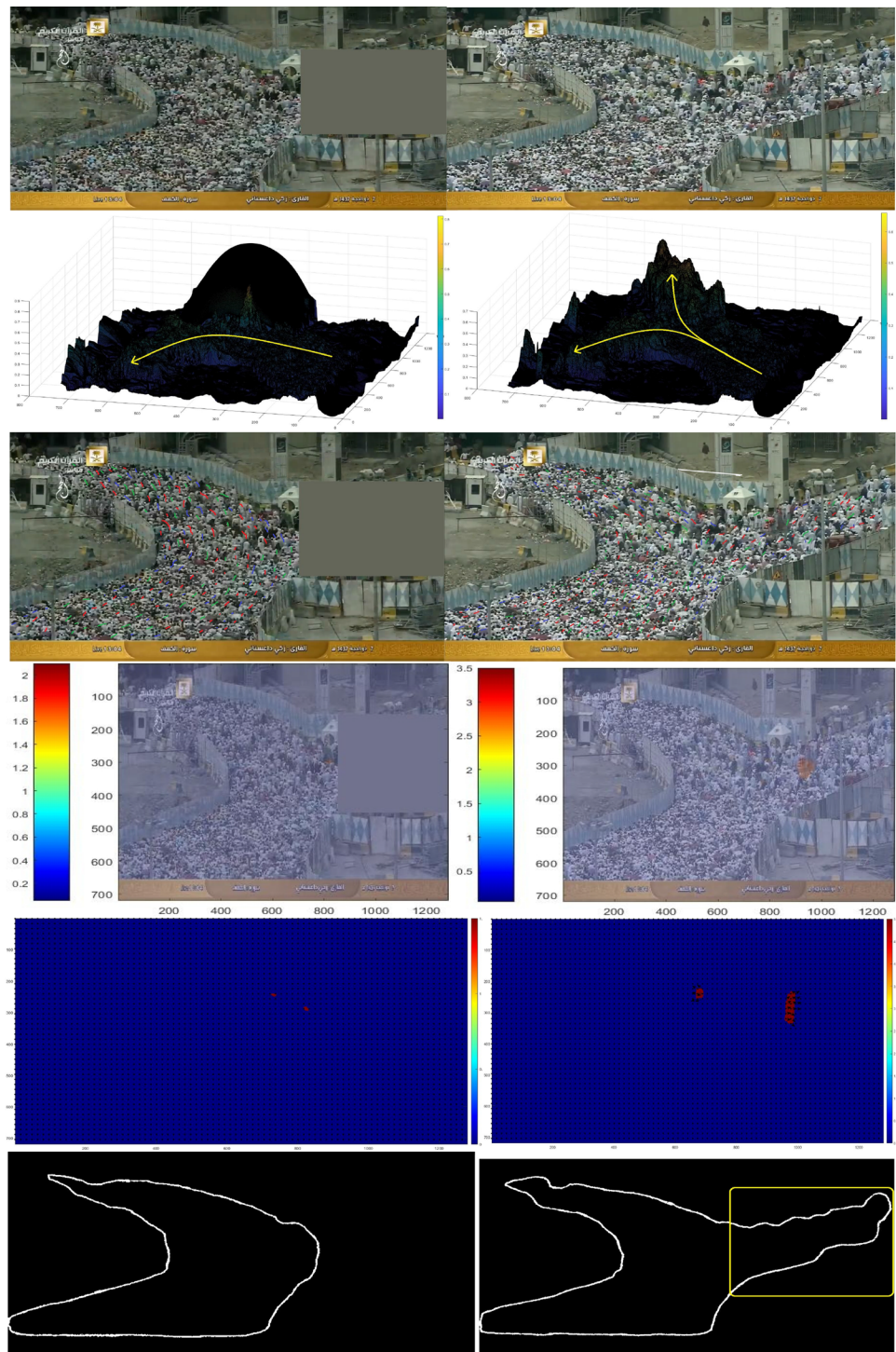
**Fig. 17** Qualitative comparison of low-level features for normal and divergence behavior at medium-density crowd. Row 1 first image is a sample of normal behavior and the second image is a sample of divergence behavior from the UMN dataset. Second row is the respective velocity magnitude maps by BM method; the circular area in the first image shows that velocity magnitude distribution is concentrated at the center for normal behavior and the second image shows that velocity is distributed across the scene for divergence behavior with more concentration toward corners. Third row shows representative trajectories from the CI method with the first image contain less significant representative trajectories in normal scene, whereas significant trajectories in divergence scene. Fourth row is the divergence descriptor map from the CDT method where low values (in blue) represent divergence areas and high values of the descriptor (in red) represent convergence areas. Fifth row is the velocity potential map from the STF method. High values of the potential function (red) represent convergence regions, and low values of velocity potential (yellow) represent divergence regions. Sixth row is the motion shape from our proposed method with clear shape difference between normal and divergence behaviors



and divergence behaviors or even increase in case of divergence, the method performance degrades significantly as the velocity features are no longer discriminative. The situation is shown in Fig. 18 second row, where the distribution of velocity magnitude is almost similar in both normal and divergent scenarios as crowd density remains high in both cases. Similarly, representative trajectories of the CI method are not significant in divergence scenarios and are like the normal scenario (Fig. 18 third row). Both CDT and STF also suffer from the same problem and could partially perform divergence behavior detection by locating the divergence regions at the crowded places containing gaps. However, if similar gaps are present in normal

behavior regions, then these methods could easily lead to false detection. Finally, the motion shape features from our method are significantly discriminative for normal and divergence behaviors compared to the existing art and can easily detect crowd divergence behavior. Since our motion shape is extracted from LCSs in the FTLE field and the FTLE method cleanly generates LCS ridges at the crowd boundaries, in turn, motion shapes are significantly different for normal and divergence scenarios. This also shows that most existing methods that were efficient in ME at low/medium density degrade their performance significantly for the ME at the high-density crowd.

**Fig. 18** Qualitative comparison of low-level features for normal and divergence behavior at the high-density crowd. Row 1 first image is a sample of normal behavior and the second image is a sample of divergence behavior from our PILGRIM dataset. Second-row first image shows velocity magnitude is uniform across the scene on a crowded area and the second image shows for the divergence scenario, the magnitude map is also uniformly distributed across the crowded region, and there is no significant difference among magnitude maps for both behaviors. Third row shows representative trajectories by the CI method are not significant for both normal and divergence crowd behaviors. The fourth-row second image shows partial divergence behavior detections by the CDT method by locating divergence regions at the gaps in high-density crowd and similar behavior is shown by the STF method in the 5th-row second image. Sixth row shows motion-shape features by our method are significantly discriminative for normal and divergence behaviors at the high-density crowd



**4.1.3 Quantitative evaluation**

**4.1.3.1 Evaluation metrics** In quantitative evaluation, we test how well a classifier performs using discriminative features discussed in the previous subsection. We treat divergence behavior detection as a binary classification problem and accuracy (ACC); the basic evaluation metric is used to

evaluate the classification performance. However, ACC is measured at a single cutoff (threshold) of class output probabilities. In our case, there are cases where divergence-like shape also appears in normal crowd behavior, and it would be important to investigate whether the model classifies test normal image as normal or divergent behavior. Therefore, we sweep the full probability range [0, 1] by setting differ-

ent thresholds and obtain receiver operating curves (ROCs). ROC is computed from true-positive rate (TPR) and false-positive rate (FPR). TPR corresponds to the proportion of divergent image samples correctly classified as divergent w.r.t all divergent image samples ( $TPR = TP / (FN + TP)$ ). FPR corresponds to the proportion of normal behavior image samples that are mistakenly considered as divergent w.r.t all normal behavior image samples ( $FPR = FP / (TN + FP)$ ). Finally, a classifier with the highest area under ROC (AUC) value most efficiently predicts the divergent class.

High ACC and AUC scores may be misleading for imbalanced datasets, unfortunately, that is the case with the majority of real benchmark videos. A large portion of videos contains normal behavior images, whereas anomaly (divergence) exists only for a short duration. It is clear from Table 2 benchmark datasets (UCF, UMN, PETS2009, and NGSIM) are skewed toward crowd normal behavior and are imbalanced. For classifier evaluation on imbalanced datasets, we compute precision (P) and recall (R) and obtain PR curves and corresponding area under the PR curve (AP). Precision and recall are defined in Eqs. (3) and (4), respectively.

$$\text{Precision} = \frac{(\text{div. sample correctly identified})}{(\text{div. sample correctly identified} + \text{normal sample incorrectly identified as divergent})} \tag{3}$$

$$\text{Recall} = \frac{(\text{div. sample correctly identified})}{(\text{div. sample correctly identified} + \text{divergence sample incorrectly identified as normal})} \tag{4}$$

Precision is a good measure to determine when the cost of normal samples incorrectly identified as divergent is high. Similarly, recall is a good measure to determine, when the high cost is associated with divergence image samples that are incorrectly identified as normal. Precision and recall have an inverse relationship, and to find the optimal balance among the two, we compute the F1 score (that is the harmonic mean of PR) as given in Eq. (5).

$$F1 = 2 \times \frac{(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}} \tag{5}$$

where the F1 score of 1 indicates the best optimal balance between P and R and 0 indicates no balance. Also, F1 score provides a better measure of incorrectly classified cases (false positive—normal classified as divergent and false negative—normal classified as divergent) compared to ACC that only focuses on true positives and true negatives.

**4.1.3.2 Comparison with state-of-the-art methods** In this experimentation, we train all methods on our large synthetic dataset containing scenes of divergence at low- as

well as high-density crowds, and evaluation is performed on real and synthetic datasets.

Figure 19 compares the accuracy on various datasets, where SYN-KAB stands for synthetic Kabbah and SYN-LVP stands for synthetic Loveparade datasets. It is clear from the bar chart that for low-density crowd datasets (UCF, UMN, PETS, and NGSIM), our method performs well compared to existing art, while our method also outperforms others in divergence detection at high-density crowd datasets (MELAN, PILGRIM, and synthetic).

ROC comparison of methods at selected datasets is shown in Fig. 20. It can be seen that from ROC plots and AUC score in Fig. 21, our method outperforms existing art in predicting divergent class not only at the low-density crowd but also divergence at the high-density crowd. Noticeably for high-density datasets (synthetic, MELAN, and PILGRIM), other methods lost the capability of separating divergent and normal classes, whereas in our methods ROC curve is still in the upper region close to the top left corner and shows it can efficiently separate divergent and normal classes. The fact is also evident from the high AUC score of our method compared to existing art in Fig. 21.

PR curves comparison is shown in Fig. 22, and the corresponding area under the PR curve (AP) bar chart is shown in Fig. 23. PR curves show that our model behaves reasonably well at low-density imbalanced datasets (UCF, UMN, PETS, and NGSIM) specifically on the UMN dataset where few existing art methods’ performances are dropped significantly.

Lastly, a comparison of the F1 score is provided in Fig. 24. It is clear from the figure that our method achieves

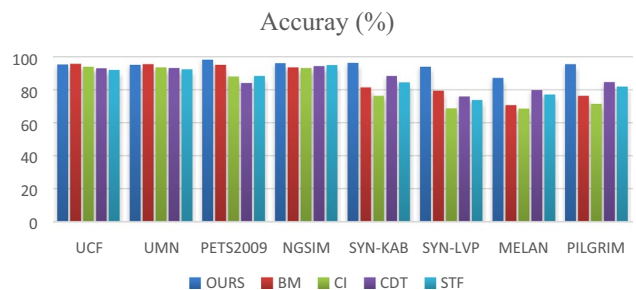
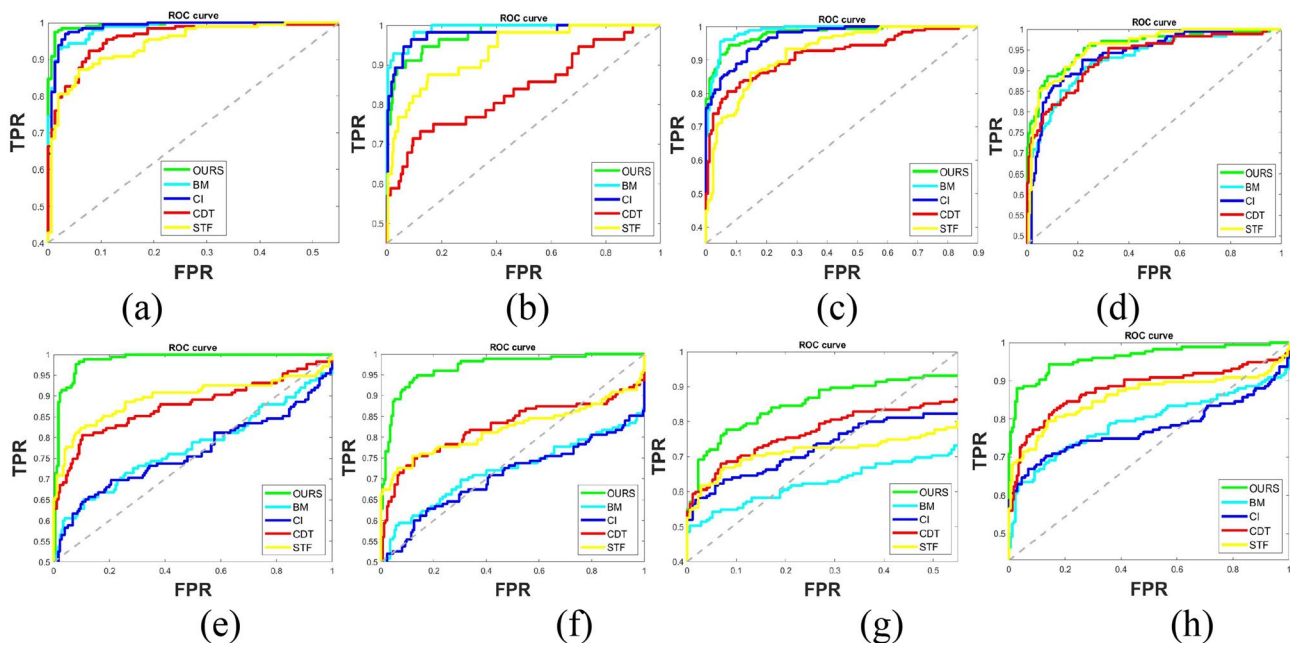
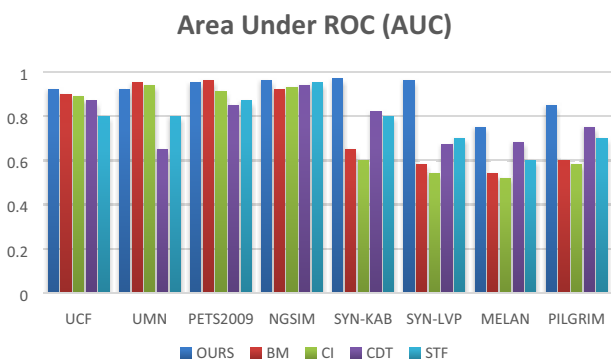


Fig. 19 Accuracy comparison for divergence detection on selected datasets





**Fig. 20** ROC curves for divergence behavior comparing with state-of-the-art methods on low- and high-density crowd datasets: **a** UCF, **b** UMN, **c** PETS2009, **d** NGSIM, **e** synthetic—kabbah, **f** synthetic Loveparade, **g** MELAN concert, **h** PILGRIM



**Fig. 21** AUC score comparison—higher the AUC score, the better the method performs in divergence detection

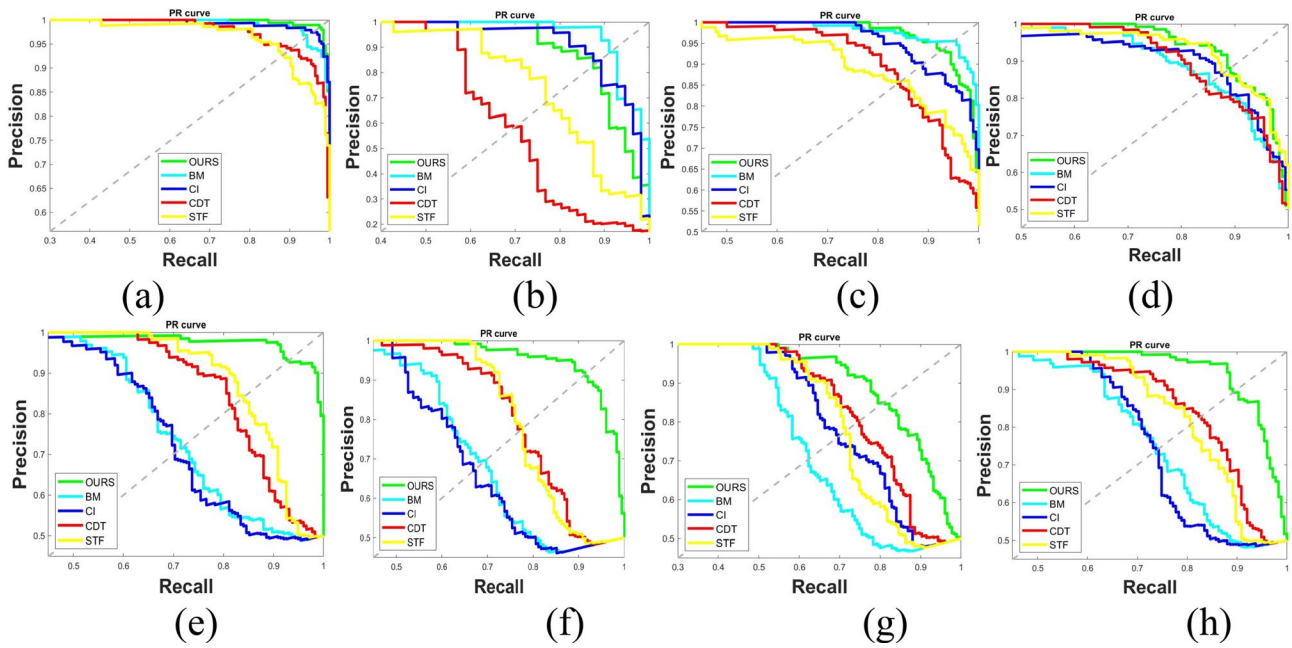
high F1 scores at all datasets compared to existing art methods. It shows that our method achieves the best optimal balance between precision and recall. It is also clear from Fig. 24 that our method better able to classify normal and divergent classes even at imbalanced (and skewed) datasets.

**4.1.3.3 Classifier performance evaluation** In this section, we thoroughly test our CNN-based classifier for model inference and generalization. Initially, we perform a quantitative test for model inference by training the classifier on one crowd dataset and evaluate on remaining datasets. Table 4 shows a confusion matrix with ACC values, where our model is trained on a dataset in a row and evaluate on all datasets in the columns.

The confusion matrix shows that our method is efficient in inferring divergence behavior learn from one video and can predict divergence in unseen videos. However, there is a slight degradation in ACC when our method is trained on a high-density crowd dataset and evaluated on low-density crowd datasets. At high-density crowd, mostly single (or few) global crowd shape(s) is obtained (as the whole crowd is acting as a single segment), whereas at low density, usually, inter-person distance is greater that divides the crowd into smaller crowd segments. Multiple crowd segments result in multiple crowd shapes within the same image (e.g., in the case of UMN, PETS, and NGSIM datasets). Multiple motion shapes in an image can confuse classifier to classify normal scene as divergent or vice versa.

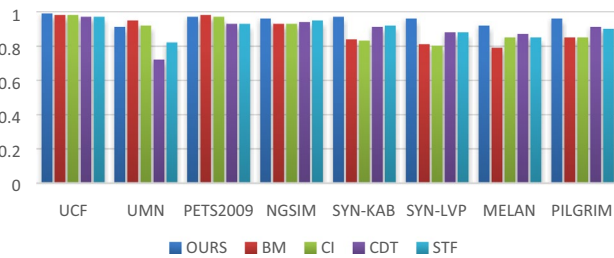
We also perform a qualitative test for model inference by plotting the CNN output probability score of every video frame.

Figure 25 shows four samples of class probability scores plotted against the number of images in a video. Figure 25a is the PETS2009 divergence behavior scenario for sequence 2 view 1. After image 50, the crowd starts diverging left, and the corresponding divergence score increases. From frame 85 to frame 110, crowd divergence is visible and after, frame number 110, people disappear from the scene and the divergence score reduced to zero at the end of the video. A similar divergence scenario is shown in Fig. 25b with a traffic dataset. Normally traffic flowing from east to west and under divergent scenario traffic start flowing south at frame 190, predicted by an increase in divergent class probability



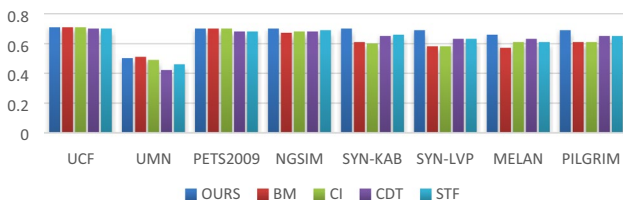
**Fig. 22** PR curve comparison **a** UCF, **b** UMN, **c** PETS2009, **d** NGSIM, **e** synthetic—kabbah, **f** SYNTHETIC Loveparade, **g** MELAN concert, **h** PILGRIM

**Area under PR curve (AP)**



**Fig. 23** Area under PR curve (AP) comparison—higher the AP value, better the method in predicting divergence class at imbalanced datasets

**F-Score**



**Fig. 24** F1 score comparison: high F1 score of our method shows it is better able to classify normal as well as divergent classes for imbalanced datasets (UCF, UMN, PETS, and NGSIM)

score. However, the assumption of multiple crowd segments at low density is not always true as is the case with the UCF scenario where the crowd is coherent over the zebra crossing

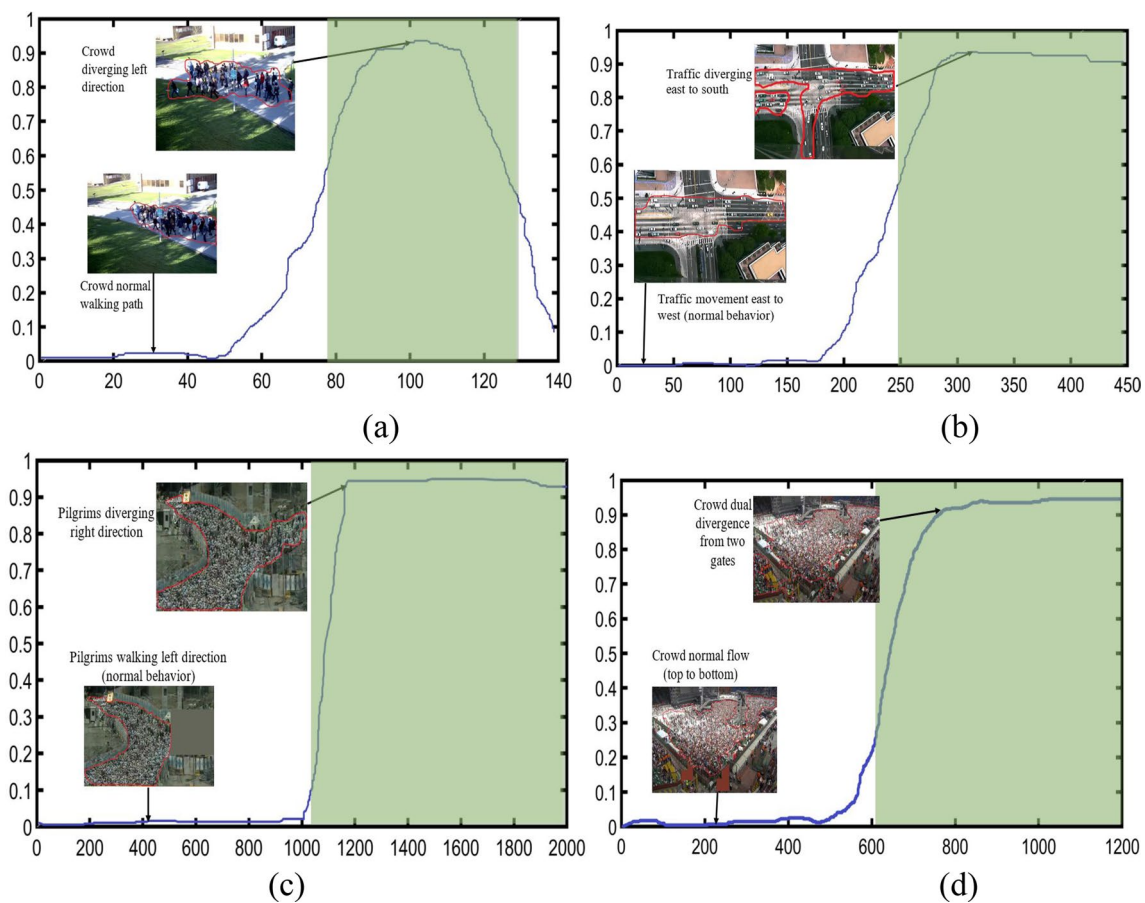
and ACC achieved is high compared to ACC at other low-density crowd datasets. Figure 25c, d demonstrates two examples of divergent class probability scores obtained at high-density crowd scenes (PILGRIM and MELAN datasets). However, there are also cases where our classifier confused due to multiple motion shapes within a single image, and in some cases, our classifier also fails due to motion shape mimicking divergence shape. Examples of these two scenarios are depicted in Fig. 26. Figure 26a demonstrates an escape activity in the UMN dataset of sequence 2. Under the normal scenario, people walk inside a square tile floor area; however, during frame 5710–5850, a person walk of the tile area for a short duration and then return inside. Although these frames are marked as normal in the ground truth, due to motion-shape change, our classifier misclassifies it as divergent due to a high divergence score. A similar event again happens between frames 6090–6150, and the divergence score goes high. Actual divergence occurs at frame 6196 where the divergence probability score is maximum.

Figure 26b demonstrates another example of a partial false positive that affects divergence score but still the classifier manages to predict correct classes. In this example, probability scores of both normal (magnetic color) and divergence (blue color) classes are plotted against each frame. It can be noticed that between frames 50–150, there is a small segment of the crowd walking off zebra crossing creating false positive and results in increasing divergence class score as well. However, the change is not that significant that it can reverse class scores, and thus, model is still

**Table 4** Confusion matrix of our CNN-based classifier performance evaluation by training on a dataset in a row and evaluate on datasets in columns

Training	Testing						
	Low-density crowd datasets				High-density crowd datasets		
	UCF (%)	PETS 2009 (%)	UMN (%)	NGSIM (%)	SYN (%)	PILGRIM (%)	CONCERT (%)
UCF	90	80	85	75	80	85	73
PETS2009	80	92	83	72	82	79	71
UMN	82	95	97	78	87	75	68
NGSIM	90	88	83	98	88	80	75
SYN	85	62	61	65	99	90	89
PILGRIM	87	59	63.2	67	92	98.5	90
CONCERT	81	63	60.5	61	93	98	99

The percentages shown are the ACC measure of each evaluation

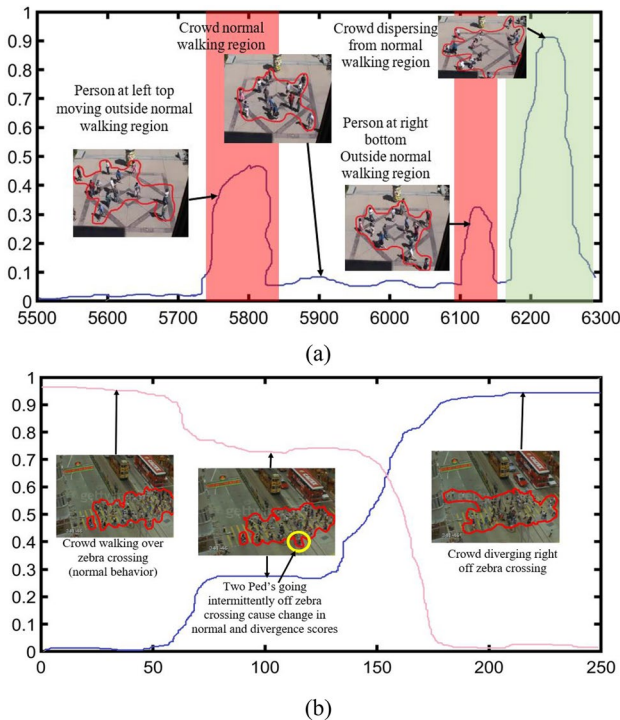


**Fig. 25** Qualitative results of our method on testing videos. Green colored windows show ground-truth divergence region, and our predicted divergence score falls within ground-truth windows. **a** and **b**

are samples of divergence detection at the low-density crowd; **c** and **d** are samples of divergence detection at the high-density crowd

able to predict images as normal behavior in this time duration. After frame 180, the crowd starts walking off the zebra crossing toward the top right side that is real divergence, and motion-shape change is significant from normal motion

shape. At this time duration, the normal class score is close to zero and the divergence class score is close to one as expected.



**Fig. 26** Failure case examples for our method for divergence detection **a** UMN escape activity depicting two false divergence detections due to few people walking off the expected ground truth normal behavior area and create divergence-like motion shapes. **b** Partial false detection scenario, from frame 50 to frame 100, few people walking off ground-truth normal zebra crossing area creating an illusion of divergence that causes the divergence class score to go high. Since the normal class score is still greater than the divergence class score, our model is still able to classify it as a normal class

#### 4.1.4 Comparison with CNN-based anomaly detection methods

The effectiveness of our method is compared with state-of-the-art CNN-based anomaly detection methods for divergence behavior detection. Selected methods include Motion information Image with CNN (MII-CNN) [52], Generative Adversarial Nets (GAN) [53], and temporal CNN pattern (TCP) [54]. We measure the accuracy of the methods, which is the percentage of correctly classified frames in comparison with the ground truth. Comparisons are performed on low-density crowd datasets (UMN and PETS2009) and high-density crowd datasets (PILGRIM and CONCERT). Table 5

**Table 5** ACC comparison of CNN-based anomaly detection methods at the low-density crowd

Dataset/method	Ours	MSI+CNN	MII+CNN	GAN	TCP
UMN	98.75		99.08	99	98.8
PETS2009	98		98.39	98.2	97.9

provides the ACC measurement of four methods on low-density crowd datasets. At low crowd density, our method performs comparably to existing state-of-the-art methods. However, there is a slight degradation by our method at the UMN dataset as the motion-shape difference among normal and divergent behaviors is less significant; hence, the classification accuracy is dropped.

Next, a comparison is performed for divergence detection at a high-density crowd. Table 6 provides the ACC measurement of four methods on high-density crowd datasets.

At high density, our method outperforms existing CNN-based anomaly detection methods. Selected methods primarily use OPF information to generate images for CNN to train. At high density, global motion information is not captured well by OPF due to noise issues, and performance is significantly degraded for existing art methods, whereas FTLE provides clean motion shape at high crowd density and our CNN classification performs efficiently.

#### 4.1.5 Divergence localization evaluation

**4.1.5.1 Parameter settings** Divergence class score value from CNN lies between 0 and 1, and the threshold at the class score is set to 0.65 (65%) to trigger the localization detection process. Residual blobs are filtered out through temporal averaging filtering and spatially through connected neighbors filtering. The number of images for temporal averaging  $N_{\text{filter}}$  is set to 10 i.e., averaging previous 10 frames OPF and FTLE fields. Spatial filtering is applied by performing an 8-connected neighbor on residual blob(s).

**4.1.5.2 Qualitative evaluation** In qualitative evaluation, we compare our proposed divergence localization scheme with two state-of-the-art divergence localization methods [7, 8]. The method in [7] estimates divergent centers by placing potential destination points over normal image sequences. For the escape scene, a set of divergent points is initialized, and the foreground velocity patches start from the divergent points and reach potential destinations that are marked as divergent centers. The method in [8] analyzes intersections of foreground velocity vectors and uses distance segmentation method and  $knn$  search to locate the divergent center. Intersections are obtained with the assumption that there are three moving objects in the neighbor of the desired diver-

**Table 6** ACC comparison of CNN-based anomaly detection methods at the high-density crowd

Dataset/method	Ours	MSI+CNN	MII+CNN	GAN	TCP
PILGRIM	99.05		97	96.8	95
CONCERT	99		96.5	97	96

gent center and straight lines of moving objects should intersect to declare the point as the divergent center.

A qualitative comparison of three methods is shown in Fig. 27. Divergent centers are marked with triangles for three methods and ground truth. Scenarios in the first two rows of Fig. 27 depict divergent center detections at the low-density crowd. Existing state-of-the-art methods accurately detect divergent centers low density, whereas our method shows few false detections at the low crowd density divergence. At low density, crowd motion shape is broken and produces many residuals after the differencing from reference motion shape. The residual blobs cause marking of false divergent centers by our method. Also, since our method analyzes variations at crowd boundaries, the divergence points detected by our method at low crowd density are slightly offset from the ground truth.

The last two images in the third row of Fig. 27 show divergence center detections at the high-density crowd. Existing art failed to detect divergence sources at the

high-density crowd. The method at [7] analyzes variations in foreground velocity patches from high (in non-escape) to low (in escape) case. However, at high density, foreground velocity concentration is the same in both behaviors; no divergence center points are reaching potential destinations.

However, the method is still able to detect divergence center point at crowd boundary as patches at boundaries experience variations in velocity. Similarly, the method in [8] shows poor performance at high-density divergence. Assumption of three moving objects and solving three straight lines equations is possible at low density but at the high-density crowd. As there can be hundreds of people in the neighbor of a divergent center, solving straight line equations for hundreds of moving objects is computationally very expensive. The method with existing model settings is unable to detect actual divergent center locations at high density. However, the model can be improved by obtaining intersections of many neighboring moving objects with reasonable compute. False detections by existing art on the high-density



**Fig. 27** Divergence localization comparison with state-of-the-art methods. Legends: red triangle [7], green triangle [8], orange triangle our method, blue triangle ground truth. The top two rows show divergence localization at low-density scenes. False divergent center

detections by our method at 2nd, 3rd image at the first row and 1st image, second row. Existing art failed to detect divergent centers at high density (last two images in the third row), while our method can accurately detect divergent centers

crowd can be seen in the last two images of the third row in Fig. 27. Our method performs well at high-density scenario and can identify divergence points at correct locations.

As mentioned earlier, our localization algorithm also identifies divergence regions (or divergence mask) evolve temporally. Figure 28 shows three examples of divergence mask detected by our algorithm. The first row is a crowd divergence sequence taken from the PETS2009 dataset, the second row shows people diverging from the normal path of walking over the zebra crossing, and the last row shows people diverging from circular Tawaf after completing seven laps of Tawaf. Images in columns show the temporal progression of divergence where images are taken every  $n\_frames$  after the time instant ( $t_d$ ) divergence is detected by CNN. Divergence mask size can be seen increasing over time indicating variations in the size of divergence and shows the direction in which divergence is leading.

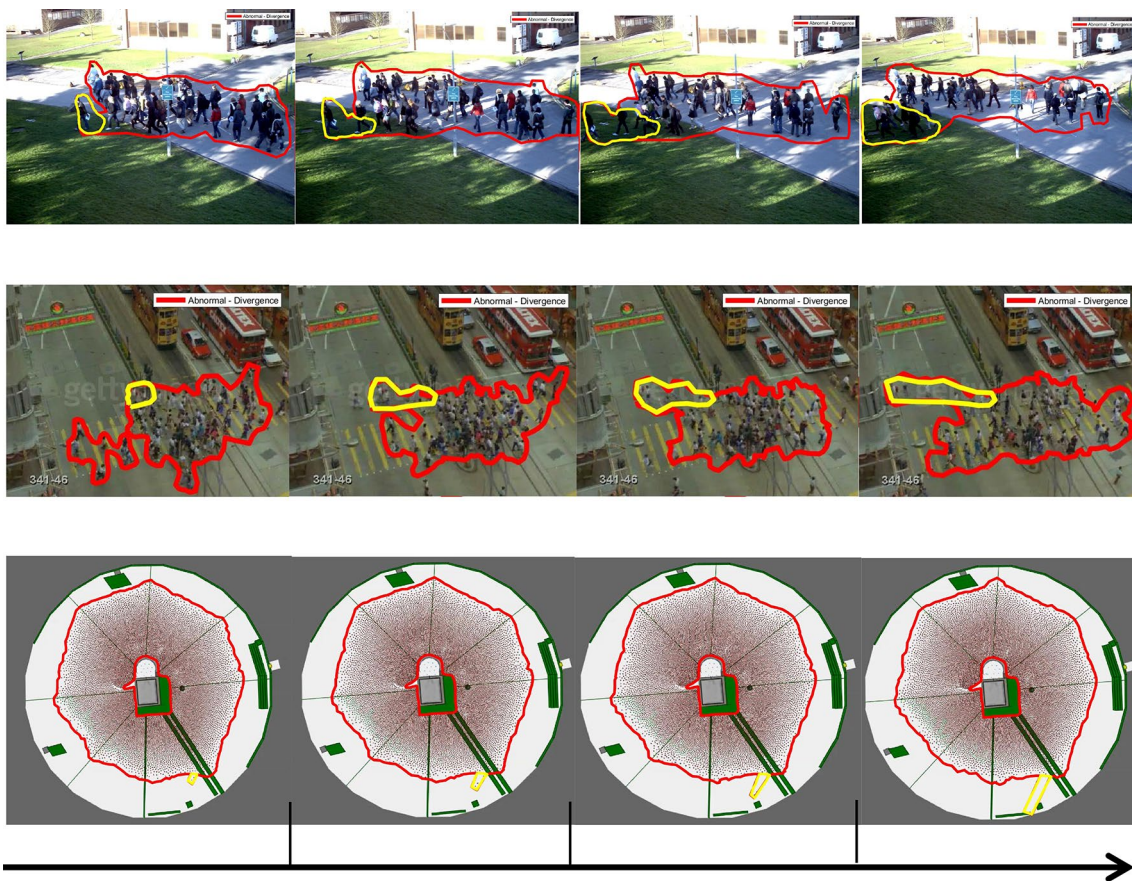
**4.1.5.3 Quantitative evaluation** As mentioned earlier, our divergence localization algorithm not only identifies diver-

gence source points but also detects region(s) of divergence. To qualitatively evaluate the performance of our algorithm, we compute Intersection over Union (IoU) between the predicted divergence region and ground-truth divergence region. Ground-truth regions are obtained by hand-labeling divergence regions at each abnormal frame. IoU score is calculated using Eq. (6).

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (6)$$

where Area of overlap is the overlap area between the predicted region and ground-truth region. Area of union is the area encompassed by both the predicted bounding box and ground-truth bounding box.

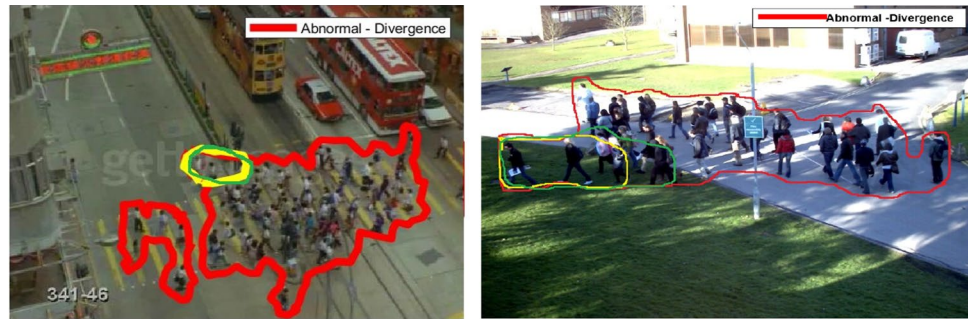
IoU score  $> 0.5$  (50% overlap) is generally considered a good prediction by algorithm [55]. Two samples of divergent regions detected overlaid with ground-truth divergence region are shown in Fig. 29.



**Fig. 28** Divergence region mask changes over time. Divergence regions: first row PETS2009; second-row UCF; and third-row Synthetic Kabbah datasets. The first column is the region at the time of divergence detection ( $t_d$ ) by CNN. The second column represents

divergence region detection  $n\_frames$  after  $t_d$ , the third column represents divergence region detection  $2*n\_frames$  after  $t_d$ , and so on. The value of  $n\_frames$  is different for each sequence

**Fig. 29** Examples of divergence region by our method (yellow) compared to ground truth (green)



**Table 7** IoU score obtained by our method for divergence regions

Dataset	IoU score
UCF	0.70
UMN	0.55
PETS2009	0.63
NGSIM	0.68
SYN-KABBAH	0.82
SYN-LOVEPARADE	0.89
CONCERT	0.76
PILGRIM	0.80

\*IoU score is averaged for dataset contains multiple sequences

The IoU scores for divergent datasets used in this work are provided in Table 7.

Again it is clear from Table 7 that at low-density datasets (UMN, PETS2009, NGSIM), divergence shape is not smooth and is broken, resulting in low IoU score, whereas IoU score at high densities is better, meaning that our algorithm is better able to detect divergent regions at high crowd densities.

## 5 Conclusion

In this work, we propose a novel method of divergence detection at the high-density crowd using MSIs combined with the power of CNN. Our approach estimates crowd motion using the FTLE method for both normal and divergent scenes. MSIs are then obtained by extracting LCS from the FTLE field and used for CNN supervised learning for behavior classification. Experimental results show that our method outperforms both manual and CNN-based state-of-the-art anomaly detection methods and achieve better accuracy. We also propose a divergence localization method with a new feature of producing localization mask for divergence size indication. Qualitative and quantitative results indicate that our localization method is effective in detecting divergence starting or source points both at low and at high crowd density.

There are a few limitations to our approach. Our method provides anomaly information at crowd boundaries only. This limits our method to the global anomaly detection, whereas at the high-density crowd, triggering events for global anomaly occur at the crowd local level termed as crowd local anomalies. There is no local crowd anomaly information provided by our method. In future work, we shall extend this method to include crowd local behavior changes to detect local anomalies that lead to global anomalies like divergence. One of the possible solutions for local anomaly detection at the high-density crowd is to perform head detection and then observe heads movements for normal/abnormal patterns.

**Acknowledgments** We thank Dr. Yasir Salih for sharing high-density crowd dataset and valuable suggestions on motion estimation at high-density crowd. The MassMotion crowd simulation software is supported by the Center for Intelligent Signal for Imaging Research (CISIR) under PO Number 3920089787/30.10.2017.

## References

1. Illiyas, F.T., Mani, S.K., Pradeepkumar, A.P., Mohan, K.: Human stampedes during religious festivals: a comparative review of mass gathering emergencies in India. *Int. J. Disaster Risk Reduct.* **5**, 10–18 (2013). <https://doi.org/10.1016/j.ijdr.2013.09.003>
2. Batty, M., Desyllas, J., Duxbury, E.: The discrete dynamics of small-scale spatial events: agent-based models of mobility in carnivals and street parades. *Int. J. Geogr. Inf. Sci.* **17**(7), 673–697 (2003). <https://doi.org/10.1080/1365881031000135474>
3. Dong, Y.-H., Liu, F., Liu, Y.-M., Jiang, X.-R., Zhao, Z.-X.: Emergency preparedness for mass gatherings: lessons of ‘12.31’ stampede in Shanghai Bund. *Chin. J. Traumatol.* **20**(4), 240–242 (2017). <https://doi.org/10.1016/j.cjtee.2016.08.005>
4. Helbing, D., Mukerji, P.: Crowd disasters as systemic failures: analysis of the Love Parade disaster. *EPJ Data Sci.* **1**(1), 1–40 (2012). <https://doi.org/10.1140/epjds7>
5. Johansson, A., Helbing, D., Al-Abideen, H.Z., Al-Bosta, S.: From crowd dynamics to crowd safety: a video-based analysis (2008) [Online]. <http://arxiv.org/abs/0810.4590>
6. Cong, Y., Yuan, J., Liu, J.: Sparse reconstruction cost for abnormal event detection. In: *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3449–3456 (2011). <https://doi.org/10.1109/CVPR.2011.5995434>

7. Wu, S., Wong, H.S., Yu, Z.: A bayesian model for crowd escape behavior detection. *IEEE Trans. Circuits Syst. Video Technol.* **24**(1), 85–98 (2014). <https://doi.org/10.1109/TCSVT.2013.2276151>
8. Chen, C.Y., Shao, Y.: Crowd escape behavior detection and localization based on divergent centers. *IEEE Sens. J.* **15**(4), 2431–2439 (2015). <https://doi.org/10.1109/JSEN.2014.2381260>
9. <https://www.worldbulletin.net/asia-pacific/stampede-at-hindu-festival-kills-18-in-south-india-h162104.html>
10. <https://www.oasys-software.com/products/pedestrian-simulation/massmotion/>
11. Horn, B.K., Schunck, B.G.: Determining optical flow. *Artif. Intell.* **17**(1981), 185–203 (1981)
12. Brox, T., Papenbergh, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. *Comput. Vis.—ECCV 2004*(4), 25–36 (2004). [https://doi.org/10.1007/978-3-540-24673-2\\_3](https://doi.org/10.1007/978-3-540-24673-2_3)
13. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. *Proc. Imaging Underst. Work.* **130**, 121–130 (1981)
14. Fortun, D., Bouthemy, P., Kervrann, C., Fortun, D., Bouthemy, P., Kervrann, C.: Optical flow modeling and computation : a survey. *Comput. Vis. Image Underst.* **134**, 1–21 (2015)
15. Lawal, I.A., Poiesi, F., Anguita, D., Cavallaro, A.: Support vector motion clustering. *IEEE Trans. Circuits Syst. Video Technol.* **27**(11), 1–1 (2016). <https://doi.org/10.1109/TCSVT.2016.2580401>
16. Cheriadat, A.M., Radke, R.J.: Detecting dominant motions in dense crowds. *IEEE J. Sel. Top. Signal Process.* **2**(4), 568–581 (2008). <https://doi.org/10.1109/JSTSP.2008.2001306>
17. Benabbas, Y., Ihaddadene, N., Djeraba, C.: Motion pattern extraction and event detection for automatic visual surveillance. *Eurasip J. Image Video Process.* **2011**, 1–15 (2011)
18. Ali, S., Shah, M.: A Lagrangian particle dynamics approach for crowd flow simulation and stability analysis. In: *IEEE Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, USA, pp. 1–6 (2007). <https://doi.org/10.1109/CVPR.2007.382977>
19. Hu, M.H.M., Ali, S., Shah, M.: Learning motion patterns in crowded scenes using motion flow field. In: *2008 19th Int. Conf. Pattern Recognit.*, pp. 2–6 (2008). <https://doi.org/10.1109/ICPR.2008.4761183>
20. Solmaz, B., Moore, B.E., Shah, M.: Identifying behaviors in crowd scenes using stability analysis for dynamical systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**, 2064–2070 (2012). <https://doi.org/10.1109/TPAMI.2012.123>
21. Chen, D.Y., Huang, P.C.: Motion-based unusual event detection in human crowds. *J. Vis. Commun. Image Represent.* **22**(2), 178–186 (2011). <https://doi.org/10.1016/j.jvcir.2010.12.004>
22. 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.* **28**(9), 1450–1464 (2006). <https://doi.org/10.1109/TPAMI.2006.176>
23. Kratz, L., Nishino, K.: Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work*, pp. 1446–1453 (2009). <https://doi.org/10.1109/CVPRW.2009.5206771>
24. Kratz, L., Member, S., Nishino, K.: Spatio-temporal motion patterns in extremely crowded scenes. *Analysis* **34**(5), 987–1002 (2012)
25. Cong, Y., Yuan, J., Tang, Y.: Video anomaly search in crowded scenes via spatio-temporal motion context. *IEEE Trans. Inf. Forens. Secur.* **8**(10), 1590–1599 (2013). <https://doi.org/10.1109/TIFS.2013.2272243>
26. Wu, Y., Ye, Y., Zhao, C.: Coherent motion detection with collective density clustering. In: *Proceedings of the 23rd ACM international conference on Multimedia—MM'15*, 2015, vol. 1, no. 1, pp. 361–370. <https://doi.org/10.1145/2733373.2806227>
27. Ali, S., Shah, M.: A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–6 (2007)
28. Zitouni, M.S., Bhaskar, H., Dias, J., Al-Mualla, M.E.: Advances and trends in visual crowd analysis: a systematic survey and evaluation of crowd modelling techniques. *Neurocomputing* **186**, 139–159 (2015). <https://doi.org/10.1016/j.neucom.2015.12.070>
29. Wu, S., Moore, B.E., Shah, M.: Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2054–2060 (2010)
30. Wang, X., Gao, M., He, X., Wu, X., Li, Y.: An abnormal crowd behavior detection algorithm based on fluid mechanics. *J. Comput.* **9**(5), 1144–1149 (2014). <https://doi.org/10.4304/jcp.9.5.1144-1149>
31. Wang, X., Yang, X., He, X., Teng, Q., Gao, M.: A high accuracy flow segmentation method in crowded scenes based on streakline. *Opt.—Int. J. Light Electron Opt.* **125**(3), 924–929 (2014). <https://doi.org/10.1016/j.ijleo.2013.07.166>
32. Mehran, R., Moore, B.E., Shah, M.: A streakline representation of flow in crowded scenes. *Lect. Notes Comput. Sci. (incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinf.)* **6313 LNCS(PART 3)**, 439–452 (2010). [https://doi.org/10.1007/978-3-642-15558-1\\_32](https://doi.org/10.1007/978-3-642-15558-1_32)
33. Pereira, E.M., Cardoso, J.S., Morla, R.: Long-range trajectories from global and local motion representations. *J. Vis. Commun. Image Represent.* **40**, 265–287 (2016). <https://doi.org/10.1016/j.jvcir.2016.06.020>
34. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11**, 3371–3408 (2010)
35. Huang, S., Huang, D., Zhou, X.: Learning multimodal deep representations for crowd anomaly event detection. *Math. Probl. Eng.* (2018). <https://doi.org/10.1155/2018/6323942>
36. Xu, J., Ren, D., Zhang, L., Zhang, D.: Patch group based bayesian learning for blind image denoising. *Lect. Notes Comput. Sci. (incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinf.)* **10116 LNCS**, 79–95 (2017). [https://doi.org/10.1007/978-3-319-54407-6\\_6](https://doi.org/10.1007/978-3-319-54407-6_6)
37. Li, X., Shen, H., Li, H., Zhang, L.: Patch matching-based multi-temporal group sparse representation for the missing information reconstruction of remote-sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **9**(8), 3629–3641 (2016). <https://doi.org/10.1109/JSTARS.2016.2533547>
38. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 886–893 (2005)
39. Barnich, O., Van Droogenbroeck, M.: ViBe: A universal background subtraction algorithm for video sequences. *IEEE Trans. Image Process.* **20**(6), 1709–1724 (2011). <https://doi.org/10.1109/TIP.2010.2101613>
40. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: *2009 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work. CVPR Work*, no. 2, pp. 935–942 (2009). <https://doi.org/10.1109/CVPRW.2009.5206641>
41. Shadden, S.C., Lekien, F., Marsden, J.E.: Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows. *Phys. D Nonlinear Phenom.* **212**(3–4), 271–304 (2005). <https://doi.org/10.1016/j.physd.2005.10.007>
42. Lipinski, D., Mohseni, K.: A ridge tracking algorithm and error estimate for efficient computation of Lagrangian coherent structures. *Chaos Interdiscip. J. Nonlinear Sci.* **20**(1), 017504 (2010). <https://doi.org/10.1063/1.3270049>
43. Peikert, R., Schindler, B., Carnecky, R.: Ridge surface methods for the visualization of Lagrangian coherent structures. *Semseg. Org* [Online]. [http://www.semseg.org/results/\\_files/2012-ICFD-PeikertEtAl-RidgeSurfaceMethods.pdf%5Cnpapers3://publication/uuid/7F60AE35-CBAF-4E7B-B59F-23A81EADD827](http://www.semseg.org/results/_files/2012-ICFD-PeikertEtAl-RidgeSurfaceMethods.pdf%5Cnpapers3://publication/uuid/7F60AE35-CBAF-4E7B-B59F-23A81EADD827)



44. Zeiler, M.D., et al.: On rectified linear units for speech processing New York University, USA Google Inc., USA University of Toronto, Canada. In: IEEE International Conference on Acoustic Speech and Signal Processing (ICASSP 2013), pp. 3–7 (2013)
45. Krausz, B., Bauckhage, C.: Loveparade 2010: automatic video analysis of a crowd disaster. *Comput. Vis. Image Underst.* **116**(3), 307–319 (2012). <https://doi.org/10.1016/j.cviu.2011.08.006>
46. Wu, S., Moore, B.E., Shah, M.: Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2054–2060 (2010)
47. U. C. F. Crowd Dataset. <https://www.crcv.ucf.edu/data/crowd.php>
48. U. M. N. Crowd Dataset. [http://mha.cs.umn.edu/proj\\_events.shtml#crowd](http://mha.cs.umn.edu/proj_events.shtml#crowd)
49. 2009 Dataset, PETS. <http://www.cvg.reading.ac.uk/PETS2009/a.html>
50. NGSIM. <https://ops.fhwa.dot.gov/trafficanalysis/tools/ngsim.htm>
51. Wu, S., Yang, H., Zheng, S., Su, H., Fan, Y., Yang, M.H.: Crowd behavior analysis via curl and divergence of motion trajectories. *Int. J. Comput. Vis.* **123**(3), 1–21 (2017). <https://doi.org/10.1007/s11263-017-1005-y>
52. Direkoglu, C.: Abnormal crowd behavior detection using motion information images and convolutional neural networks. *IEEE Access* **8**, 80408–80416 (2020). <https://doi.org/10.1109/ACCESS.2020.2990355>
53. Ravanbakhsh, M., Nabi, M., Sangineto, E., Marcenaro, L., Regazzoni, C., Sebe, N.: Abnormal event detection in videos using generative adversarial nets. In: Proceedings—International Conference on Image Processing, ICIP, vol. 2017, pp. 1577–1581 (2018). <https://doi.org/10.1109/ICIP.2017.8296547>
54. Ravanbakhsh, M., Nabi, M., Mousavi, H., Sangineto, E., Sebe, N.: Plug-and-play CNN for crowd motion analysis: an application in abnormal event detection. In: Proceedings—2018 IEEE Winter Conference on Applications of Computer Vision, WACV 2018, vol. 2018, pp. 1689–1698 (2018). <https://doi.org/10.1109/WACV.2018.00188>
55. Ahmed, F., Tarlow, D., Batra, D.: Optimizing expected intersection-over-union with candidate-constrained CRFs. In: Proc. IEEE Int. Conf. Comput. Vis., vol. 2015 Inter, pp. 1850–1858 (2015). <https://doi.org/10.1109/ICCV.2015.215>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Muhammad Umer Farooq** received B.S. in Electrical Engineering from COMSATS Institute of Information Technology (2004) and masters from Universiti Teknologi PETRONAS (UTP) in 2013. He has more than ten years of industry job experience and worked in areas of mixed signal circuit design and development, digital design using FPGA (Xilinx/Altera), firmware development of various DSP

processors, etc. Currently, he is working at Intel and pursuing his Ph.D. at the Center for Intelligent Signal and Imaging Research (CISIR), UTP. His research interests include behavior detection at the high-density crowd and deep learning.



**Mohamad Naufal M. Saad** received the masters degree from the Ecole Nationale Supérieure d'Ingénieurs de Limoges, France, and the Ph.D. degree in telecommunication from the Université de Limoges, France, in 2005. He is currently an Associate Professor with the Electrical and Electronic Engineering Department, Universiti Teknologi Petronas (UTP), Malaysia. He is a Core Research Member with the Center for Intelligent Signal and Imaging Research, UTP. His research interests include neurosignal processing,

medical imaging, and communication.



**Sultan Daud Khan** is currently an Associate Professor in the Department of Computer Science, National University of Technology, Pakistan. He has published several papers in conferences and journals such as AVSS, IVCNZ, ICGIP, Neurocomputing, Journal of Cellular Automata, and IEEE Access. His research interests include crowd analysis, action recognition and localization, object detection, visual tracking, multi-camera, and airborne surveillance using deep learning techniques. He

received the B.Sc. (Hons) degree in Computer Engineering from the University of Engineering & Technology in 2005, M.Sc. (Hons) in Electronics & Communication Engineering from Hanyang University, South Korea, in 2010, and the Ph.D. degree in Computer Science from the University of Milano-Bicocca 2016. He is an active reviewer of prestigious journals such as Neurocomputing, Pattern Recognition, IEEE IET signal processing, ACM Multimedia, IEEE Access, and ACM TOMM. He received the best reviewer award from Pattern Recognition in 2017.