

Research Article

Group Abnormal Behaviour Detection Algorithm Based on Global Optical Flow

Yu Hao ¹, Ying Liu,¹ Jiulun Fan,¹ and Zhijie Xu²

¹School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China

²School of Computing and Engineering, University of Huddersfield, Huddersfield HD1 3DH, UK

Correspondence should be addressed to Yu Hao; haoyu@xupt.edu.cn

Received 30 January 2021; Revised 20 April 2021; Accepted 29 April 2021; Published 5 May 2021

Academic Editor: Wei Wang

Copyright © 2021 Yu Hao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abnormal behaviour detection algorithm needs to conduct behaviour analysis on the basis of continuous video inclination tracking, and the robustness of the algorithm is reduced for the occlusion of moving targets, the occlusion of the environment, and the movement of targets with the same colour. For this reason, the optical flow information between RGB (red, green, and blue) images and video frames is used as the input of the network in view of group behaviour. Then, the direction, velocity, acceleration, and energy of the crowd were weighted and fused into a global optical flow descriptor. At the same time, the crowd trajectory map is extracted from the original image of a single frame. Following, in order to realize the detection of large displacement moving target and solve the problem that the traditional optical flow algorithm is only suitable for the detection of displacement moving target, a video abnormal behaviour detection algorithm based on the double-flow convolutional neural network is proposed. The network uses two network branches to learn spatial dimension information and temporal dimension information, respectively, and uses short- and long-time neural network to model the dependency relationship between long-time video frames, so as to obtain the final behaviour classification results. Simulation test results show that the proposed method can achieve good recognition effect on multiple datasets, and the performance of abnormal behaviour detection can be significantly improved by using interframe motion information.

1. Introduction

Abnormal behaviour detection is one of the important and challenging research hotspots, which is based on the basis of video image processing such as scene understanding and visual target tracking [1, 2]. For a specified video or surveillance image sequence, abnormal crowd behaviour detection is to extract specific information representing the abnormal behaviour of the crowd in the video sequence, such as population density and group behaviour characteristics, and perform classification [3–5]. Detecting abnormal crowd behaviours plays a very important role in the field of computer vision [6–8].

Extracting useful information from massive surveillance videos and detecting abnormal behaviours and events in the video require a large number of workers to keep a high

degree of attention on the surveillance pictures for a long time [9–11]. However, relying solely on manual detection methods can easily lead to false alarms and missed detections [12]. Therefore, how to extract useful information from massive surveillance videos and improve the recognition accuracy of emergencies and abnormal behaviours has a wide range of economic and application value in the field of security and social safety [13–16]. Abnormal behaviour detection refers to the initial frame that can classify the event and find the abnormal behaviour in time when there is an abnormality in a video. In order to effectively distinguish the normal events and abnormal events in the video, it is necessary to extract the relevant features from the video sequence and classify them. In traditional feature extraction methods, researchers often use temporal and spatial features, such as directional gradient histogram feature, optical flow

histogram feature, dynamic texture feature, and social force model, to model the motion patterns of video targets. At present, with the wide application and development of the deep neural network in industry and academia, it has achieved high accuracy and good effect in speech recognition, natural language processing, and computer vision. Therefore, more and more fields begin to use deep neural networks to solve the problem of abnormal behaviour detection in video. Although these methods have achieved good results, the abnormal behaviour detection algorithm needs to conduct behaviour analysis on the basis of continuous video inclination tracking. For the occlusion of moving target, the occlusion of environment, and the movement of the same colour target, the robustness of the current algorithm is reduced.

Based on above, this paper takes the optical flow information between RGB images and video frames as the input of the network. Then, the extracted crowd movement direction, velocity, acceleration, and energy are weighted and fused into a global optical flow descriptor. At the same time, the motion trajectory map is extracted based on the single-frame original image. After that, two network branches are used to learn spatial and temporal information, and long- and short-term neural networks are used to model the dependence between long-term video frames, to obtain the final behaviour classification results.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 introduces the proposed methods, including optical flow methods and two-stream convolutional neural networks. Section 4 reports the experiments and results. Section 5 concludes our work.

2. Related Works

Abnormal behaviour detection mainly includes feature extraction, feature fusion, and behaviour classification. In the process of feature extraction, people can select appropriate manual extraction of low-level features and use the deep learning method to extract high-level features based on low-level features. The extracted features are then fused to form a more complete spatiotemporal feature that represents a moving target. Finally, according to the extracted features, an appropriate classifier is designed to judge the behavioural characteristics of the represented population and output the correct behaviour detection results [17–19].

Mass detection methods can be divided into different categories according to different standards. This article divides the literature into three categories according to the research purpose, such as group behaviour analysis, mass behaviour analysis, and anomaly detection and location. The three types of population detection methods are described below with detailed introduction.

2.1. Group Behaviour Analysis. Group behaviour analysis is aimed at the group and analyses the behaviour by detecting and tracking the group. In [20], template matching is used to track individuals, and a Voronoi diagram is generated in each frame. Use the Voronoi diagram to determine the time

evolution of some sociological and psychological parameters, such as personal spatial distance. Through the analysis of these parameters, the group is formed and divided into normal or abnormal. In [21], a two-layer tracking target method is proposed. The first layer tracks the active area based on low-level operations and generates a set of features. The second layer uses a Bayesian network-based statistical model to consistently label the detected fragments. In [22], the feature histogram of each frame is extracted to express their group connectivity and motion characteristics. Then, these features are extracted by using detection and tracking, and the features are represented by bag of words. Finally, SVM is trained to classify video clips for event recognition.

2.2. Mass Behaviour Analysis. Mass behaviour analysis is the first step in anomaly detection research, and anomaly detection is inseparable from behaviour analysis. In [23], a target detection method based on motion estimation is proposed. In motion estimation, block matching, optical flow, and Gabor filtering techniques are used to implement the HS method. When the calculated stop time exceeds the threshold, and the position is not where you normally want to stop, the system will issue an alarm. In [24], a variant of the counter flow detection for subway stations is proposed. By filtering the obtained optical flow motion vector, the noise of the motion trajectory construction is reduced, and then the motion trajectory is used to detect the reverse flow. In the literature [25], a method for detecting abnormalities in dense mass is proposed. Firstly, optical flow is used to track to obtain trajectories, and then trajectories that are close to each other in space and have similar motion directions are clustered. Finally, manually analyse the movement and detect abnormalities. In [26], the feature-tracking algorithm is used to count the moving mass. Track the mass through the KLT algorithm, then use the connection graph to combine these features, and count the combined features to calculate the number of objects. The literature [27] is an unsupervised method for detecting independent movement in a mass. The feature tracker is used to generate trajectories, and the unsupervised Bayesian clustering method is used to cluster the trajectories, so that each cluster corresponds to a different independent movement in the mass. The purpose of the study is to detect abnormal movements of individuals in low-density populations. The literature [28] uses a set of particle trajectories generated by optical flow to locate the region of interest in the scene. Use the characteristic values and angles of particles to determine the behaviour types in crowded scenes. In the literature [29], a method of tracking shallow particles to avoid blockage was proposed to determine the abnormality of the crowd. In the literature [30], a method of dividing individuals in a group based on Bayesian is proposed. First, a 3D human body model is used to extract the foreground, and then the shape, height, camera model, head candidate, foreground object, and other features of the person are integrated into the Bayesian framework based on the Markov chain Monte Carlo probability model. In [31], a monitoring system was developed to estimate the degree of

congestion in subway stations. Through the background subtraction algorithm of Kalman filter, the neural network is trained for foreground features. Moreover, integrate the neural network into the fuzzy decision rules to get an overall neurofuzzy classifier for judging the degree of congestion.

2.3. Anomaly Detection and Location. People pay more and more attention to anomaly detection, and there are more and more methods. In the literature [32], a method for estimating sudden changes and abnormal motion changes was proposed. First, generate a motion heat map to use as the foreground of the video. Then, use optical flow to detect features and track them. Calculate a set of statistical measures and determine thresholds based on these measures to make decisions. In the literature [33], the social force model is used to detect and locate anomalies. The algorithm first uses the optical flow method to extract the particle trajectory. Then, use the social force model to calculate the force flow for each pixel. Finally, based on a fixed threshold likelihood estimation, normal and abnormal are marked. In the literature [34], the force field model is used to describe the clustering behaviour with attributes such as direction, location, and group size. The group attribute that appears suddenly is marked as an abnormal event. In [35], an abnormal event detection for occluded scenes is proposed. The algorithm first uses the average displacement method to segment the video into regions and then uses the shape matching method to match the model with the video segment. The literature [36] uses the principle of thermal imaging to detect pedestrian posture in a mass. After background subtraction and head detection, a number of weak classifiers combined with the human body model are used to detect abnormal poses. Model-based anomaly detection methods [37–40] have received widespread attention due to their high success rate. However, most of these methods require learning and training and the model is complex.

3. Abnormal Behaviour Detection Algorithm

3.1. Abnormal Behaviour Description and Algorithm Framework. Normally, the direction and speed of crowds are similar. However, when an abnormal event occurs, people will run away quickly due to fear to avoid potential danger. However, the abnormal behaviour of the crowd has the characteristics of fast movement speed, sudden increase in acceleration, and obvious concentration of movement in a certain direction or balance in multiple directions, large movement range, large pace, panic expression, and chaotic trajectory. Among them, the calculation of characteristics such as speed, acceleration, direction, and motion amplitude is relatively simple and can be expressed by optical flow. The extraction of features such as steps and expressions is more complicated. In order to reduce the complexity of the proposed method, this paper extracts the characteristics of the moving target's speed, acceleration, direction, movement

amplitude, and trajectory to detect abnormal crowd behaviour.

Traditional abnormal behaviour detection algorithms only use RGB images as the input of the network, without considering the hidden motion information in the video sequence. In order to overcome this problem, this paper proposes a video abnormal behaviour detection algorithm based on dual-stream convolutional neural network. The algorithm first takes the optical flow information between the RGB image and the video frame as the input of the network. Then, the extracted crowd movement direction, velocity, acceleration, and energy are weighted and fused into a global optical flow descriptor. At the same time, the motion trajectory map is extracted based on the single frame original image. Finally, the global optical flow descriptor and trajectory map are input into DCNN to detect abnormal crowd behaviour. The algorithm framework is shown in Figure 1.

3.2. LK Optical Flow Algorithm. Optical flow is the instantaneous speed of the pixels observed by the moving object from space to the imaging plane, which is caused by the movement of the moving object itself, the movement of the monitoring device, or the joint movement of the two. It uses the correlation between time-domain pixels and the mapping between two adjacent frames to calculate the instantaneous velocity of the object between the current frame and the previous frame.

The optical flow method has good temporal and spatial characteristics. It can detect independent moving objects in the crowd under unknown prior information scenes and accurately calculate their movement speed, so it can be used to describe the movement information such as the speed and direction of the moving target.

The Lucas–Kanade (LK) optical flow algorithm is a two-frame difference optical flow estimation algorithm. It was proposed by Bruce D. Lucas and Takeo Kanade, so it is called LK optical flow algorithm. Among them, the LK optical flow algorithm only needs to specify a set of feature points with a certain characteristic for tracking calculation, so it is relatively stable and reliable.

The LK algorithm originated in 1981 and was originally used to find dense optical flow. After adding feature points, it is often used to find sparse optical flow. The original two basic assumptions did not have a good solution, so another assumption, the “spatial consistency assumption,” was added to solve this problem. Assuming that adjacent pixels have similar motion, that is, in a certain area around the target pixel, each pixel has the same optical flow vector.

The optical flow in a neighbourhood is minimized by the weighted sum of squares:

$$\sum \lambda^2(a)(F_a \bullet \alpha + F_b \bullet \beta + F_c). \quad (1)$$

Among them, $\lambda^2(a)$ is the weight of the window, so that the weight in the neighbourhood is larger than the surrounding ones.

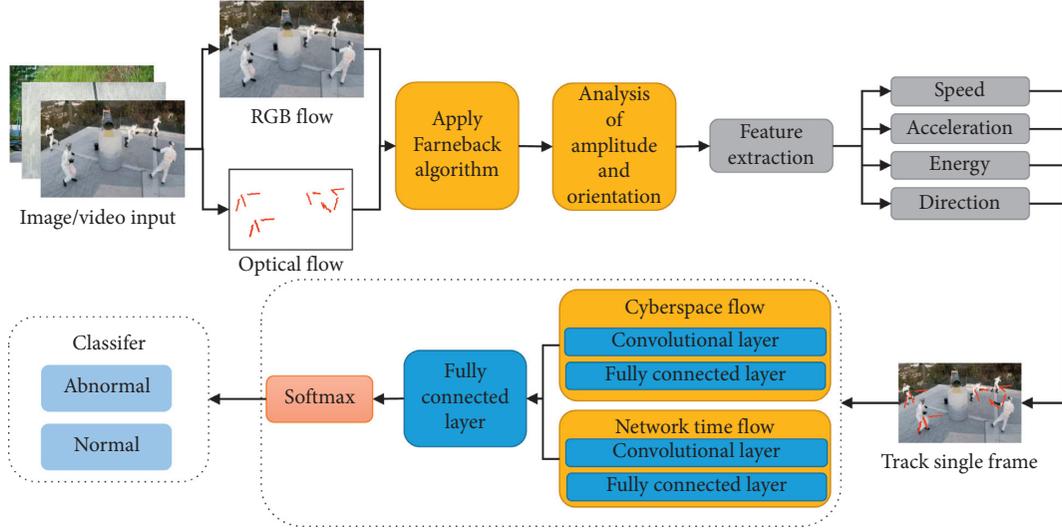


FIGURE 1: Abnormal behaviour detection algorithm based on global optical flow.

The specific steps of the algorithm are as follows:

- (1) First, a Gaussian pyramid is established for each frame. As the number of pyramid layers' increases, the resolution of the image gradually decreases.
- (2) Calculate optical flow. Calculate continuously from the top layer, by minimizing the error and calculating the optical flow in the top image:

$$Of = \sum \sum (F(a, b) - H(a + w_a, b + w_b))^2. \quad (2)$$

The displacement of each layer is

$$w_n = \frac{w}{2^n}. \quad (3)$$

Among them, w is the displacement size of the original image and n is the number of layers.

- (3) The input of each layer of image is the output result of the previous layer; connect it to the end, and calculate the optical flow value of each layer in sequence:

$$Of_{-v_{n-1}} = 2 * (Of_{-v_n} + w_n). \quad (4)$$

- (4) It can be seen that the acquisition of the optical flow value is the superposition of the optical flow vectors of all layers.
- (5) Calculate from top to bottom along the pyramid, repeat the estimation operation, and get the optical flow size of the bottom image as

$$Of_{-v} = (Of_{-v_0} + w_0). \quad (5)$$

The schematic diagram visually shows the simple implementation process of the LK optical flow algorithm, as shown in Figure 2. The advantage of the pyramid LK optical flow method is that the optical flow value calculated for each time is relatively small, and the result is to amplify all the

values cumulatively. In this way, a smaller neighbourhood window can handle larger pixel motions.

3.3. Trajectory Single-Frame Image Extraction. Optical flow information represents the instantaneous information of motion, while the trajectory contains continuous motion information. When an emergency occurs, sudden changes in speed, acceleration, direction, and energy caused by abnormal crowd behaviour can be represented by optical flow. However, the chaotic and intersecting crowd movement trajectories are easily overlooked. In order to improve the performance of abnormal behaviour detection, this paper considers adding motion trajectory information in a single-frame image.

The detection of crowd abnormal behaviour in medium- and high-density places has problems such as intergroup occlusion and small individual target size. The use of Kalman filtering and YOLO network are susceptible to pedestrian occlusion or collision, which makes it impossible to obtain complete individuals. The target further reduces the target tracking performance, as shown in Figures 3(a) and 3(b). In response to the above problems, this paper regards individuals as moving particles and uses the KLT feature point-tracking algorithm to track the moving particles in the group. In this paper, Harris corner extraction algorithm is used in the KLT feature point-tracking algorithm to obtain stable and reliable tracking performance. Different from the traditional multitarget tracking algorithm, the proposed algorithm comprehensively considers the problems of intergroup occlusion and small individual target size and has a better detection effect for crowded people with more serious occlusion, as shown in Figure 3(c).

In normal group movement, pedestrians follow pedestrians with the same direction of travel in a self-organizing manner; that is, the trajectories are similar. Moreover, because there is a safe distance between pedestrians, the trajectories will not cross in a short time, as shown in Figure 4(a). However, when an abnormal event occurs, the

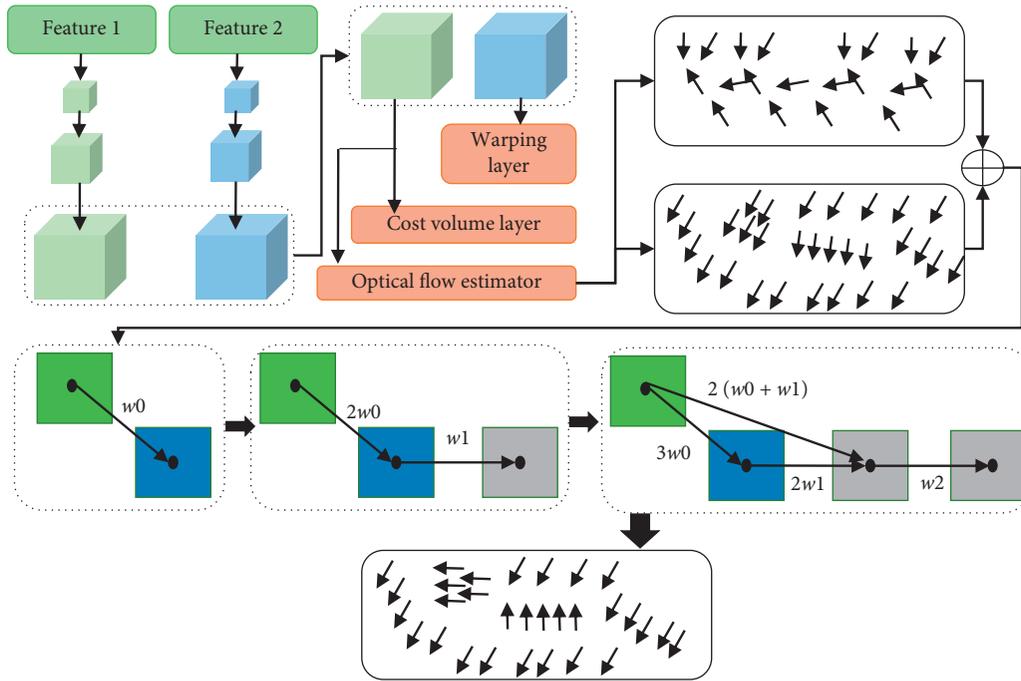


FIGURE 2: Pyramid LK optical flow diagram.



FIGURE 3: Comparison of multitarget tracking algorithms: (a) Kalman filter, (b) YOLO network, and (c) KLT.

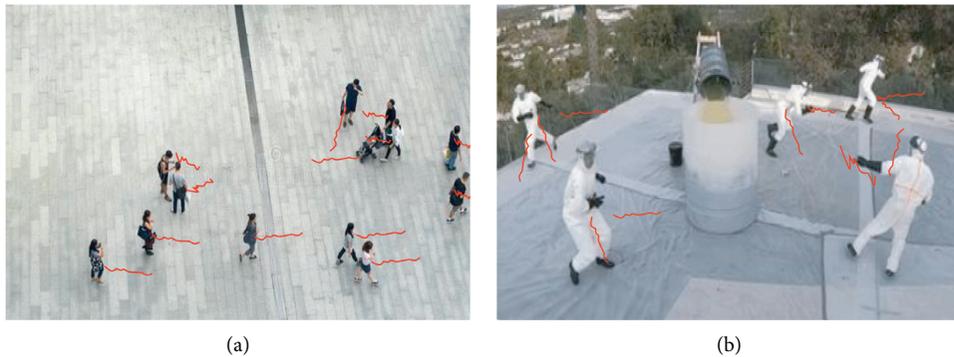


FIGURE 4: Crowd movement trajectory under normal and abnormal conditions: (a) normal scene; (b) abnormal scene.

pedestrian deviates from the expected trajectory due to panic and escapes in different directions at a speed different from the normal traveling speed. At this time, the trajectory is

chaotic and easy to collide with others, causing the trajectory to cross, as shown in Figure 4(b). In summary, the crowd trajectory helps distinguish normal and abnormal

behaviours. Therefore, the proposed method adds trajectory information to a single-frame image to improve the performance of crowd abnormal behaviour detection.

3.4. Two-Stream Convolutional Neural Network. CNN automatically learns the complex features of the image in the input CNN for image classification and recognition by using hierarchical training operations. Compared with manually generated features, this process can achieve higher efficiency and better performance. Because video sequences have more time sequence concepts than images, they introduce more information for time-dependent recognition and detection tasks. The traditional method generally divides the video into single frames and then trains and learns each frame of image, then sums and averages the confidences of all the obtained prediction values, and finally classifies and recognizes the extracted feature maps. However, this method only uses the appearance information of the video, and it is easy to pay attention to irrelevant and unimportant information. It cannot be classified accurately when the feature difference is large.

By imitating the human visual process, based on processing the video space information, understanding the video time information, while considering the space and time domain, this paper proposes the DCNN network. For the middle-to-high-density population, the size of individual targets is small and there are correlations between moving targets. If the size of the input image is greatly reduced, the correlation between more motion information and moving targets is easily lost. However, large-size input images easily lead to a large amount of calculation, and there is a problem of nonconvergence, which affects the performance of abnormal behaviour detection. In summary, this article has a uniform size of 256×256 for the input image. In order to effectively use the global optical flow descriptor and the trajectory single frame image information, retain the correlation of moving targets, and appropriately reduce the amount of calculation and control overfitting and underfitting, the first layer of the network convolution kernel is 7×7 . The purpose is to extract dynamic information such as colour, texture, and trajectory. The size of the pooling layer is 2×2 , and the maximum pooling is used to reduce repetitive information and retain important feature point information of the moving target. The second layer of convolution kernel is 5×5 to ensure that the key feature point information is not lost during feature extraction. The subsequent three layers of convolution kernels are all 3×3 in size, and more abstract and advanced features are extracted. Finally, design three fully connected layers. Since there are only two types of output, the number of nodes is two. Due to the small training sample size, dropout is added to the fully connected layer to prevent network overfitting and enhance network generalization performance.

In the softmax layer or convolutional layer, spatial and temporal information fusion is realized, and the classification function is realized only through the spatial part, which is insensitive to time and does not make good use of time information. However, the time information in the

abnormal behaviour detection based on the video sequence is the key clue and cannot be ignored. This paper merges after convolution, does not cut off the spatial network after the fusion, and continues the transmission of time and space network. After the fully connected layer, it merges again to realize the pixel correspondence of the space and time information. The network structure can be obtained as shown in Figure 5. The fusion process does not cause model parameters to be too complicated and does not cause loss of model performance. At the same time, complete spatio-temporal features can be formed, thereby improving the performance of crowd abnormal behaviour detection.

The Relu function can enhance the nonlinear characteristics of the convolutional layer and even the entire CNN and can speed up the convergence speed. Therefore, this article uses the Relu function. In addition, cross entropy can be used as a loss function to avoid slow training speed. The cross entropy loss is shown in the following formula:

$$\text{loss} = \sum a * \log(b) + (1 - a)\log(1 - b). \quad (6)$$

Among them, variable a is the true value of the sample and b is the predicted value of the sample.

4. Results and Discussion

4.1. Datasets. The network training samples in this paper mainly come from the UCSD dataset, Avenue dataset, and Behave dataset:

- (1) Ped1 and Ped2 datasets in UCSD: the Ped1 dataset contains 34 training video clips and 36 test video clips, while the Ped2 dataset has 16 training video clips and 12 test video clips. In these datasets, the training video clips contain only normal behaviour, while the test clips not only have normal behaviour but also mix with a lot of abnormal behaviour. In these two datasets, the main background is public pedestrian roads, so cycling, roller skating, and the presence of vehicles on nearby pedestrian roads are regarded as abnormal events.
- (2) Avenue dataset: the dataset contains 16 normal event training videos and 21 test videos. These videos are obtained in the real world, including normal and abnormal events. Walking on the sidewalk is a normal event, while abnormal events include running, abnormal gait or going in the wrong direction, and packet loss.
- (3) Behave dataset: surveillance video clips contain two perspectives in which people interact with each other, and capture data at a rate of 25 frames per second. The original team members developed several basic categories, namely: free gathering of people, separation of people, chasing, fighting, and running together. Among them, 40 video clips containing only normal behaviour are cut as training datasets, and 25 video clips containing normal behaviour and abnormal behaviour are used as test datasets.

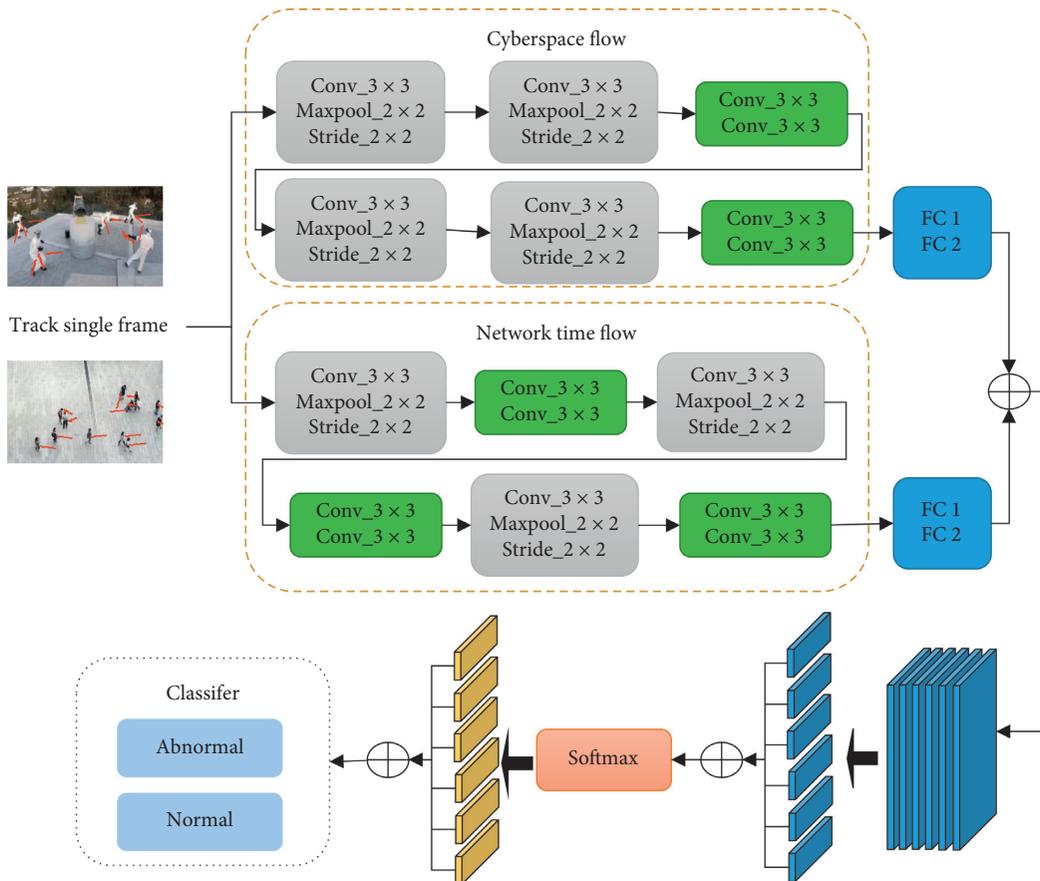


FIGURE 5: DCNN network structure.

The above is the dataset used in this experiment, the training set only contains normal behaviour fragments, and the test set mixes abnormal behaviour fragments with normal behaviour fragments, in which part of the training set is divided into parts for verification. The specific dataset distribution is shown in Table 1.

4.2. Effectiveness of Multimodal Input. AUC and EER are indicators to evaluate model performance. Since the network uses RGB images and optical flow images as inputs, we use different inputs to calculate the AUC of the network to obtain the true positive rate and false positive rate, thereby obtaining EER. It can be seen from Figure 6 that the multimode input mode proposed in this paper is better than other input modes. In addition, the RGB feature and the optical flow feature were used as input for testing separately. From Figures 6(a) and 6(b), it can be seen that the network that only uses optical flow as input is in the abnormal appearance of the UCSD Ped1. It performs well on the dataset because RGB features are more sensitive to appearance. Nevertheless, the results of other datasets show that the network with multiple inputs has higher AUC and smaller EER than the single-input network. In short, multiinput models make up for their respective shortcomings and outperform the latest methods on public datasets.

TABLE 1: Experimental dataset description.

Dataset	Ped1	Ped2	Avenue	Behave
Training set	13000	12242	15210	15731
Validation set	1000	1100	1300	1300
Testing set	5000	4000	4630	3023

In addition, RGB features and optical flow features are used as inputs to get regularity scores. Figure 7 shows the abnormal score curves of optical flow feature input, RGB feature input, and the fusion of the two as input for the same test video segment. The experimental results clearly show that the model obtained by fusing optical flow features and RGB features can more effectively detect whether there are abnormal behaviours in the test video. Compared with multifeature input, due to the complexity of the environment when detecting abnormal events, single-feature input will have a certain error, while multifeature input integrates RGB features and optical flow features, which can get better results on the dataset.

4.3. Performance Analysis of Abnormal Behaviour Detection. The ROC curve on the Avenue dataset is shown in Figure 8. The figure shows an intuitive comparison of the detection results of this method and the other three methods at the

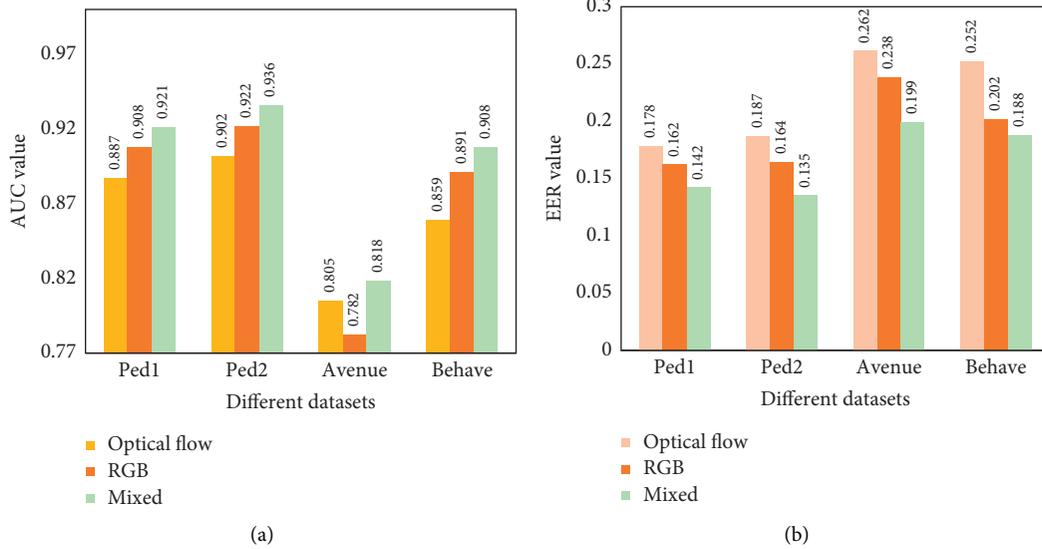


FIGURE 6: AUC and EER of different input modes: (a) AUC comparison of different inputs for the same algorithm; (b) EER comparison of different inputs for the same algorithm.

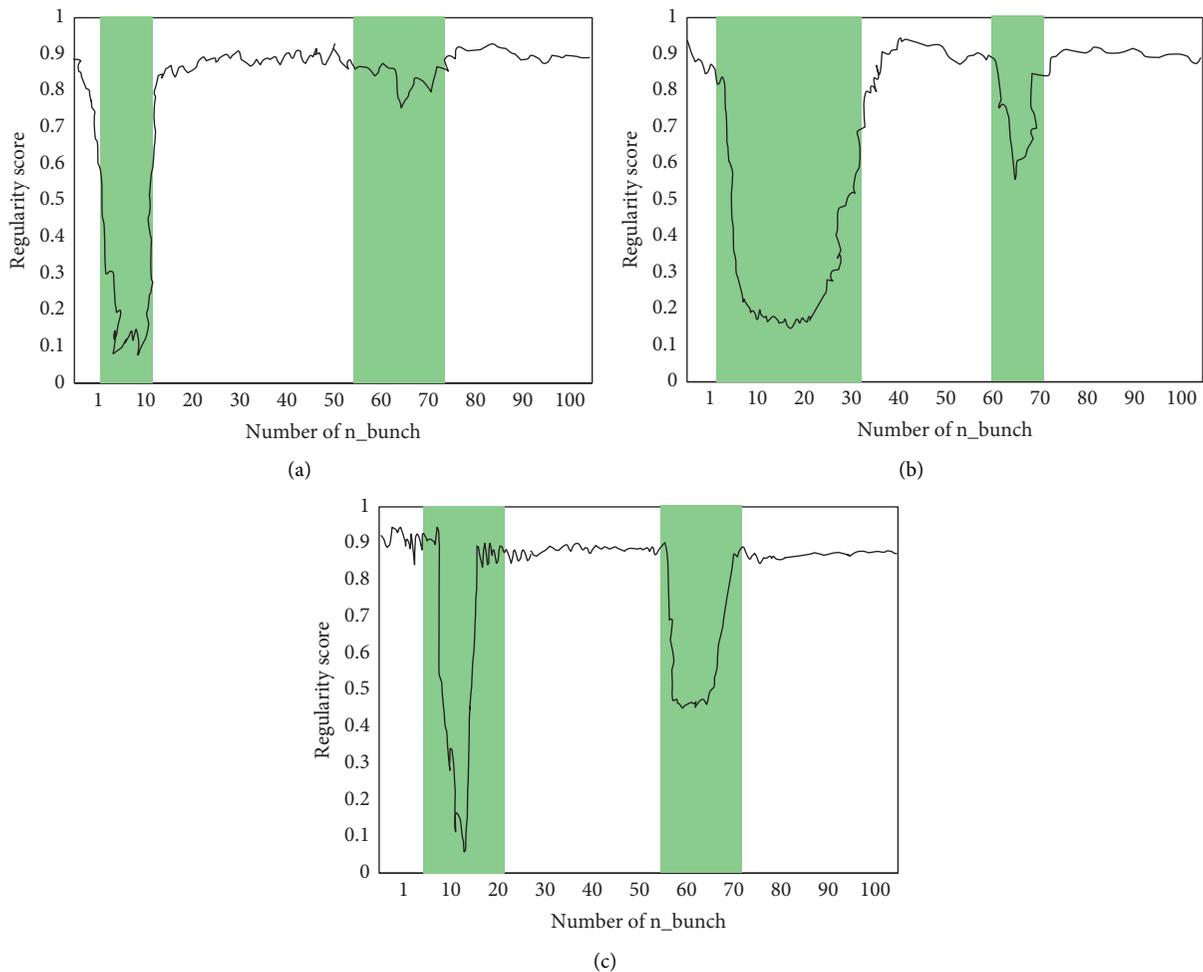


FIGURE 7: The regularity score curve of the same data with different inputs: (a) only use optical flow information as input; (b) only RGB information as input; (c) mixed information as input.

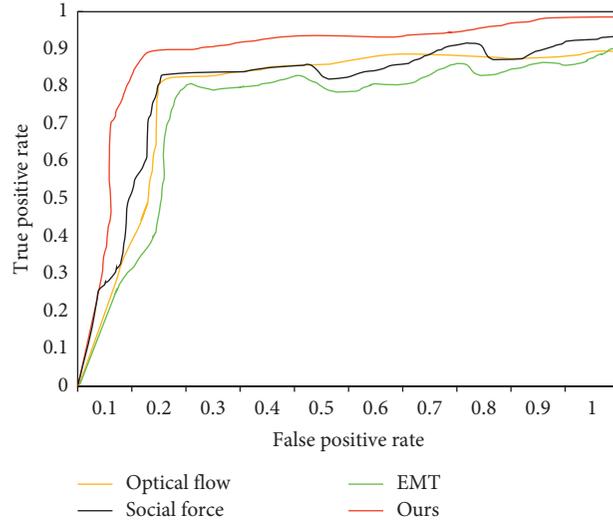


FIGURE 8: ROC curve on the Avenue dataset.

TABLE 2: AUC and EER for global abnormal behaviour detection in the Avenue dataset.

Method	AUC	EER
Optical flow	0.839	0.278
Social force	0.927	0.136
EMT	0.962	0.135
Ours	0.976	0.128

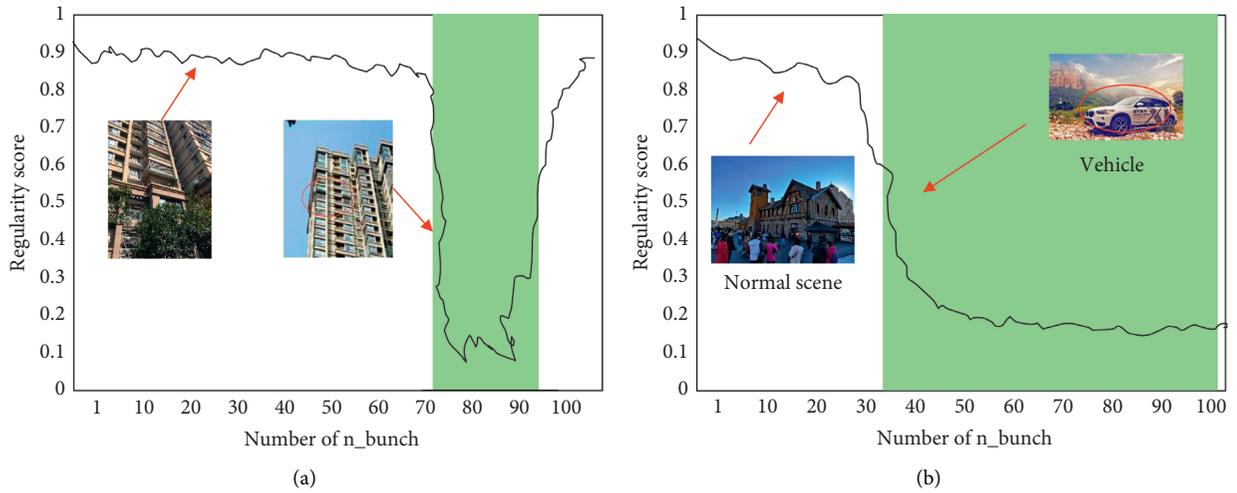


FIGURE 9: Continued.

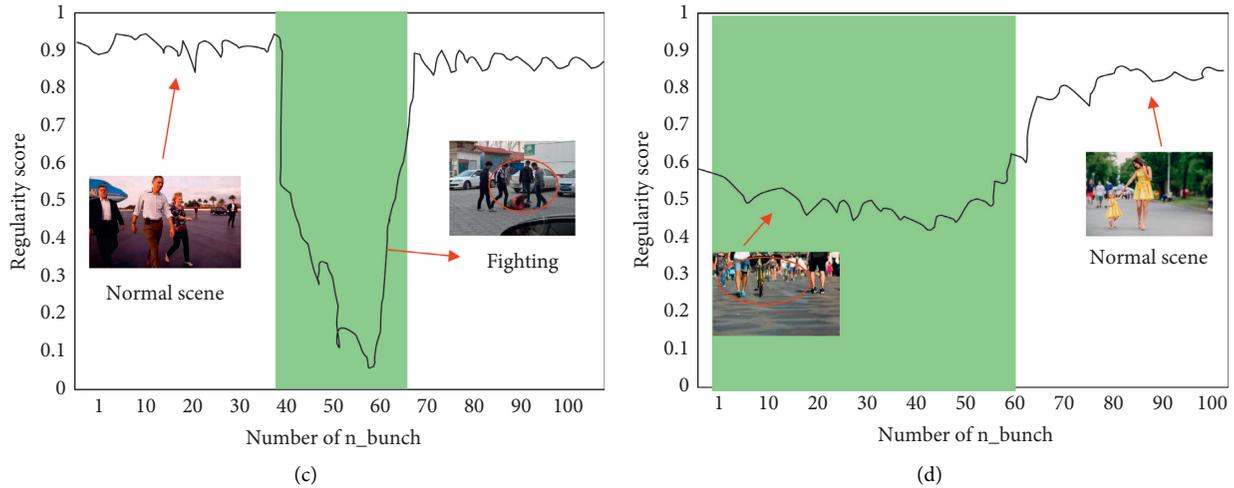


FIGURE 9: The regularity score curve of abnormal fragments: (a) abnormal fragment 1; (b) abnormal fragment 2; (c) abnormal fragment 3; (d) abnormal fragment 4.

frame level. The other three methods are optical flow method [41], method based on social force model [42], and energy model with threshold method (EMT) [43]. It can be seen from Figure 8 that the optical flow method only uses the size and intensity information of the optical flow and the global anomaly will be lost; the social force model has an effect on the individual's hope based on the optical flow. The resultant force of force and social interaction force is calculated, and there is a misdetection caused by the maximum local correlation. The experimental results can be drawn that the method proposed in this article has certain advantages over the other three methods. The method proposed in this paper can suppress the interference of surrounding environmental factors, and the algorithm is more stable and effective than other methods.

Table 2 shows a comparison of the AUC value and EER value including the method in this article and the other three methods. Observing the data in the observation table, the method proposed in this paper has the largest AUC value and the smallest EER value, indicating that the method in this paper has both high accuracy and low misjudgement rate.

In order to observe the effect of the algorithm more intuitively, the experimental results are visualized, and the normal behaviour score curve of the video to be tested is drawn, as shown in Figure 9. The experimental results show that using the method in this paper, abnormal event can be correctly detected in surveillance videos in public places. The green area indicates when the actual abnormal behaviour occurred. The model only learns normal behaviour events, but no abnormal events. When an abnormality occurs in the

video segment to be tested, the reconstruction error will increase and the regularity score will decrease.

5. Conclusion

Crowd abnormal behaviour detection technology is one of the research hotspots in the field of vision, and it is an important part of the field of intelligent surveillance. It is widely used in the smart security of airports, shopping malls, schools, and communities. With the continuous deepening of research on crowd abnormal behaviour detection algorithms, significant progress has been made. However, it is still a challenging task to perfect the detection technology of abnormal behaviour in complex scenes. In order to improve the accuracy and robustness of the crowd abnormal behaviour detection algorithm, for the low performance of crowd abnormal behaviour detection caused by complex environment, this paper proposes a crowd abnormal behaviour detection algorithm based on global optical flow. First, the optical flow information between the RGB image and the video frame is used as the input of the network. After that, the motion trajectory map is extracted based on the single frame original image. Finally, a video abnormal behaviour detection algorithm based on the dual-stream convolutional neural network is proposed. Experimental results show that the proposed method can achieve better recognition results on multiple datasets, and the use of interframe motion information can significantly improve the performance of abnormal behaviour detection.

Although the algorithm in this paper can achieve better recognition effect on multiple datasets by using interframe motion information, the algorithm only works with two

frames. Our future work is to further improve the performance of abnormal behaviour detection by simultaneously utilizing the information between multiple frames of images.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Key Research and Development Program of Shaanxi, a Research of Crowd Abnormal Gathering Behaviors Detection in Surveillance Videos (2019GY-054), and Shaanxi Smart City Technology Project of Xianyang, Crowd Congestion Detection System Based on Spatiotemporal Texture (2017k01-25-5).

References

- [1] A. Al-Dhamari, R. Sudirman, and N. H. Mahmood, "Transfer deep learning along with binary support vector machine for abnormal behavior detection," *IEEE Access*, vol. 8, pp. 61085–61095, 2020.
- [2] X. Hu, J. Dai, Y. Huang et al., "A weakly supervised framework for abnormal behavior detection and localization in crowded scenes," *Neurocomputing*, vol. 383, pp. 270–281, 2020.
- [3] S. Benmakrelouf, C. St-Onge, N. Kara, H. Tout, C. Edstrom, and Y. Lemieux, "Abnormal behavior detection using resource level to service level metrics mapping in virtualized systems," *Future Generation Computer Systems*, vol. 102, pp. 680–700, 2020.
- [4] J. Wang, N. Wang, L. Li, and Z. Ren, "Real-time behavior detection and judgment of egg breeders based on YOLO v3," *Neural Computing and Applications*, vol. 32, no. 10, pp. 5471–5481, 2020.
- [5] C. Direkoglu, "Abnormal crowd behavior detection using motion information images and convolutional neural networks," *IEEE Access*, vol. 8, pp. 80408–80416, 2020.
- [6] X. Gao, "Abnormal behavior detection and warning based on deep intelligent video analysis for geriatric patients," *Journal of Medical Imaging and Health Informatics*, vol. 11, no. 2, pp. 345–352, 2021.
- [7] Q. Cheng, Y. Zhou, Y. Feng, and Z. Liu, "An unsupervised ensemble framework for node anomaly behavior detection in social network," *Soft Computing*, vol. 24, no. 9, pp. 6421–6431, 2020.
- [8] S. K. Jarraya, M. Masmoudi, and M. Hammami, "A comparative study of Autistic Children Emotion recognition based on Spatio-Temporal and Deep analysis of facial expressions features during a Meltdown Crisis," *Multimedia Tools and Applications*, vol. 80, no. 1, pp. 83–125, 2021.
- [9] J. Hu, X. Zhang, and S. Maybank, "Abnormal driving detection with normalized driving behavior data: a deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 6943–6951, 2020.
- [10] C.-T. Yang, J.-C. Liu, E. Kristiani, M.-L. Liu, I. You, and G. Pau, "NetFlow monitoring and cyberattack detection using deep learning with ceph," *IEEE Access*, vol. 8, pp. 7842–7850, 2020.
- [11] O. Ye, J. Deng, Z. Yu, T. Liu, and L. Dong, "Abnormal event detection via feature expectation subgraph calibrating classification in video surveillance scenes," *IEEE Access*, vol. 8, pp. 97564–97575, 2020.
- [12] T. Wang, M. Qiao, A. Zhu, G. Shan, and H. Snoussi, "Abnormal event detection via the analysis of multi-frame optical flow information," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 304–313, 2020.
- [13] Y. Yu and N. Bian, "An intrusion detection method using few-shot learning," *IEEE Access*, vol. 8, pp. 49730–49740, 2020.
- [14] K.-E. Ko and K.-B. Sim, "Deep convolutional framework for abnormal behavior detection in a smart surveillance system," *Engineering Applications of Artificial Intelligence*, vol. 67, pp. 226–234, 2018.
- [15] S. Xie, X. Zhang, and J. Cai, "Video crowd detection and abnormal behavior model detection based on machine learning method," *Neural Computing and Applications*, vol. 31, no. 1, pp. 175–184, 2019.
- [16] J. Wang and L. Xia, "Abnormal behavior detection in videos using deep learning [J]," *Cluster Computing*, vol. 22, no. 4, pp. 9229–9239, 2019.
- [17] Y. Hu, "Design and implementation of abnormal behavior detection based on deep intelligent analysis algorithms in massive video surveillance," *Journal of Grid Computing*, vol. 18, no. 2, pp. 227–237, 2020.
- [18] Q. Ke, S. Zeng-Guo, Y. Liu, W. Wei, M. Woźniak, and R. Scherer, "High-resolution SAR image despeckling based on nonlocal means filter and modified AA model," *Security and Communication Networks*, vol. 2020, Article ID 8889317, 8 pages, 2020.
- [19] W. Wei, Q. Ke, J. Nowak, M. Korytkowski, R. Scherer, and M. Woźniak, "Accurate and fast URL phishing detector: a convolutional neural network approach," *Computer Networks*, vol. 178, Article ID 107275, 2020.
- [20] P. Tian and Q. Donglian, "Outlier detection method based on compressed time series and voronoi diagram for power data," *Electric Power Construction*, vol. 38, no. 5, pp. 105–110, 2017.
- [21] J. Jiang, F. Ni, Q. Dong, L. Yao, and X. Ma, "Investigation of the internal structure change of two-layer asphalt mixtures during the wheel tracking test based on 2D image analysis," *Construction and Building Materials*, vol. 209, pp. 66–76, 2019.
- [22] X.-H. Chen and J.-H. Lai, "Detecting abnormal crowd behaviors based on the div-curl characteristics of flow fields," *Pattern Recognition*, vol. 88, pp. 342–355, 2019.
- [23] Y. Li and Y. Zhang, "Robust infrared small target detection using local steering kernel reconstruction," *Pattern Recognition*, vol. 77, pp. 113–125, 2018.
- [24] Z. Wang, J. Zhong, and Z. Pan, "Research of cooling load & chiller selection in metro station," *International Journal of Low-Carbon Technologies*, vol. 15, no. 4, pp. 506–512, 2020.
- [25] J. Ma, Y. Dai, Y. Dai, and K. Hirota, "A survey of video-based crowd anomaly detection in dense scenes," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 21, no. 2, pp. 235–246, 2017.
- [26] X. Wei, J. Du, M. Liang, and L. Ye, "Boosting deep attribute learning via support vector regression for fast moving crowd counting," *Pattern Recognition Letters*, vol. 119, pp. 12–23, 2019.
- [27] H. Ullah, M. Uzair, M. Ullah, A. Khan, A. Ahmad, and W. Khan, "Density independent hydrodynamics model for

- crowd coherency detection,” *Neurocomputing*, vol. 242, pp. 28–39, 2017.
- [28] S. D. Khan, “Congestion detection in pedestrian crowds using oscillation in motion trajectories,” *Engineering Applications of Artificial Intelligence*, vol. 85, pp. 429–443, 2019.
- [29] J. Zhang, G. Ma, Z. Dai, R. Ming, X. Cui, and R. She, “Numerical study on pore clogging mechanism in pervious pavements,” *Journal of Hydrology*, vol. 565, pp. 589–598, 2018.
- [30] L. Cao, Y. Wang, B. Zhang, Q. Jin, and A. V. Vasilakos, “GCHAR: an efficient group-based context-aware human activity recognition on smartphone,” *Journal of Parallel and Distributed Computing*, vol. 118, pp. 67–80, 2018.
- [31] F. S. Hänseler, J. P. A. van den Heuvel, O. Cats, W. Daamen, and S. P. Hoogendoorn, “A passenger-pedestrian model to assess platform and train usage from automated data,” *Transportation Research Part A: Policy and Practice*, vol. 132, pp. 948–968, 2020.
- [32] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, “Soft + Hardwired attention: an LSTM framework for human trajectory prediction and abnormal event detection,” *Neural Networks*, vol. 108, pp. 466–478, 2018.
- [33] K. Xu, X. Jiang, and T. Sun, “Anomaly detection based on stacked sparse coding with intraframe classification strategy,” *IEEE Transactions on Multimedia*, vol. 20, no. 5, pp. 1062–1074, 2018.
- [34] R. Alessandri, P. C. T. Souza, S. Thallmair, M. N. Melo, A. H. De Vries, and S. J. Marrink, “Pitfalls of the martini model,” *Journal of Chemical Theory and Computation*, vol. 15, no. 10, pp. 5448–5460, 2019.
- [35] S. Amraee, A. Vafaei, K. Jamshidi, and P. Adibi, “Abnormal event detection in crowded scenes using one-class SVM,” *Signal, Image and Video Processing*, vol. 12, no. 6, pp. 1115–1123, 2018.
- [36] A. Al-Sheary and A. Almagbile, “Crowd monitoring system using unmanned aerial vehicle (UAV),” *Journal of Civil Engineering and Architecture*, vol. 11, no. 11, pp. 1014–1024, 2017.
- [37] A. Deng and B. Hooi, “Graph neural network-based anomaly detection in multivariate time series,” in *Proceedings of the 2021 35th AAAI Conference on Artificial Intelligence*, Vancouver, Canada, February 2021.
- [38] R. C. Ripan, I. H. Sarker, S. M. M. Hossain et al., “A data-driven heart disease prediction model through K-means clustering-based anomaly detection,” *SN Computer Science*, vol. 2, no. 2, pp. 1–12, 2021.
- [39] S. Tariq, H. Farooq, A. Jaleel, A.-U. Rehman, and S. A. Wasif, “Anomaly detection with particle filtering for online video surveillance,” *IEEE Access*, vol. 9, pp. 19457–19468, 2021.
- [40] H. D. Nguyen, K. P. Tran, S. Thomassey, and M. Hamad, “Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management,” *International Journal of Information Management*, vol. 57, Article ID 102282, 2021.
- [41] Y. Z. Hsieh and Y. L. Jeng, “Development of home intelligent fall detection IoT system based on feedback optical flow convolutional neural network,” *IEEE Access*, vol. 6, pp. 6048–6057, 2017.
- [42] R. Chaker, Z. A. Aghbari, and I. N. Junejo, “Social network model for crowd anomaly detection and localization,” *Pattern Recognition*, vol. 61, pp. 266–281, 2017.
- [43] H. Wu, Y. Hao, and J.-H. Weng, “How does energy consumption affect China’s urbanization? New evidence from dynamic threshold panel models,” *Energy Policy*, vol. 127, pp. 24–38, 2019.