Journal of Artificial Intelligence and Data Mining (JAIDM), Vol. 10, No. 2, 2022, 171-183.

Shahrood University of Technology

Journal of Artificial Intelligence and Data Mining (JAIDM)
Journal homepage: http://jad.shahroodut.ac.ir

Research paper

Abnormal Behavior Detection over Normal Data and Abnormal-Augmented Data in Crowded Scenes

Vahid Fazel Asl, Babak Karasfi* and Behrooz Masoumi

Department of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran.

Article Info

Article History:
Received 08 August 2021
Revised 13 November 2021
Accepted 26 January 2022

DOI:10.22044/jadm.2022.11061.2253

Keywords:
Surveillance Video Systems, Abnormal Event Detection, Data Augmentation, Crowd Behavior Analysis, Cycle Generative Adversarial Networks.

**Corresponding author:*
karasfi@qiau.ac.ir. (B. Karasfi).

Abstract

In this article, we consider the problems of the abnormal behavior detection in a high-crowded environment. One of the main issues in the abnormal behavior detection is the complexity of the structure patterns between the frames. In this work, the social force and optical flow patterns are used in order to prepare the system for training the complexity of the structural patterns. The cycle generative adversarial net (cycle GAN) system is used to train the behavioral patterns. The two models of normal and abnormal behavioral patterns are used in order to evaluate the accuracy of the system detection. In the case of the abnormal patterns used for training, due to the lack of this type of behavioral pattern, which is another challenge in detecting the abnormal behaviors, the geometric techniques are used to augment the patterns. If the normal behavioral patterns are used for training, there is no need to augment the patterns since the normal patterns are sufficient. Then using the cycle GANs, the normal and abnormal behavior trainings are considered separately. This system produces the social force and optical flow pattern for the normal and abnormal behaviors on the first and second sides. We use the cycle GAN system both to train the behavioral patterns and to assess the accuracy of the abnormal behavior detection. In the testing phase, if the normal behavioral patterns are used for training, the cycle GAN system should not be able to reconstruct the abnormal behavioral patterns with a high accuracy. In addition, if the abnormal behavioral patterns are used for training, the cycle GAN system should be able to reconstruct the abnormal behavioral patterns with a high accuracy. In each one of the above cases, it follows that the cycle GAN system could detect the abnormal behaviors with a high accuracy. In the experimenal section, the results from the normal and abnormal behavior trainings are compared with the other methods. The experimental results obtained from the databases are related to the abnormal behavior, showing that the method used has better results than the other similar methods in this field. When the system is trained for the normal behaviors, the accuracy obtained is 98.9% in the frame-level in the database ped1, 95.4% in the database ped2, and 75.8% in the pixel-level in the database ped1. In addition, when the system is trained for the abnormal behaviors, the accuracy obtained is 98.4% in the frame-level in the database ped1, 94.2% in the database ped2, and 73.9% in the pixel-level in the database ped1.

1. Introduction

Today, one of the exciting and essential subjects in video surveillance is to detect the abnormal behaviors in a high-crowded environment. Establishing security in this environment has always been one of the critical concerns. Accordingly, cameras have been installed in the streets, stadiums, shopping malls, parks, etc. Most of the time, the possibility of such behavior occurrence in the environment is remote; thereby, detection of such behaviors requires a high accuracy. Therefore, for detecting such phenomena, the humans' workforce and manual intervention will not be efficient, and automatic detection is necessary when such phenomena occur. In the recent years, the researchers have presented some models and algorithms for the abnormal behavior detection. These models have resolved the abnormal behavior detection challenges such as the difficulty of network training [1], individuals' occlusion [2-4], low resolution of the video, and modeling population behavior [2], various movements in the population [5, 6], problems related to the sparse population [7], and shortage of the visual features [8]. There are two major challenges in detecting the abnormal behaviors that have been extensively researched in the recent years, and these challenges have not yet been fully resolved. First, the lack of sufficient data has caused problems for network training in the abnormal behavior detection. When the network is not well-trained, the detection operation is not performed well in the test stage. It is very challenging for the systems that use deep methods of learning [9, 10]. The second challenge in detecting the abnormal behaviors is the complexity of the structure patterns between the frames that have caused problems in identifying the abnormal behaviors. In other words, when the network is not properly trained for the desired pattern of abnormal behavior, which may be complex, the detection operation is not performed well in the test stage.

In this work, the data augmentation techniques and the cycle GAN system [11] are used in order to solve these challenges; this network will be trained for the existing behaviors. However, training of the normal and abnormal behaviors in these networks could be performed separately. In the training normal behavior, the network training will be done based on the normal behaviors due to the shortage of anomalous data. Then in the test stage, the system diagnoses the reconstruction error of abnormal behaviors. In other words, this model will train only the normal behaviors. As a result, this system cannot reconstruct the abnormal behaviors in the test stage; therefore, identifying the abnormal behaviors will be done. In training the abnormal behaviors, first, the abnormal behaviors will be augmented using the data augmentation techniques. Then we will train the abnormal behaviors in the system, and as a result, this system can reconstruct the abnormal behaviors in the test stage; therefore, identifying the abnormal behaviors will be done.

In this paper, first and in the state of training the normal behavior, there is no need for data augmentation due to the sufficiency of normal patterns. Still, in training the abnormal behaviors due to the insufficiency of the abnormal behavioral patterns, the abnormal behaviors are augmented using the geometric techniques. Then we will train the social force and optical flow model of the constructed frames. The purpose of producing the social force and optical flow models is better training of the networks based on temporal features for identifying the complexities existing in the behaviors.

In the following and the second section, some issues from the previous studies including the abnormal behavior detection and structure of cycle GAN system will be presented. In the third section, the details of the proposed method for the abnormal behavior detection will be described. In the fourth stage, the results are presented, and in the fifth stage, the conclusion will be presented.

2. Related Works

In the recent years, the abnormal behavior detection methods have faced challenges, and many researchers have tried to solve them. The important challenges include the complexity of understanding the environmental behaviors [12], occlusion [13], storage space and lighting changes [14], tampered images [15], multi-scale objects [16], population movement changes, complex calculations movement of people [17], and providing all the information related to the activity of people [18].

In the recent years, the researchers have proposed the methods to resolve each one of these challenges. In the following, the qualitative criteria are proposed in order to evaluate these methods, and then each one of the methods is evaluated by these criteria. The qualitative criteria are accuracy, occlusion management, background complexity monitoring, specificity, and sensitivity [19]. Table 1 shows the evaluation results of these methods.

Accuracy: this criterion is included in different standards such as ROC, AUC, and EER. The ROC curve is a performance criterion for the difficulties of classification called the communication curve. This curve draws the TPR amount versus the FPR amount so that the y-curve includes the TPR amount and the x-curve includes the FPR amount. Like the ROC criterion, the AUC criterion is used for the difficulties of classification. This criterion shows the degree or measure of separation. This criterion refers to this point that a model with more capability of separability is a better model. For example, it could separate the normal classes from the abnormal classes with a high accuracy. The EER criterion represents the percent of frames with a correct classification, and this criterion is computed in two pixels and frame levels.

Occlusion management (OM): When the fixed and static objects emerge between the moving objects and the cameras, occlusion occurs. In order to prevent a false detection and a factitious warning, supervision on this subject is essential. Based on occlusion, some parts of the objects may not subtend in front of the camera or have a standard section with other things in the picture. Accordingly, the processing operations on the item will not run well.

Background complexity monitoring (BCM): The existence of tiny and large objects in the image, the inappropriate distance of the camera from the individuals, the clutter of the background and the presence of silhouette could be mentioned as the background complexities.

Specificity: This criterion shows the capability of the normal behavior processing.

Sensitivity: This criterion shows the capability of the abnormal behavior processing.

After introducing the evaluation criteria, the methods used to solve the challenges in the recent years will be evaluated based on the evaluation criteria. In [20], using the CNN network, in [21], through extracting the spatial-temporal structure-based features, and in [22], using the SMMIP method and reducing dimensions of traits through the PCA method, the abnormal behaviors have been detected with a high accuracy, and the results obtained show that these methods are appropriate for complicated environments, and they support the partial occlusion. H.Chebi *et al*. [23] by using the ANN network and A.E.Gunduz *et al*. [17] through utilizing the ORB and HMM techniques have tried to eliminate the high occlusion environments, and their findings show that these methods are appropriate for the environments that have many occlusions, and somehow support the environments with a high complexity. The results obtained show that paper 23 is more accurate in the abnormal behavior detection than paper 17. In [24], by considering the motion patterns of the objects and using the HMOFP method, in [25] by extracting and separating things based on the features (movement, shape, and region) and using the CNN and LSTM network and SVM classifier, and in [26], by considering the motion patterns of the objects and using the optical flow method and SVM classifier, the abnormal behaviors have been the detected with high accuracy. The results obtained show that these methods are not appropriate for the environments with increased complexity and occlusion. X.Li *et al*. [27] have used the technique of extracting volume, traffic speed, and optical flow, and the HMM methods to detect the traffic and abnormal behaviors, and the result show the high accuracy of these methods. This method is not appropriate for the environment with a high occlusion but can be used in the environments with a complicated background.

Table 1. Evaluation of existing methods in SBBA system according to qualitative criteria.

Environment	Operations	Method	ACC	OM	BCM	Specificity	Sensitivity
Crowded	Detection	[20-22]	HIGH	MEDIUM	HIGH	LOW	HIGH
		[23]	HIGH	HIGH	MEDIUM	LOW	HIGH
		[17]	MEDIUM	HIGH	MEDIUM	LOW	HIGH
		[24-26]	HIGH	LOW	MEDIUM	LOW	HIGH
		[27]	HIGH	LOW	HIGH	MEDIUM	HIGH

In the recent years, the high-performance methods have been proposed in order to solve the challenges of the abnormal behavior detection but now the two challenges of the complexity of the structural patterns between the frames and the lack of sufficient training data for the abnormal behaviors are important challenges, and many studies have been proposed to solve it.

For the abnormal behavior detection in a high-crowded environment, finding these behaviors' structure and movement patterns is essential due to the instability of the behavior type. In the recent years, the researchers have done various studies in this case. Wang *et al*. [21] have used the spatial-temporal structure model and the public and local event description.

Finding the movement and structure patterns in the frames, if there are unexpected objects or different walking styles, etc., will cause some challenges. Using the SMMIP structure and the Gaussian mixed model [22] have solved this challenge. In this model, the threshold value used for the abnormal behavior detection and the provided model will grow increasingly by upgrading the parameters. In some environments, diagnosing a specific behavior entirely different from the behaviors in the ambiance is a challenge. Manfredi *et al*. [28] have resolved this challenge. First, they obtained the path information related to each movement through the inner population densities, and classified the paths and behaviors using SVM. In [2], Chalker *et al*. have used the social network in order to solve the inherent problem of modeling the population behavior. In other words, first, the video scenes are divided into the spatial and temporal levels using the window-based theory. Then the diagnosis of behavior in a high-crowded environment will be considered using a local social network. Various movements in the ambiance cause the complexity of the abnormal behavior detection in the perimeter. In order to solve this challenge, analysis of the spatial-temporal movement patterns [5] can identify the abnormal movements and sudden changes.

In contrast to the introduced methods, the method presented in this work has used the temporal characteristics like the social force and optical flow in order to generate its model and identify the structural feature of the abnormal behaviors accurate. Since in this work we used two models of temporal pattern, the system train features will have jump, speed, etc. This issue causes an accurate training of the behavioral patterns, which is very effective in a high-accuracy detection.

Another critical challenge in the abnormal behavior detection–that has attracted significant attention in the recent years is the non-existence of the appropriate data for network training for the abnormal behavior detection [1, 9, 10, 22, 24, 26, 29-31]. In order to diagnose the abnormal behaviors in the high-crowded environments, only normal events are accessible. Therefore, implementing the semi-unsupervised method is a very effective way in this field [32]. HAMDI *et al*. have proposed a two-stream fully convolutional network architecture in order to solve this problem. In addition, a complete description of the shapes and movements that may happen in the scene was provided.

Chang *et al*. [9] have used the normal data training in order to solve the challenges of the abnormal data shortage. The structure was an auto encoder-decoder network and, in it, the way of reconstructing the normal data was trained. In the test stage, rebuilding the data is considered for the abnormal behavior detection. Using a cycle GAN system [10], the normal patterns can be considered. In this system, two GAN systems have been used. The generator of the first network produces the optical flow, and the generator of the second network produces a related frame. In the test stage, the abnormal behavior detection will be done based on the pattern reconstruction.

In this work, in contrast to the introduced methods, the augmentation techniques were used to produce the abnormal patterns. When a sufficient pattern for a specific behavioral training is provided for the network, the network identifies accurately the desired behavior. In the introduced methods, they often use the normal behavioral training to detect the abnormal behaviors, while in this work, the abnormal behavior training was used to detect the abnormal behaviors. After this stage, the network training will be done based on the normal and abnormal patterns (separately), and in the test stage, the detection operation will be done according to the pattern reconstruction. In the cycle generative adversarial nets with short-form, the cycle GAN implies a malicious network made up of two GAN networks for diagnosing the actual images from the fake ones. In each network, the data is generated by a generator (G). They are then analyzed using a discriminator (D) that is the actual data or it has been produced by G (the fake one) [33]. D discriminates the areas using the space of characteristic. There is no need for SVM in this method anymore, which is considered advantageous for diagnosis of the acceleration. In each network, the component G produces a fake pattern for the corresponding network according to the noise component Z. The component D must diagnose the usual and counterfeit patterns (produced by G) and send the result to the connected network. G knows to build which pattern in the next stage so that the component D cannot identify it. This stage progressed as the component D can diagnose the actual images from the fake ones in a good proportion. In [34], a cycle GAN has been provided, in which image training has been considered peer-to-peer. In other words, there are two components, G and D, in this network. Each one of the components of G has the duty of producing the fake images of the inter-connected network, and each one of the components of D identifies the actual images from the fake ones. The structure used in this work is similar to the one provided form the article [34]. In this work, a cycle GAN has been used for the abnormal behavior detection. In the cycle GAN, the generator of the first network produces a fake social force related to the sequential frames for the second network, and the second network generator undertakes the duty of making a fake optical flow related to the sequential frames for the first network.

3. Proposed Method

In this section, the details of the proposed method are provided. As one can see in Figure 1, in the stage of system training, first, the training data is provided. Due to the sufficiency of the normal

patterns, there is no need for a typical pattern augmentation but the abnormal patterns should be augmented. Then we will prepare the social force and optical flow pattern of the constructed frames.
After this stage and providing the training data, the cycle GAN is used to train the normal and abnormal data, and ultimately, the normal and abnormal training model will be obtained. Finally, in the test stage, according to the training model of the network–either normal or abnormal–the detection accuracy of the system will be considered.
In the following sections, first, it is discussed how to reconstruct and augment the abnormal frames. Then the structure of the cycle GAN for training the normal and abnormal data will be addressed. Then in the test stage, the system that has been prepared with normal patterns will be tested using the abnormal patterns. In this experiment, the greater the error of reconstructing the abnormal patterns, the more accurate the system will be.
To the contrary, the system trained with the abnormal patterns will be tested using the abnormal patterns. In this experiment, if the error of reconstructing the abnormal patterns is smaller, the system will have a more accuracy.

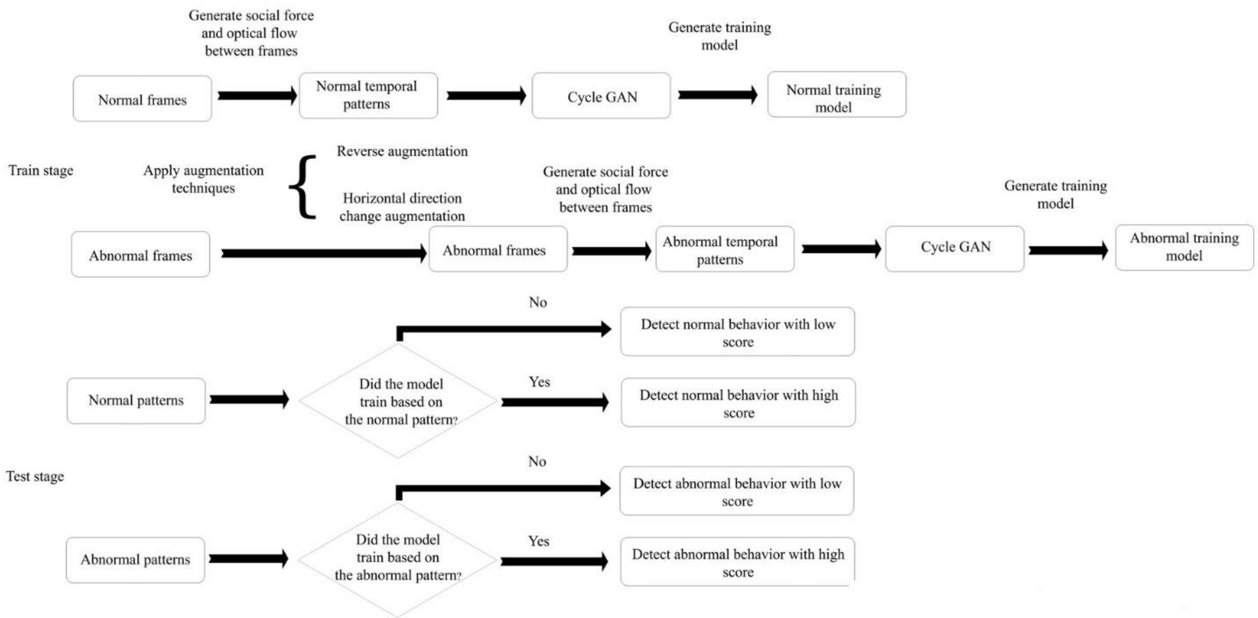


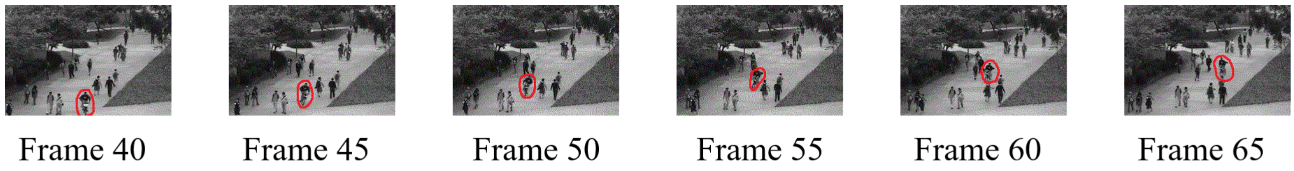
Figure 1. Diagram of proposed method.

Augmenting abnormal images

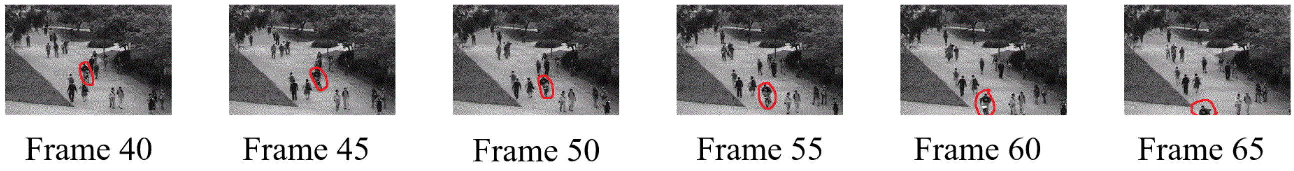
These behaviors should be trained in order to detect the abnormal behaviors to discriminate this type of behavior from the normal behavior. One of the critical requirements for the exact training of these behaviors is having sufficient data for training. Since the abnormal behaviors rarely happen in the environment and the existing databases also have insufficient data on the abnormal behaviors, we will encounter appropriate and enough data for training these behaviors. The image augmentation techniques help to augment the incomplete data using different geometrical methods. These methods include:

1. The augmentation applies changes to the pixel intensity.
2. The augmentation flips the video images horizontally and vertically.
3. The augmentation applies mirror transformations.
4. The augmentation applies the geometrical transformations.
5. The augmentation applies a collection of boosts.
6. The augmentation performs time transformations.
7. The augmentation adds noise to the images.

For example, of the best augmentation in terms of performance, we can point at the video images reversing and direction change of the video images. In Figure 2(a), an instance of frames related to reverse augmentation results, and Figure 2(b) shows an example of the frames related to the results of the horizontal direction change augmentation.



(a)



(b)

Figure 2. a) An example of abnormal image frames after applying reverse augmentation; b) An example of abnormal image frames after applying horizontal direction change augmentation.

The images were taken from the dataset ped1, and were augmented using the rotation and reverse techniques. As one can observe in Figure 2(a), the frames are the reverse of mainframes, and in Figure 2(b), the frames are in the status of 180-degree rotation than the mainframes. Thus each one of these frames differs from the primary frames. Thus using the main frames and the frames produced by different augmentations, we can have a massive volume of abnormal images and samples used for training in the GAN network[35].

Cycle GAN

After augmentation of the abnormal frames, we used the framework from paper [11] for training the normal and abnormal behaviors in the high-crowded scenes. As one can see in Figure 3, this network has two GAN networks in the cycle form, patterns of optical flow [36], and social force [37], and has been used for training the normal and abnormal behaviors.

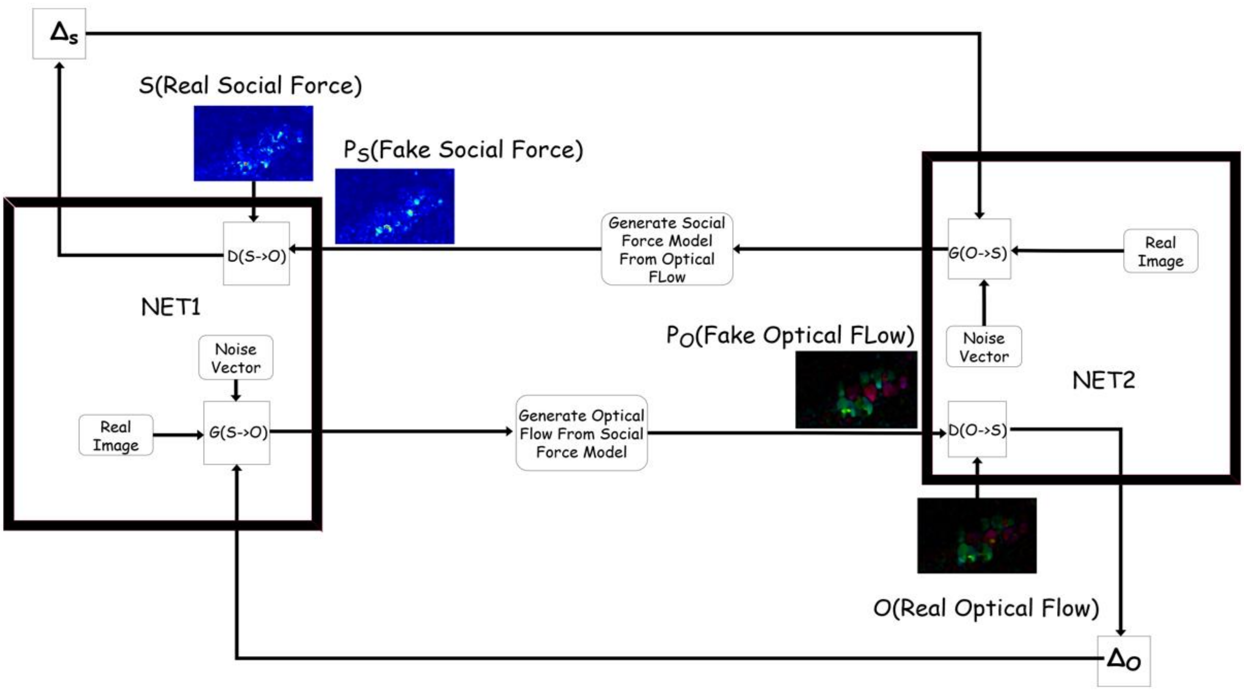


Figure 3. Generator of first network (left) generates an optical flow image from social force, and generator of second network (right) produces a social force image from optical flow.

In the left GAN, there are two components, G and D. G of this network is responsible for generating the optical flow images of the second network that corresponds to the social force, and is used as the fake images of the second network. In the right GAN network, there are two components, G and D. G of this network is responsible for generating the social force images of the first network that corresponds to the optical flow, and is used as the fake images of the first network. D in the first network is responsible for considering the actual social force patterns and the unnatural social force generated by the second network, and gives the results as a feedback to the second network. D in the second network is responsible for considering the real optical flow patterns and the fake optical flow generated by the first network, and gives a feedback to the first network. In this network, S_t was considered as a social force obtained using the frames F_t and F_{t+1}, and O_t as the optical flow obtained from the frames F_t and F_{t+1}. Thus we will have two networks: 1) $N^{S\text{->}O}$, and this network is responsible for generating an optical flow from the relevant social force, and 2) $N^{O\text{->}S}$, and this network is responsible for generating a social force from the relevant optical flow. As it was mentioned earlier, each one of these networks will have two components, G and D.

The first component of this network is the component G, in which has x (real image) and z (noise vector) are the input, and the output will be an image p = G (x, z). In the first network ($N^{S\text{->}O}$), the variable x is a social force image (x = S_t), and the input p is a reconstructed image from the corresponding optical flow that is generated by the second network. This variable p is considered as a fake social force.

The second component of this network is component D that takes two images (x, y) and (x, p), and the output indicates the probability that two images are similar or, in other words, real.

The two components, G and D, are trained using the adversarial and reconstruction error [10]. The reconstruction error means that the network has generated real images well. For example, formula 1 obtains the reconstruction image.

$$L_A(x,y) \;=\; \| y - G(x,z) \| \tag{1}$$

As one can see in formula 1, y means the real image and G is the image generated from the noise vector. The adversarial error means that component D had to guess the deviances of the G network well. When the hostile error is small, it implies that component D could distinguish the real images from the fake ones; in other words, it recognized all the deviances of component G, and it could determine the real images well. The adversarial error can be obtained from formula 2.

$$L_R(G,D) = E_{x,y \sim p_{data}(X)}[\log D_Y(x,y)] +$$

$$E_{x \sim p_{data}(X), z \in Z}[\log(1 - D_Y(G(x,z))] \tag{2}$$

In formula 2, D has the duty of distinguishing the generated images by component G from the real images. Thus in formula 2 and the first part of it, the pictures are caused without noise but the illustrations are generated with noise in the second part.

These scenarios also happen in the second network. In other words, in the second network ($N^{O\text{->}S}$), x is an optical flow image (x = O_t), and is a kind of real image. The variable p is a reconstructed image from the relevant social force generated by the first network, and is considered a fake optical flow. Component D in this network also takes the two images (x, y) and (x, p), and the output indicates the probability that the two images are similar or, in other words, actual.

In the test stage, it is expected that the system can detect the abnormal behavior patterns by giving a social force and/or an optical flow. In other words, when the normal sample has been trained to the system, it cannot reconstruct the abnormal images, and *vice versa*; when the abnormal patterns have been prepared for the design, it can reconstruct the abnormal images well.

As one can see in Figure 3, the social force and the optical flow patterns are constructed well. In the first network and the second network, the abnormal patterns have been used in the training stage. The patterns used are related to the biker passing through people, and have also been considered an the abnormal image. The exact reconstruction of the image indicates that the network has detected the abnormal images well, and its reason is that the network has been trained for an abnormal behavior well. At the time of the test, in order to know the detection accuracy of the abnormal behaviors, the reconstructed images with the main images will be compared. In other words, the original social force and generated social force pattern are compared together in the first network. In the second network, the primary optical flow and the generated optical flow patterns are compared together. That means that we will have $\Delta_s = S - p_s$, in which S is the primary social force, and p_s is the social force generated by the second network. In addition, we will have $\Delta_o = O - p_o$, in which O is the main optical flow, and p_o is the optical flow generated by the first network.

4. Experiments

In this section, the UCSD and UMN datasets estimate the measures for considering the method of abnormal behavior detection and the details of implementation, and the results will be provided. The Linux operating system with 8 GB SSD RAM, core i5 intel CPU, and 1GB Nvidia Geforce has been used for implementation. The experiment includes two stages: 1) The normal behavior training model is used as an experiment such that by reconstructing the abnormal behaviors, the accuracy of this type of model is considered. 2) The abnormal behavior training model is used as an experiment such that by reconstructing the abnormal behaviors, the accuracy of this type of model is considered.

4.1. GANs setup

In our experiments, the Adam optimization was used for stochastic gradient descent for the training model in the networks $N^{S\text{->}O}$ and $N^{O\text{->}S}$. The training data is split into batches of size 128, and all frames are resized to 256 × 256 pixels. Each network is trained for 80 epochs.

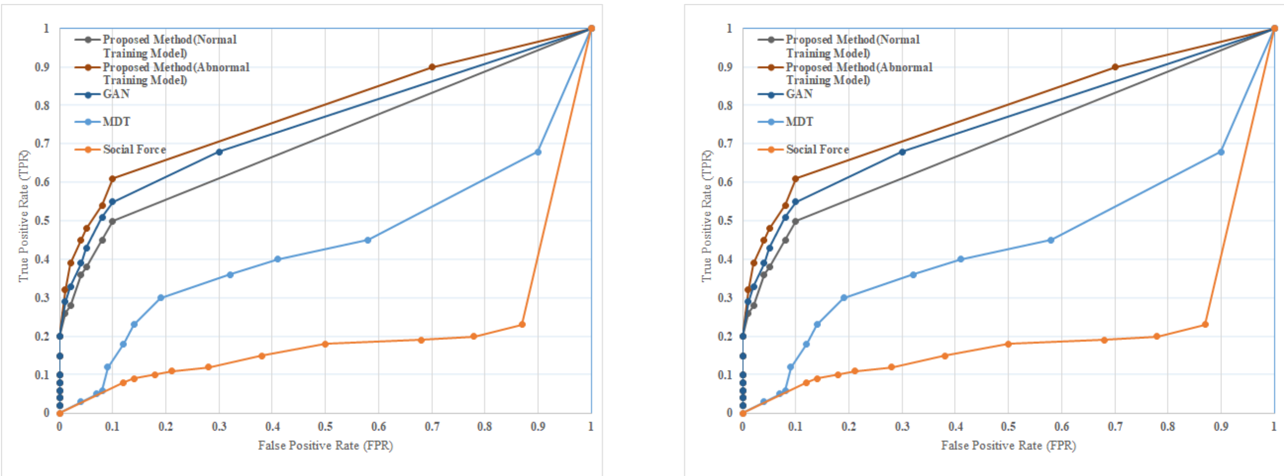
4.2. Dataset

We use two standard datasets: the UCSD dataset [38] and the UMN dataset [37]. The UCSD data collection includes the recorded video clips with the fixed camera, and these cameras are installed at a height, overlooking the path of people's walking. In this environment, the dense population changes every moment. In this environment, the abnormal behaviors are generated based on the individuals' abnormal movement patterns and/or moves of some vehicles in the path of pedestrians. The common anomalies in this environment include biking, skating, passing a small wagon or wheelchair in the crowd. This dataset is divided into two categories, ped1 and ped2. Ped1 includes the frames with resolution 158 * 238, and ped2 includes the frames with a resolution of 240 * 360 pixels. Ped1 consists of 34 clips, and ped2 consists of 16 video clips for training; 36 clips from ped1 and 12 video clips from ped2 for the experiment. The UMN dataset comprises the videos of 11 different scenarios of an escape event in 3 other indoor and outdoor scenes.

Each video consists of an initial part of the normal behavior, and ends with sequences of the abnormal behavior. In both datasets, all frames are resized to 256 * 256 pixels.

4.3. Estimated measures and results of experiments

In this work, two measures introduced in [39] including frame-level and pixel-level measures were used for estimating the abnormal detection accuracy. In both of the two measurements, true positive (TP) and false positive (FP) rates were used. Existence of the abnormal patterns and non-existence of the abnormal patterns were considered positive and negative.



(a) (b)

Figure 4. Comparison of abnormal behavior detection methods using ROC curve in UCSD database. a) ROC curve in frame-level b) ROC curve in pixel-level.

The values of true positive and false positive are based on the two following measures:

Frame-level measure: this measure indicates which frames include the abnormal behaviors. This measure compares the real abnormal frames with the frames that have been considered abnormal at the test time, and the true-positive and false-positive rates are obtained accordingly. The method used in this work was compared with the detection domain methods of the abnormal behaviors in the recent years, and the results obtained indicate that the method used has a better performance compared to the other methods. The quantitative results of this comparison by using the EER (Equal Error Rate) and AUC (Area Under Curve) measures are demonstrated in Table 2. You can also see the ROC curve–a combination of FPR and TPR–in Figure 4(a). The proposed method was also evaluated on the UMN dataset in Table 3 using the frame-level measure.

Pixel–level measure: this measure indicates which pixels include abnormal behaviors. This measure compares real abnormal pixels with pixels that have been considered abnormal at the time of the test, and true-positive and false-positive rates are obtained accordingly. In other words, it can be said that true-positive occurs when the frame is positive and or at least 40 percent of its pixels are considered abnormal [40]. It can be said that in this measure, we will have abnormal detection accuracy as local. The results of this measure and the ROC curve for this measure are demonstrated in Table 2 and Figure 4 (b), respectively. Table 2 reports the results of the method used in this paper with the detection domain methods of the abnormal behaviors in recent years and results indicate that the used method has better performance compared to other methods. It should be mentioned that the average accuracy and error is demonstrated in Table 2, but Figure 4 shows details of diagnosis accuracy and error.

Table 2. Result comparison of method used in this work with results of approaches that have been used in recent years on UCSD database; results o methods have been taken from article [41]. PM-ATM (Proposed Method-Abnormal Training Model), PM-NTM (Proposed Method-Normal Training).

Method	**Ped1 (frame-level)**		**Ped1 (pixel-level)**		**Ped2 (frame-level)**	
	AUC	**EER**	**AUC**	**EER**	**AUC**	**EER**
MPPCA [42]	59%	40%	20.5%	81%	69.3%	30%
Social force (SF) [37]	67.5%	31%	19.7%	79%	55.6%	42%
SF + MPPCA [38]	68.8%	32%	21.3%	71%	61.3%	36%
SR [43]	—	19%	45.3%	54%	—	—
MDT [38]	81.8%	25%	44.1%	58%	82.9%	25%
Detection at 150fp [39]	91.8%	15%	63.8%	43%	—	—
Plug-and-Play CNN [44]	95.7%	8%	64.5%	40.8%	88.4%	18%
AMDN (double fusion) [41]	92.1%	16%	67.2%	40.1%	90.8%	17%
GAN [10]	97.4%	8%	70.3%	35%	93.5%	14%
PM-ATM	**98.4%**	**7%**	**73.9%**	**33%**	**94.2%**	**13%**
PM-NTM	**98.9%**	**6%**	**75.8%**	**32%**	**95.4%**	**12%**

Table 3. Result comparison of method used in this work with results of approaches that have been used in recent years on UMN database; results of techniques have been taken from article [40]. PM-ATM (Proposed Method-Abnormal Training Model), PM-NTM (Proposed Method-Normal Training).

Method	AUC
optical-flow [37]	0.84
SFM [37]	0.96
Sparse reconstruction [43]	0.97
Commotion [45]	0.98
Plug-and-Play CNN [44]	0.98
Global optical flow [46]	0.98
GAN [10]	0.99
AUC maximization [47]	0.99
PM-ATM	**0.993**
PM-NTM	**0.995**

Figure 5 presents the values calculated for the equal error rate (EER) with respect to the Ped1 and Ped2 subsets. In the Ped1 subset, the average EER for the proposed method in the case that we used the model of abnormal behaviors is 1% lower than that for GAN, 9% lower than that for AMDN, 1% lower than that for the Plug-and-Play CNN, 18% lower than that for MDT, 25% lower than that for SF+MPPCA, 24% lower than that for the social force, and 33% lower than that for MPPCA.

In the Ped2 subset, the average EER for the proposed method in the case that we used the model of abnormal behaviors is 1% lower than that for GAN, 4% lower than that for AMDN, 5% lower than that for the Plug-and-Play CNN, 12% lower than that for MDT, 23% lower than that for SF + MPPCA, 29% lower than that for the social force, and 17% lower than that for MPPCA.

As mentioned in this article, the two models of normal and abnormal training models were used to evaluate the accuracy of system detection. The results of the previous sections show that the normal training model is slightly more accurate than the abnormal training model for detecting the abnormal behaviors.

In Figure 6, for example, we used an abnormal training model to detect the abnormal behaviors. Figure 6 demonstrates that the method used can identify and reconstruct the abnormal patterns well.

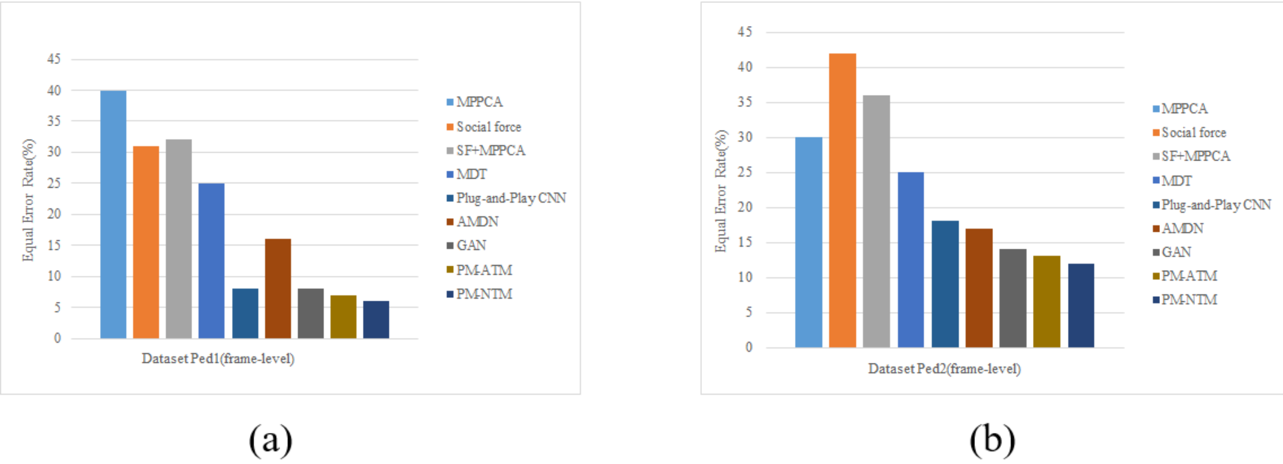


Figure 5. UCSD dataset: comparison of frame-level performance (equal error rate) with different methods. a) Dataset Ped1 b) Dataset Ped2. PM-ATM (Proposed Method-Abnormal Training Model), PM-NTM (Proposed Method-Normal Training).

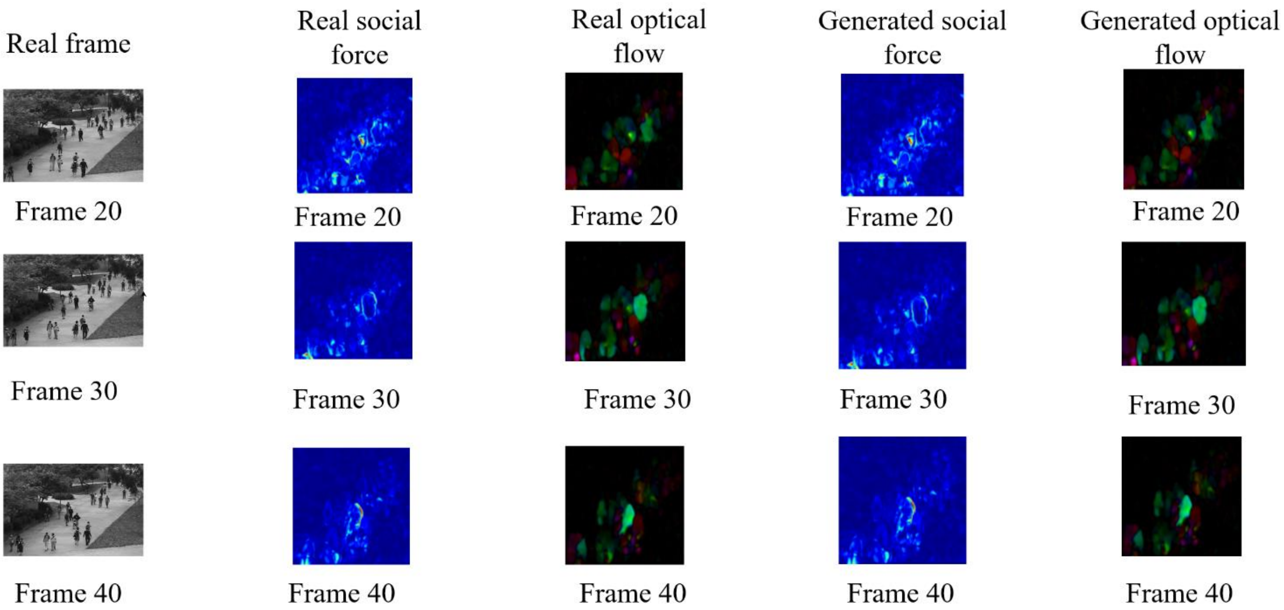


Figure 6. Examples of frame, social force, optical flow images, and abnormal behaviors detection.

4.4. Result evaluation

As one can observe in Figure 4, Table 2, and Table 3, the provided method has had acceptable results than the similar methods. Important points that are kept from the results include: 1) if a normal training method is used for the test, we will have a better accuracy, and this issue is due to the comprehensive patterns and regular data. Also using the movement models like social force and optical flow, this case causes a better behavior pattern training. On the one hand, if the unusual training method is used for the test, the detection accuracy will increase due to the network training using the augmented abnormal patterns obtained from the geometrical techniques. In comparison with the case that we used the normal training model, we have a lower accuracy, and this is due to the comprehensive patterns and the normal data than the abnormal data. 2) Regarding the test time, the normal training model has a relative superiority over the abnormal training model in this paper's system. Still, it should be mentioned that in the abnormal training model, fewer patterns are available, which is one of the main challenges in the research works. Therefore, on the one hand, the abnormal training model results are slightly different from the normal training model. 3) On the other hand, if an abnormal training model is used for the test, the results obtained are better than the results provided in paper [10]. Its reason is that in using the temporal pattern for training in the network, the results are much better than the spatial-temporal model. In paper [10], the spatial-temporal model (row frame and optical flow images) of normal behavior has been used. Still, it has a lower accuracy than the abnormal training model provided in this paper, which has used the abnormal patterns. 4) In paper

[10], blobs will be created that incorrectly lead to the detection patterns due to generating optical flow from row frames. In thiswork, using the temporal patterns and generating optical flow from the relevant social force, such a problem will not be created.

5. Conclusion

In this work, the main problems of the abnormal behavior detection including the shortage of training data and the complexity of the behavioral patterns were considered. The data augmentation techniques including the geometrical methods were used in order to compensate for the lack of abnormal data. For the training behaviors and the normal and abnormal patterns, a cycle GAN network was used, and the first and second network was trained using the social force patterns of the normal and abnormal behaviors and the optical flow of the normal and abnormal behaviors, respectively. In the test stage, since the network was trained based on the normal and abnormal patterns, the results obtained indicated that the abnormal behaviors would be detected with a high accuracy if we used the normal training model in the test stage. The experimental results on the standard database of the abnormal behavior diagnosis show that the technique used in this paper has a high performance and accuracy than the other works performed in the recent years. In this work, two measures, frame–level and pixel–level, were used. As a future work, we intend to perform the operations of identifying the abnormal behaviors as well; in other words, in addition to the abnormal behavior detection, the type of behavior will also be estimated.

Refrences

[1] M. Ravanbakhsh, E. Sangineto, M. Nabi, and N. Sebe, "Training Adversarial Discriminators for Cross-channel Abnormal Event Detection in Crowds," *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1896-1904, 2019.

[2] 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.

[3] A. Samdurkar, Sh. Kamble, Ni. Thakur, and A. Patharkar, "Overview of Object Detection and Tracking based on Block Matching Techniques," *Second International Conference on Research in Intelligent and Computing in Engineering*, pp. 313-319, 2017.

[4] Li. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X.Liu, and M. Pietikäinen, "Deep Learning for Generic Object Detection: A Survey," *International Journal of Computer Vision*, Vol. 128, No. 2, pp. 261-318, 2020.

[5] T. A. Mostafa, J. Uddin, and M. H. Ali, "Abnormal event detection in crowded scenarios," *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)*, pp. 1-6, 2017.

[6] M. Bertini, A. D. Bimbo and L. Seidenari, "Multi-scale and real-time non-parametric approach for anomaly detection and localization," *Computer Vision and Image Understanding*, Vol. 116, No. 3, pp. 320-329, 2012.

[7] C. Lu, J. Shi, W. Wang, and J. Jia, "Fast Abnormal Event Detection," *International Journal of Computer Vision*, Vol. 127, No. 8, pp. 993-1011, 2019.

[8] J. Wang and Z. Xu, "Spatio-temporal texture modelling for real-time crowd anomaly detection," *Computer Vision and Image Understanding*, Vol. 144, pp. 177-187, 2016.

[9] Y. S. Chong and Y. H. Tay, "Abnormal Event Detection in Videos using Spatiotemporal Autoencoder," *Computer Vision and Pattern Recognition*, pp. 189-196, 2017.

[10] M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, "Abnormal event detection in videos using generative adversarial nets," *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 1577-1581, 2017.

[11] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2242-2251, 2017.

[12] 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.

[13] M. Paul, S. M. E. Haque, and S. Chakraborty, "Human detection in surveillance videos and its applications-a review," *EURASIP Journal on Advances in Signal Processing*, Vol. 2013, No. 1, pp. 176, 2018.

[14] C. Hemalatha, S. Muruganand, and R. Maheswaran, "A Survey on Real Time Object Detection, Tracking and Recognition in Image Processing," *International Journal of Computer Applications*, Vol. 91, No. 16, pp. 38-42, 2014.

[15] T. Anbu, M. M. Joe, and G. Murugeswari, "A comprehensive survey of detecting tampered images and localization of the tampered region," *Multi-media Tools and Applications*, Vol. 80, No. 2, pp. 2713-2751, 2021.

[16] Y. Xiao, Zh. Tian, J. Yu, Y. Zhang, Sh. Liu, Sh. Du and X. Lan, "A review of object detection based on deep learning," *Multimedia Tools and Applications*, Vol. 79, No. 33, pp. 23729-23791, 2020.

[17] A. E. Gunduz, C. Ongun, T. T. Temizel, and A. Temizel, "Density aware anomaly detection in crowded scenes," *IET Computer Vision*, Vol. 10, No. 5, pp. 374-381, 2016.

[18] D. Shehab and H. Ammar, "Statistical detection of a panic behavior in crowded scenes," *Machine Vision and Applications*, Vol. 30, No. 5, pp. 919-931, 2019.

[19] S. Ezatzadeh and M. R. Keyvanpour, "ViFa: an analytical framework for vision-based fall detection in a surveillance environment," *Multi-media Tools and Applications*, Vol. 78, No. 18, pp. 25515-25537, 2019.

[20] C. Vishnu, D. Singh, C. K. Mohan, and S. Babu, "Detection of motorcyclists without helmet in videos using convolutional neural network," *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 3036-3041, 2017.

[21] J. Wang and Z. Xu, "Crowd anomaly detection for automated video surveillance," *6th International Conference on Imaging for Crime Prevention and Detection (ICDP-15)*, pp. 1-6, 2015.

[22] D. Dawei, Q. Honggang, H. Qingming, Z. Wei, and Z. Changhua, "Abnormal event detection in crowded scenes based on Structural Multi-scale Motion Interrelated Patterns," *2013 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1-6, 2013.

[23] H. Chebi and D. Acheli, "Dynamic detection of anomalies in crowd's behavior analysis," *2015 4th International Conference on Electrical Engineering (ICEE)*, pp. 1-5, 2015.

[24] A. Li, Z. Miao, Y. Cen, and Q. Liang, "Abnormal event detection based on sparse reconstruction in crowded scenes," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1786-1790, 2016.

[25] K. Vignesh, G. Yadav, and A. Sethi, "Abnormal Event Detection on BMTT-PETS 2017 Surveillance Challenge," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2161-2168, 2017.

[26] T. Wang, J. Chen, and H. Snoussi, "Online Detection of Abnormal Events in Video Streams," *Journal of Electrical and Computer Engineering*, Vol. 2013, pp. 1-12, 2013.

[27] X. Li, Y. She, D. Luo, and Zh. Yu, "A Traffic State Detection Tool for Freeway Video Surveillance System," *Procedia-Social and Behavioral Sciences,* Vol. 96, pp. 2453-2461, 2013.

[28] M. Manfredi, R. Vezzani, S. Calderara, and R. Cucchiara, "Detection of static groups and crowds gathered in open spaces by texture classification," *Pattern Recognition Letters,* Vol. 44, pp. 39-48, 2014.

[29] M. Sabokrou, M. Fayyaz, M. Fathy, Z. Moayed, and R. Klette, "Deep-anomaly: Fully convolutional neural network for fast anomaly detection in crowded scenes," *Computer Vision and Image Understanding*, Vol. 172, pp. 88-97, 2018.

[30] X. Zong, Y. Chen, A. Liu, R. Li, S. Liu, H. Yu, and M. Tan, "Abnormal Event Detection in Video based on Sparse Representation," *2020 15th International Conference on Computer Science and Education (ICCSE)*, pp. 649-653, 2020.

[31] C. Spampinato, S. Palazzo, P. D'Oro, D. Giordano, and M. Shah, "Adversarial Framework for Unsupervised Learning of Motion Dynamics in Videos," *International Journal of Computer Vision*, Vol. 128, No. 5, pp. 1378-1397, 2020.

[32] S. Hamdi, S. Bouindour, K. Loukil, H. Snoussi, and M. Abid, "Two-streams Fully Convolutional Networks for Abnormal Event Detection in Videos," *12th International Conference on Agents and Artificial Intelligence*, pp. 514-521, 2020.

[33] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, Sh. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Proceedings of the 27th International Conference on Neural Information Processing Systems*, Vol. 2, pp. 2672–2680, 2014.

[34] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5967-5976, 2017.

[35] E. Pejhan and M. Ghasemzadeh, "Multi-Sentence Hierarchical Generative Adversarial Network GAN (MSH-GAN) for Automatic Text-to-Image Generation," *Journal of AI and Data Mining*, Vol. 9, No. 4, pp. 475-485, 2021.

[36] Th. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High Accuracy Optical Flow Estimation based on a Theory for Warping," *In: Pajdla T.. Matas J. (eds) Computer Vision - ECCV 2004. ECCV 2004. Lecture Notes in Computer Science*, Vol. 3024, pp. 25-36, 2004.

[37] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 935-942, 2009.

[38] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1975-1981, 2010.

[39] C. Lu, J. Shi, and J. Jia, "Abnormal Event Detection at 150 FPS in MATLAB," *2013 IEEE International Conference on Computer Vision*, pp. 2720-2727, 2013.

[40] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Trans Pattern Anal Mach Intell*, Vol. 36, No. 1, pp. 18-32, 2014.

[41] D. Xu, Y. Yan, E. Ricci, and N. Sebe, "Detecting anomalous events in videos by learning deep representations of appearance and motion," *Computer Vision and Image Understanding*, Vol. 156, pp. 117-127, 2017.

[42] J. Kim and K. Grauman, "Observe locally, infer globally: A space-time MRF for detecting abnormal

activities with incremental updates," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921-2928, 2009.

[43] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," *CVPR 2011*, pp. 3449-3456, 2011.

[44] M. Ravanbakhsh, M. Nabi, H. Mousavi, E. Sangineto, and N. Sebe, "Plug-and-Play CNN for Crowd Motion Analysis: An Application in Abnormal Event Detection," *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1689-1698, 2018.

[45] H. Mousavi, M. Nabi, H. Kiani, A. Perina, and V. Murino, "Crowd Motion Monitoring using Tracklet-based Commotion Measure," *IEEE International Conference on Image Processing(ICIP)*, pp. 2354-2358, 2015.

[46] Y. Hao, Y. Liu, J. Fan, and Z. Xu, "Group Abnormal Behaviour Detection Algorithm based on Global Optical Flow," *Complexity*, Vol. 2021, pp. 12, 2021.

[47] A. Feizi, "Hierarchical detection of abnormal behaviors in video surveillance through modeling normal behaviors based on AUC maximization," *Soft Computing*, Vol. 24, No. 14, pp. 10401-10413, 2020.

تشخیص رفتار غیرعادی بر روی داده‌های نرمال و داده‌های غیر نرمال تقویت شده در صحنه‌های پرازدحام

وحید فاضل اصل، بابک کرسفی* و بهروز معصومی

دانشگاه آزاد اسلامی ،واحد قزوین، دانشکده مهندسی کامپیوتر و فناوری اطلاعات.

ارسال ۲۰۲۱/۰۸/۰۸؛ بازنگری ۲۰۲۱/۱۱/۱۳؛ پذیرش ۲۰۲۲/۰۱/۲۶

چکیده:

در این مقاله به بررسی مشکلات تشخیص رفتار غیر نرمال در محیط‌های پرجمعیت می‌پردازیم. یکی از مسائل اصلی در تشخیص رفتار غیر نرمال، پیچیدگی الگوهای ساختار بین فریم‌ها است. در این مقاله از الگوهای نیروی اجتماعی[1] و جریان نوری[2] به منظور آماده سازی سیستم برای آموزش پیچیدگی الگوهای ساختاری استفاده شده است. سیستم شبکه متخاصم مولد چرخه[3] برای آموزش الگوهای رفتاری استفاده می‌شود. برای ارزیابی دقت تشخیص سیستم از دو مدل الگوهای رفتاری نرمال و غیر نرمال استفاده می‌شود. در مورد الگوهای غیر نرمال مورد استفاده برای آموزش، به دلیل عدم وجود این نوع الگوی رفتاری که چالش دیگری در تشخیص رفتارهای نابهنجار است، از تکنیک‌های هندسی برای تقویت الگوها استفاده می‌شود. اگر از الگوهای رفتاری نرمال برای آموزش استفاده شود، نیازی به تقویت الگوها نیست زیرا الگوهای نرمال کافی هستند. در مرحله آزمایش، اگر از الگوهای رفتاری نرمال برای آموزش استفاده شود، سیستم شبکه متخاصم مولد چرخه نباید بتواند الگوهای رفتاری غیر نرمال را با دقت بالایی بازسازی کند. علاوه بر این، اگر از الگوهای رفتاری غیر نرمال برای آموزش استفاده شود، سیستم شبکه متخاصم مولد چرخه باید بتواند الگوهای رفتاری غیر نرمال را با دقت بالایی بازسازی کند. نتایج تجربی به دست آمده از پایگاه‌های اطلاعاتی مربوط به رفتار غیر نرمال نشان می‌دهد که روش مورد استفاده نتایج بهتری نسبت به سایر روش‌های مشابه دارد.

کلمات کلیدی: سیستم های نظارت تصویری، تشخیص رویدادهای غیر نرمال، تقویت سازی داده ها، تجزیه و تحلیل رفتار جمعیت، شبکه های متخاصم مولد چرخه.

[1] social force
[2] optical flow
[3] cycle GAN