International Journal on Applications in Engineering and Technology
Volume 8 : Issue 2 : May 2022, pp 129 –138 www.aetsjournal.com ISSN (Online) : 2455 - 0523

ANOMALY DETECTION IN BANKING USING A VIDEO SURVEILLANCE SYSTEM

ELANGOVAN.T ,KARTHIK RAJA .B,SANJAYKUMAR.M,PRAKASH.P

[1,2,3] *Undergraduate student, Department of Computer Science and Engineering,*
Paavai College of Engineering
[4] *Assistant Professor, Department of Computer Science and Engineering,*
Paavai College of Engineering

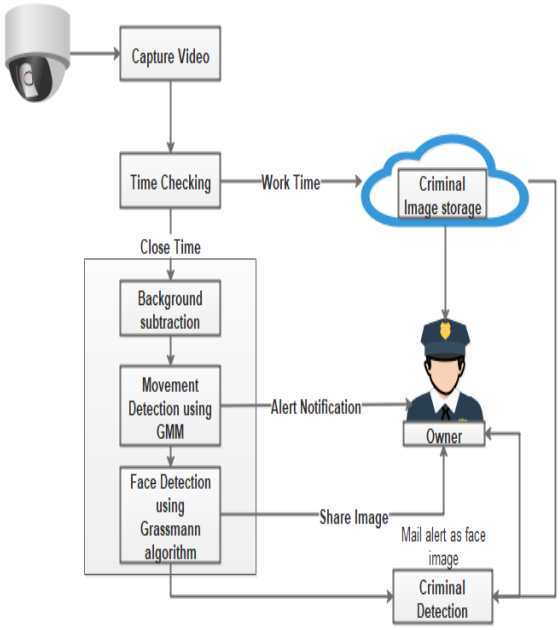
Abstract: - The job of finding frames from a video sequence that reflect occurrences that do not conform to expected behaviour is known as anomaly identification in videos, and it is a difficult problem due to anomalies' ambiguous and unbounded qualities. Video anomaly detection approaches based on deep neural networks have advanced significantly with the advancement of deep learning. Frame reconstruction and frame prediction are the two basic approaches used by existing technologies. Reconstruction-based methods are limited in their application due to the high generalisation capacity of neural networks. Recently, prediction-based anomaly detection systems have demonstrated superior performance. When they can't promise fewer prediction errors for ordinary events, however, their performance decreases. We propose a novel future frame prediction model for anomaly detection based on a bidirectional retrospective generation adversarial network (BR-GAN) in this research. To completely mine the bidirectional temporal information between the predicted frame and the input frame sequence, we offer a bidirectional prediction paired with a retrospective prediction method to predict a future frame with greater quality for normal events. then For appearance (spatial) restrictions, an adversarial loss is used with an intensity and gradient loss between the anticipated and actual frames. Furthermore, we propose a sequence discriminator composed of a three-dimensional (3D) convolutional neural network to capture the long-term temporal relationships between frame sequences composed of predicted frames and input frames; this network is critical in maintaining the predicted frames' motion (temporal) consistency for normal events. Such appearance and motion limitations aid future frame prediction for typical events, allowing the prediction network to discern between normal and abnormal patterns with ease. Extensive tests on benchmark datasets show that our method beats most existing state-of-the-art methods, proving the method's usefulness in detecting anomalies.

***Key Words*:** Detection, Banking, 3D, BRGAN, Neural Network

I. INTRODUCTION

The ability to detect anomalous activity in video surveillance is critical to maintaining public safety. Video surveillance systems have increasingly been extended over our entire society, and the amount of surveillance video data has drastically expanded. Manual detection of odd occurrences from vast surveillance video data is both time consuming and inefficient. As a result, research on automatic detection algorithms for anomalous behaviour in video surveillance has exploded. It is, nevertheless, a difficult undertaking since abnormal events are vague and limitless, making it impossible to accurately characterise abnormal events. Furthermore, abnormal events are uncommon, making it challenging to obtain aberrant samples from a huge amount of surveillance video data for algorithm learning. In video surveillance, however, capturing routine events is much easier.As a result, semi-supervised learning is one of the most often used approaches for anomaly identification, where only normal data is provided in the training set.A popular way for determining whether an abnormal event occurs is to exploit normal patterns based on their appearance and motion in the training set. An abnormal pattern is defined as one that is not compatible with the usual patterns.Anomaly detection approaches that are semi-supervised can be loosely classified into two categories: methods based on reconstruction and methods based on prediction. Hand-crafted appearance and motion features are used in the earlier reconstruction approaches [2], [3], [5], [6], [10]. Then, by learning those features, a dictionary is learned to sparsely encode all normal occurrences with minor reconstruction mistakes. The traits that correspond to anomalous occurrences may cause a high reconstruction error during testing. However, optimising sparse coefficients takes a long time, and the efficiency of anomaly detection is restricted by hand-crafted features, therefore dictionary learning suffers from restrictions for these two reasons. With the advancement of deep learning, various studies [7–9], [11–13], [24], [25] have employed deep neural networks to

extract deep features automatically rather than hand-crafted features.



Hasan et al., for example, proposed using an auto-encoder to reconstruct normal events with minor reconstruction errors in [9].

Deep neural networks, on the other hand, have excellent generalisation capabilities, therefore certain anomalous frames can be recreated with minor errors. The main principle behind recently proposed prediction-based methods is to create a model that can accurately forecast the future frame of regular events, however the unpredictable nature of abnormal events would result in huge prediction errors. In this paper, we present a new future frame prediction framework for anomaly detection dubbed a bidirectional retrospective generative adversarial network (BR-GAN) that overcomes these restrictions. The BR-GAN is made up of a generator, a frame discriminator, and a sequence discriminator that are all made up of 3-dimensional (3D) convolutional neural networks that can be trained from start to finish. BR-general GAN's architecture is depicted in Fig. 1. The following characteristics are present in this framework: 1) It can fully explore the bidirectional mapping relationship among video frame sequences to more accurately establish the mapping model from a few frames in the past to a future frame for normal events; 2) the motion constraint can be performed from the perspective of longterm temporal consistency to ensure that the predicted frame and the real frame are identical; and 3) In terms of mobility, the l frame is compatible with regular events.

We present a bidirectional prediction method paired with a retrospective prediction method to fully mine the bidirectional mapping relationship across video frame sequences. The term "bidirectional prediction" refers to the generator's ability to anticipate both forward and backward.The essential notion of retrospective prediction is that if the previously forecast frame is realistic, the frame that is projected again using the previously predicted frame as input should be realistic as well. We also combine a frame discriminator to detect whether the image produced by the generator is real or false, which inhibits the generator from producing blatantly fraudulent images, which improves the generator's robustness and the quality of the forecast frames.

Furthermore, we present a sequence discriminator built of 3D convolutional neural networks to collect long-term temporal information across video frame sequences in order to make the predicted frame consistent with the real object in motion. The term "bidirectional prediction" refers to the generator's ability to anticipate both forward and backward.

The essential notion of retrospective prediction is that if the previously forecast frame is realistic, the frame that is projected again using the previously predicted frame as input should be realistic as well. Forward prediction is predicting a future frame by observing a few frames in the past, while backward prediction is predicting a future frame by viewing a few frames in the past a future frame by viewing previous frames . We also combine a frame discriminator to detect whether the image produced by the generator is real or false, which inhibits the generator from producing blatantly fraudulent images, which improves the generator's robustness and the quality of the forecast frames.

Furthermore, we present a sequence discriminator built of 3D convolutional neural networks to collect long-term temporal information across video frame sequences in order to make the predicted frame consistent with the real object in motion. . The relevant input frames are fed into the sequence discriminator to increase the generator's resilience and the quality of the predicted frames, and then the authenticity assessment loss is utilised to constrain the motion. The generator U-Net predicts the next frame based on its historical observation in the testing phase, as shown in Fig. 2. Because abnormal occurrences are unpredictable, a substantial difference between the projected and actual future frames suggests the presence of possible abnormal events in the frame. To conclude, the following are the major contributions of our study.

1) For anomaly detection, we offer a unique future frame prediction framework based on a bidirectional retrospective generative adversarial network that can be trained end-to-end.

2) To completely explore the bidirectional mapping relationship between video frame sequences, we offer a bidirectional prediction approach paired with a retrospective prediction method, and our method provides more accurate future frame prediction of normal events.

3) We propose a sequence discriminator made up of 3D convolutional neural networks to capture the long-term temporal links between frame sequences made up of predicted frames and input frames, which are important for maintaining the motion consistency of predicted frames in typical occurrences.

II. CONNECTED WORK

A vast number of research projects [1]–[13], [15], [18]–[23], [26], [32]–[36], [40], [48], [49], [53]–[55] have been dedicated to tackling the video anomaly detection problem throughout the last few decades. Traditional methods, deep learning-based approaches, and other classic methods can be used to categorise all of these methods.

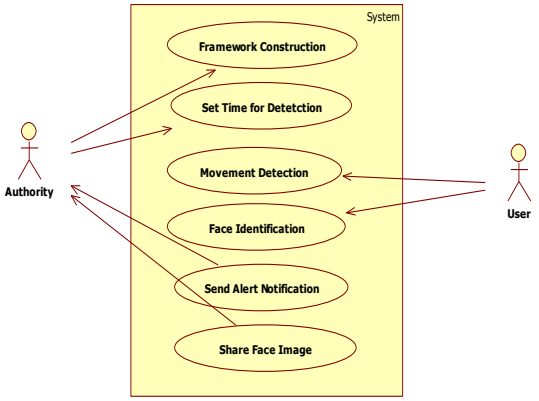
A. DETECTION OF TRADITIONAL ANOMALY

Traditional anomaly detection methods construct feature space primarily using handcrafted features or features learned from the training set using traditional machine learning techniques, then characterise the distribution of normal or anomaly scenarios, and finally identify isolated clusters or outliers as abnormal. In classical anomaly detection, statistical models and sparse coding are prominent modelling strategies.

1) ANOMALY DETECTION METHODS BASED ON STATISTICAL MODELS

The typical patterns are usually represented by dynamic trajectory features [56], [57] of an item in early work based on statistical models. Hu et al. [57], for example, suggested a method for automatically learning motion patterns for anomaly detection using trajectory data clustered hierarchically using spatial and temporal information.

Piciarelli et al. developed an online trajectory cluster technique for the detection of anomalous occurrences in [58], and the proposed method constructs the clusters as the tracking system collects data. However, in crowded environments, trajectory features based on object tracking can fail, making these approaches ineffective in complicated scenes.



2) ANOMALY DETECTION METHODS BASED ON SPARSE CODING

The early sparse coding-based approaches [2], [3], [6], [10], [48], [49] commonly learn a dictionary from hand-crafted characteristics to reconstruct normal events with minor reconstruction errors, whereas aberrant occurrences are flagged as abnormal. Cong et al., for example, established the sparse reconstruction cost (SRC) over the normal lexicon to assess the normality of testing samples in [2].

Based on the online sparse reconstruction of query signals from an atomically learnt event dictionary, Zhao et al. introduced a fully unsupervised dynamic sparse coding strategy for recognising uncommon occurrences in films in [3].

B. ANOMALY DETECTION BASED ON DEEP LEARNING

Deep learning algorithms have had tremendous success in recent years in a variety of domains, including image classification [27], [60], object identification [68], [69], and video retrieval [61], [62]. Many deep learning-based algorithms for detecting anomalies in videos have been proposed.

In general, these techniques can be split into two groups: methods based on reconstruction and methods based on prediction

1) ANOMALY DETECTION METHODS BASED ON RECONSTRUCTION

The reconstruction-based method is comparable to the sparse coding-based method that was used previously. Due to the superiority of deep features [9], [16], [25], [27], [53] over prior hand-crafted features [28], [29], [50], deep features have been included in reconstruction in recent years, resulting in advancements in anomaly identification. Wu et al. presented a two-stream neural network to extract spatial-temporal fusion features (STFF) in hidden layers in [21], and then used a fast sparse coding network (FSCN) to generate a normal dictionary based on the STFF. Hasan et al. [9] learned the temporal regularity in the videos using the extracted feature as input to a fully connected neural network-based autoencoder.

2) ANOMALY DETECTION METHODS BASED ON PREDICTION

Video prediction algorithms have advanced fast in recent years because to their ability to learn the internal representation of a video from large amounts of unlabeled data, and they have a wide range of applications in robot decisionmaking, autonomous driving, and video interpretation. The purpose of the video prediction task is to forecast future frames from the video's previous frames. Mathieu et al. introduced a multi-scale network with adversarial training to create future frames given a sequence of input frames in [14].D'Avino et al. [30] proposed a recurrent autoencoder based on an LSTM that detects video fraud by modelling temporal dependences between patches from a succession of input frames.

C. OTHER TRADITIONAL METHODS

Sun et al. [34] proposed an unsupervised model dubbed online growing neural gas (online GNG) for learning surveillance scenes by changing the learning parameters on a continuous basis. To detect anomalous behaviours, Chaker et al. [35] used a social network to represent the local and global behaviour patterns.

By hierarchically modelling normal patches using deep features and Gauss distributions, Sabokrou et al. [51] proposed a DNN-based cascade classifier for fast anomaly identification.Wu et al. developed a revolutionary deep oneclass (DeepOC) neural network in another paper [20], which can learn a one-class classifier and compact features using convolutional neural networks at the same time. Pang et al. [4] recently published a paper describing a novel end-to-end approach to unsupervised video anomaly detection using self-trained ordinal regression.

III. PROPOSED METHOD

The anomaly detection approach based on future frame prediction usually involves studying video data of normal behaviour to train a future frame prediction network that can accurately anticipate the next frame of normal behaviour based on the previous frames. Because abnormal occurrences are unpredictable, the future frame generated by the prediction network can result in higher prediction errors, allowing abnormal events to be discovered. Currently, certain studies [15], [32], and [38] use a method to detect anomalous events based on the prediction of future frames.These algorithms' promising performance demonstrates the promise of prediction-based methods for anomaly detection.

A. ARCHITECTURES OF NETWORKS

The BR-GAN is made up of a 3D convolutional neural network generator, frame discriminator, and sequence discriminator. Figure 1 depicts the general architecture of BR-GAN.

GENERATOR 1

As the generator of our prediction framework, we adopt the U-Net [39] network design. U-Net is a symmetrical down-sampling and up-sampling network that has been transformed from a fully convolutional network. The U-Net network is seen in Fig. 3. The left half comes first.A down-sampling component of the network is made up of pooling layers and numerous convolutional layers, which are used to extract features from a video sequence as input The left and right halves of the network are symmetrical.is a component that generates an up-sampled image.by gradually increasing the size of the image inferred from the feature resolution in space

2) DISCRIMINATOR OF FRAME

The frame discriminator's job is to compare the predicted future frame with the actual future frame to see if the input comes from a real distribution or is generated by the generator, which can help improve the generator's robustness and the quality of the predicted frame by playing a dynamic game with it. In our work, we use the patch GAN as the frame discriminator, as described in [37]. Patch GAN differs from traditional GAN discriminators in that it generates a matrix in which each element decides whether the patch is real or artificial rather than mapping an input picture to a single scalar output in the [0, 1] range.

3) DISCRIMINATOR OF SEQUENCES

A sequence discriminator is used to improve the predicted frame's temporal consistency by determining if the frame sequence contains false frames or not, allowing the generator to forecast video frames that are consistent with the genuine sequence in the temporal connection.

We create a sequence discriminator using a four-layer 3D convolutional network, which is inspired by the work in [52].

B. GENERATOR RETROSPECTIVE BIDIRECTIONAL PREDICTION

The implementation process of the generator's bidirectional retrospective prediction is detailed in this subsection. The generator, frame discriminator, and sequence discriminator are represented by G, DF, and DS, respectively, for clarity. G begins by doing bidirectional prediction, which includes both forward and backward predictions. The input sequences for forward and backward prediction are defined mathematically as follows:

Xs:e = xs, xs+1,..., xe1, xe s.t. s e s.t. s e s.t. s e s.t. s e s.t. s e s.t. s e s.t. s e s.t. s (1)

Xs+1:e+1 = xe+1, xe, xe, etc., xs+2, xs+1 s.t. s e (2), where xi R 2 denotes a frame image and s and e are the start and end frame indices . Specifically, the input sequence containing x 0 e+1 and x 0 s is defined as follows:

X R s+1:e+1 = x 0 e+1 e+1 xs+1:e

(3) X R s:e = x 0 s xs+1:e X R s:e = x 0 s xs+1:e X R s:e = x 0

(4) X R s+1:e+1 is sorted in reverse chronological order and input into G to anticipate the previous frame xs once more.

CONSTRAINTS ON APPEARANCE AND MOTION

Multiple prediction frames are created during the bidirectional and retrospective prediction process. Following the work of [14], intensity loss and gradient loss are used to limit the appearance of the predicted frames, bringing them closer to actual frames. To ensure that all pixels in RGB space are comparable, the intensity loss penalises pixel discrepancies between the anticipated and real frames, while the gradient loss penalises gradient variations between them to sharpen the predicted frame. The intensity loss function is specifically specified as follows:

I pair s,e f1(p, g) Lint = X (p, g) Lint = X (p, g) Lint = X (p, g) Lint = X (p, g) (5)

where f1(,) is the L2 distance between two pictures, which is defined as follows: (6). I pair s,e is a collection of picture pairs defined in (7).

f1(p, g) = kp gk 2 2 f1(p, g) = kp gk 2 2 f1(p, g) = (6)

I pair s,e = (xs, x 0 s), (xs, x 00 s), (x 0 s, x 00 s), (x 0 s, x 00 s), (x 0 e+1, x 00 e+1), (x 0 e+1, x 00 e+1), (x 0 e+1, x 00 (7)

The L2 distance of six pairs of images in I pair s,e is minimised using the intensity loss function. To minimise forward prediction errors, the pair (xe+1, x 0 e+1) is employed, which comprises of the forward prediction frame and the actual frame.

The generator can fully use the bidirectional mapping relationship between the frame sequences composed of anticipated frames and input frames thanks to these six pairs of constraints.

In addition, the gradient loss function is as follows:

I pair s,e f2(p, g) (8) f2 = X i,j Lgd = X (p,g) I pair s,e f2(p, g)

pi,j, pi1,j, pi2,j, pi3,j, pi4,j, pi5,j, pi − gi,j gi1,j gi2,j gi3,j gi4,j gi5,j gi 1 \s+

pi,j1 = pi,j2 = pi,j3 = pi,j4 = pi,j5 = pi, − gi,j1 gi,j2 gi,j3 gi,j4 gi,j5 gi,j1 (9)

where I j are the video frame's spatial indexes and f2(,) are the gradient differences between two pictures.

D. COMPETITIVE TRAINING

In the realm of picture and video generation, the GAN offers a wide range of applications [41]–[46]. A GAN is usually made up of a generator and a discriminator. The discriminator recognises the frames that the generator creates, while the generator creates frames that trick the discriminator.

The discriminator and generator are alternately updated to train the model. BR-GAN is made up of a generator G and two discriminators DF and DS in our suggested technique. We employ alternate training to optimise our model BR-

GAN, which is similar to training a traditional GAN. The specific training process is explained below.

Training G. The goal of training G is to make the frame predicted by it be consistent with the actual frame in appearance and motion by observing the past few frames. Ideally, both DF and DS would classify the predicted frame as 1, where 1 represents real labels

The following loss functions are imposed from DF and DS:

LMSE(DF (x)i,j, 1) L G adv DF = X x X i,j 1 2 LMSE(DF (x)i,j, 1) (10)

L G adv DS = X X 1 2 LMSE(DS (X), 1) (11) in (10), where =x 0 s, x 00 s, x 0 e+1, x 00 e+1, and I j are the spatial patch indexes. LMSE is an MSE function with the following definition:

LMSE(y, y) = (y y) 2 (12), where y is a number between 0 and 1 and y is a number between 0 and 1. is defined as follows in (11)

Xs:e x 0 e+1, x 0 s Xs+1:e+1, x 00 s Xs+1:e x 00 e+1,

x 00 s Xs+1:e x 00 e+1, x 00 s Xs+1:e x 00 e+1 (13)

in which there are four groups of From DF and DS, the following loss functions are imposed:

LMSE(DF (x)i,j, 1) LMSE(DF (x)i,j, 1) LMSE(DF (x DF = X x X i,j 1 L G adv 2 LMSE(DF (x)i,j, 1) LMSE(DF (x)i,j, 1) LMSE(DF (x) (10)

DS = X X 1 L G adv 2 LMSE(DS (X), 1) (11) in (10), where the spatial patch indexes are =x 0 s, x 00 s, x 0 e+1, x 00 e+1, and I j. LMSE is an MSE function that is defined as follows:

LMSE(y, y) = (y y) 2 (12), where y is a positive integer and y is a negative integer. (11) is defined as follows:

= Xs:e x 0 e+1, x 0 s Xs+1:e+1, x 00 s Xs+1:e x 00 e+1, x 00 s Xs+1:e x 00 e+1, x 00 s Xs+1:e x 00 e+1, x 00 s Xs+1:e x 00 e+1, x 00

DS's training The purpose of DS training is to put a true sequence Xs:e+1 in class 1 and a fake sequence in (13) in class 13.

class 0, where 0 and 1 are the same as in the previous specification.Similarly, when training DS, the G and DF parameters are used.L is a sequence adversarial loss function that is fixed.

DS composed adverb

The MSE is imposed in the following manner:

LDS adv(X, X) = adv(X, X) = adv(X, X) =X

X=Xs:e+1,X=Xs:e+1,X=Xs:e+1,X=Xs:e+1,

LMSE(DS (X), 1) LMSE(DS (X), 1) LMSE(DS (X

LMSE(DS (X), 0) LMSE(DS (X), 0) LMSE(DS (X), 0) (17)

E. OBJECTIVE PURPOSE

There are three aspects to the total objective functions: generation loss, frame adversarial loss, and sequence adversarial loss. The generation loss function is created by combining the appearance constraint, motion restrictions, and adversarial loss during training G.

F. DETECTION OF ANOMALY

1) DATA TESTING ANOMALY DETECTION

Only the generator G is utilised to anticipate future frames during the testing phase. Generator G has learned to predict future frames of normal behaviour based on the previous few frames by learning from a vast amount of normal video data. When anomalous behaviour occurs, such as when a car unexpectedly comes on the sidewalk, the accompanying forecast frame will have a big inaccuracy. Anomaly detection is based on the difference between the expected and actual frames. We utilise the peak signal-to-noise ratio (PSNR) to assess the quality of the predicted images, as described in [14]. T is the total number of frames in a standard video clip. Then, based on experimental data, we determined a suitable threshold k, which is around 85 percent of the PSNRave. When the PSNR of a video frame falls below a predetermined threshold, we consider the frame abnormal. An abnormal event will last for numerous consecutive frames, as described in [48] for the event-level assessment criterion; if more than one frame is detected as abnormal and the position of the frame is inside the ground truth range, it is termed an abnormal event. If the anomalous frame appears in the nonground truth range, however, it is considered a false alarm. In addition, in order to limit the amount of noise.

EXPERIMENTS

In this section, we first go through the datasets, assessment metric, and training details that we discussed in Section A. The experimental results of the proposed method on four common datasets are then shown in Section B, along with comparisons to the performance of existing methods. Then, in Section C, we conduct thorough ablation investigations to examine the capabilities of various components of our technique. Then, in Section D, we validate our method's performance in real-world applications, and in Section E, we examine the benefits and drawbacks of our suggested method. In Section F, the running time is examined.

EXPERIMENTAL SETUP (A)

2) DATABASES

We put our strategy to the test on four commonly known datasets in the field of video anomaly detection. (1) The CUHK Avenue (Avenue) dataset [10] is made up of 16 training videos and 21 testing videos totaling 30652 frames. The CUHK Avenue dataset contains 47 aberrant events, including throwing objects, loitering, and running. (2) UCSD Pedestrian 1 (Ped1) dataset [33], which contains 34 training and 36 testing movies with 40 aberrant events. All of these unusual occurrences include vehicles such as bicycles and automobiles. (3) UCSD Pedestrian 2 (Ped2) dataset [33]: this dataset contains 16 training films and 12 testing videos, each with 12 aberrant events of the same type.

2) METRIC FOR EVALUATION

The FIGURE 6 is evaluated using the Receiver Operation Characteristic (ROC) curve and the Area Under the Curve (AUC). On the four datasets, we compared FFP [15] and our approach in terms of frame-level ROC curves. Our proposed method's effectiveness; these are extensively utilised as performance metrics for classification tasks. A higher AUC value indicates greater anomaly detection performance. We employ frame-level AUC for performance evaluation in this paper to ensure comparability across different techniques. In addition, the ROC curve's equal error rate (EER) is employed as a performance evaluation statistic.

3) DETAILS ON TRAINING

To train the BR-GAN, we first reduce the length of the input frame sequence L to 4, resize each frame to 256 256, then normalise the values of the pixels in all frames to [1, 1], as described in [15].

The network's parameters are then optimised using the Adam optimizer, which is based on the stochastic gradient descent approach. The batch size has been set to four. The generator's learning rate is set to 0.0002, while the frame discriminator and sequence discriminator's learning rates are also set to 0.00002. The grid search approach is used to select a set of ideal hyperparameters for balancing the losses of the four datasets.

B. RESULTS OF EXPERIMENTATION

1) COMPARISON WITH PREVIOUS TECHNIQUES In Table 1, we compare our method to a dozen state-of-the-art methods. On the four datasets, our technique clearly outperforms practically all known methods. Despite the fact that the approaches provided in [38], [32] include prediction and reconstruction branches, their AUC still falls short of ours.

2) QUALITATIVE RESULTS

In this section, we give various qualitative outcomes to further demonstrate the efficacy of our strategy.

Figure 8 depicts the outcomes of our prediction framework's future frame prediction as well as the prediction error map for normal and abnormal events. For normal events, the projected future frames tend to be close to the actual future frame, resulting in a reduced residual map error.

C. ABLATION STUDY

To validate the effectiveness of our method, we conduct ablation experiments on the components and loss functions of the proposed method in this part.

1) BIDIRECTIONAL RETROSPECTIVE PREDICTION VS UNIDIRECTIONAL PREDICTION

We use three alternative prediction methods to test the effectiveness of bidirectional retrospective prediction: bidirectional retrospective prediction (BR-Prediction), bidirectional prediction (B-Prediction), and unidirectional prediction (U-Prediction). On the Avenue and Ped2 datasets, Table 2 provides the AUC of these three techniques.

2) SEQUENCE DISCRIMINATOR VS OPTICAL FLOW

We replace the sequence discriminator in BR-GAN with the method of constraining the optical flow as in FFP [15] to conduct comparative experiments to show that the sequence discriminator in our proposed method can

also play a vital role in motion constraints compared to the method of extracting optical flow.

3) IMPACT OF HYPERPARAMETERS

The critical hyperparameters for balancing the four loss terms of Lint, Lgd, L G adv DF, and L G adv DS in our method are int, gd, adv DF, and adv DS, respectively. We use the Avenue, Ped1, Ped2, and ShanghaiTech datasets to investigate the effects of these hyperparameters and do multiple trials. The key loss items used to optimise the generator are lint and lgd, and their loss values are on the same order of magnitude, therefore we set them to 1 by default. L G adv DF and L G adv DS are two identical loss items that are coupled with the frame and sequence discriminators, respectively, to help optimise the generator.

4) ANALYSIS OF LOSSES

The performance of the approach is heavily reliant on the loss functions. We run ablation experiments on the Avenue dataset to see how different loss functions affect our technique. We gradually remove the four loss terms Lint, Lgd, L frame adv, and L seq adv. The AUC and score gap are used to assess how different loss terms affect anomaly detection performance. A huge score gap and a high AUC suggest higher anomaly detection performance.

V. CONCLUSION

We present a new future frame prediction framework based on a bidirectional retrospective generative adversarial network (BR-GAN) that can be trained end-to-end for anomaly detection in video in this research. The prediction network may more fully exploit the bidirectional mapping link between video frame sequences by combining bidirectional prediction with retrospective prediction, resulting in more accurate future frame prediction of normal events. Furthermore, we present a sequence discriminator made up of 3D convolutional neural networks that can capture the long-term temporal relationship between frame sequences made up of predicted and input frames, which is critical for maintaining predicted frame motion consistency. Extensive tests on four benchmark datasets show that our method beats most existing state-of-the-art methods, demonstrating that our method is effective in detecting anomalies.

REFERENCES

[1]. D. Xu, Y. Yan, E. Ricci, and N. Sebe, ‘‘Detecting anomalous events in videos by learning deep representations of appearance and motion,’’ Comput. Vis. Image Understand., vol. 156, pp. 117–127, Mar. 2016.
[2]. Y. Cong, J. Yuan, and J. Liu, ‘‘Sparse reconstruction cost for abnormal event detection,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2011, pp. 3449–3456.
[3]. B. Zhao, L. Fei-Fei, and E. P. Xing, ‘‘Online detection of unusual events in videos via dynamic sparse coding,’’ in Proc. CVPR, Jun. 2011, pp. 3313–3320.
[4]. G. Pang, C. Yan, C. Shen, A. van den Hengel, and X. Bai, ‘‘Self-trained deep ordinal regression for end-to-end video anomaly detection,’’ 2020, arXiv:2003.06780. [Online]. Available: http://arxiv.org/abs/2003.06780
[5]. 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, Jan. 2014.
[6]. J. K. Dutta and B. Banerjee, ‘‘Online detection of abnormal events using incremental coding length,’’ in Proc. AAAI Conf. Artif. Intell., 2015, pp. 3755–3761.
[7]. J. R. Medel and A. Savakis, ‘‘Anomaly detection in video using predictive convolutional long short-term memory networks,’’ 2016, arXiv:1612.00390. [Online]. Available: http://arxiv.org/abs/1612.00390
[8]. Y. S. Chong and Y. H. Tay, ‘‘Abnormal event detection in videos using spatiotemporal autoencoder,’’ in Proc. Int. Symp. Neural Netw., vol. 10262. Long Beach, CA, USA: Springer, Dec. 2017, pp. 189–196.
[9]. M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, ‘‘Learning temporal regularity in video sequences,’’ in Proc. CVPR, Jun. 2016, pp. 733–742.
[10]. C. Lu, J. Shi, and J. Jia, ‘‘Abnormal event detection at 150 FPS in MATLAB,’’ in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2013, pp. 2720–2727.
[11]. W. Luo, W. Liu, and S. Gao, ‘‘Remembering history with convolutional LSTM for anomaly detection,’’ in Proc. IEEE Int. Conf. Multimedia Expo (ICME), Jul. 2017, pp. 439–444.
[12]. H. Park, J. Noh, and B. Ham, ‘‘Learning memory-guided normality for anomaly detection,’’ in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 14360–14369.
[13]. J. T. Zhou, J. Du, H. Zhu, X. Peng, Y. Liu, and R. S. M. Goh, ‘‘Anomalynet: An anomaly detection network for video surveillance,’’ IEEE Trans. Inf. Forensics Security, vol. 14, no. 10, pp. 2537–2550, Oct. 2019.
[14]. M. Mathieu, C. Couprie, and Y. LeCun, ‘‘Deep multi-scale video prediction beyond mean square error,’’ in Proc. Int. Conf. Learn. Represent., 2016, pp. 1–14.
[15]. W. Liu, W. Luo, D. Lian, and S. Gao, ‘‘Future frame prediction for anomaly detection—A new baseline,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6536–6545.
[16]. C. Xia, F. Qi, and G. Shi, ‘‘Bottom–up visual saliency estimation with deep autoencoder-based sparse reconstruction,’’ IEEE Trans. Neural Netw. Learn. Syst., vol. 27, no. 6, pp. 1227–1240, Jun. 2016.
[17]. A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. V. D. Smagt, D. Cremers, and T. Brox, ‘‘FlowNet: Learning optical flow with convolutional

networks,'' in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 2758–2766.

[18]. R. T. Ionescu, S. Smeureanu, B. Alexe, and M. Popescu, ''Unmasking the abnormal events in video,'' in Proc. ICCV, Oct. 2017, pp. 2914–2922.

[19]. P. Wu, J. Liu, F. Shen, Y. Shi, Y. Sun, F. Shao, Z. Wu, and Z. Yang, ''Not only look, but also listen: Learning multimodal violence detection under weak supervision,'' in Proc. Eur. Conf. Comput. Vis., 2020, pp. 322–339.

[20]. P. Wu, J. Liu, and F. Shen, ''A deep one-class neural network for anomalous event detection in complex scenes,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 7, pp. 2609–2622, Jul. 2020.

[21]. P. Wu, J. Liu, M. Li, Y. Sun, and F. Shen, ''Fast sparse coding networks for anomaly detection in videos,'' Pattern Recognit., vol. 107, pp. 1–30, Nov. 2020.

[22]. J. Kim and K. Grauman, ''Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2009, pp. 2921–2928. [23] Y. Chang, Z. Tu, W. Xie, and J. Yuan, ''Clustering driven deep autoencoder for video anomaly detection,'' in Proc. Eur. Conf. Comput. Vis., 2020, pp. 329–345.

[23]. J. Li, X. Mei, D. Prokhorov, and D. Tao, ''Deep neural network for structural prediction and lane detection in traffic scene,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 3, pp. 690–703, Mar. 2017.

[24]. A. Stuhlsatz, J. Lippel, and T. Zielke, ''Feature extraction with deep neural networks by a generalized discriminant analysis,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 23, no. 4, pp. 596–608, Apr. 2012.

[25]. W. Luo, W. Liu, and S. Gao, ''A revisit of sparse coding based anomaly detection in stacked RNN framework,'' in Proc. ICCV, Oct. 2017, pp. 341–349.

[26]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''ImageNet classification with deep convolutional neural networks,'' in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

[27]. N. Dalal and B. Triggs, ''Histograms of oriented gradients for human detection,'' in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., Jun. 2005, pp. 886–893.

[28]. N. Dalal, B. Triggs, and C. Schmid, ''Human detection using oriented histograms of flow and appearance,'' in Proc. 9th Eur. Conf. Comput. Vis. Vienna, Austria: Springer, May 2006, pp. 428–441. [30] D. D'Avino, D. Cozzolino, G. Poggi, and L. Verdoliva, ''Autoencoder with recurrent neural networks for video forgery detection,'' in Proc. IS&T Int. Symp. Electron. Imag., Media Watermarking, Secur. Forensics, Jan. 2017, pp. 92–99.

[29]. A. Munawar, P. Vinayavekhin, and G. De Magistris, ''Spatio-temporal anomaly detection for industrial robots through prediction in unsupervised feature space,'' in Proc. IEEE Winter Conf. Appl. Comput. Vis., Santa Rosa, CA, USA, May 2017, pp. 1017–1025.

[30]. M. Ye, X. Peng, W. Gan, W. Wu, and Y. Qiao, ''AnoPCN: Video anomaly detection via deep predictive coding network,'' in Proc. 27th ACM Int. Conf. Multimedia, Oct. 2019, pp. 1805–1813.

[31]. V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, ''Anomaly detection in crowded scenes,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 1975–1981.

[32]. Q. Sun, H. Liu, and T. Harada, ''Online growing neural gas for anomaly detection in changing surveillance scenes,'' Pattern Recognit., vol. 64, pp. 187–201, Apr. 2017.

[33]. R. Chaker, Z. Al Aghbari, and I. N. Junejo, ''Social network model for crowd anomaly detection and localization,'' Pattern Recognit., vol. 61, pp. 266–281, Jan. 2017.

[34]. M. Ravanbakhsh, M. Nabi, E. Sangineto, L. Marcenaro, C. Regazzoni, and N. Sebe, ''Abnormal event detection in videos using generative adversarial nets,'' in Proc. Int. Conf. Image Process. (ICIP), Sep. 2017, pp. 1577–1581.

[35]. P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, ''Image-to-image translation with conditional adversarial networks,'' in Proc. CVPR, Jul. 2017, pp. 5967–5976.

[36]. Y. Tang, L. Zhao, S. Zhang, C. Gong, G. Li, and J. Yang, ''Integrating prediction and reconstruction for anomaly detection,'' Pattern Recognit. Lett., vol. 129, pp. 123–130, Jan. 2020.

[37]. O. Ronneberger, P. Fischer, and T. Brox, ''U-net: Convolutional networks for biomedical image segmentation,'' in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent., 2015, pp. 234–241.

[38]. W. Liu, W. Luo, Z. Li, P. Zhao, and S. Gao, ''Margin learning embedded prediction for video anomaly detection with a few anomalies,'' in Proc. 28th Int. Joint Conf. Artif. Intell., Aug. 2019, pp. 3023–3030.

[39]. X. Liang, L. Lee, W. Dai, and E. P. Xing, ''Dual motion GAN for futureflow embedded video prediction,'' in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 1762–1770. 107856 VOLUME 9, 2021 Z. Yang et al.: BR-GAN for Anomaly Detection in Videos

[40]. V. C. Tim Salimans, I. Goodfellow, and W. Zarema, ''Improved techniques for training GANs,'' Proc. Adv. Neural Inf. Process. Syst., 2018, vol. 19, no. 1, pp. 1–9.

[41]. S. Mahdizadehaghdam, A. Panahi, and H. Krim, ''Sparse generative adversarial network,'' in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), Oct. 2019, pp. 3063–3071.

[42]. J. Li, J. Jia, and D. Xu, ''Unsupervised representation learning of imagebased plant disease with deep convolutional generative adversarial networks,'' in Proc. 37th Chin. Control Conf. (CCC), Jul. 2018, pp. 1–16.

[43]. C. Vondrick, H. Pirsiavash, and A. Torralba, ''Generating videos with scene dynamics,'' in Proc. Adv. Neural Inf. Process. Syst. Conf., 2016, pp. 613–621.

[44]. F. Liu, L. Jiao, and X. Tang, ''Task-oriented GAN for PolSAR image classification and clustering,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 9, pp. 2707–2719, Sep. 2019.

[45]. PyTorch. Accessed: Jun. 2021. [Online]. Available: https://github. com/pytorch/

[46]. Y. Cong, J. Yuan, and J. Liu, ''Abnormal event detection in crowded scenes using sparse representation,'' Pattern Recognit., vol. 46, no. 7, pp. 1851–1864, 2013.

[47]. X. Zhu, J. Liu, J. Wang, C. Li, and H. Lu, ''Sparse representation for robust abnormality detection in crowded scenes,'' Pattern Recognit., vol. 47, no. 5, pp. 1791–1799, 2014.

[48]. S. Wu, B. E. Moore, and M. Shah, ''Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2010, pp. 2054–2060.

[49]. M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette, ‘‘Deep-cascade: Cascading 3D deep neural networks for fast anomaly detection and localization in crowded scenes,’’ IEEE Trans. Image Process., vol. 26, no. 4, pp. 1992–2004, Apr. 2017.

[50]. J. Wu, C. Zhang, T. Xue, W. T. Freeman, and J. B. Tenenbaum, ‘‘Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling,’’ in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 82–90.

[51]. D. Xu, E. Ricci, Y. Yan, J. Song, and N. Sebe, ‘‘Learning deep representations of appearance and motion for anomalous event detection,’’ in Proc. Brit. Mach. Vis. Conf., 2015, pp. 8.1–8.12,

[52]. G. Pang, C. Shen, L. Cao, and A. van den Hengel, ‘‘Deep learning for anomaly detection: A review,’’ 2020, arXiv:2007.02500. [Online]. Available: http://arxiv.org/abs/2007.02500

[53]. P. Wu and J. Liu, ‘‘Learning causal temporal relation and feature discrimination for anomaly detection,’’ IEEE Trans. Image Process., vol. 30, pp. 3513–3527, 2021.

[54]. D. Makris and T. Ellis, ‘‘Learning semantic scene models from observing activity in visual surveillance,’’ IEEE Trans. Syst., Man, Cybern. B, Cybern., vol. 35, no. 3, pp. 397–408, Jun. 2005.

[55]. W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, ‘‘A system for learning statistical motion patterns,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, no. 9, pp. 1450–1464, Sep. 2006.

[56]. C. Piciarelli and G. L. Foresti, ‘‘On-line trajectory clustering for anomalous events detection,’’ Pattern Recognit. Lett., vol. 27, no. 15, pp. 1835–1842, Nov. 2006.

[57]. A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, ‘‘Robust real-time unusual event detection using multiple fixed-location monitors,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 3, pp. 555–560, Mar. 2008.

[58]. K. He, X. Zhang, S. Ren, and J. Sun, ‘‘Deep residual learning for image recognition,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2016, pp. 770–778.

[59]. X. Nie, W. Jing, C. Cui, C. J. Zhang, L. Zhu, and Y. Yin, ‘‘Joint multiview hashing for large-scale near-duplicate video retrieval,’’ IEEE Trans. Knowl. Data Eng., vol. 32, no. 10, pp. 1951–1965, Oct. 2020.

[60]. X. Nie, B. Wang, J. Li, F. Hao, M. Jian, and Y. Yin, ‘‘Deep multiscale fusion hashing for cross-modal retrieval,’’ IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 1, pp. 401–410, Jan. 2021.

[61]. R. Cai, H. Zhang, W. Liu, S. Gao, and Z. Hao, ‘‘Appearance-motion memory consistency network for video anomaly detection,’’ in Proc. AAAI Artif. Intell., May 2021, vol. 35, no. 2, pp. 938–946.

[62]. M.-I. Georgescu, A. Barbalau, R. T. Ionescu, F. S. Khan, M. Popescu, and M. Shah, ‘‘Anomaly detection in video via self-supervised and multi-task learning,’’ 2020, arXiv:2011.07491. [Online]. Available: http://arxiv.org/abs/2011.07491

[63]. Y. Lu, F. Yu, M. Reddy, and Y. Wang, ‘‘Few-shot scene-adaptive anomaly detection,’’ in Proc. Eur. Conf. Comput. Vis., vol. 12350, Oct. 2020, pp. 125–141.

[64]. S. Smeureanu, R. T. Ionescu, M. Popescu, and B. Alexe, ‘‘Deep appearance features for abnormal behavior detection in video,’’ in Proc. Int. Conf. Image Anal. Process., vol. 10485, Oct. 2017, pp. 779–789.

[65]. R. T. Ionescu, S. Smeureanu, M. Popescu, and B. Alexe, ‘‘Detecting abnormal events in video using narrowed normality clusters,’’ in Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV), Jan. 2019, pp. 1–17.

[66]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ‘‘You only look once: Unified, real-time object detection,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.

[67]. S. Ren, K. He, R. Girshick, and J. Sun, ‘‘Faster R-CNN: Towards realtime object detection with region proposal networks,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

BIOGRAPHIES

ELANGOVAN.Tis an undergraduate student, department of computer science and engineering in paavai college of engineering.

KARTHIK RAJA.B is an undergraduate student, department of computer science and engineering in paavai college of engineering.

SANJAY KUMAR.Mis an undergraduate student, department of computer science and engineering in paavai college of engineering.

PRAKASH.P is an Assistant Professor, department of computer science and engineering in paavai college of engineering.