International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Anomaly Detection in Surveillance Videos Using Hybrid Deep Learning Model DBNSSGAN

Ms. R. Mariswari[1], Dr. V. Narayani[2]

Submitted: 07/05/2024 **Revised**: 20/06/2024 **Accepted**: 27/06/2024

Abstract: Urban planners and academics are influenced by the contemporary notion of smart cities to create modern, secure, and sustainable infrastructure that offers a respectable standard of living to its inhabitants. In order to improve citizen safety and well-being, video surveillance cameras have been installed to meet this demand. Even with today's scientific advancements in technology, abnormal event detection in CCTV footage and surveillance video remains difficult and time-consuming for humans to complete. Surveillance videos that contain anomalous events are automatically identified by video anomaly detection. The ability to determine whether a video contains anomalous events has improved in previous efforts. Since the development of deep learning methods, researchers have become interested in automatic video surveillance. The task of video anomaly detection, can be approached as a semi-supervised learning problem because to the strong bias in the datasets towards normal samples. The widely used reconstruction techniques solely use regular images to train the network. Assuming that the network cannot precisely recreate anomalous regions, these approaches identify anomalous occurrences by comparing the input with the reconstructed image. These approaches, however, have a significant drawback in that the anomaly zones are not sufficiently generic. This issue narrows the difference between the reconstructed and anomalous input images, which decreases the capacity to detect anomalies. In this paper the semi supervised Generative Adversarial Networks (SSGAN) is combined with Deep belief network (DBN) in detecting the abnormal events in surveillance video which greatly improves the quality of reconstruction and classifies the anomaly effectively. The outcomes are compared with the most advanced deep learning methods using two well-known surveillance data sets.

Keywords: *Deep learning, video anomaly detection, Surveillance videos, Anomaly detection, Deep belief network*

Introduction

Globally, both public and private security systems use video surveillance as a common technology. The volume of data generated by video monitoring systems these days necessitates an automatic analysis that could be handled by intelligent surveillance systems. In addition to enhancing public safety, video surveillance is essential for deterring crime and safeguarding a particular area [1]. In criminal proceedings, the recorded surveillance footage is frequently utilized as evidence. Anomalies are peculiar, context-dependent circumstances. Surveillance camera abnormalities occasionally represent innocuous situations, such as someone walking in the wrong way or running into someone else. A violent crime or a serious accident, for example, could put lives in danger, which is where anomalies may occur.

It is essential to identify and discover abnormalities like violence as soon as possible in order to deter crime and lower the crime rate. For people to manually search through this video data looking for violent incidents would be inefficient and time-consuming. A manual, labor-intensive method's efficiency may also be lowered by human error. Consequently, there is a great demand for effective and automatic techniques to identify violent or unusual activity, particularly in surveillance footage.

Deep learning has had an impact on intelligent video surveillance anomaly detection (AD) in recent years [2]. Because autoencoder models can perform semi-supervised training, they have been a stalwart in this discipline [3]. The performance of AD system has been greatly increased by autoencoders (AE), Convolutional Neural Networks (CNNs) and other deep learning architectures [4-6]. However, for systems with constrained computational resources, the majority of suggested architectures have astronomical, unaffordable, or insufficient reaction times. The demand for more effective models that may produce acceptable outcomes for the AD assignment in video surveillance is what inspired this paper.

In order to automate the anomaly detection process, current research in the subject focuses on applying deep learning approaches, primarily in three categories: end-to-end anomaly score learning, learning regular representations of normal data, and generic feature extraction. The superior dimensionality reduction

[1] Research Scholar,Reg No: 21211282282011 Department of Computer Science, St. Xavier's College (Autonomous) Palayamkottai -Tirunelveli.
Email: marissiram547@gmail.com
Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627 012, Tamilnadu.
[2] Assistant Professor, Department of Computer Science, St. Xavier's College (Autonomous) Palayamkottai-Tirunelveli
Email:narayaniv1979@gmail.com
Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627 012, Tamilnadu.

capabilities of deep neural networks in comparison to well-known linear techniques like random projections [7] and principal component analysis (PCA) [8] and iii) their simplicity in end-to-end feedforward implementation. On the other hand, splitting apart anomaly detection and feature extraction typically results in less than ideal performance. In order to reduce dimensionality, these methods typically combine AE/convolutional AE (AEs/CAEs) [9] and generative adversarial networks (GANs) with well-proven classification techniques. Keeping in this mind and to reduce the computation time the GAN and Deep Belief Network is combined in this work for automatic anomaly detection which perform both feature extraction and classification in a single hybrid model.

This work uses GAN and DBN deep learning techniques to handle the problem of violence detection. The work makes the following specific contributions:

Suggest a deep learning-based method (GAN and DBN) for classifying videos as having normal or violent content. Because of its computing efficiency, our approach is useful in practical applications.

Present a fully functional system that implements the proposed methods for automated violence detection with two benchmark dataset.

Outperforms a number of cutting-edge techniques for violence detection on widely used video classification criteria. Furthermore, excellent classification accuracy can retain with our method.

Related Works:

Deep learning methods provide the efficient result in anomaly detection. Among traditional methods this section shows the hybrid novel deep learning methods applied on UCSD and CUHK Avenue datasets with their evaluation result.

Pushpajit Khaire, Praveen Kumar offers a CNN-BiLSTM [10] autoencoder architecture that is multi-modal and semi-supervised DL based, with the goal of detecting abnormal events in important surveillance environments such as bank ATMs. The framework's ability to be trained using only incorrectly labeled typical video examples makes it noteworthy. We use the power of transfer learning by using a compact pretrained CNN to extract pertinent video properties, therefore significantly reducing the computing cost of training and detection. Avenue and UCFCrime2Local, two benchmark video anomaly datasets from the real world, are used to test the suggested methodology.

For real-time VAD, Singh et al. [11] suggest a Constrained Generative Adversarial Network (CVAD-GAN). The fine-grained features that CVAD-GAN learns from normal video frames are improved when white Gaussian noise is added to the input video frame with confined latent space. Furthermore, in order to comprehend the larger context of intricate video scenes in real-time, the skip-connection and dilated convolution layers maintain the information across layers. A greater Area Under Curve (AUC) score is obtained using our suggested method. On the UCSD Peds1, UCSD Peds2, and CUHK Avenue datasets, CVAD-GAN obtains an AUC score of 98.0%, 97.8%, and 94.0%.

Ullah proposed a framework to extracts both spatial and temporal information from surveillance videos in order to identify anomalous events. It works in two stages: first, it extracts spatial features using an effective backbone CNN model, and then it uses these features to pass through to a transformer-based model in order to determine the long-term temporal relationships between different complex surveillance events. By feeding the backbone model's features into a sequential learning model, which uses temporal self-attention to create an attention map, the suggested framework is able to efficiently learn spatiotemporal features and identify anomalous occurrences.

The goal of Ning et al.'s [13] Memory-enhanced Appearance-Motion Consistency (MAMC) architecture is to comprehend intricate appearance-motion consistency patterns in video data in order to identify anomalies. Developing an Appearance-Motion Fusion (AMF) module is the first step in this approach. Its goal is to generate a dependable scene dynamics representation that captures appearance-motion consistency. Subsequently, the memory module processes the consistency data, enhancing the capacity to discriminate between normal and anomalous events. The AUCs of 96.7%, 87.6%, and 71.5% is achieved using three benchmark dataset.

Proposed Methodology

This research provides a deep belief network-based multi-modal, semi-supervised DL models for detecting anomalous occurrences. The system is notable since it only needs poorly tagged typical video examples to be trained. The power of semi-supervised GAN is used to remove crucial video features, hence reducing the computing cost of training and detection. A distinct dataset of surveillance images is provided since there was no publicly available surveillance dataset. Using the gathered dataset, the suggested framework is examined. The suggested framework may be able to spot irregularities in actual surveillance situations, which can include both indoor and outdoor settings.This study uses a three-phase approach: the input image is first scaled using the min max scaler, and then U-Net is used to capture local and global temporal data for the purpose of identifying odd areas in the image. In the end, the anomaly is successfully classified by the suggested semi-supervised GAN classifier, which is based on deep belief networks.

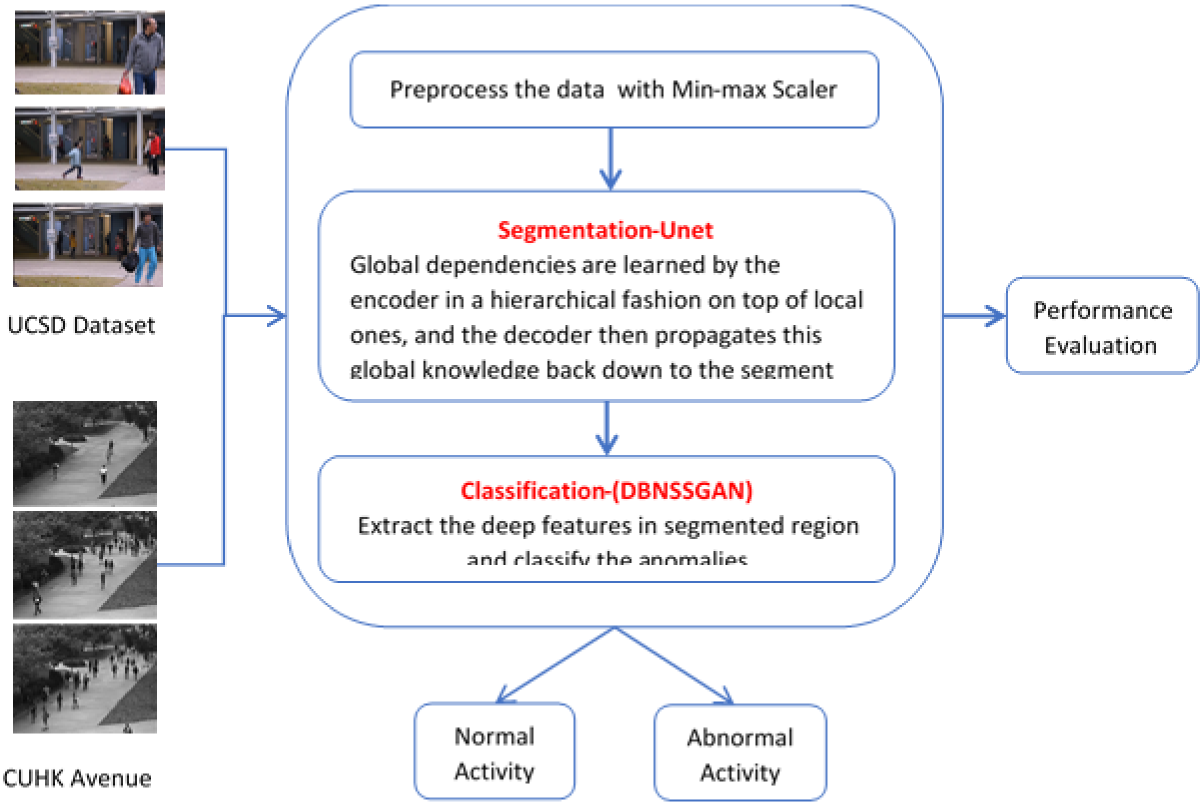

Fig 1. Architecture of DBNSSGAN Anomaly Detection model

Dataset

UCSD Dataset

In Ped1, there are 16 test and 34 training videos in the UCSD dataset [14]. Ped1's picture dimensions are 238 × 158 pixels. This dataset contains anomalous events related to wheelchairs, cars, skating, and cycling.

The Chinese University of Hong Kong (CUHK) Avenue

This dataset has 21 testing films and 16 training videos at a resolution of 25 frames per second [15]. It is a common occurrence for an object's size to decrease with increasing distance from the video surveillance system recording it.

Preprocessing

The first step involved preprocessing the medical data using the min-max scalar approach. With the original data, Min-Max normalization applies a linear transformation. Scale each feature to a specified range in order to transform it. In order to place each feature in the training set within the specified range [0, 1], this estimator scales and translates each feature separately. The attributes D's minimal and maximal values are indicated by the symbols B_{min} and B_{max}. A Min-Max Scalar (S) has the following definition.

$$S = \frac{B - B_{min}}{B_{max} - B_{min}} \tag{1}$$

This can be applied to the anomalous dataset using the Python MinMaxScaler function. Rather of decreasing the significance of anomalies, MinMaxScaler reduces them linearly within a specified range, where the largest value corresponds to the maximum and the smallest value to the minimum.

U Net segmentation

The two paths that make up U-Net are in charge of localizing and classifying every object in a frame [16]. The first pathway, called the encoder, uses max-pooling layers, rectified linear units (ReLU), and repeated convolution to down sample data in order to extract features. The recovered feature maps are re-sampled using convolution, concatenation, and ReLU layers to obtain localization information in the second pathway, known as the decoder. The U-Net design efficiently extracts location and segmentation data through the encoder and decoder pathways, allowing extremely accurate anomaly segmentation models to be trained at a minimal cost.

Encoder:

It is also known as the contracting network, or encoder network. The first question this network attempts to answer is "what" is in the image by learning a feature map of the input image. A U-Net does not contain completely connected layers at the end since the output we need is a mask the same size as our input image rather than the class label. Other than that, it is similar to any classification task we carry out with convolutional neural networks.

There are four encoder blocks in this encoder network. Every block has a Relu activation function after two convolutional layers with a 3*3 kernel size and appropriate padding. A max pooling layer with a 2*2 kernel size receives this as input. The bottleneck layer is located between the encoder and decoder networks. The layer at the bottom is this one. It is composed of two convolutional layers, then Relu.

Decoder:

Decoding resample feature maps to match the dimensions of the original image. Using skip connections, this network takes the feature map from the bottleneck layer and creates a segmentation mask. The object's location within the image is the second question the decoder network attempts to answer. There are four decoder blocks in it. Each block begins with a transpose convolution with a kernel size of 2*2, denoted in the diagram as up-conv. The matching skip layer connection from the encoder block is concatenated with this output. Next, a Relu activation function is applied after the use of two convolutional layers with a kernel size of 3*3.

Deep belief network_ Semi supervised GAN (DBN_SSGAN)

The most widely used deep learning technique is the deep belief network (DBN), which Hinton originally proposed. Compared to previous algorithms, this one discovers the optimal settings faster and picks up new information quicker. The core elements of a conventional DBN are a logistic regression layer and an unsupervised learning module based on restricted Boltzmann machines (RBMs).

A DBN is produced via layer-wise training of the RBM, a popular stochastic neural network. Two layers make up the RBM: a layer of binary-valued neurons and a hidden layer of Boolean neurons. Despite being between layers, the connections between the neurons within a layer are not symmetrical or bidirectional.

The probability distribution between the two levels in layer-wise setups is found using the energy function of the configuration, which is provided in (2). The probability distribution can then be expressed using the following equation.

$$Ef(a,b) = -\sum_{m=1}^{z_a} x_m a_m - \sum_{n=1}^{z_b} y_n b_n - \sum_{m=1}^{z_a}\sum_{n=1}^{z_b} b_n W_{n,m} b_m \quad (2)$$

$$\mathrm{pd} = \frac{\mathrm{e}^{-\mathrm{Ef(a,b)}}}{\sum_{\mathrm{a}}\sum_{\mathrm{b}} \mathrm{e}^{-\mathrm{Ef(a,b)}}} \quad (3)$$

$W_{n,m}$ neurons make up the visible layer, while b_nBoolean hidden neurons make up the hidden layer. The two layers are separated by the weight matrices b_m and b_n. x_m and $y_n$are the biases for the two layers. The activation probability functions are then expressed using an equation.

$$pd(a_m = 1|b) = sig\left(\alpha_m + \sum_{n=1}^{i_b} W_{n,m} b_m\right) \quad (4)$$

$$pd(b_m = 1|a) = sig\left(y_m + \sum_{m=1}^{i_a} W_{n,m} a_n\right) \quad (5)$$

sig() is another way to represent the logistic sigmoid function. This is supported by the pre-training principles since the layer biases and weight matrices can be learned without supervision. The data's peculiarities are too intricate for a single concealed RBM. A DBN, which is constructed by stacking layers of RMBs in a hierarchical fashion and concluding with a logistic regression layer, can progressively extract deep features from the input dataset. The training data are used as inputs to pre-train the DBN's initial RBM so that it can operate independently.

Once the weight matrix and bias parameters are established, the output of the first RBM is selected to be the input for the second RBM. The unseen layers of the first two RBMs are then repeatedly trained to produce a new RBM using the same procedure. The next step is to carefully monitor the network's training while superimposing a thorough predictor (such a logistic regression layer) on top of it. After implementing the aforementioned steps, the back-propagation (BP) approach is used to fine-tune the parameters of the trained network.

In recent years, GANs have been widely used in both image processing [15–16] and natural language processing [14–16]. The cornerstone of Generative Adversarial Networks (GAN) was laid by game theory.

The GAN framework consists of the discriminator and generating models. The generator model transforms a latent vector from a known distribution into a data space that it receives as input. The discriminator model makes an effort to differentiate between a fake sample from the generator and a real sample from the data space.

Let q be the samples of the D_d data distribution and l be the latent vector in S_d sampled from the noise distribution $N(x)$, and. Let $A(l;\theta_A)$ be the generator model; A is a differentiable function with parameters θ_A that is represented by a deep belief network.

Likewise, suppose that the discriminator model is represented by a deep belief network $B(q;\ \theta_B)$, and that the scalar output of this network indicates the likelihood that q originates from D_d instead of $A(l)$. $log[B(q)]$ and $log\left(1 - B\left(A(l)\right)\right)$are thus maximized by B during training. B is trained to trick the discriminator by simultaneously minimizing$log(1 - B(A(l)))$. By optimizing the following objective function, the two models act as opponents in a two-player min-max game:

Let $B(q;\ \theta_B)$), be the deep belief network that represents the discriminator model in a similar way. The scalar output of this network indicates the probability that p comes from D_d rather than $A(l)$. In order to maximize $log[B(q)]$ and $log\left(1 - B\left(A(l)\right)\right)$, B is so trained. Moreover, B is simultaneously trained to minimize minimize $log\left(1 - B\left(A(l)\right)\right)$in order to fool the discriminator. Stated otherwise, this is a two-player min-max game in which two models fight to maximize the previously given goal function:

$$\min_{A} \max_{B} E(A,B) = \underset{q\sim D_d}{\mathrm{V}}\left[\log\left(B(i)\right)\right] + \underset{l\sim N(x)}{\mathrm{V}}\left[\log\left(1 - B\left(A(l)\right)\right)\right] \quad (6)$$

As long as A and B have sufficient capacity, the training criterion allows the data generating distribution to converge to a genuine data distribution D_d, according to Goodfellow et al. [17]. The min-max objective function, however, necessitates the identification of a Nash equilibrium, which can include a non-convex function with high dimensional and continuous parameters. Given that gradient descent methods are employed in GAN training, it usually fails to converge when attempting to find the Nash equilibrium.

Now let's look at a typical K class classifier model, where a given data point (p) is mapped to one of the M possible outputs. One such model, $p_m(y|q, 1 \leq\ y \leq \mathrm{K})$., yields the class probability corresponding to each class. In supervised learning, the model is trained by reducing the cross-entropy between the true label and the prediction distribution $V_{q,y\sim D_d}[logP_{model}(y|q)]$.

Semi-supervised learning can be used to improve any such traditional classifier by only adding fresh unlabeled data generated by the GAN generator. This unlabeled data can be utilized as a new class (y = K + 1) for unsupervised learning. In supervised scenarios, the discriminator can be compared to the B standard classifier $B_{ss}(q) = P_{model}\ (y|q, 1 \leq\ y \leq \mathrm{K})$.. While q in the original GAN function reflects the likelihood that q is real, $P_{model}(y = K + 1|q)$ in unsupervised learning is equivalent to the likelihood that q is a fake. The objective of semi-supervised learning is therefore

$$\min_{A} \max_{B} \underset{q\sim D_d}{\mathrm{V}}\left[\log\left(B(i)\right)\right] + \underset{l\sim N(x)}{\mathrm{V}} \log\left(1 - B\left(A(l)\right)\right) + \alpha_s \underset{q,y\sim D_d}{\mathrm{V}} \log\left(B_{ss}(p)\right) \quad (7)$$

The hyperparameter α_s is incorporated to maintain equilibrium between the supervised and unsupervised losses.

Result and discussion

In this section, the proposed DBNSSGAN model performance is assessed on two anomaly benchmarks and compares it with the state-of-the-art techniques Python serves as the foundation for our suggested model. Sklearn and Tensorflow libraries are used to implement the testing and training procedures. Accuracy, precision, recall, and F1 score are taken into account while assessing the DBNSSGAN's quantitative performance.

The anomaly detection system will be implemented using Python 3.7.5 and will adhere to the following specifications: Windows 10 PC with an Intel i3-core and 2 GB of RAM. Table 2 provides the performance metrics values for the current model, and Table 3 displays a comparison with the existing model.

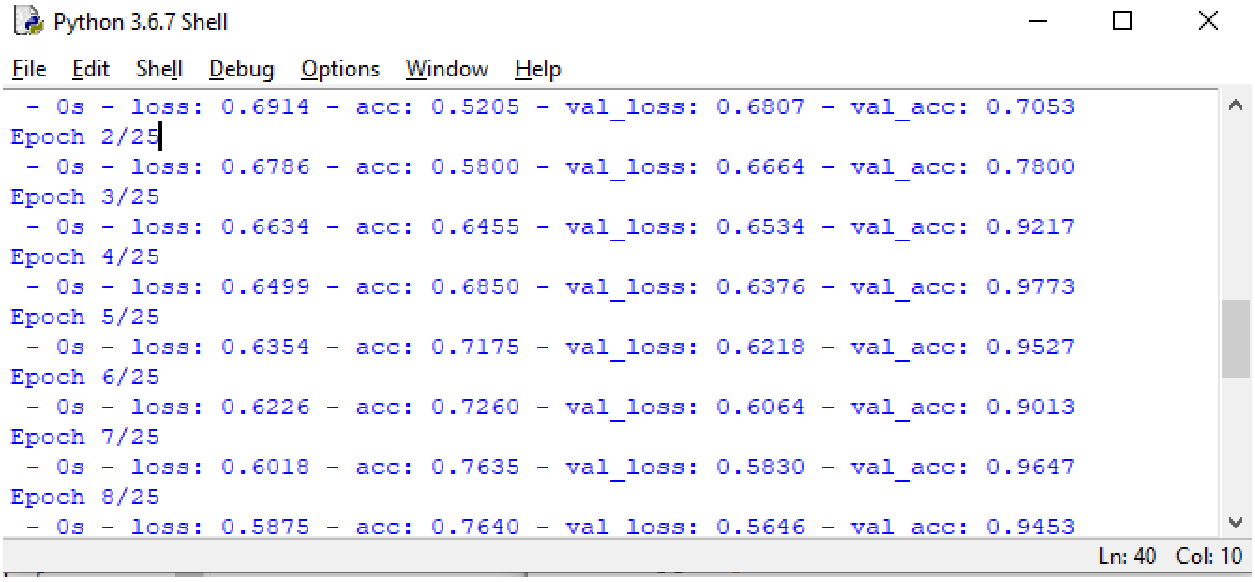

Fig 1 python simulation workout

Fig 2. Sample images of CUHK Avenue dataset

Fig 3 Sample images of UCSD dataset

Performance metrics calculation for USCD dataset

$$Precision = \frac{1458}{1458 + 27} = 98.84$$

$$Recall = \frac{1458}{1458 + 161} = 98.51$$

$$F1\ score = \frac{1458}{1458 + \frac{1}{2}(22 + 17)} = 98.68$$

$$Accuracy = \frac{1458 + 1503}{1458 + 1503 + 22 + 17} = 98.70$$

Performance metrics calculation for CUHK dataset

$$Precision = \frac{1024}{1024 + 5} = 97.24$$

$$Recall = \frac{1024}{1024 + 29} = 99.15$$

$$F1\ score = \frac{1024}{1024 + \frac{1}{2}(5 + 29)} = 98.36$$

$$Accuracy = \frac{1024 + 942}{1024 + 942 + 5 + 29} = 98.30$$

Table 2: performance of DBNSSGAN

Method	**Dataset**	**Precision**	**Recall**	**F1score**	**Accuracy**
DBNSSGAN	USCD	98.84	98.51	98.68	98.70
DBNSSGAN	CUHK	97.24	99.15	98.36	98.30

Table 3: DBNSSGAN comparison with existing model

Methods	USCD	CUHK
DBN	96.98	96.54
CNN	97.05	97.00
GAN	94.94	94.03
SSGAN	97.45	97.72
DBNSSGAN	98.70	98.30

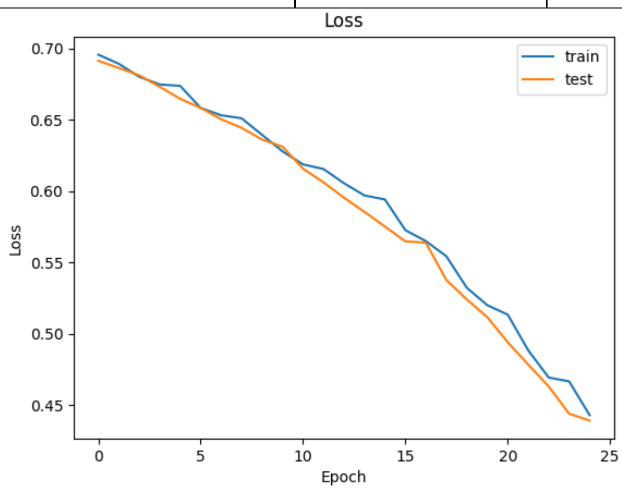

Figure 4. Loss Report of DBNSSGAN on USCD dataset

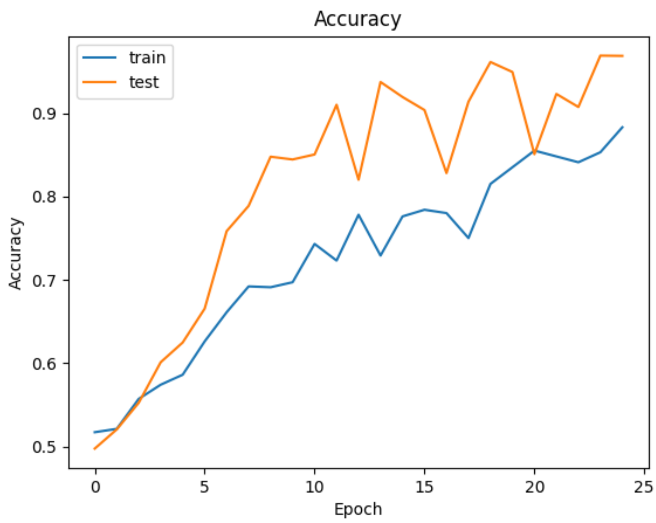

Fig 5. Accuracy Report of DBNSSGAN on USCD dataset

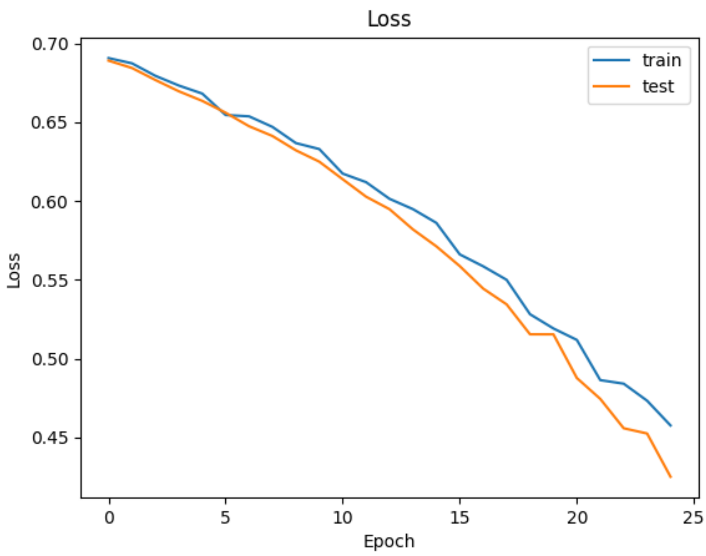

Fig 6 Loss Report of DBNSSGAN on CHUK dataset

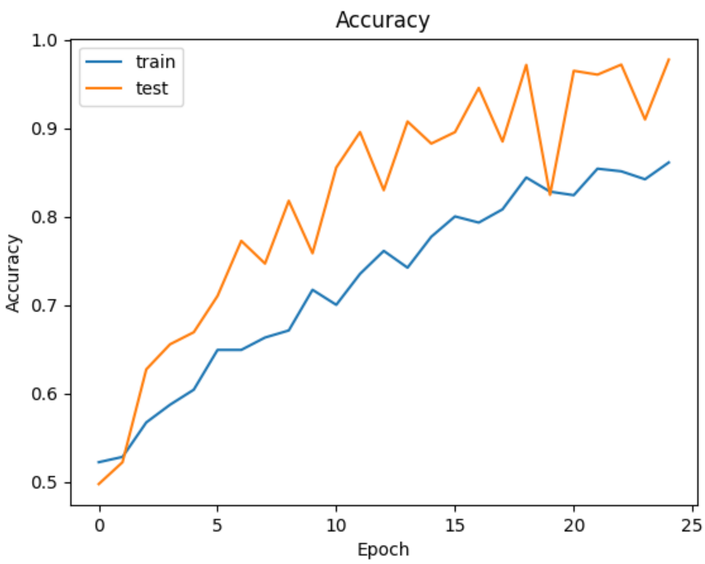

Fig 7 Accuracy of DBNSSGAN on CHUK dataset

It has been observed that our framework infers anomalous frames, particularly in frames with a high anomaly score, when an individual is throwing objects, sprinting, or walking in the incorrect direction. The loss and accuracy of the proposed model is computed and shown in above figures. From the figure 3 and 4, the accuracy of the semi supervised classification is 98.80% and 98.73% on USCD and CHUK dataset. Following these experimental results, it is clear that applying DBNSSGAN with the Anomaly data benefits anomaly detection in real time application.

Conclusion

The uncommon and unpredictable character of anomalous events in real-world circumstances makes them difficult to detect in recordings. A video anomaly detection method using Unet segmentation and Min-Max scaler is presented in this paper under the name DBNSSGAN. Both the USCD and CUHK benchmark datasets are widely used for evaluation purposes. For anomalous events with anomalous behaviors and objects, the suggested framework is appropriate. On the USCD and CHUK datasets, the semi-supervised classification accuracy is 98.80% and 98.73%, respectively. These experimental findings demonstrate that using DBNSSGAN in conjunction with anomaly data enhances anomaly detection for real-time applications. In order to confirm whether these events are genuinely aberrant or merely infrequent normal occurrences, our future work will concentrate on continual learning of unknown events.

References

[1] Piza, E.L., Welsh, B.C., Farrington, D.P. and Thomas, A.L., 2019. CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis. *Criminology & public policy*, *18*(1), pp.135-159.

[2] Chriki, A., Touati, H., Snoussi, H. and Kamoun, F., 2021. Deep learning and handcrafted features for one-class anomaly detection in UAV video. *Multimedia Tools and Applications*, *80*, pp.2599-2620.

[3] Yan, S., Shao, H., Min, Z., Peng, J., Cai, B. and Liu, B., 2023. FGDAE: A new machinery anomaly detection method towards complex operating conditions. *Reliability Engineering & System Safety*, *236*, p.109319.

[4] Ribeiro, M., Lazzaretti, A.E. and Lopes, H.S., 2018. A study of deep convolutional auto-encoders for anomaly detection in videos. *Pattern Recognition Letters*, *105*, pp.13-22.

[5] Siddalingappa, R. and Kanagaraj, S., 2021. Anomaly detection on medical images using autoencoder and convolutional neural network. *International Journal of Advanced Computer Science and Applications*, (7).

[6] Yin, C., Zhang, S., Wang, J. and Xiong, N.N., 2020. Anomaly detection based on convolutional recurrent autoencoder for IoT time series. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *52*(1), pp.112-122.

[7] Mao, J. and Jain, A.K., 1995. Artificial neural networks for feature extraction and multivariate data projection. *IEEE transactions on neural networks*, *6*(2), pp.296-317.

[8] San Martin, G., López Droguett, E., Meruane, V. and das Chagas Moura, M., 2019. Deep variational auto-encoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis. *Structural Health Monitoring*, *18*(4), pp.1092-1128.

[9] Zaied, S., Toumi, A. and Khenchaf, A., 2018, March. Target classification using convolutional deep learning and auto-encoder models. In *2018 4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)* (pp. 1-6). IEEE.

[10] Khaire, P. and Kumar, P., 2022. A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments. *Forensic Science International: Digital Investigation*, *40*, p.301346.

[11] Singh, R., Sethi, A., Saini, K., Saurav, S., Tiwari, A. and Singh, S., 2024. CVAD-GAN: Constrained video anomaly detection via generative adversarial network. *Image and Vision Computing*, p.104950.

[12] Ullah, W., Hussain, T., Ullah, F.U.M., Lee, M.Y. and Baik, S.W., 2023. TransCNN: Hybrid CNN and transformer mechanism for surveillance anomaly detection. *Engineering Applications of Artificial Intelligence*, *123*, p.106173.

[13] Ning, Z., Wang, Z., Liu, Y., Liu, J. and Song, L., 2024. Memory-enhanced appearance-motion consistency framework for video anomaly detection. *Computer Communications*, *216*, pp.159-167.

[14] http://www.svcl.ucsd.edu/projects/anomaly/dataset.html

[15] https://www.cse.cuhk.edu.hk/leojia/projects/detectabnormal/dataset.html

[16] Jiang, J., Zhu, J., Bilal, M., Cui, Y., Kumar, N., Dou, R., Su, F. and Xu, X., 2022. Masked swin transformer unet for industrial anomaly detection. *IEEE Transactions on Industrial Informatics*, *19*(2), pp.2200-2209.

[17] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2020. Generative adversarial networks. Communications of the ACM, 63(11), pp.139-144.