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

Deep Learning Based Integrated Model using Deep Belief Network with Semi Supervised GAN for Anomaly Detection in Surveillance Video

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

Submitted: 15/10/2023 **Revised**: 13/12/2023 **Accepted**: 25/12/2023

Abstract: In the modern era of smart cities, video surveillance has grown in importance. In order to monitor infrastructure property and ensure public safety, numerous surveillance cameras are installed in both public and private spaces. Large volumes of video data are produced by these surveillance cameras, making it impossible for a human observer to manually watch these hours-long recordings every day and find any unwelcome or unusual activity. The process of identifying abnormal behavior involves figuring out what behavior is different from normal. In a monitoring paradigm, these events will range from kidnapping to traffic accidents, violence to war, and so forth. Because anomalous occurrences occur often, video anomaly detection via video surveillance is a challenging scientific endeavor. This paper offers an AD-DBNSSGAN, a multi-modal semi-supervised deep learning system based on deep belief networks and GANs for identifying anomalous instances in critical surveillance situations. The framework is significant since it can be trained with only poorly tagged normal video or image samples. We contributed a unique dataset of surveillance photos because there was no public surveillance dataset available. The gathered dataset is used to test the suggested framework. The suggested framework may be used to detect abnormalities in real-world surveillance locations both indoors and outdoors, and the results demonstrate that it can produce results that are competitive with other cutting-edge techniques.

Keywords: *Video surveillance, Video anomaly detection, Deep belief networks, Anomaly detection, deep learning, semi supervised GAN.*

Introduction

The use of video surveillance has grown in importance and popularity in the current era of smart cities. Many surveillance cameras are placed in public and private locations, such as schools, ATMs, offices, train platforms, traffic lights and other places, in order to monitor infrastructure and ensure public safety [1]. A significant amount of data is continuously produced by these surveillance cameras. Manually keeping an eye on these lengthy live video streams and spotting any unusual activity or unwelcome incident takes a great deal of time and effort for a human observer [2]. By automatically identifying these unusual occurrences in films, monitoring labor can be greatly decreased. Thus, video anomaly detection is a crucial computer vision study topic [3].

The method of automatically identifying anomalous events in pictures or videos using computer vision algorithms is known as anomaly detection. The odd or anomalous events are frequently dependent on the surroundings, subjective, and contextual in character. Among the difficult problems in video anomaly identification are unwanted incidents like fire explosions, stampedes, accidents in public spaces, and irregular human actions including fights, robberies, and breaking traffic laws [4]. Finding these unusual occurrences in public areas helps protect infrastructure and save lives. Hence, the primary goal of video-based surveillance systems is the automatic and precise real-time detection of such events.

Deep learning-based techniques have recently been taken into consideration for cutting down on time complexity without sacrificing detection performance, as low computational complexity and high detection accuracy are critical for surveillance video analysis. Convolutional neural networks (CNNs) are being used in computer vision more and more because of their remarkable performance in image recognition tasks [5], [6]. CNNs are predicted to become more widely used in the future due to their rapid evolution in various academic domains. These learning algorithms still have a lot of room to expand given the abundance of big data and the rapid increase in processing power.

This research provides a Deep belief network based multi-modal semi-supervised deep learning system to identify anomalous occurrences. By removing significant video features, we take advantage of the capabilities of semi-supervised GAN to drastically lower the computational complexity of training and detection.

[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.

Since there was no public surveillance dataset accessible, we submitted a unique dataset of surveillance photos. The proposed framework is tested using the collected dataset. The proposed framework has the potential to identify anomalies in real-world surveillance scenarios, encompassing both indoor and outdoor environments.

The Contributions of the present study are

This article provides a semi-supervised deep learning model with a deep belief network and GAN framework (AD-DBNSSGAN) for real-time video anomaly identification in surveillance footage utilizing UCF-Crime data. We believe this to be the first semi-supervised deep learning system for anomaly identification in videos.

Training for the proposed semi-supervised AD-DBNSSGAN system only requires video samples with weak labels. Furthermore, no abnormal video samples are needed for the training process to identify abnormalities. Because of this, the suggested architecture gets around the drawbacks of using imbalanced normal and anomalous data for training.

The suggested framework is tested using additional real-world benchmark datasets, such as UCFCrime, which was gathered in various settings. The outcomes were competitive with those of other cutting-edge methods. Experiments demonstrate that it may be used for anomaly detection in real-world surveillance locations in both indoor and outdoor contexts.

Literature Review

It is challenging to identify different anomalous events from films because there aren't many of them and there aren't many significant datasets for them. This section describes the several approaches that are now in use for utilizing deep learning techniques to identify anomalous occurrences. The research was driven to produce an efficient method for identifying anomaly action from video by outlining the methods' benefits and difficulties in detecting abnormalities.

Using images from CCTV, Nischita Waddenkery and Shridevi Soma [7] devised an effective method for identifying theft activities. The recently introduced method performs crime identification by utilizing videos that are taken from exterior security cameras. After that, frame sequence extraction and bounding box segmentation are applied to the videos. After video summarization, the acquired frames are used for criminal detection using the Adam-Dingo_Deep maxout network, which classifies the theft as either normal or an incident.

Maryam Qasim and Elena Verdu [8] create an automated method that uses a simple recurrent unit (SRU) and a deep convolutional neural network (CNN) to identify anomalies in videos. While the SRU gathers temporal characteristics, the ResNet architecture uses the incoming video frames to extract high-level feature representations. The expressive recurrence and highly parallelized implementation capabilities of the SRU improve the accuracy of the video anomaly detection system. the UCF-Crime dataset was used to examine the ResNet50 + SRU model. With 91.24% accuracy, this study was able to distinguish between typical and abnormal behavior.

Shao et al [9] offers COVAD, a unique technique for detecting anomalies in videos that primarily concentrates on the area of interest inside the video rather than the full video. Based on an autoencoded CNN and a coordinated attention mechanism, our proposed COVAD technique can efficiently extract meaningful items and their dependencies from the video. Our system can more accurately forecast the future motion and appearance of objects in a video by using the memory-guided video frame prediction network that is currently in use.

Table 1: An overview of existing work

S.No	Reference	Model	Accuracy
1	[7]	Adam Dingo_Deep Maxout Network (ADDMN)	94.5%
2	[8]	ResNet50 + SRU	91.24%
3	[9]	COVAD	96.5%
4	[10]	Light weight neural network (LWNN)	95.72%
5	[11]	IBaggedFCNet (IBFCNet)	92.06
6	[12]	A3DConvNet	91%
7	[13]	KFCRNet	91%

While learning all segments at once is vital, WATANABE et al.'s [10] analysis of the current, effective methods over the past few years revealed that high accuracy can be attained without regard to the

temporal order of the segments. To automatically extract information from all input segments that are crucial for evaluating what is normal or abnormal, suggest a simple model with a self-attention mechanism. Despite using a neural network with 1.3% trainable parameters, our approach performs better than previous approaches on the UCF-Crime dataset.

Zahid et al [11] Introduced IBaggedFCNet, a bagging framework that uses ensembles' strength for robust classification to find abnormalities in videos. Our method looks at the state-of-the-art Inception-v3 image classification network and does not require segmenting the video before extracting features, which can lead to inconsistent segmentation outcomes and a large memory footprint. We demonstrate improvement by empirical means on several benchmark datasets, with the UCF-Crime dataset being the most notable.

Ansari et al [12] proposed the use of 18 convolutional operations in A3DConvNet, a 15-layer deep deep neural network, to efficiently analyze video input and generate spatiotemporal data. These properties are utilized by the integrated dense layer for an effective learning process, and the output layer, the softmax layer, is responsible for labeling the sequences. As a follow-up to this paper, we have also developed a dataset of video clips representing aberrant human behaviors at megastores/shops.

Two keyframe-based algorithms, EvoKeyNet and KFCRNet, were proposed by Shoaib et al [13]. to detect violence in big video collections. The suggested categorization models, EvoKeyNet and KFCRNet, make use of feature extraction from ideal keyframes. Evolutionary algorithms are used by EvoKeyNet to choose the best feature attributes, whereas an ensemble of LSTM, Bi-LSTM, and GRU models with a voting method is used by KFCRNet. Our main contributions, which tackle the problem of violence detection in dynamic surveillance scenarios, include the creation of effective keyframe selection techniques and classification models. The suggested models achieve 91% accuracy (UCF-Crime) and 91% computational efficiency (ShanghaiTech), respectively, outperforming the current techniques in these areas.

Methods

This paper focuses on combining two deep learning mechanisms offer a novel anomaly detection method that is mostly developed to attain the better performance. The architecture of our proposed method is illustrated in figure 1. The data is taken from the ucf crime data and the input image is initially processed with Gaussian filter and the processed image is segmented using the Unet segmentation the crime region are extracted in this phase and the deep features of those segmented area are extracted through the auto encoder feature extractor. Eventually the anomaly present in the input data is predicted and classified using the hybrid model of semi supervised GAN and deep belief network.

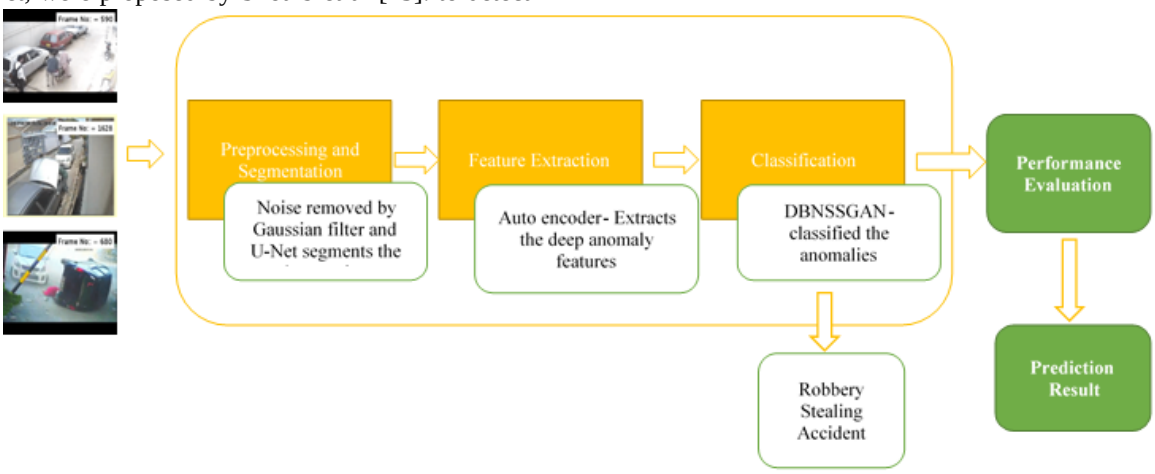

Fig 1. Architecture of the Proposed DBNSSGAN Anomaly Detection model

Dataset

Our technique is evaluated on a new large-scale dataset University of Central Florida crime (UCF-Crime). It consists of unedited surveillance recordings that cover seven abnormalities that occur in real life: abuse, assault, robbery, shooting, theft, and vandalism. Because of their substantial bearing on public safety, these anomalies have been chosen. The 1000 images are collected from the Surveillance System under the 7 classes. Totally 7000 images are trained and 200 images for each class tested (totally 1400 images) for the anomaly detection. The sample images of each class is given in table 2.

Table 2: sample data from each class in UCF-CRIME data

Abuse

Assault

Robbery

Stealing

Shooting

Vandalism

Accident

Preprocessing

Due to the input size's flexibility, it can be adjusted to fit all of the algorithms employed in this study. For data normalization, we used the min-max scaling strategy in addition to regular scaling. Normalization keeps the

model from becoming stale and is essential for learning on neural network.

The following is a definition of a Min-Max Scalar (MMS).

$$MMS = \frac{DS - DS_{min}}{DS_{max} - DS_{min}} \quad (1)$$

DS represents a data sample in a column, the value of DS_{min} and DS_{max} denotes the minimum and maximum data sample in the column. The numbers after Min-Max normalization fall between [0, 1], which is occasionally a desirable quality for input [16].

The data around a mean with a single standard deviation are normalized using conventional scaling which is represented by the following equation.

$$cs = \frac{DS - \mu}{\sigma} \quad (2)$$

U Net segmentation

A neural network architecture called U-net was primarily developed for image segmentation [1]. Two pathways make up the fundamental framework of a U-net design. The first path is the contraction path, which is sometimes referred to as the analytic path or the encoder. It functions similarly to a standard convolution network and offers categorization data. The second is an expansion path, sometimes referred to as the synthesis path or the decoder. Comprised of concatenations and up-convolutions using characteristics from the contracting pathThis augmentation allows the network to gather localized classification information. The extension path further increases the resolution of the output, allowing it to be sent to a final convolutional layer for complete image segmentation. The resultant network has a u-shaped form since it is nearly symmetrical.

As previously stated, there are two components to the U-net network: The first one employs a standard CNN architecture and is called the contracting path. Every block along the contracting path is made up of two consecutive 3 × 3 convolutions, a max-pooling layer, and a ReLU activation unit. There are multiple iterations of this arrangement. The second section of U-net, referred to as the expanded route, is where the innovation lies, using 2 × 2 up-convolution, each stage upsamples the feature map. Subsequently, the upsampled feature map is concatenated and cropped using the feature map from the appropriate layer in the contracting path.

The following steps use ReLU activation and two consecutive 3 × 3 convolutions. To construct the segmented image and reduce the feature map to the required number of channels, a last step includes applying an extra 1 × 1 convolution. Since pixel features near the edges provide the least contextual information and must be deleted, cropping is required. This produces a u-shaped network and, more crucially, spreads contextual information throughout the network, enabling it to use context from a broader overlapping area to segment items inside an area. The following represents the network's energy function:

$$R = \sum w(i) \log\left(x_{k(i)}\,(i)\right) \quad (3)$$

where x_k represents the final feature map after applying the pixel-wise SoftMax function, which is defined as

$$x_k = exp(A_k(i))/\sum_{k^1}^{K} \exp\,(A_k(i)') \quad (4)$$

and A_k indicates the channel's activation function.

Deep belief network_ Semi supervised GAN (DBN_SSGAN)

Deep belief network (DBN) is the most popular DL methods; it was first proposed by Hinton. This algorithm finds the ideal settings more quickly than the others and is quick to learn new information. A logistic regression layer and an unsupervised learning module based on restricted Boltzmann machines (RBMs) are the fundamental components of a traditional DBN.

Layer-wise training of a well-known stochastic neural network, known as the RBM, creates a DBN. The RBM consists of two layers: one hidden layer of Boolean neurons and another layer of binary-valued neurons. The connections amongst neurons within a layer are not bidirectional or symmetrical, despite being between layers.

Layer-wise configurations use the energy function of the configuration, given in (2), to discover the probability distribution that exists between the two levels. This leaves us with the following equation to express the probability distribution.

$$Ef(a,b) = -\sum_{m=1}^{z_a} x_m a_m - \sum_{n=1}^{z_b} y_n b_n - \sum_{m=1}^{z_a}\sum_{n=1}^{z_b} b_n W_{n,m} b_m \quad (5)$$

$$pd = \frac{e^{-Ef(a,b)}}{\sum_a \sum_b e^{-Ef(a,b)}} \quad (6)$$

In the hidden layer, there are b_n Boolean hidden neurons, where as $W_{n,m}$, neurons make up the visible layer. The

weight matrices that divide the two layers are b_m and b_n. The biases for the two layers are x_m and y_n.

After that, an equation is created to express the activation probability functions.

$$pd(a_m = 1|b) = sig\left(\alpha_m + \sum_{n=1}^{i_b} W_{n,m} b_m\right) \quad (7)$$

$$pd(b_m = 1|a) = sig\left(y_m + \sum_{m=1}^{i_a} W_{n,m} a_n\right) \quad (8)$$

The logistic sigmoid function is also represented as sig(). The pre-training principles support this since the weight matrices and layer biases can be taught without supervision. The idiosyncrasies of the data are too complex for a single hidden RBM. Deep features from the input dataset can be gradually extracted by a DBN, which is built by stacking layers of RMBs in a hierarchical manner and ending with a logistic regression layer. The DBN's first RBM is pre-trained to function as an independent RBM using the training data as inputs.

The output of the first RBM is chosen to be the input for the second RBM once the weight matrix and bias settings are determined. Then, using the same process, the invisible layers of the previous two RBMs are repeatedly trained to create a new RBM. The last stage involves overlaying a comprehensive predictor (such a logistic regression layer) over the network and closely monitoring its training. Following the application of the previously stated stages, the trained network's parameters are slightly altered by fine tuning using the back-propagation (BP) technique.

Natural language processing [15] and image processing [16-17] have both made extensive use of GANs in recent years. It was game theory that provided the foundation for Generative Adversarial Networks (GAN).The discriminator and generative models make up the GAN framework. The generator model takes a data space as input and translates a latent vector from a known distribution into it. The discriminator model attempts to discern between a genuine sample from the data space and a phony sample from the generator.

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.

Similarly, let a deep belief network $B(q;\ \theta_B)$ represents the discriminator model with a scalar output representing the probability that q comes from D_d rather than $A(l)$. So B is trained to maximize $log[B(q)]$ and $log(1 - B(A(l)))$. Simultaneously, B is trained to minimize $log(1 - B(A(l)))$ to fool the discriminator. This means that the two models compete against each other in a two-player min-max game that optimizes the following objective function.:
Similarly, let the discriminator model be represented by a deep belief network $B(q;\ \theta_B)$), where the scalar output denotes the likelihood that p originates from $D_d$rather than $A(l)$. Thus, B is trained to maximize $log[B(q)]$ and $log\left(1 - B\left(A(l)\right)\right)$. In order to trick the discriminator, B is simultaneously taught to minimize $log\left(1 - B\left(A(l)\right)\right)$. In other words, two models compete against one another in a two-player min-max game while maximizing the aforementioned objective function:

$$\min_A \max_B E(A,B) = \mathop{\mathrm{V}}_{q\sim D_d}\left[\log\left(B(i)\right)\right] + \mathop{\mathrm{V}}_{l\sim N(x)}\left[\log\left(1 - B\left(A(l)\right)\right)\right]$$

According to Goodfellow et al. [18], the training criterion enables the data generating distribution to converge to a real data distribution D_d, provided that A and B have adequate capacity. Finding a Nash equilibrium, however, is necessary for the min-max objective function and may include a non-convex function with continuous and high dimensional parameters. GAN typically fails to converge when looking for the Nash equilibrium since it is taught using gradient descent approaches.

Let's now examine at a typical K class classifier model, which maps a given data point (p) to one of the M potential outputs. The class probability that corresponds to each class is produced by such a model, $p_m(y|q, 1 \leq y \leq K)$. By lowering the cross-entropy between the true label and the prediction distribution$V_{q,y\sim D_d}[logP_{model}(y|q)]$, the model is trained in supervised learning.

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. The discriminator in the supervised scenario is comparable to the B standard classifier $B_{ss}(q) = P_{model}(y|q, 1 \leq y \leq K)$. $P_{model}(y = K + 1|q)$ in unsupervised learning is equivalent to the likelihood that q is a fake, while q in the original GAN function represents the likelihood that q is real. Next, the goal of semi-supervised learning is

$$\min_{A} \max_{B} \; \mathop{\mathrm{V}}_{q \sim D_d} [\log(B(i))] + \mathop{\mathrm{V}}_{l \sim N(x)} \log\left(1 - B(A(l))\right) + \alpha_s \mathop{\mathrm{V}}_{q, y \sim D_d} \log(B_{ss}(p)) \quad (9)$$

In order to balance the supervised and unsupervised losses, the hyperparameter α_s is included.

Result and discussion

This section explains the experimental results of the developed strategy and the performance is examined in relation to several current methodologies. Additionally, the metrics that are utilized to assess the method are explained in depth. Python 3.7.5 will be used to implement the recently introduced crime detection system on a system that meets the following requirements: Intel i3-core Windows 10 PC with 2 GB RAM. The performance metrics values of present model are given in table 3 and the comparison with existing model is shown in table 4. The performances metrics are listed below with obtain true positive, true negative, false positive and false negative values.

$$Precision = \frac{1472}{1472 + 8} = 98.38$$

$$Recall = \frac{1472}{1472 + 25} = 98.99$$

$$F1\ score = \frac{1472}{1472 + \frac{1}{2}(8 + 25)} = 98.69$$

$$Accuracy = \frac{1472 + 1495}{1472 + 1495 + 8 + 25} = 98.69$$

Table 3: performance of DBNSSGAN

Method	Precision	Recall	F1score	Accuracy
DBNSSGAN	98.38	98.99	98.69	98.69

Table 4: DBNSSGAN comparison with existing model

Methods	Accuracy
DBN	93.42
CNN	96.75
GAN	96.94
SSGAN	97.20
DBNSSGAN	98.69

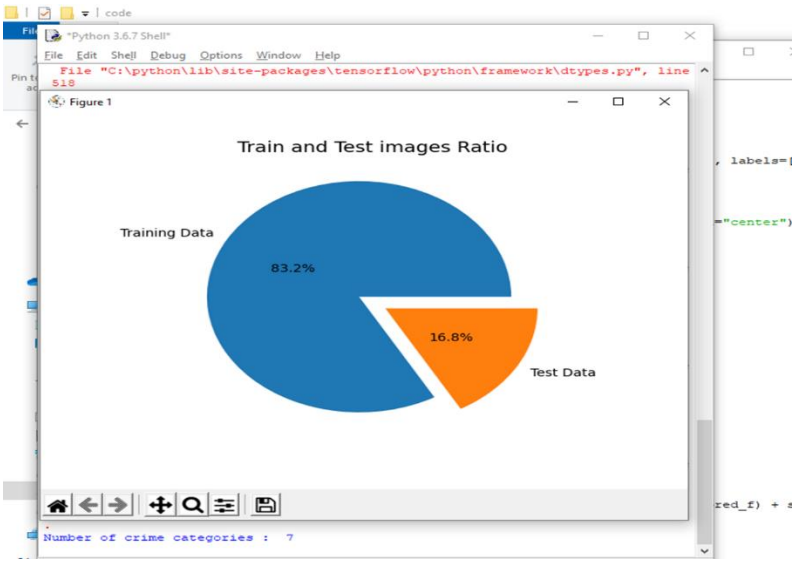

Fig 1. Ratio of Train and Test data

The performance of the suggested network for detecting crime events is assessed using metrics, including training and testing accuracy. The dataset split into train (83.2%) and test data (16.8%) as illustrated in figure 2. The frame count for each class is represented in the figure 3.

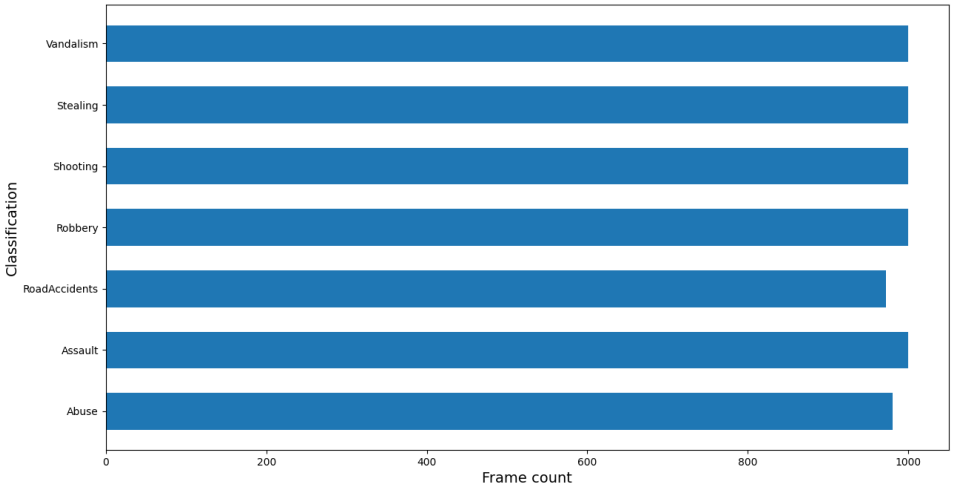

Fig 2 Frame count in each class

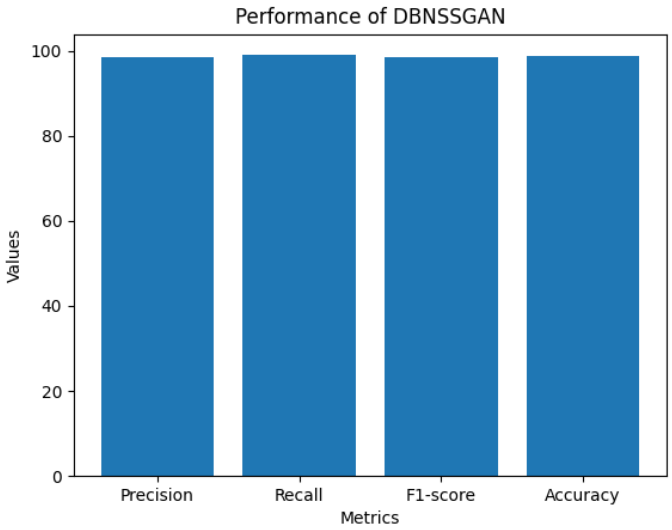

Fig 3 Performance metrics value of proposed classifier

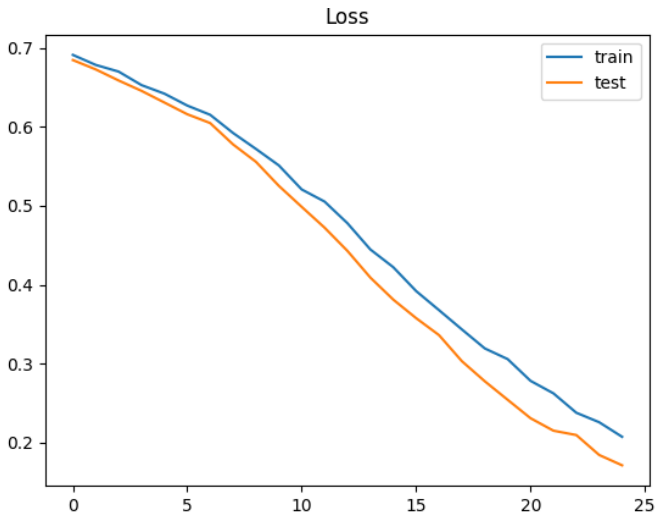

Fig 4 Loss value of proposed classifier at each iteration

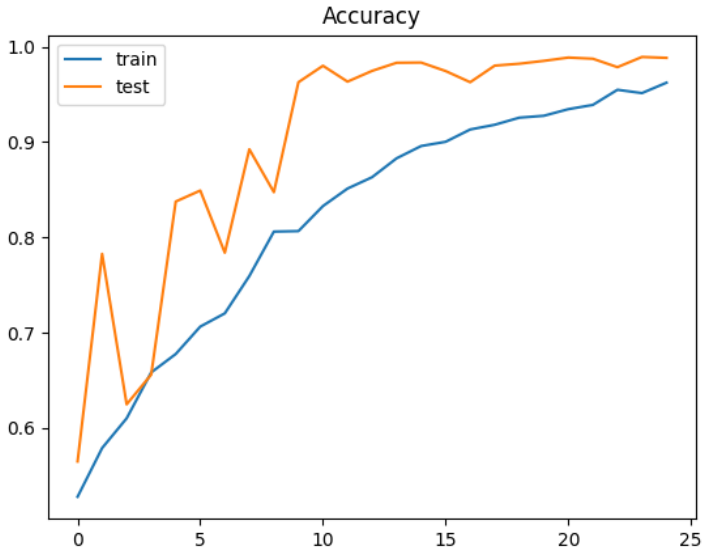

Fig 5 Accuracy value of proposed classifier at each iteration

Comparison with Prior study

The proposed anomaly detection model is compared with the existing work as mentioned in the literature review and the result of the comparison is illustrated in figure 6. Compared to the existing study the present model achieved the highest accuracy on the crime dataset. As mentioned in literature (Table 1) the COVAD model achieved 96.5% accuracy. But our model achieved the better performance of 99.206% training accuracy and 98.83% validation accuracy.

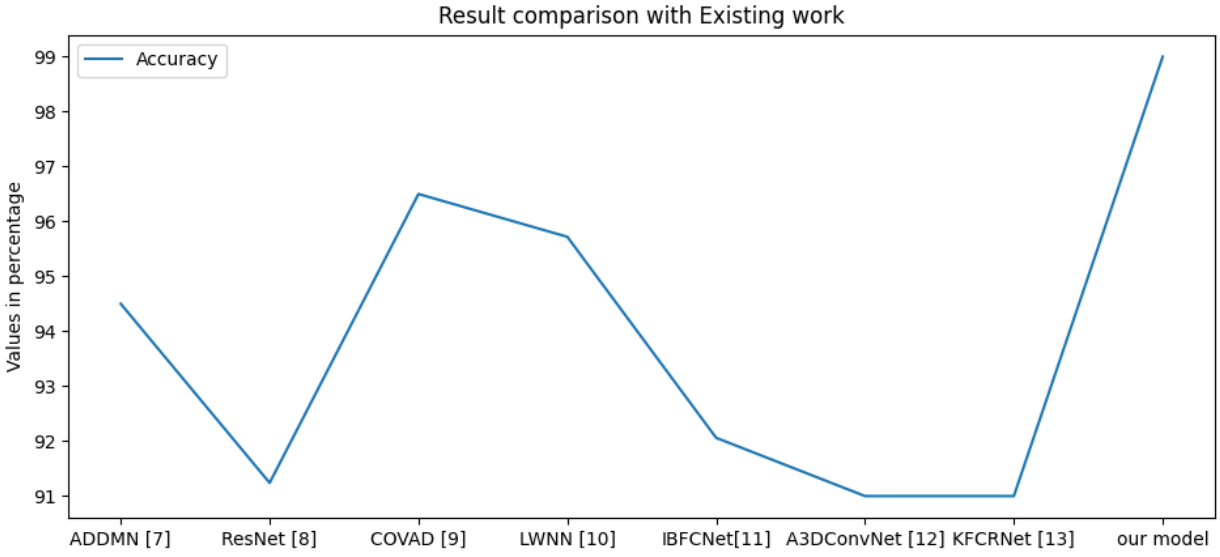

Fig 6. Accuracy comparison with existing work

From the result analysis it is clear that the proposed model achieved the great performance with respect to training and testing the model. The performance (accuracy and loss) is evaluated with different iteration and their results are shown in figure 4 and 5. In addition the model is compared with prior methods surveyed in literature. Nevertheless, our approach surpasses the other current approaches and obtains the second-highest value. Furthermore, over 98% accuracy is obtained, suggesting that an anomaly detector with sufficient practicality can be taught.

Conclusion

Large-scale surveillance camera exploitation has resulted in the everyday collection of enormous volumes of data. Nevertheless, the enormous volumes of gathered data make it extremely difficult to manually identify crimes. Deep learning and computer vision technologies open up a world of research possibilities in the field of real-world issue solutions, like anomaly detection. A deep learning architecture with a semi-supervised basis is suggested for detecting anomalies in crime datasets. This research implements three phase: first the min max scaler is

applied to scale the input image then applies U-Net to record local and global temporal data in order to identify unusual portions in an image. The proposed model achieved higher accuracy in both training and validation with 99.206% and 98.83% respectively. Eventually the proposed deep belief network based semi supervised GAN classifier classifies the anomaly effectively. More effective optimization algorithms may be used in the future to improve the classifier's effectiveness, and more deep learning techniques may be used for classification. In the future, identifying crimes under various weather conditions may be a project.

References

[1] Y. Tang et al, "Integrating prediction and reconstruction for anomaly detection", Pattern Recogn. Lett., (2020)

[2] Nayak, R., Pati, U.C. and Das, S.K., 2021. A comprehensive review on deep learning-based methods for video anomaly detection. *Image and Vision Computing*, *106*, p.104078.

[3] M. Ribeiro et al, "A study of deep convolutional auto-encoders for anomaly detection in videos", Pattern Recogn. Lett. (2018)

[4] Yogameena, B. and Nagananthini, C., 2017. Computer vision based crowd disaster avoidance system: A survey. *International journal of disaster risk reduction*, *22*, pp.95-129.

[5] Khan, A., Sohail, A., Zahoora, U. and Qureshi, A.S., 2020. A survey of the recent architectures of deep convolutional neural networks. *Artificial intelligence review*, *53*, pp.5455-5516.

[6] Ramirez, I., Cuesta-Infante, A., Pantrigo, J.J., Montemayor, A.S., Moreno, J.L., Alonso, V., Anguita, G. and Palombarani, L., 2020. Convolutional neural networks for computer vision-based detection and recognition of dumpsters. *Neural Computing and Applications*, *32*, pp.13203-13211.

[7] Waddenkery, N. and Soma, S., 2023. Adam-Dingo optimized deep maxout network-based video surveillance system for stealing crime detection. *Measurement: Sensors*, *29*, p.100885.

[8] Qasim, M. and Verdu, E., 2023. Video anomaly detection system using deep convolutional and recurrent models. Results in Engineering, 18, p.101026.

[9] Shao, W., Rajapaksha, P., Wei, Y., Li, D., Crespi, N. and Luo, Z., 2023. COVAD: Content-oriented video anomaly detection using a self-attention based deep learning model. *Virtual Reality & Intelligent Hardware*, *5*(1), pp.24-41.

[10] Watanabe, Y., Okabe, M., Harada, Y. and Kashima, N., 2022. Real-World Video Anomaly Detection by Extracting Salient Features in Videos. *IEEE Access*, *10*, pp.125052-125060.

[11] Zahid, Y., Tahir, M.A., Durrani, N.M. and Bouridane, A., 2020. Ibaggedfcnet: An ensemble framework for anomaly detection in surveillance videos. *IEEE Access*, *8*, pp.220620-220630.

[12] Ansari, M.A., Singh, D.K. & Singh, V.P. Detecting abnormal behavior in megastore for crime prevention using a deep neural architecture. *Int J Multimed Info Retr* **12**, 25 (2023).

[13] M. Shoaib, A. Ullah, I. A. Abbasi, F. Algarni and A. S. Khan, "Augmenting the Robustness and Efficiency of Violence Detection Systems for Surveillance and Non-Surveillance Scenarios," in *IEEE Access*, vol. 11, pp. 123295-123313, 2023.

[14] Pires, I.M., Hussain, F., Garcia, N.M., Lameski, P. and Zdravevski, E., 2020. Homogeneous data normalization and deep learning: A case study in human activity classification. *Future Internet*, *12*(11), p.194.

[15] Thirumagal, E. and Saruladha, K., 2021. GAN models in natural language processing and image translation. In Generative adversarial networks for image-to-image translation (pp. 17-57). Academic Press.

[16] L. Yu, W. Zhang, J. Wang, and Y. Yu, ''SeqGAN: Sequence generative adversarial nets with policy gradient,'' in Proc. AAAI, 2016, pp. 2852–2858.

[17] A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, and J. Yosinski, ''Plug & play generative networks: Conditional iterative generation of images in latent space,'' in Proc. Comput. Vis. Pattern Recognit., 2017, pp. 3510–3520.

[18] 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.