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

# Real time Video Anomaly Detection using Deep Belief Network with Semi Supervised GAN

Journal of Electrical Systems

***Abstract: -*** In recent years, real-time video anomaly detection has emerged as a crucial task in various applications, including surveillance, security, and smart cities. Big data is created by video streams that are recorded as the number of surveillance cameras increases. It has become imperative to analyze the video streams gathered from those traffic surveillance cameras in order to identify unusual, suspicious events and various harmful activities, as it is not feasible to view, evaluate, and understand the contents of these films with human labor.The article introduced the Deep Belief Network (DBN) in conjunction with a Semi-Supervised Generative Adversarial Network (GAN) to effectively detect anomalies in video streams. Our method utilizes the unsupervised learning capabilities of DBNs to capture complex patterns from normal video frames while the semi-supervised GAN aids in generating realistic examples of both normal and anomalous behaviors. By integrating these two frameworks, we achieve improved feature extraction and robust anomaly detection. Our approach is evaluated on real time datasets collected from the CCTV footages from news and demonstrate significant improvements in detection accuracy and compared to existing methods. Our findings suggest that this hybrid model is a good alternative for dynamic environments where anomalies may regularly occur because it not only improves performance but also adapts well to real-time applications.

***Keywords:*** Deep learning, Surveillance video, video anomaly detection, Generative Adversarial Network.

## INTRODUCTION

Video anomaly detection (VAD) plays a pivotal role in ensuring security and safety across various domains, including surveillance systems, transportation monitoring, and smart city infrastructures [1]. The increasing volume of video data generated in these environments necessitates efficient and accurate methods to identify unusual events that deviate from expected behavior. Traditional approaches [2] often rely on handcrafted features and supervised learning, which can be limited by the need for extensive labeled data and may struggle with dynamic environments where anomalies are infrequent or diverse [3-5].

Recent advancements in deep learning have shown promise in addressing these challenges. Convolutional neural networks (CNNs) [6,8] and recurrent neural networks (RNNs) [7] in particular are two deep learning models that have demonstrated impressive capacity to extract intricate spatial and temporal patterns from video data [9]. However, the effectiveness of these models is often contingent upon the availability of large labeled datasets, which can be both time-consuming and expensive to curate [10].

To address these challenges, this paper explores a hybrid deep learning-based approach for real-time video anomaly detection [11]. By leveraging the strengths of state-of-the-art neural architectures, we aim to create a robust and

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

efficient framework that can adapt to varying contexts and operational conditions. Our approach combines methods that improve feature extraction with semi-supervised learning capabilities, so the model can learn from both labeled and unlabeled data.

Deep Belief Networks (DBNs), known for their ability to learn hierarchical representations of data, can effectively capture the underlying patterns of normal activities in video sequences [12]. However, while DBNs excel in unsupervised feature learning, they often require additional techniques to enhance their robustness in detecting anomalies.

To bridge this gap, we propose a hybrid approach that combines DBNs with a Semi-Supervised Generative Adversarial Network (GAN). The semi-supervised framework allows the model to leverage both labeled and unlabeled data, thereby improving the learning process [13]. The GAN component generates realistic samples, facilitating the identification of anomalous behaviors by enriching the training set and providing a diverse range of scenarios.

In this paper, we detail our proposed methodology, highlighting its effectiveness in real-time video anomaly detection. Using benchmark datasets, experiments are conducted to validate our approach, demonstrating significant improvements in detection accuracy and operational efficiency. Our results underscore the potential of combining DBNs and semi-supervised GANs as a robust solution for real-time anomaly detection in dynamic video environments.

**Contributes Of The Study**

- Introduce a novel hybrid approach that combines Deep Belief Networks with Semi-Supervised Generative Adversarial Networks (DBNSSGAN), leveraging the strengths of both architectures to improve anomaly detection performance.
- Utilizes DBNs for unsupervised feature extraction, enabling the model to effectively learn complex patterns from normal video frames, which enhances the representation of typical behaviors.
- By integrating semi-supervised learning through GANs, our approach effectively utilizes both labeled and unlabeled data, thereby reducing reliance on large amounts of annotated samples and improving generalization to unseen anomalies.
- Optimize our model for real-time applications, demonstrating that our approach maintains high detection accuracy while achieving the necessary speed for immediate deployment in dynamic environments

**Related works**

The combination of Deep Belief Networks and semi-supervised GANs represents a promising direction for advancing video anomaly detection. By integrating these powerful techniques, our approach aims to enhance both the accuracy and efficiency of anomaly detection in real-time scenarios, addressing the key challenges faced in dynamic environments.

**Recent Advanced Frameworks and Architectures in Deep Learning for VAD**

**Convolutional Neural Networks (CNNs):** Many studies leverage CNNs for spatial feature extraction, often in combination with RNNs or LSTMs to capture temporal dynamics. Recent works explore novel architectures, such as ResNet and DenseNet, to improve accuracy.

Amin et al [14] examines two models that are based on the J. DCNN and J. QCNN proposals. The two suggested models are built using the chosen layer combinations and ideal parameters, yielding better results. Using the highly challenging UNI-Crime and UCF-Crime datasets, the suggested models are evaluated. The developed architectures' performance is assessed ten times over, with a validation value of 0.7.

In order to provide accurate item identification and localization inside each frame, NischitaWaddenkery and Shridevi Soma combine multibook detector (ESSD) to enable object detection in video footage. Therefore, theft incidents are efficiently detected and classified by the suggested MSAC-SOA. Ultimately, the control room receives

a theft detection alarm message. Two datasets are used to evaluate this method: real-time video data and the UCF crime sample for theft detection.

The proposed method performs better than previous approaches in terms of computing times, demonstrating its efficacy and suitability for a range of real-world movies with various configurations.

**Autoencoders:** Variants like convolutional autoencoders are frequently used to reconstruct video frames. Anomalies are identified through high reconstruction error. Recent research has introduced hybrid models combining autoencoders with attention mechanisms.

Viet-Tuan Le and Yong-Guk Kim [15] have designed a network that combines spatial and temporal branches to effectively handle both data types. This network utilizes an unsupervised residual autoencoder architecture, which includes a deep CNNencoder and a multi-stage channel attention (MSST) based decoder. It leverages temporal features through a temporal shift approach and extracts contextual dependencies using channel attention modules. The system's performance was assessed using three common benchmark datasets.

**Generative Adversarial Networks (GANs):** GANs are gaining traction for their ability to model normal behaviors, with anomalies detected by discrepancies between generated and real data. Techniques like conditional GANs have shown promise in enhancing performance.

Singh [16] suggests using a Constrained GAN (CVAD-GAN) to perform VAD in real time. 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. In comparison to the current state-of-the-art VAD approaches.

For real-time video anomaly detection (VAD), Rituraj Singh [17] proposes an Attention-guided Generator with Dual Discriminator GAN (A2D-GAN). This framework employs an encoder-decoder architecture, where the generator network incorporates MSST in decoder and encoder. Adversarial learning, through noise and video frame reconstruction, enhances the generalization of the generator network. Furthermore, the dual discriminators in A2D-GAN have distinct roles: one differentiates between real and reconstructed video frames, while the other distinguishes between real and reconstructed noise.

**Transformer Models:** The emergence of transformers in VAD has led to significant improvements in capturing long-range dependencies and global context in videos. Models like ViT and TimeSformer are increasingly applied to VAD tasks.

Longkai Sui and Yongguo Jiang [18] added the Fast Fourier Transform (FFT) to the Transformer model to help it learn typical data patterns and better detect periodic patterns and complicated correlations in multivariate data. This will help the model detect anomalies in Argo data more accurately. Experiments on the Argo dataset and three public datasets show that the improved model performs better than the original model. This illustrates the potential of FFT in multidimensional data anomaly identification as well and offers fresh perspectives on how to handle anomaly detection difficulties in intricate datasets found in the real world.

Nazia Aslam and Maheshkumar H. Kolekaroffers TransGANomaly [19], a revolutionary technique to anomaly detection, which is a generative adversarial network (GAN) based on video vision transformer (ViViT). The suggested framework is a video frame predictor that was only adversarially trained on typical video data. A ViViT network serves as the GAN's generator, obtaining 3D input tokens from the short videos. The generator uses previous sequences to forecast the next frame. Subsequently, the discriminator of the model receives both the original and anticipated frames for binary classification.

## PROPOSED METHODOLOGY

This article proposes a novel methodology for real-time video anomaly detection by integrating a Deep Belief Network (DBN) with a Semi-Supervised Generative Adversarial Network (GAN). The DBN is utilized for effective feature extraction from video frames, capturing the intricate patterns and temporal dependencies in the data. Subsequently, a Semi-Supervised GAN is employed to distinguish between normal and anomalous patterns in the extracted features. The GAN leverages both labeled and unlabeled data, enhancing its ability to detect anomalies even with limited labeled samples. Our approach aims to provide a robust and scalable solution for real-time

anomaly detection, capable of adapting to diverse and dynamic environments. The effectiveness of our suggested strategy is demonstrated by experimental findings on benchmark datasets, showing significant improvements in anomaly detection accuracy and processing speed compared to existing techniques. This methodology holds promise for enhancing the reliability and responsiveness of video surveillance systems, contributing to improved safety and security in various real-world applications.

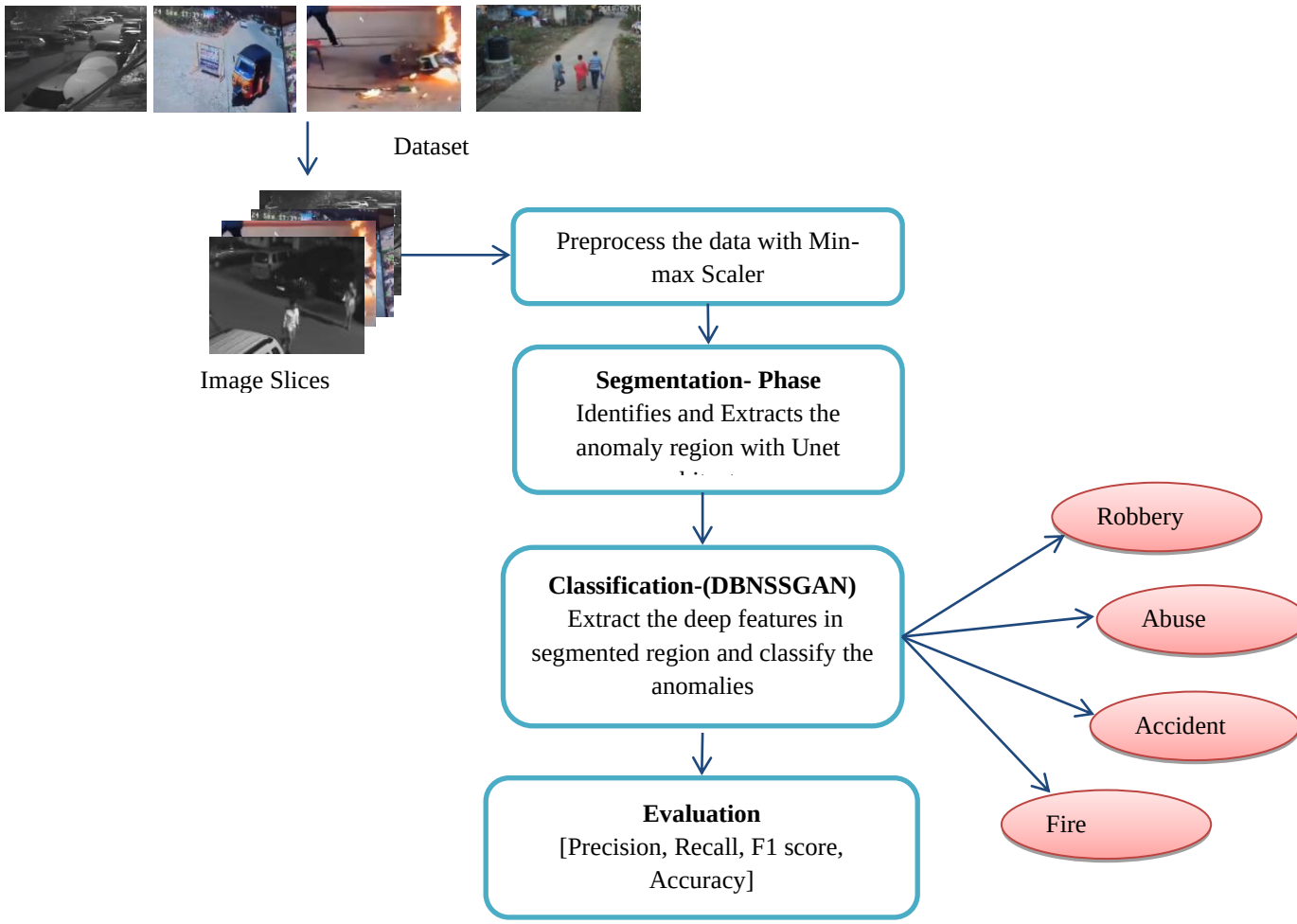


**Figure 1. Proposed video anomaly detection framework**

**Preprocessing**

In the realm of video anomaly detection, preprocessing steps such as normalization are crucial to ensure the effective performance of machine learning models. One widely used normalization technique is the Min-Max Scaler, which scales the pixel values of images to a specified range, typically [0, 1]. This article discusses the importance of normalization, the principles behind the Min-Max Scaler, and its application in preprocessing CCTV footage images for anomaly detection tasks.

The Min-Max Scaler is a simple yet effective normalization technique. It scales each pixel value of an image to a specified range, typically [0, 1], by using the following formula:

$$O' = \frac{O - O_m}{O_n - O_m} \quad (1)$$

Where $O$ denotes the original pixel value $O'$scaled pixel value. $O_m$and$O_n$ are the minimum and maximum pixel values in the image, respectively.

**Dataset**

The real time CCTV footages are collected from the new channels like india today,NDTV news Hindustan Times, etc., Data is collected for four different categories: Robbery, Accident, abuse and fire. The Video sequence contains with normal and abnormal events with these categories which are collected and framed train and test dataset with 400 and 150 images respectively. The proposed model is evaluated with this dataset for anomaly detection in surveillance videos.

**Robbery**

**Accident**

**Fire**

**Abuse**

**Figure 1. Sample input from each classes**

**Segmentation**

An important task in computer vision is image segmentation, which involves breaking a picture up into sections or segments that have significance. One of the most effective and widely used architectures for this task is U-Net, which has shown remarkable performance in various segmentation challenges, particularly in biomedical imaging, anomaly detection etc.

Olaf Ronneberger et al introduced U-Net []. The U-shaped architecture, which consists of an expanding path (decoder) and a contracting path (encoder), gives rise to the moniker "U-Net".

**Encoder**

The encoder adheres to the standard architecture of a convolutional network. It repeatedly applies two 3x3 convolutions each followed by ReLU and a 2x2 max pooling operation with a stride of 2 for downsampling. This process aims to capture the image context by gradually reducing its spatial dimensions while increasing the depth of the feature maps.

**Convolutional Layers**: These layers perform a sequence of convolutions on the input image, extracting features and progressively reducing the spatial dimensions.

**Max Pooling**: Max pooling operations downsample the feature maps, effectively reducing the spatial resolution and increasing the receptive field of the following layers.

**Decoder**

The expansive path consists of a 2x2 convolution (also known as a "up-convolution") that halves the number of feature channels once the feature map has been upsampled.

This process helps in precise localization and enables the network to combine coarse and fine features effectively.

**Upsampling**: Upsampling operations increase the spatial resolution of the feature maps, bringing them back to the original input size.

**Concatenation**: Concatenating feature maps from the contracting path with those from the expansive path using skip connections gives prior layers' context and details.

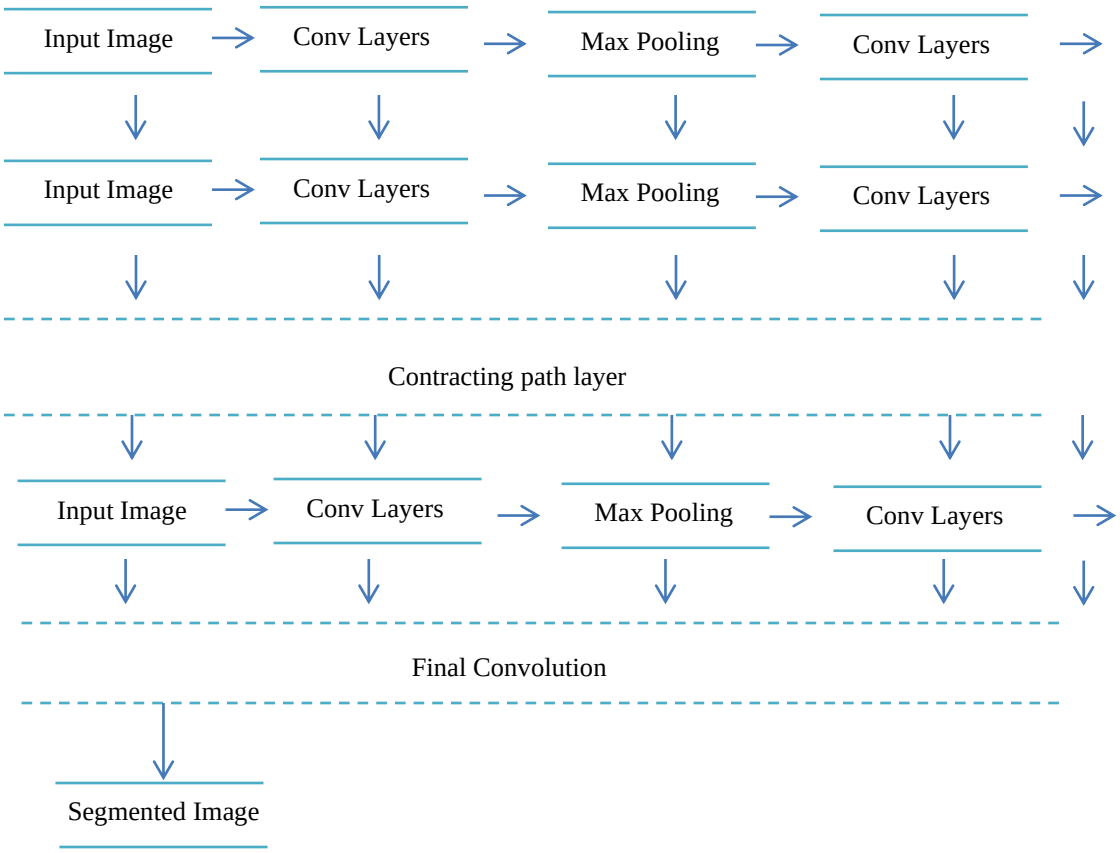


**Figure 2. Unet structure in image segmentation**

**Convolutional Layers**: Additional convolutions refine the upsampled feature maps, improving the segmentation accuracy.

Detecting anomalies involves identifying events or objects that deviate from the norm. Leveraging U-Net for segmentation in video anomaly detection enhances the ability to pinpoint precise areas of interest, thus improving the detection accuracy and providing more meaningful insights.

**DBNSSGAN**

A Deep Belief Network (DBN) is a deep learning model consisting of multiple layers of latent variables, typically in the form of Restricted Boltzmann Machines (RBMs) or autoencoders. DBNs are used for unsupervised learning, feature extraction, and dimensionality reduction. DBNs are typically trained in a layer-wise manner. Each RBM is trained individually before moving to the next layer. This approach helps in efficiently learning the parameters and initializing the weights for fine-tuning.

Generator generates synthetic samples that mimic the distribution of normal data, enhancing the training set. Discriminator distinguishes between real and synthetic samples while also classifying input frames as normal or anomalous.

**Pre-training**: Each RBM layer is trained in an unsupervised manner using Contrastive Divergence (CD).

**Fine-tuning:** The entire network is fine-tuned using backpropagation to improve the learned representations.

Energy Function in Restricted Boltzmann Machine is represented by the following equation

$$ef(a,b) = \sum_x a_x r_x - \sum_y b_y s_y - \sum_y \sum_y a_x b_x W_{x,y}$$

$a$denotesvisible units, b hidden units, r and s represents the bias for and b. w denotes the weight.The proposed architecture integrates a Deep Belief Network (DBN) with a Semi-Supervised Generative Adversarial Network (SSGAN). The system is designed to extract features from video frames using the DBN and then leverage these features within the SSGAN for anomaly detection.

Semi-Supervised Generative Adversarial Networks (SGANs) are a variation of the traditional GAN architecture that integrates both supervised and unsupervised learning elements. By utilizing a small amount of labeled data alongside a large amount of unlabeled data, SGANs enhance model performance. This approach is particularly beneficial in situations where labeled data is limited.

**Adversarial Training:**Like traditional GANs, SGANs involve a minimax game between the generator and the discriminator. The generator aims to create data that the discriminator cannot differentiate from real data, while the discriminator attempts to accurately identify both real and fake data.

**Semi-Supervised Learning:** The discriminator is trained not only to classify real vs. fake data but also to predict the class labels of real data. This is done by incorporating labeled data into the training process.

**Handling Sparse Anomalies:** In many video anomaly detection scenarios, anomalous events are rare. SGANs can effectively utilize a small number of labeled anomalous examples along with abundant unlabeled normal data to improve detection accuracy.

**Reducing False Positives:** The adversarial training process helps in refining the model, reducing the likelihood of false positives by generating realistic normal patterns that the discriminator learns to differentiate from anomalies.

Features extracted by the DBN are fed into the SSGAN for further processing. The SSGAN uses these features to train the generator and discriminator in an adversarial setting, enhancing the model's robustness. The discriminator assigns anomaly scores to input frames based on their deviation from the learned normal patterns.

**Result and discussion**

Evaluating the performance of a real-time video anomaly detection system using Deep Belief Networks (DBNs) and Semi-Supervised Generative Adversarial Networks (GANs) involves assessing several parameters to ensure the models' effectiveness and efficiency.

**Performance metrics calculation for Realtime dataset**

$$Precision = \frac{6286}{6286 + 194} = 98.19$$

$$Recall = \frac{6286}{6286 + 16} = 99.66$$

$$F1\ score = \frac{6286}{6286 + \frac{1}{2}(194 + 16)} = 96.71$$

$$Accuracy = \frac{6286 + 6488}{6286 + 6488 + 194 + 16} = 98.16$$

**Table 2: performance of DBNSSGAN**

| Method | Dataset | Precision | Recall | F1score | Accuracy |
|---|---|---|---|---|---|
| DBNSSGAN | Real time data | 98.19 | 99.66 | 96.71 | 98.16 |

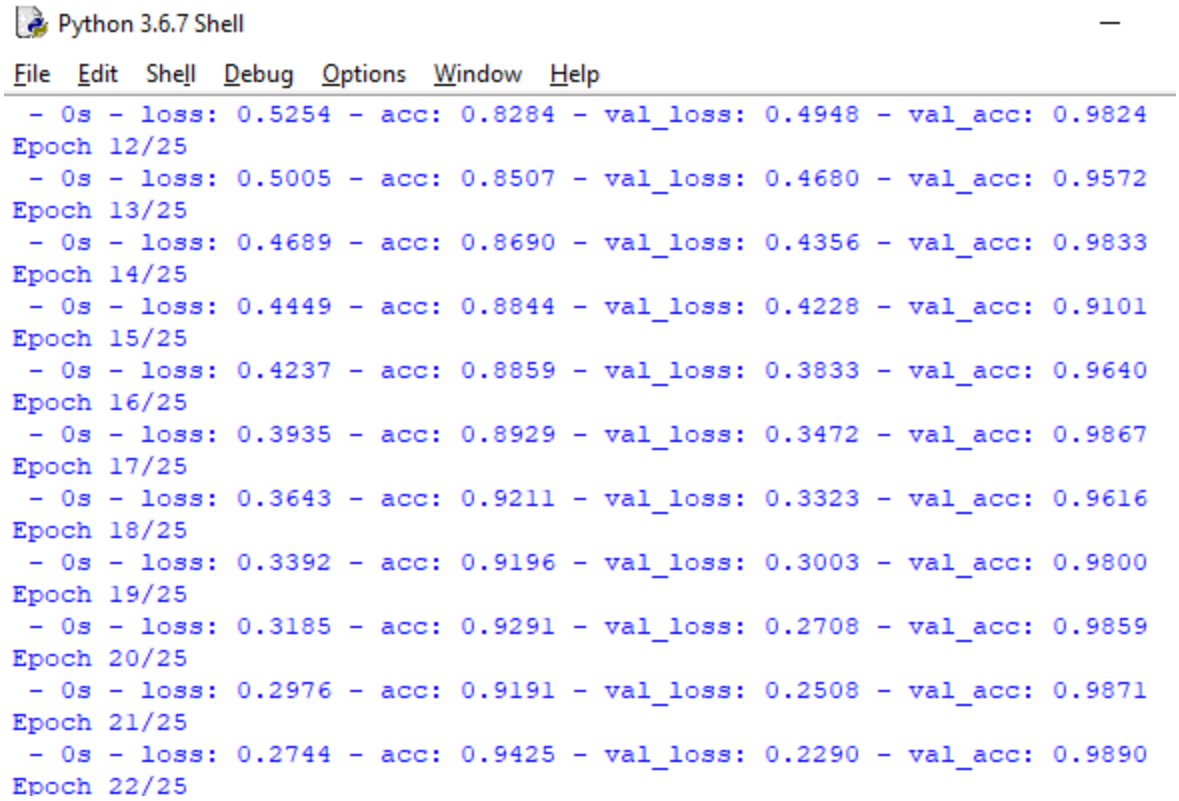


Figure 3 python simulation result

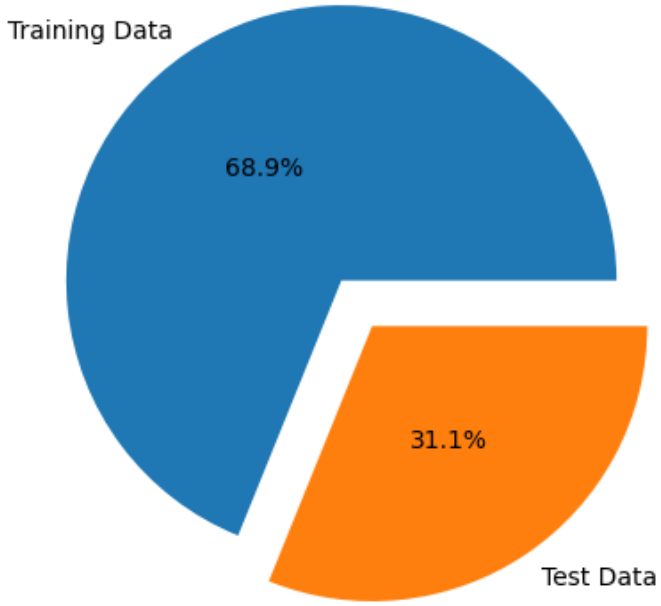


**Figure 4. Class distribution of the dataset**

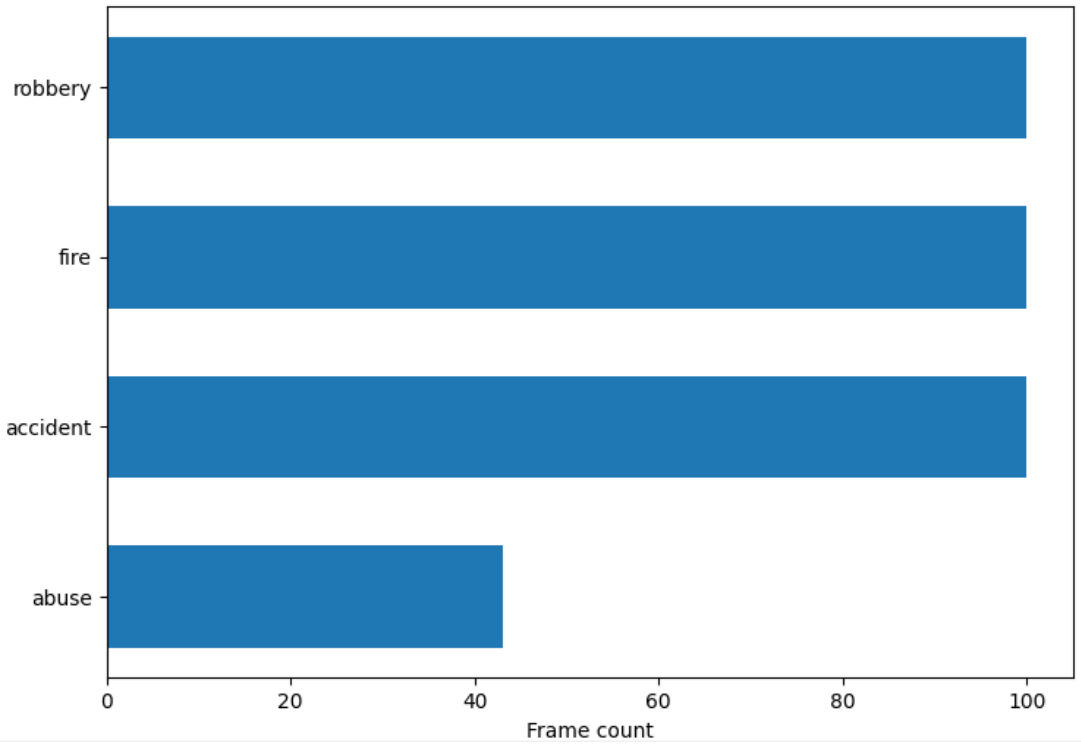


**Figure 5.Frame count for each class**

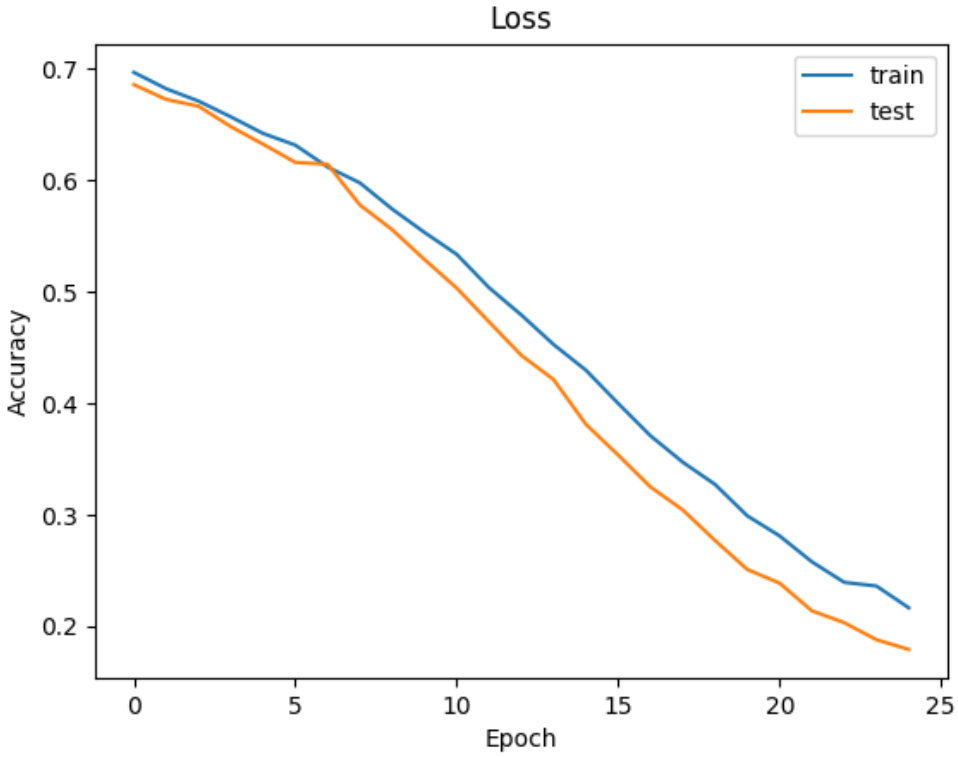


**Figure 6.Train and test lossof applied dataset**

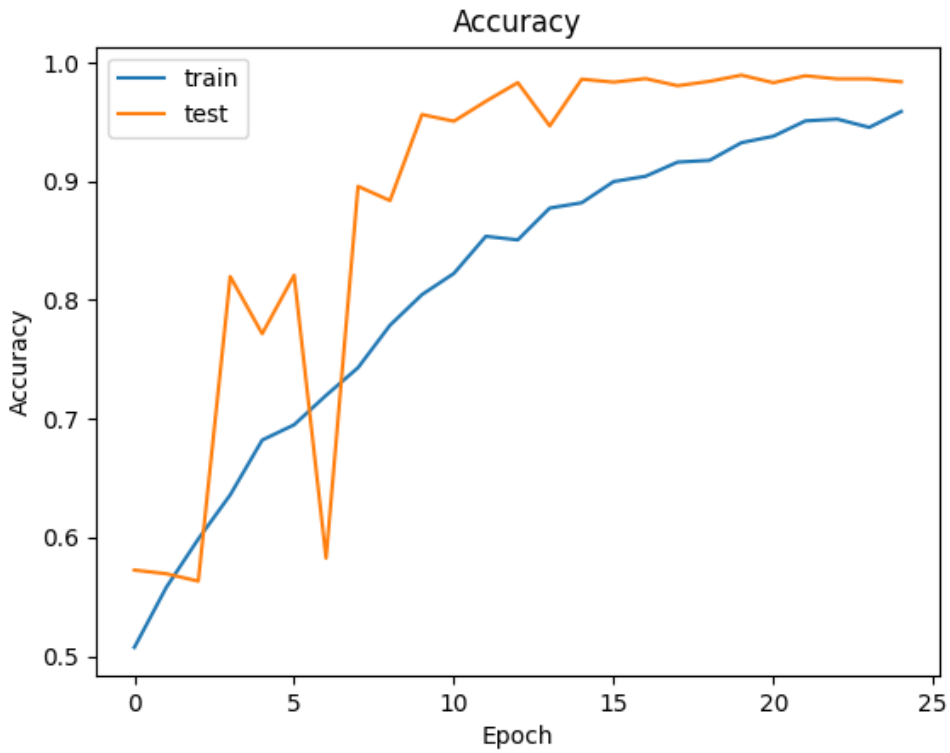


**Figure 7. Train and test Accuracy of Real time data**

The experimental result of the proposed method is given in figure 4 to 8. The proposed model obtained highest accuracy for real time data. The DBNexcel at learning hierarchical features from video frames. They can capture both low-level features, such as edges and textures, and higher-level features, such as shapes and objects, which are essential for identifying anomalies that might not be apparent in the raw data.Similarly SSGANs enhance this by providing a robust feature representation through adversarial training, which helps in identifying subtle anomalies.The combination of DBNSSGAN provides 98.16% accuracy on real time. These experimental findings demonstrate that using DBNSSGAN in conjunction with anomaly data enhances anomaly detection for real-time applications.

## DISCUSSION

The DBN effectively captured significant features from video frames, facilitating the GAN's ability to distinguish between normal and anomalous events. Using the DBN for feature extraction reduced the computational complexity, making the real-time inference more efficient. The GAN demonstrated the capability to learn from both labeled and unlabeled data, leveraging the semi-supervised approach to improve anomaly detection performance. Training stability was achieved by fine-tuning hyperparameters, including the learning rate, batch size, and the number of epochs. The discriminator's accuracy in distinguishing real from generated frames improved with training, indicating successful learning.

## CONCLUSION

In this study, we explored an innovative approach to real-time video anomaly detection by integrating Deep Belief Networks (DBNs) with Semi-Supervised Generative Adversarial Networks (GANs). The primary objectives were to effectively extract relevant features from video frames using DBNs and to leverage the semi-supervised capabilities of GANs for accurate anomaly detection. U-Net enhances video anomaly detection by providing precise segmentation, which isolates regions of interest and improves detection accuracy. The integration of Deep Belief Networks with Semi-Supervised GANs presents a viable approach to real-time video anomaly detection. The DBN effectively reduces the dimensionality of video frames, while the GAN distinguishes between normal and anomalous

events. Despite challenges related to data quality and computational requirements, the experimental setup demonstrated promising results, paving the way for further research and development in this domain. As the field of video analysis continues to evolve, advanced segmentation techniques like U-Net will be pivotal in developing robust and reliable anomaly detection systems.This study can be further enhanced by incorporate advanced data augmentation techniques to increase the diversity and robustness of the training data.

REFERENCES


[1] Patrikar, D.R. and Parate, M.R., 2022. Anomaly detection using edge computing in video surveillance system. *International Journal of Multimedia Information Retrieval*, *11*(2), pp.85-110.

[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] Pawar, K. and Attar, V., 2019. Deep learning approaches for video-based anomalous activity detection. *World Wide Web*, *22*(2), pp.571-601.

[4] Tian, Y., Pang, G., Chen, Y., Singh, R., Verjans, J.W. and Carneiro, G., 2021. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 4975-4986).

[5] Berroukham, A., Housni, K., Lahraichi, M. and Boulfrifi, I., 2023. Deep learning-based methods for anomaly detection in video surveillance: a review. *Bulletin of Electrical Engineering and Informatics*, *12*(1), pp.314-327.

[6] Kumaran, S.K., Dogra, D.P., Roy, P.P. and Mitra, A., 2018. Video trajectory classification and anomaly detection using hybrid CNN-VAE. *arXiv preprint arXiv:1812.07203*.

[7] Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M. and Baik, S.W., 2021. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. *Multimedia tools and applications*, *80*, pp.16979-16995.

[8] Girisha, S., Pai, M.M., Verma, U., Pai, R.M. and Shreesha, S., 2021, December. Anomaly detection using classification CNN models: A video analytic approach. In *TENCON 2021-2021 IEEE Region 10 Conference (TENCON)* (pp. 923-928). IEEE.

[9] Baradaran, M. and Bergevin, R., 2024. A critical study on the recent deep learning based semi-supervised video anomaly detection methods. *Multimedia Tools and Applications*, *83*(9), pp.27761-27807.

[10] Habeb, M.H., Salama, M.A. and Elrefaei, L.A., 2023. Video Anomaly Detection using Residual Autoencoder: A Lightweight Framework. *Mansoura Engineering Journal*, *49*(2), p.10.

[11] Ramoliya, D. and Ganatra, A., 2023. Insights of Deep Learning-Based Video Anomaly Detection Approaches. In *Intelligent Communication Technologies and Virtual Mobile Networks* (pp. 663-676). Singapore: Springer Nature Singapore.

[12] Kishore, D.R., Suneetha, D., Ghantasala, G.P. and Sankar, B.R., 2022. Anomaly Detection in Real-Time Videos Using Match Subspace System and Deep Belief Networks. In *Multimedia Computing Systems and Virtual Reality* (pp. 151-170). CRC Press.

[13] Ouali, Y., Hudelot, C. and Tami, M., 2020. An overview of deep semi-supervised learning. *arXiv preprint arXiv:2006.05278*.

[14] Amin, J., Anjum, M.A., Ibrar, K., Sharif, M., Kadry, S. and Crespo, R.G., 2023. Detection of anomaly in surveillance videos using quantum convolutional neural networks. *Image and Vision Computing*, *135*, p.104710.

[15] Waddenkery, N. and Soma, S., 2024. An efficient convolutional neural network for detecting the crime of stealing in videos. *Entertainment Computing*, *51*, p.100723.

[16] Le, VT., Kim, YG. Attention-based residual autoencoder for video anomaly detection. *ApplIntell* **53**, 3240–3254 (2023).

[17] 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*, *143*, p.104950.

[18] Sui, L. and Jiang, Y., 2024. Argo data anomaly detection based on transformer and Fourier transform. *Journal of Sea Research*, *198*, p.102483.

[19] Aslam, N. and Kolekar, M.H., 2024. TransGANomaly: Transformer based Generative Adversarial Network for Video Anomaly Detection. *Journal of Visual Communication and Image Representation*, *100*, p.104108.