Video Anomaly Detection using Generative Adversarial Network

Chitte Anil
Department of Data Science
Institute of Aeronautical Engineering
Hyderabad, India
chitte.anil1215@gmail.com

Ratan Varsha
Department of Data Science
Institute of Aeronautical Engineering
Hyderabad, India
ratanvarsha321@gmail.com

Thummala Sugathri
Department of Data Science
Institute of Aeronautical Engineering
Hyderabad, India
sugatrithummala20@gmai.com

Manemoni Aravind
Department of Data Science
Institute of Aeronautical Engineering
Hyderabad, India aravindkumar0895@gmail.com

Abstract—Automatic abnormal event detection and recognition is the need of today's surveillance scene because public safety is a growing concern these days. The subject is still open for study in the present due to its utility and complexity. Because every individual has a different definition for abnormality, automatically identifying aberrant events is a task not so easy to accomplish. What is normal in one set of circumstances may be viewed as abnormal in another. In a surveillance film of large crowds, automatic anomaly identification is difficult because of traffic and excessive occlusion. This thesis study attempts to provide the answer for this use case by using machine learning techniques, making it possible to eliminate human resources for monitoring the records of the surveillance system for any type of activity that is unexpected. We are going to develop a new anomaly detection model based on GAN, which can train models to learn how to create a high-dimensional picture space and also learn how to extract the latent space of the video from its context. The residual Autoencoder architecture of the generator is coupled with a two-stream deep convolutional encoder that can realize both temporal and spatial data, and a decoder with multi-stage channel attention. Furthermore, we propose an approach to further enhance the performance of the GAN model by knowledge transfer between different datasets in order to reduce training time and improve the generalization capability of the model. We benchmark our model against the state-of-theart methods in current usage on four benchmark datasets using a variety of assessment metrics. The results of empirical study prove that, as compared to methods in use now, our network performs favourably on the dataset.

Index Terms—Anomaly, Generator, Discriminator, Surveillance, Traffic, Detection, GAN (Generative Adversarial Network), spatio-temporal, Convolution.

I. Introduction

CCTV cameras are everywhere, helping keep us safe. But watching all the footage is tough for people. So, smart computer programs are being made to spot unusual stuff in the videos, so we don't have to watch everything ourselves .Anomaly detection involves recognizing patterns or occurrences that differ from typical behavior or anticipated results. These anomalies can take many forms and often depend on the specific context. While algorithms for anomaly detection have achieved notable accuracy under certain conditions, their performance can still be influenced by factors such as lighting variations, movement direction, object motion speed, occlusions, and the presence of similar object movements.[2] . Additionally, not every anomaly event could be covered in the training video because certain anomalies happen infrequently or never at all. A local discrepancy between the generated and real photos is employed during testing to identify potential anomalies[4]. The following issues impact how challenging it is to detect an anomaly present in the video:

1)The complexity of anomalies are highly dependent on the circumstances. While running in a park is accepted as usual, running in a bank would be considered weird.

2) The cost and computational challenge of handling the spatial and temporal properties of the video data during a single video frame. Actually, the algorithm ought to be able to issue the warning as soon as possible with a respectable accuracy and false alarm rate when an abnormality occurs.

3) A wide range of varied environmental factors, including crowd density, complicated backgrounds, illumination, object occlusions, and object shadows. The goal of this paper is to create a unique anomaly detection model using Generative Adversarial Networks (GANs).

This research focuses on designing an innovative anomaly detection model leveraging generative adversarial networks (GANs). The proposed model is trained to simultaneously understand the construction of a high-dimensional image space and identify the latent space based on the contextual information of videos.

II. Literature Review

Algorithms that are currently applied for anomaly detection in videos are Gaussian Mixture Models (GMM), where it employs background subtraction to identify anomalies by the foreground objects that are highly different from the background model. Conv-LSTM end-to-end trainable composite networks that can predict the progression of a video sequence provided with a limited number of input frames[14].a temporal faux anomaly synthesis using only normal data to synthesize artificial abnormalities.

[10] The researchers introduced a "multi-scale U-Net" architecture for unsupervised video anomaly detection, utilizing a GAN framework. To enhance the generator network's performance during training and testing, they incorporated Shortcut Inception Modules (SIMs) and residual skip connections.

[7] In this approach, GAN-based training was employed, where a U-Net generator was trained to predict the next video frame from the current frame combined with an encoded motion

descriptor. A scoring mechanism was used to compare predicted and actual frames, along with a strategy for tracking multiple anomalies within a single video.

[1] The study proposed a novel method for video anomaly detection called S²-VAE. This method includes two key neural networks: the Stacked Fully Connected Variational Autoencoder (SF-VAE), which is a shallow generative network designed to model data distribution like a Gaussian mixture, and the Skip Convolutional VAE (SC-VAE).

[12] The researchers designed a GAN model incorporating super-resolution and self-attention mechanisms. The generator, structured as an autoencoder, utilized dense residual networks and self-attention components. A relativistic discriminator with a self-attention input was also introduced. Additionally, by integrating optical flow and gradient differences, the model improved video frame predictions for normal events.

[16] The VALD-GAN method was presented as an innovative solution for video anomaly detection. This approach leverages GAN's representational strength using a latent discriminator framework, ensuring the latent space adheres to a predefined distribution. Experimental results demonstrated that VALD-GAN significantly improved anomaly discrimination capabilities.

III. Exisiting System

Video anomaly detection is a significant field in computer vision that focuses on identifying abnormalities or unexpected events within video sequences. Traditional methods for addressing this problem have relied on statistical techniques, machine learning algorithms, and deep learning frameworks. However, these approaches do not involve the use of Generative Adversarial Networks (GANs).

Statistical methods represent the problem in the detection of anomalies and are indeed some of the most widely used statistical methods include Gaussian mixture models GMMs model the distribution of pixel values over time to enable the visible anomalies when observed the data diverge considerably from expected statistical patterns other prevalent techniques include background subtraction which separates moving objects from the background and then detect those behaviors or events which do not conform to existing norms

The optical flow analysis is another method of detecting anomalies which analyzes the movement of pixels between the frames. If any sudden change in optical flow may be relevant to anomalous activities and throws hints toward those anomalies usage of methods like k-means clustering would enable recognition of outliers of data making events not trending as usual.Although these traditional approaches have merits, GANs are among the best solutions for video anomaly detection. GANs characterize very complex data distributions well and generate realistic synthetic data that closely follows normal patterns with training on this synthetic data. GANs can effectively distinguish anomalous behavior for better detection accuracy because GANs are adversarial in nature this ensures that they could learn rich feature representations thus allowing them to catch subtle anomalies in a case that would most probably be ignored by other methods.

There are numerous approaches to video anomaly detection however methods developed based on statistical methods machine learning algorithms and deep learning models the GAN model has an extremely attractive benefit its capability for learning complex data distributions that makes the one crowned at the top in the anomaly-detection process which leads to indicating great choice in this very significant research direction.

IV. Methodology

GAN Architecture

A Generative Adversarial Network (GAN) is a machine learning framework consisting of two interconnected neural networks that engage in a competitive process. These networks work in tandem to enhance performance and improve the quality of generated outputs by employing deep learning methods. Operating predominantly in an unsupervised manner, GANs simulate a zero-sum game, where the advancement of one network inherently challenges the other.

The architecture of a GAN includes two main components: the Generator and the Discriminator. The Generator, typically implemented as a convolutional neural network (CNN), generates synthetic data resembling real-world examples. Meanwhile, the Discriminator, often modelled as a network capable of distinguishing real data from synthetic, is designed to evaluate the authenticity of the input. The Generator's primary goal is to produce data so convincing that the Discriminator cannot reliably identify it as artificial, thereby achieving realistic data generation.

In this setup, the Generator is tasked with generating false data, while the Discriminator is trained to differentiate between authentic examples and those crafted by the Generator. If the Discriminator quickly identifies the Generator's fake data—such as an image that doesn't look like a real human face—the Generator is penalized. As both networks engage in this adversarial process, the Generator gradually learns to produce real and convincing outputs, where Discriminator improves at detecting fabricated data. For example, a GAN can be trained to produce images of human faces that do not belong to any actual person.

The typical workflow of a GAN can be broken down into the following steps:

1.The Generator receives random input values and uses them to create an image.

Spatio-Temporal Frame Extraction

Generative Adversarial Networks (GANs) are specialized models designed to effectively analyze and integrate both spatial features and temporal dynamics in video data simultaneously. These models utilize 2D convolutional layers to process video frames across both spatial and temporal dimensions. By combining the adversarial training framework of GANs with techniques for capturing dynamic changes over time, the model can handle complex video data. The architecture is built around two main components: the generator, responsible for producing synthetic data sequences, and the discriminator, tasked with distinguishing authentic sequences from those generated artificially. As the generator learns to produce realistic sequences by capturing both spatial details and temporal dynamics, the discriminator becomes better at identifying fake sequences. By extracting frames using this method, the model ensures that both the spatial and temporal characteristics of the data are taken into account.

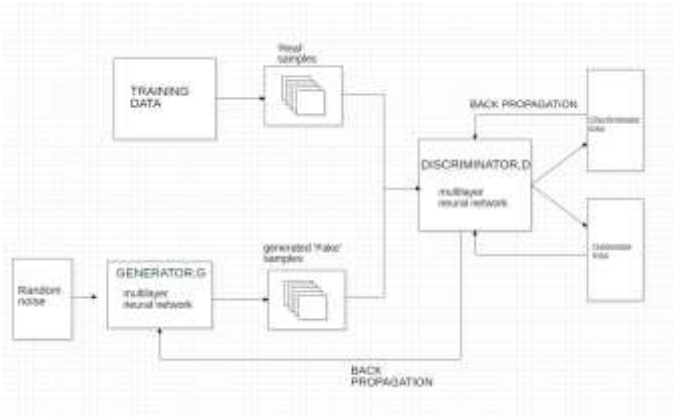


Fig. 1: Architecture of GAN

2.The Discriminator evaluates this generated image alongside real images from a dataset.

3.The Discriminator assigns a probability to each image, determining whether it believes the image is real (closer to 1) or fake (closer to 0).

This process establishes a feedback loop: the Discriminator receives feedback from real data, and the Generator receives feedback from the Discriminator. Over time, both networks improve their performance.

3D-convolution

Convolutions play a key role in enabling the model to track how objects and scenes change over time, which is essential for distinguishing between typical and abnormal behavior in videos. They can learn and identify intricate motion patterns, such as walking, running, or abrupt movements, which may signal anomalies. By analyzing multiple frames simultaneously, convolutions provide context around each frame, enhancing the model's capacity to detect subtle anomalies that might go unnoticed when examining frames in isolation.

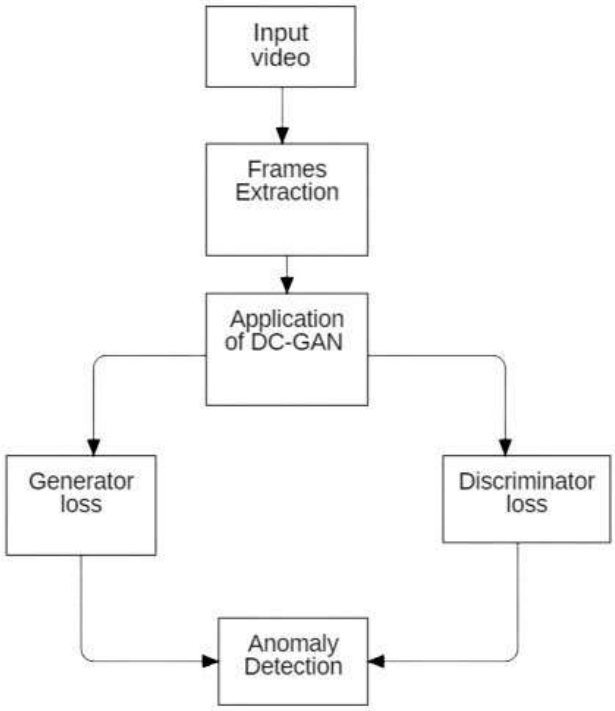
Fig. 2: Data flow diagram

V. IMPLEMENTATION

The traffic dataset used for anomaly spotting typically comprises videos captured from various traffic surveillance cameras installed at different intersections and road segments.This dataset comprises a wide range of traffic situations including varying times of day weather conditions and traffic densities.It includes instances of normal traffic flow as well as various anomalies like accidents sudden stops lane violations no helmet anomalous behavior and pedestrian intrusions each video sequence is labeled with metadata including the timestamp location and type of anomaly if any, so much rich and heterogeneous dataset thus forms a robust basis of training and testing the machine learning model designed for identifying oddity in traffic videos that would be presented.

A.Extracting frames from input video:

- Load video data from the specified folder.
- Resize and normalize frames.
- Convert video data into sequences of frames.



B.Defining the GAN Architecture:

- Generator: Defines the generator with transpose convolutional layers to generate video sequences.
- Discriminator: Defines the discriminator with convolutional layers to discriminate between real and generated video sequences.

C.Training the GAN:

- Initialize the generator and discriminator models.
- Define optimizers and loss function of generator and discriminator.
- Create a Data Loader for batching the video data and upsample the video using generator
- In each epoch, update the discriminator and generator alternately.

D.Calculating the loss:

Fig. 3: Upsampling and Downsampling

Upsampling refers to increasing the resolution or size of data, commonly used in the generator network of a GAN to create high-resolution images from low-dimensional noise. The generator starts with a low-dimensional input (typically noise) and uses upsampling layers to produce a high-dimensional output like an image.

Down sampling reduces the size or resolution of data, typically used in the discriminator network to distil features from high-resolution images and simplify the data for classification. This method uses convolutions with a stride greater than 1 to reduce the spatial dimensions while maintaining learned filters. Down sampling helps efficiently analyze image details, enabling the

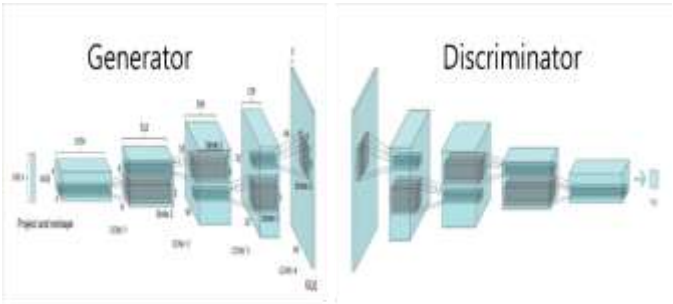
discriminator to detect subtle differences between real and generated images.
These operations are essential for balancing the generator and discriminator in GANs to create high-quality, realistic images.

VI. Results and Discussions

The results are obtained after the entire process and calculation of the score on comparison with the ground truth of the image and reconstruction errors. For normal (non-anomalous) videos, the reconstruction error will be low. For anomalous videos, the reconstruction error will be significantly higher.
By setting a suitable threshold on the reconstruction error, you can effectively classify videos as normal or anomalous. Videos with reconstruction errors above the threshold will be flagged as anomalies. Anomalies are detected and marked as red while the non-anomalies are marked as green. Not wearing a helmet is an anomaly which is shown for the persons who are not wearing the helmet. Consequently, their number plates are also marked with red.
The process starts with extraction of frames from the video on which the GAN is implemented and then the whole process is initiated.

- For the discriminator, calculate loss for real and fake data, backpropagate, and update the parameters.
- For the generator, calculate loss for fake data (labeled as real), backpropagate, and update the parameters.

E.Anomaly Detection: During inference, generate videos using the generator.Compare the generated videos with rea]l videos using reconstruction error.The algorithm marks videos with high reconstruction error as anomalies.



After the successful extraction of frames, generator loss and discriminator loss are calculated for each epoch, and with the help of the losses, the accuracy is determined.

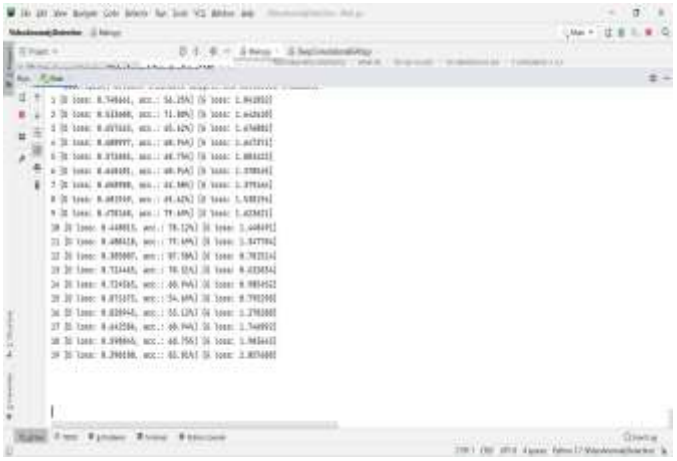


Fig. 5: Generator and Discriminator Loss for each epoch

The overall accuracy of the algorithm is hence calculated, and the discriminator loss and generator loss are generated.

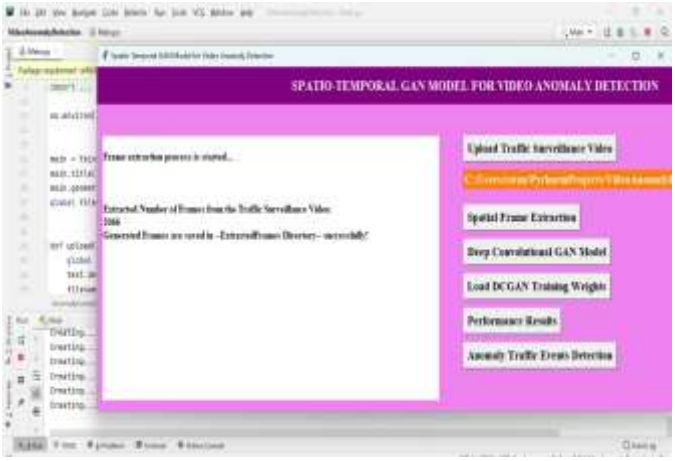


Fig. 4: Frame Extraction of input

Fig. 6: Accuracy score of Anomaly detection

The final result is obtained highlighting the anomaly that is not wearing a helmet.

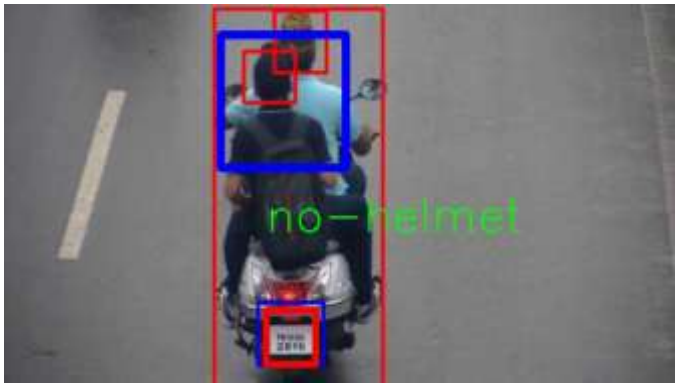


Fig. 7: Anomaly detection - No helmet

VII. Conclusion

Basically GAN for video oddity spotting is a useful technique in detecting anomalies within daily life since training is not dependent on annotated oddity data GAN-based oddity spotting will be applied in a situation where labeled anomalies are scarce or non-existent it is an unsupervised model due to the fact that good generalization to various settings is attained during training usually GAN-based anomaly prediction tends to strong against noise and fluctuations from input data applications requiring fast oddity spotting within video streams are quite capable of GAN-based oddity spotting it can execute in real time or nearly real time with optimized implementations this work helps us strengthen our knowledge about the methodology in the design of a machine learning model concerning the analysis stage we become aware that realistic use cases are very important and also explain technical deep knowledge in developing the machine learning model we have utilized the innovative spatiotemporal DCGAN algorithm that is particularly made for anomaly spotting in different types of traffic video datasets because this makes the model self sufficient due to the implicit temporal shift mechanism used in learning the motion information we added more structural comparability restrictions to ensure that in the case of regular events the predicted image would match the ground truth therefore based on higher anomaly scores images can be classified as anomalous results of thorough testing demonstrate the power of our approach our model outperforms state-of the-art models designed specifically for anomaly spotting for regular non-anomalous videos the reconstruction error-the difference between original and generated videos-will be minimal when video anomaly spotting will be used for anomalous videos the reconstruction error will be much larger you can classify videos as normal versus anomalous by using the gan effectively using generator loss along with discriminator loss some videos are detected to flag as anomalies

VIII. Future Enhancements

Most anomaly-detecting GANs are unconditional-that is they depend on no conditions or contextual information most anomalies are condition-dependent-for example a value may be anomalous in one context but ordinary in another conditioning both generation and detection on additional variables such as time context or metadata could prove useful for capturing condition-sensitive anomalies attention layers in DCGAN could pay more attention to the more relevant parts in data to further aggressively identifying anomalous in specific conditions although GAN-based anomalous spotting has proven effective in most cases it is often computationally expensive and not feasible for real-time applications in the future GANs trained should be able to adapt to new data without retraining the whole model which would be the most required for environments with dynamic data distributions GANs themselves are prone to adversarial attacks where small perturbations of data lead to incorrect oddity spotting or undetected anomalies methods for adversarial training can be used to make GAN-based techniques anomaly detection stable to adversarial attacks GAN-based anomaly detection fails to detect complex anomalies sometimes as resulting from imperfections in generators or discriminators GANs have a large potential role in deep oddity spotting in various industries

References

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

[2] B. R. Kiran, D. M. Thomas, and R. Parakkal, "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos," J. Imag., vol. 4, no. 2, p. 36, 2018.

[3] Shin, W., Cho, SB. (2018). CCTV Image Sequence Generation and Modeling Method for Video Anomaly Detection Using Generative Adversarial Network. In: Yin, H., Camacho, D., Novais, P., Tallon-´ Ballesteros, A. (eds) Intelligent Data Engineering and Automated Learning – IDEAL 2018.

[4] T. Wang *et al*., "Generative Neural Networks for Anomaly Detection in Crowded Scenes," in *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1390-1399, May 2019

[5] Srijan Das, Rui Dai, Michal Koperski, Luca Minciullo, Lorenzo Garattoni, Francois Bremond, Gianpiero Francesca; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 833-842

[6] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, Anton van den Hengel; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1705-1714

[7] Nguyen, K.-T., Dinh, D.-T., Do, M. N., Tran, M.-T. (2020). Anomaly Detection in Traffic Surveillance Videos with GAN-based Future Frame Prediction. Proceedings of the 2020 International Conference on Multimedia Retrieval.

[8] Hyunjong Park, Jongyoun Noh, Bumsub Ham; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 14372-14381

[9] A. Markovitz, G. Sharir, I. Friedman, L. Zelnik-Manor, S. Avidan, Graph embedded pose clustering for anomaly detection, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10536–10544

[10] S. Saypadith and T. Onoye, "An Approach to Detect Anomaly in Video Using Deep Generative Network," in *IEEE Access*, vol. 9, pp. 150903150910, 2021,

[11] Chen, Dongyue, et al. "NM-GAN: Noise-modulated generative adversarial network for video anomaly detection." *Pattern Recognition* 116 (2021): 107969.

[12] W. Zhang, G. Wang, M. Huang, H. Wang and S. Wen, "Generative Adversarial Networks for Abnormal Event Detection in Videos Based on Self-Attention Mechanism," in *IEEE Access*, vol. 9, pp. 124847-124860, 2021, doi: 10.1109/ACCESS.2021.3110798.

[13] W. Zhang, et al., "A Dynamic Convolutional Generative Adversarial Network for Video Anomaly Detection," *Arabian Journal for Science and Engineering*, vol. 48, no. 2, pp. 2075-2085, 2021.

[14] R. Nayak, U. C. Pati, and S. K. Das, "A comprehensive review on deep learning-based methods for video anomaly detection,"Image Vis. Comput., vol. 106, Feb. 2021, Art. no. 104078

[15] Zhang, Wei, et al. "A Dynamic Convolutional Generative Adversarial Network for Video Anomaly Detection." *Arabian Journal for Science and Engineering* 48.2 (2021): 2075-2085.

[16] Singh, R., Sethi, A., Saini, K. *et al.* VALD-GAN: video anomaly detection using latent discriminator augmented GAN. *SIViP* 18, 821–831

[17] Lee, Chang-Ki, Yu-Jeong Cheon, and Wook-Yeon Hwang. "Studies on the GAN-based anomaly detection methods for the time series data." *IEEE Access* 9 (2021): 73201-73215.

[18] Chang, Y., Tu, Z., Xie, W., Luo, B., Zhang, S., Sui, H., Yuan, J. (2022). Video anomaly detection with spatio-temporal dissociation. Pattern Recognition, 122(2021).

[19] Li, Daoheng, et al. "Context-related video anomaly detection via generative adversarial network." *Pattern Recognition Letters* 156 (2022): 183-189.

[20] Fan, Y., Wen, G., Xiao, F. *et al.* Detecting Anomalies in Videos using Perception Generative Adversarial Network. *Circuits Syst Signal Process* 41, 994–1018 (2022).