https://doi.org/10.2352/EI.2023.35.3.MOBMU-364

Generative Adversarial Networks (GANs) and Object Tracking (OT) for Vehicle Accident Detection

Taraka Rama Krishna Kanth Kannuri[1], *Kirsnaragavan Arudpiragasam*[1], *Klaus Schwarz*[1,3], *Michael Hartmann*[1], *Reiner Creutzburg*[1,2]

[1] *SRH Berlin University of Applied Sciences, Berlin School of Technology, Ernst-Reuter-Platz 10, D-10587 Berlin, Germany*
Email: klaus.schwarz@srh.de, reiner.creutzburg@srh.de

[2] *Technische Hochschule Brandenburg, Department of Informatics and Media, IT- and Media Forensics Lab, Magdeburger Str. 50, D-14770 Brandenburg, Germany*
Email: creutzburg@th-brandenburg.de

[3] *University of Granada, Faculty of Economics and Business, P.° de Cartuja, 7, ES-18011 Granada, Spain*

Keywords: Anomaly detection (AD); Attention connection; Generative Adversarial Networks GANs; Bidirectional-Generative Adversarial Network (Bi-GAN); Deep- Sort; YOLO-R (You Only Learn One Representation); Adversarial error; Skip connection

Abstract

Accident detection is a complex task in computer vision (CV) because of various anomalies, occlusions, and objects that change over time in video footage. Unlike many other CV challenges, accident detection is not solely based on image content but is also affected by the motion and appearance of objects in the scene. In recent years, researchers have explored various deep learning (DL) techniques for anomaly detection (AD), including multimodal approaches, a combination of image reconstruction and optical flow, object detection (OD), object recognition, machine learning (ML) with DL methods and generative adversarial networks (GANs). However, none of the studies has combined object tracking (OT) and image generation to detect anomalies. This study proposed a novel approach for traffic accident detection that combines OT and image generation using GANs with variations such as skip and attention connections. Initially, manual inception separated anomalies and non-anomalies frames from the video. After that, background removal (BR) techniques were applied to reduce background variability in the image. Then, OD is performed using YOLO-R (You Only Learn One Representation) and OT using the DeepSort model. Finally, the Kalman filter and GANs are utilized to calculate the distance error metric and adversarial error for AD in surveillance videos. The proposed algorithm in this study detects traffic accidents in various scenarios using the GAN model with skip connections and OT, achieving the highest accuracy among the proposed models. This approach demonstrates the effectiveness of combining OT and image generation for accident detection in video surveillance.

INTRODUCTION

Anomalies are data points or patterns in data that do not fit into a defined representation of typical behavior [1]. Besides, with the recent advancements in CV, DL, and ML techniques [2], AD in surveillance videos has become an important research area. However, AD can be a supervised learning problem for which collecting a large amount of labeled data takes time and effort [3, 4, 5]. Researchers have recently applied various ML and DL multimodal techniques to detect anomalies in live streaming, including analyzing traffic accidents, speed limits, license plates, etc., to improve road safety [6, 7]. Identifying anomalies is finding patterns in the training data that are not previously visible is the traditional way to solve the problem in videos. Most current techniques are based on end-to-end trained, deep neural networks that require a large number of training data.

In AD, methods must simulate data, which can be complicated and multidimensional [8]. GANs [9] have been successfully used to represent such complicated and high-dimensional data, especially in real images. However, most methods can only be used with homogeneous scene datasets; traffic datasets are identical. Such models need to be explicitly trained on video from each scene in which they are used and could be better for real-time applications. As a result of such model's improvement in surveillance technologies has led to the widespread of AD, which is used in a variety of areas including, IT, network intrusion analysis [10, 11, 12], medical diagnostics [13], financial fraud prevention [14, 15], manufacturing quality control [16, 17], marketing, social media analysis, and many others [18]. GANs have shown great promise in various applications [19], including image synthesis, fake image classification, and abnormal frame prediction in the video. By training a GAN on data (images), the model can learn to generate new data that closely matches the normal data. Also, anomalies can be identified by measuring the distance between new data points and the non-anomalies data. However, applying GANs to AD is not without its challenges. For example, GANs may generate data points similar to the actual image but not precisely the same as the image, leading to false positives or false negatives. Additionally, GANs may be prone to overfitting the training data, which can limit their ability to generalize to unseen data. Despite these challenges, AD with GANs has the potential to be robust and flexible in detecting anomalies in complex videos and is an active area of research in ML.

LITERATURE REVIEW

Handcrafted features-based AD

AD identifies unusual events or observations in a dataset that may indicate an underlying problem or issue. It is essential in various fields, including surveillance, fraud detection, and cybersecurity. In the early days of research on AD, handcrafted features were often used to represent the data and detect anomalies. Handcrafted features are manually designed by experts and tailored to the specific characteristics of the data. An example of handcrafted features for AD is using statistical measures [20, 21] such as mean, median, and standard deviation to detect deviations from the norm. These features are simple to calculate and easy to interpret, but they need to be improved in their ability to capture more complex patterns in the data.

Another handcrafted feature AD is trajectory-based detection [22] analyses the movement of objects in a video to detect anomalous behavior. This approach is widely used in various applications such as security video surveillance, traffic monitoring, and crowd behavior analysis. One of the main advantages of this approach is that it is relatively easy to implement, as it does not require complex ML algorithms. However, the accuracy of this approach can be limited by the quality of the handcrafted features used to represent the objects in the video and the ability of the model to capture the underlying patterns of normal behavior effectively. The limitations of the above method have led to the use of handcrafted spatiotemporal features to model motion patterns for AD. The following approach involves extracting low-level appearance features and motion cues such as color, texture, and optical flow. These features are then used to model motion activity patterns. This is a straightforward approach that has been widely used in the field of video AD [21, 20, 22]. However, it has some limitations, such as being sensitive to various noises in the video and requiring manual feature engineering, which can be time-consuming and may not capture the essential features of AD.

The Gaussian Mixture Model (GMM) is often used as a classifier. The GMM is a probabilistic model that assumes that the data follow a mixture of several Gaussian distributions. The GMM can be used to learn the underlying distribution of the data and classify new data points as either normal or anomalous based on their probability under the learned distribution. The GMM [23] is generally effective at detecting anomalies in low-dimensional data but can cause problems with high-dimensional data. In addition, the GMM requires careful initialization and can be sensitive to the choice of hyperparameters, such as the number of mixture components and the covariance structure. Despite these limitations, the GMM remains a popular choice for handcrafted features-based AD due to its simplicity and interpretability. However, in the Two-step approach for video AD, the author initially used the Spatial-Temporal Interest Point (STIP) detector to detect the region of interest in the video. In the later step, they extracted appearance and motion features from the regions of interest. Also, the author used the Histogram of Gradient (HOG) as the appearance feature descriptor and the Histogram of Optical Flow (HOF) as the motion feature descriptor. STIP allows the method to focus on the regions of the video that are most relevant for detecting anomalies [24]. By combining appearance and motion features, this method can capture the visual appearance and motion patterns of the objects in the video. Although this method relies on hand-crafted features, which may not capture the essential features of AD, it requires a lot of manual feature engineering and can be time-consuming. Additionally, this method may not be robust to various video noises such as camera jitters, illumination variations, and occlusions.

Another approach based on handcrafted features is model-based detection [25]. The study discussed that a statistical model of normal behavior is created, and the deviations from this model are identified as anomalies. The following approach has been applied in network intrusion detection, fraud detection, and manufacturing process monitoring. One of the main advantages of this approach is that it can be very accurate, as the model can capture the underlying patterns of normal behavior in detail. However, this approach can be computationally intensive and requires much construction and optimization of a complex statistical model. In addition, it can be challenging to generalize this approach to new data, as the model needs to be re-trained and fine-tuned for each new dataset. Another study discussed a popular approach for AD is using image features, such as edge detection and texture analysis, to detect anomalies in an image. These features are often used in surveillance applications to detect unusual activity or objects on the scene. Despite the popularity of handcrafted features, they require significant domain expertise to design and may only apply to some AD tasks. The features are usually not scalable to large datasets and do not exploit the power of modern computational hardware. Yet, handcrafted features are sensitive to noise and variations in the data. Also, they are not robust to changes in the environment or the appearance of the objects being detected. They cannot capture complex patterns in the data, resulting in poor performance on AD tasks.

Deep learning-based AD

DL techniques have become famous for AD due to their ability to learn complex patterns and features directly from the data. DL models can be trained using large amounts of annotated data and automatically learn to extract the most relevant features for a given task. Moreover, DL models are scale-invariant and robust to noise and variations in the data, which makes them suitable for handling complex and dynamic environments. One of the main challenges of using DL for AD is the availability of annotated data. Annotating large amounts of data is time-consuming and costly; annotating all the data needed to train a DL model is often not feasible. Moreover, researchers have proposed various approaches for semi-supervised or unsupervised AD using DL to address this problem. In recent years, there has been a significant advancement in the area of AD using ML and DL techniques such as GAN, Encoder-Decoder (ED) [26], and Autoencoder (AE). These models have shown promising results in a variety of applications. They can potentially improve accuracy and efficiency to detect anomalous patterns that may not have been explicitly defined or labeled in the training data like UCSD Ped1 and Ped2 [27, 28]. More recent methods have employed DL techniques such as CNN and RNN to automatically learn features from the video data to overcome these limitations. The following methods have shown promising results in detecting video anomalies and have become popular in the field.

Autoencoder (AE) is a popular technique for AD by learning to reconstruct a given input. The model is trained using standard data; the instance is considered abnormal if the reconstructed output does not match the standard data input. The LSTM-ED model

[29, 30, 31] is one type of AE designed explicitly for time series data and uses reconstruction error to detect anomalies. However, LSTMs must capture spatial features making them less effective for video AD. To address this limitation, the Convolutional Autoencoder (CAE) was introduced; it is essential for video AD because it can capture 2D image structures due to the sharing of weights between all locations of the input image. Another study discussed the Convolutional LSTM (ConvLSTM), which combines the strengths of LSTMs and CNN to simulate spatiotemporal [32] correlation by using convolution layers (CL) instead of fully connected layers. By combining the temporal dynamics of LSTMs with the spatial features captured by CNNs, ConvLSTMs are better suited for video AD than traditional LSTMs.

Consequently, another study proposed a crossed u-net model to improve AD's accuracy and computational time in surveillance videos. The model uses two u-net-based subnets and allows the output of each third layer of the systolic channel to be combined with the result of the corresponding layer of the second u-net. The second subnet's corresponding layer's output is used as the next layer's input. In addition, the cascading sliding window method is introduced to determine the difference between each patch in the image, selecting patches based on which ones show the most change to determine the anomaly score. Another author has discussed [33] that handles large amounts of data and learns complex relationships between the input features and the output labels. Large amounts of data needed to be annotated to train the model. A GAN is a DL model also used for AD in such cases. It consists of two neural networks: a generator and a discriminator. The GAN seeks to generate data that is indistinguishable from actual data, unlike the u-net model. This can be used to identify anomalies by flagging data that the GAN cannot generate. GAN can also predict future images by training them on a dataset of past images and then using the trained model to generate new images that resemble those in the training set. This process is often referred to as ”image prediction.” Moreover, the intensity gradient loss is calculated as the difference between the future and predicted images after using a skip connection (GAN) to predict the future image. Also, the optical flow loss is calculated as the difference between the actual frame and the predicted image and between the ground truth frame and the expected frame. Adversarial learning is used to determine the accuracy of predicting the future image. DL techniques can improve AD's accuracy and efficiency. Still, more research is needed to develop robust and scalable approaches for handling real-world data.

Research gap

This research uses YOLO-R and DeepSort for OD and OT, respectively. YOLO-R allows for accurately detecting objects in video frames, while DeepSort provides efficient and robust OT over time. Combining these two techniques can help reduce false positive and false negative rates in accident detection. Using a GAN model for image generation is another crucial aspect of vital research. The GAN can generate synthetic images that closely resemble the actual video frames. This can help to improve the performance of accident detection by providing the model with additional training data and increasing its robustness to different types of accidents. Removing the background from video frames is also essential to this research. By reducing the number of objects and variability in the images, this preprocessing step can help to improve the accuracy of accident detection by allowing the model to focus on the most relevant information in the images.

Overall, the combination of these techniques in this research aims to address the limitations of previous studies and improve the performance of accident detection in surveillance videos.

METHODOLOGY

Accident detection using OT with GAN

AD is a complex area of research that has recently focused on using DL methods to identify unusual events. However, the success of these methods depends on factors such as the quality of the data, the biases of the algorithms, and the limitations of current technology. Furthermore, multiple objects and variations in the background can make it difficult to track objects and make decisions. To address these challenges, the proposed method combines OT with a GAN to detect accidents in surveillance videos by utilizing the correlation between appearances and movements. The GAN model is used to generate fake images that look realistic, while the OT algorithm is used to identify and track objects in the video. This approach allows for more efficient and accurate detection of anomalies in the footage, as it considers both the visual appearance and movement of the objects.

The proposed model begins as shown in figure 1 by separating videos of vehicle accidents from the UCF-Crime and Shanghai Tech datasets, manually categorizing them as anomalous or non-anomalous. Next, the backgrounds of all frames are removed using a pre-trained u^2 - net model. In the following stage, the GAN model is developed with various variants as shown in figure 2, such as without a skip connection, with a skip connection, and with a skip and attention connection. Additionally, the YOLO-R model is assigned to detect objects in the frames, and the DeepSort algorithm is applied to track the objects. Finally, the error metrics from the GAN and DeepSort algorithm are compared to make a final decision. If the output is positive, an accident has occurred; otherwise, no accident is detected.

Data preparation

The quality of the data is crucial for the validation of any DL or ML models. To ensure the research data is reliable, video data from cameras must be acquired under different weather and daylight conditions. The videos collected should contain both non-accident scenes and accident scenes. After reviewing previous research papers, many different types of AD datasets have been developed by other researchers.

However, not all these datasets are publicly available. Among the publicly available datasets used for AD, the street scene dataset contains vehicle accident videos and other anomalies, Shanghai Tech contains crowd counting and other anomalies, and the UCF-Crime dataset contains around twelve different types of anomaly events, including accident videos, as shown in table 1.

Therefore, accident videos collected from these datasets are used to create the anomalies and non - anomalies datasets for this study, as shown in figure 3.

Background removal (BR)

Removing the background from the video involves converting the video into individual frames and then using a pre-trained

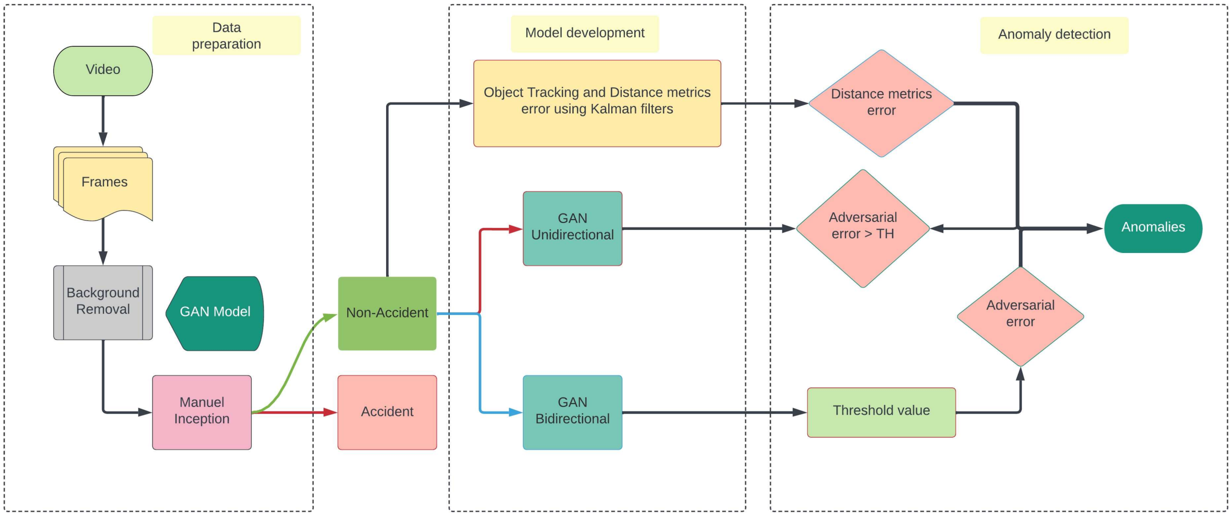

Figure 1: Accident detection algorithm overview

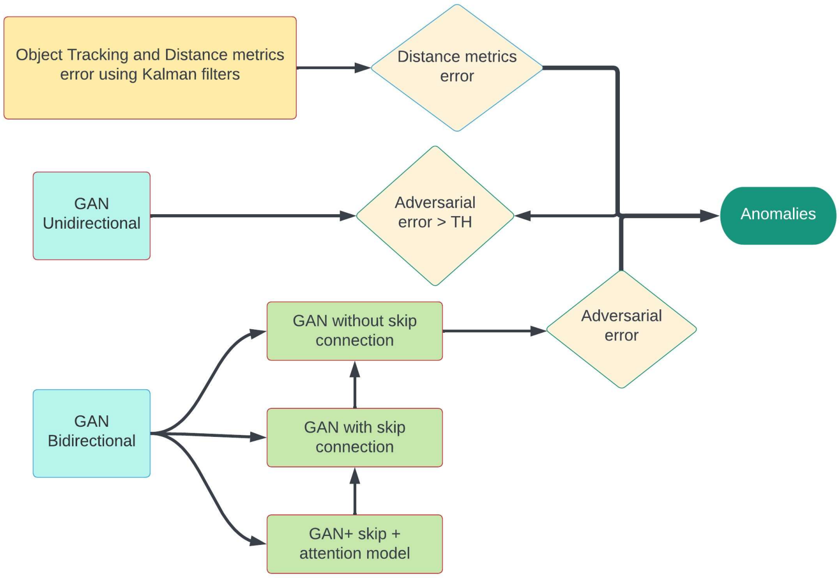

Figure 2: Comparison of errors and decision-making in the identification of anomalies

u^2 model to process those frames. Then, a bitwise operation as in figure 6 is performed between the output of the u^2 network and the input image to obtain the same number of channels as the input image. This results in the background being removed from the image. Once the background is removed, the remaining objects in the frame can be tracked using the YOLO-R model and a DeepSort algorithm. The final step is to compare the error metric received by each model variant to identify the anomaly. For example, the error metrics of the different variants, such as the adversarial error from the GAN model and Mahalanobis distance from the DeepSort algorithm, were determined. Finally, both error metrics are compared to get the final output.

OT model

YOLO is a popular OD algorithm that uses a CNN to predict object bounding boxes and class probabilities directly from full images in one pass. YOLO-R uses a single CNN to learn the image's representations of objects and backgrounds.

DeepSort is an OT algorithm that combines the Kalman filter and the Hungarian algorithm to track multiple objects in a video. It uses a CNN-based feature extractor, such as YOLO-R, to extract features from the input image. Later, it uses the Kalman filter to predict the object's location. Regarding OT, YOLO-R can be combined with DeepSort, a real-time OT algorithm that uses a Kalman filter to track objects in a video. The Mahalanobis

Dataset	UCF-Crime [93]	Street Scene [94]	Shanghai Tech [55]
Abnormal events	13.8M	205	130
Frames	-	203,257	317,398
Examples of anomalies types	abuse, arrest, arson assault, burglary, explosion, fighting, and accident	jaywalking, a biker on the sidewalk, a skateboarder in a bike lane, a biker outside the lane - a pedestrian reversing direction, and a person sitting on a bench	chasing, brawling in a sudden motion, and wrong detection

Table 1: Accident detection dataset with AD

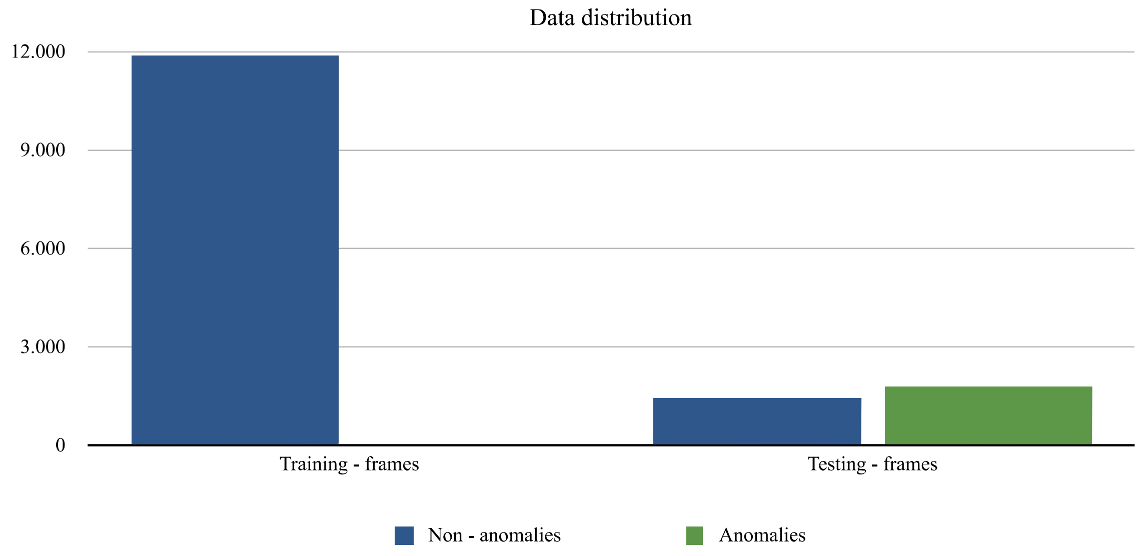

Figure 3: Distribution of the training and testing dataset classes

distance measures the distance between the predicted and actual object positions and is a valuable metric for determining the accuracy of the OT algorithm.

State-of-the-Art

GAN without skip connection

In a GAN architecture, as in figure 7, the generator network creates realistic fake data, while the discriminator network tries to distinguish between real and fake data. In a GAN architecture without skip connections, the generator network does not use shortcuts or connections between layers to allow information to pass through. This architecture typically consists of several dense layers, followed by an actuation function (AF) and dropout layers. The generator inputs a random noise vector of length 256 and produces a realistic image. Conversely, the discriminator takes an image as input and outputs a single value indicating whether the image is real or fake. The generator model consists of four dense layers, with batch normalization and ReLU AF between each layer. The first dense layer of the generator takes the random noise vector as input to produce a tensor. The final layer uses a tanh AF to produce the output image. At the same time, the discriminator model also consists of four dense layers with ReLU AF and dropout regularization between each layer.

The loss function for training the GAN is binary cross-entropy, a common loss function for binary classification tasks. The generator is trained to minimize the cross-entropy loss between the discriminator's output on the generator's output images and a vector of ones. In contrast, the discriminator is trained to

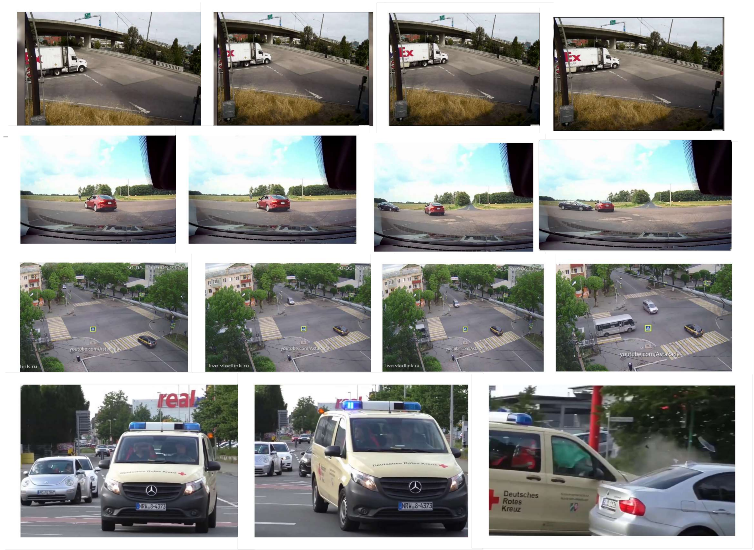

Figure 4: Non-anomalies sample images from various surveillance video footage

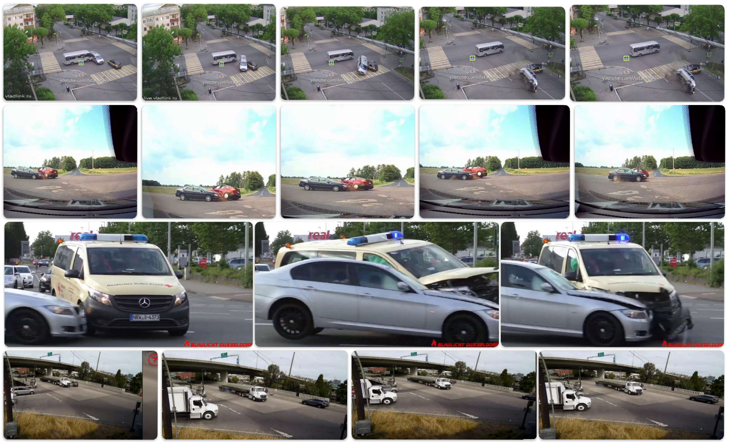

Figure 5: Anomalies sample images from various video footage after the manual inspection

minimize the cross-entropy loss between its output on real and the generator's output images. The dense layer is responsible for a random noise vector of fixed length. Using a dense layer, the random noise vector can be transformed into a tensor that can be reshaped into a 2D or 3D tensor, which subsequent CLs can then process. The AF, such as ReLU, introduces non-linearity to the network and allows it to learn more complex functions. Dropout layers prevent overfitting by randomly dropping out neurons during training. The generator network takes in a random noise vector as input and passes it through the dense layer to generate a fake image. The output of the generator network is then fed into the discriminator network, which tries to classify the image as real or fake.

Model Training

The model's training process followed specific parameters and techniques to ensure the robustness and accuracy of the model. In addition, the GAN is trained using the Adam optimizer with a learning rate of e^{-5}. The exact RGB image was used as both input and output, and the model was trained using 30% of the available images for both steps per epoch and validation steps. LR reduction techniques were used as callbacks to optimize the model's performance. The model was evaluated using test data to avoid overfitting and to analyze the training and validation loss behavior. The results were also manually examined, and if necessary, the model was retrained with different parameters to achieve satisfactory results.

GAN with skip connection

Usually, GAN base architecture may not be as effective as skip connections. The reason is that skip connections allow information to pass through from one layer to another, which can help to reduce the risk of vanishing gradients. Adding skip connections to the generator allows the generator to reuse features

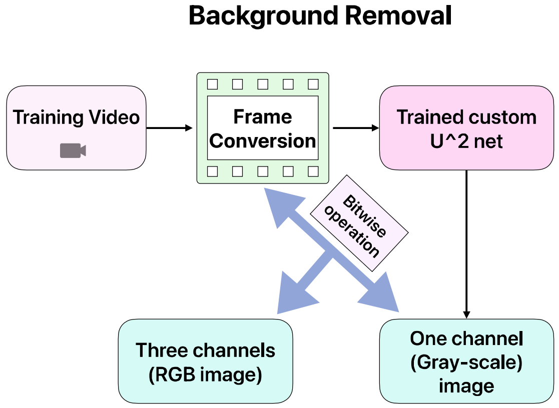

Figure 6: BR technique

learned from previous layers and improve the quality of the generated images. On the other hand, adding skip connections to the discriminator can help it to learn more efficiently, as it can use information from previous layers to better discriminate between real and fake images.

GAN model with skip connections, additional connections are added between layers to enable the gradient to propagate quickly between the input and output layers. More specifically, a GAN model with skip connections has an architecture where the generator has skip connections that directly connect the input of each layer to the output of the last layer, creating a shortcut path for the gradient to propagate. These connections allow the model to more easily learn the underlying distribution of the data, which can lead to better performance and faster training times.

GAN with skip connection and attention connection

In the GAN with skip and attention connections, attention was added to both the generator and discriminator models. Specifically, a self-attention layer was added to the generator model before each skip connection. This layer allows the model to pay attention to different parts of the input and helps improve the quality of the generated images. After each CL in the discriminator model, a similar self-attention layer was added. This layer helps the discriminator focus on essential parts of the image and improves its ability to distinguish between real and fake images.

The architecture starts with a dense layer, and The next layer is a batch normalization layer that normalizes the activation of the previous layer. It can help stabilize training by reducing the internal covariate shift. The next layer applies a leaky ReLU AF. The leaky ReLU is a ReLU function variant that allows small negative values to pass through the AF. It can help avoid the "dying ReLU" problem, where a significant fraction of the neurons become inactive during training. After the AF, the reshape layer is used to reshape the dense output. Next, the attention layer is used to apply an attention mechanism to the feature map. As it can help the generator focus on the image's essential regions and improve the generated images' overall quality. The subsequent three layers are convolutional transpose layers that upsample the feature map by a factor of 2. After each convolutional transpose layer, a batch normalization layer and a leaky ReLU activation function are applied. Additionally, an attention layer is used after each convolutional transpose layer to improve further the generated images' quality.

With attention and skip connections, this model architecture can generate high-quality images with sharp details and improved global coherence compared to a standard GAN model.

RESULTS

Background removal (BR)

After separating the images into anomalous and non-anomalous categories, a sample non-anomalous image is used to train a model. The u^2 model is then applied to this image to remove the background, resulting in a single-channel mask image. However, this black-and-white image can make detecting objects and generating the image complex. To solve this, a bitwise operation is performed on the original image and the mask image to obtain an RGB image without the background.

The next step is to train a GAN model using different variations. A GAN model without skip connections is one variation that uses dense layers and architecture for image generation and the discriminator. Another variation is a GAN model with skip connections, which uses a dense layer and architecture with added connections to transfer information between the generator dense layer block in figure 8. A third variation is a GAN model with skip and attention connections, which uses CL and architecture with added connections and attention mechanisms to focus on essential features during the transfer. These variations are trained and compared to determine the best model for AD.

Object detection (OD) and tracking (OT)

Figure 9 shows to evaluate the performance of the OT; the Mahalanobis distance error metric is used. This error metric is calculated based on the difference between the predicted position of an object and its actual position. The Mahalanobis distance error metric can be combined with the adversarial errors from the GAN model to decide whether an image contains an anomaly, represented in figure 10.

GAN without skip connection

The generator in the GAN without a skip connection comprises a series of the dense layer that takes in a random noise vector as input and generates an output image.

The discriminator in the GAN without a skip connection comprises a series of CL that takes in an image as input and outputs a scalar value representing the probability that the input image is real. While training the model, a few hyperparameters, such as learning rate, feature maps, and optimizer, are essential to define. As the generator and discriminator are optimized using an optimizer like Adam. The number of training iterations and the batch size can also be adjusted to change the model's performance.

Model Training

The GAN model without a skip connection was trained using a set of key hyperparameters, including the feature maps, optimizer, and learning rate. The feature maps were set to 256,512,754,512 for the generator. The Adam optimizer was used, and the initial learning rate was set to e^{-5}.

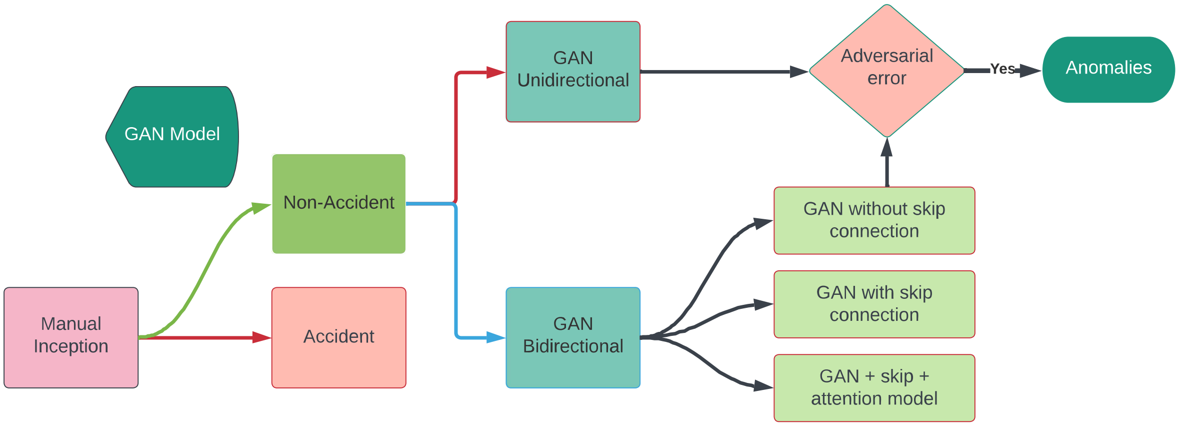

Figure 7: State-of-the-Art model

Adversarial error analysis

The GAN model without skip connections was used to generate new non-abnormal images, and the adversarial error value of each frame was plotted as a normal distribution to determine the threshold value. When analyzing a new video, the frames were generated using the model, and if the error value exceeded the threshold, the frame was considered an accident frame. This method was used to identify the video's accident frames between frames 77 and 107, as shown in figure 13. However, it's worth noting that the model may also falsely predict some accident frames as non-accident scenarios and vice versa.

GAN with skip connection

A new version known as the GAN with skip connection was introduced to enhance the performance of the previous GAN model. This new model addresses the limitations of the prior model by incorporating a skip connection, which helps capture the temporal characteristics while generating frames. The optimal feature map size was also studied to build the best model. The results showed that the model performed better when the feature map size started at 128 and ended at eight than when it started at 256 and ended at 8. However, it was observed that the accuracy of the model using a 256-feature map was not consistent compared to using a 128-feature map.

GAN with attention connection

The GAN with the skip connection and attention model has improved the accuracy of the prior models with a more robust performance. The attention connection helps to focus on the dominant temporal features between the generator and discriminator parts of the GAN, passing these features through the skip connection. This, combined with the output of the OT, results in a complete understanding of the situation and reduces missing behavior. The error analysis of this model still incorrectly predicts a few non-incident frames as accidents, but the results are better than the previous models. The confusion matrix after combining the distance shows that this model has higher overall accuracy and improved F1 score.

Finally, the comparison between the various variants of the GAN model was conducted to determine which model performed best in accuracy and F1 score. The results showed that the GAN combined with the skip connection plus attention plus OT algorithms and the GAN combined with the skip connection plus attention connection plus OT algorithms had similar results. However, the GAN with the skip connection plus attention plus OT algorithms obtained higher F1 scores and accuracy values.

Despite this improvement, the model still predicted some incident frames as non-incident. This is likely due to the difficulty in defining an incident's start and end frames, which makes it challenging for the model to classify all incident frames accurately. Adding the attention connection improved the performance of the GAN model, reducing overfitting and increasing accuracy and F1 scores. However, it's possible that further improvements could be made to the model by exploring other techniques, such as fine-tuning the model's parameters or using a different architecture.

Conclusion and Outlook

This study presents a novel method for AD that combines the correlation of patterns and movements in surveillance videos. This approach helps identify events such as accidents, abuse, and abductions. The proposed method aims to improve security and public order by utilizing surveillance videos in law enforcement, traffic, and environmental monitoring. The proposed method involves three main steps: BR, OT, OD, and image generation. The BR step eliminates variations and unwanted objects in the image to improve accuracy. OD and OT are used to identify the accident and track the objects involved, such as a vehicle with a human or a vehicle with a motorbike. Image generation is done using a GAN, with different variants, such as skip connections and attention connections, to enhance the temporal features of the image and prevent problems such as disappearing or exploding gradients.

The study also introduces the concept of patch loss and anomalous adversarial loss, which are effective for identifying defects and integrating them with other losses from previous work to perform joint learning. The weighting of each loss is inves-

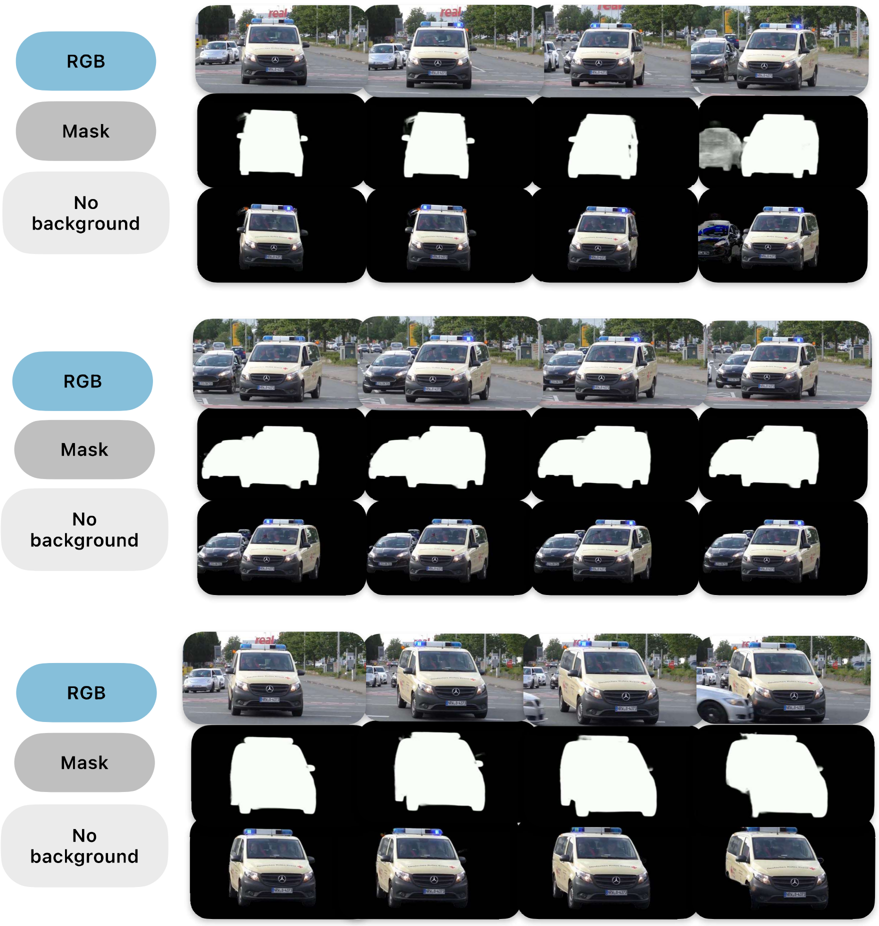

Figure 8: BR output of the bitwise operation

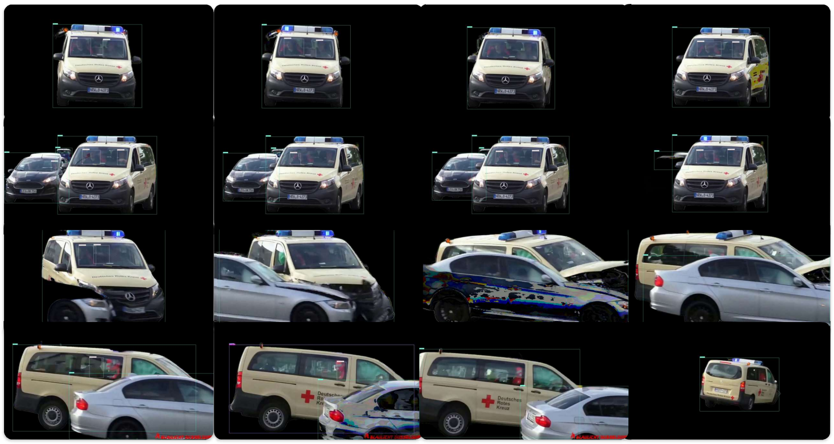

Figure 9: OT (DeepSort) output

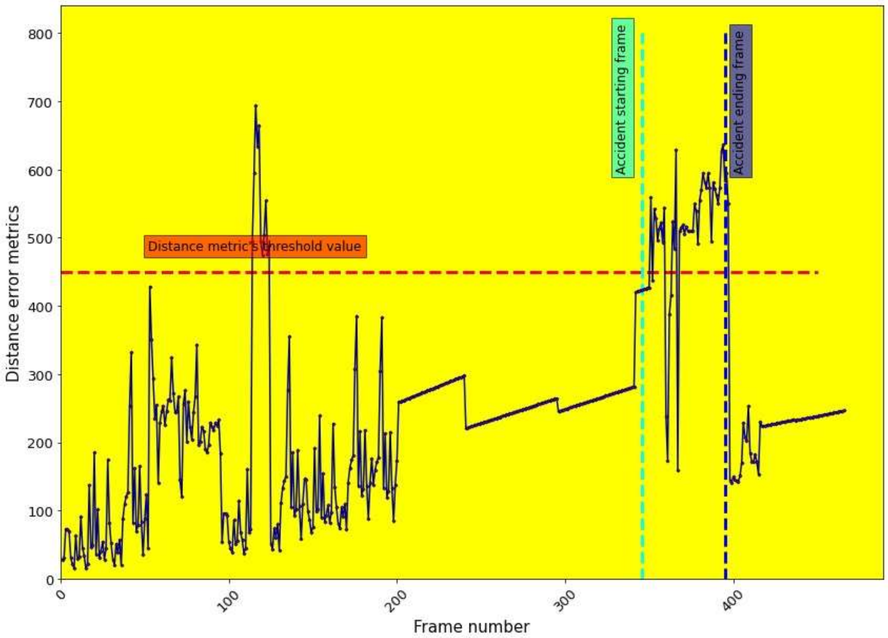

Figure 10: Mahalanobis distance error metric from DeepSort

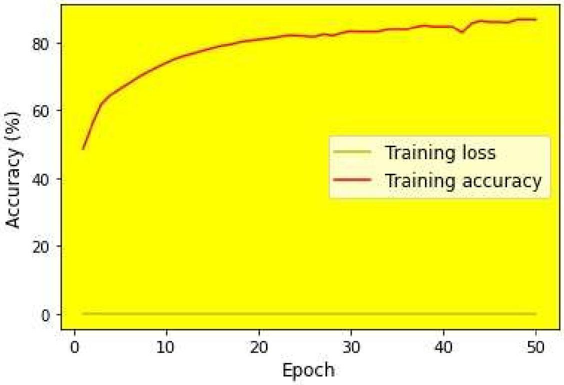

Figure 11: First training sample feature map begins with 128 and ends with 512

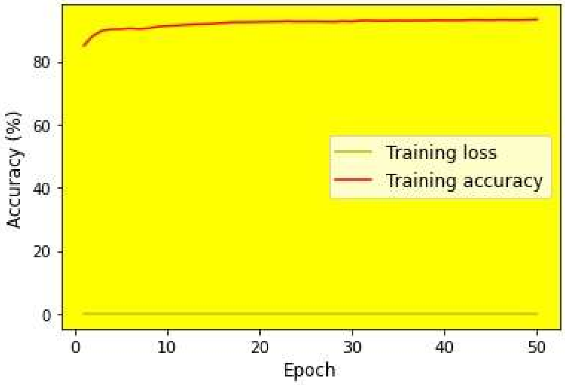

Figure 12: Second training sample feature map begins with 512 and ends with 754

tigated using a grid search to determine their contribution to the algorithm's overall performance. The use of DeepSort for OT and GAN for image generation allows the study to evaluate the effectiveness of BR methods for image generation and OT, the ability to identify accident scenes with vehicle classes, and the impact of different features on image generation. The study also uses the Mahalanobis error distance metric from DeepSort and the adversarial error from GAN to detect anomalies.

In conclusion, the study provides a systematic approach for developing an automatic crash detection algorithm that can detect vehicle classes involved in crashes and suggests future improvements such as incorporating adversarial features into the GAN model, replacing DeepSort with recurrent techniques, and training the model under different weather and daylight conditions to enhance the performance. This approach can help improve public order and security by detecting anomalous events in surveillance videos.

REFERENCES

[1] Chandola, V., Banerjee, A., and Kumar, V., "Anomaly detection: A survey," *ACM Comput. Surv.* **41** (jul 2009).

[2] Patcha, A. and Park, J.-M., "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer networks* **51**(12), 3448–3470 (2007).

[3] Feng, J., Liang, Y., and Li, L., "Anomaly detection in videos using two-stream autoencoder with post hoc interpretability," *Computational Intelligence and Neuroscience* **2021** (2021).

[4] Wan, B., Jiang, W., Fang, Y., Luo, Z., and Ding, G., "Anomaly detection in video sequences: A benchmark and computational model," *IET Image Processing* **15**(14), 3454–3465 (2021).

[5] Injadat, M., Salo, F., Nassif, A. B., Essex, A., and Shami, A., "Bayesian optimization with machine learning algorithms towards anomaly detection," in [*2018 IEEE global commu-*

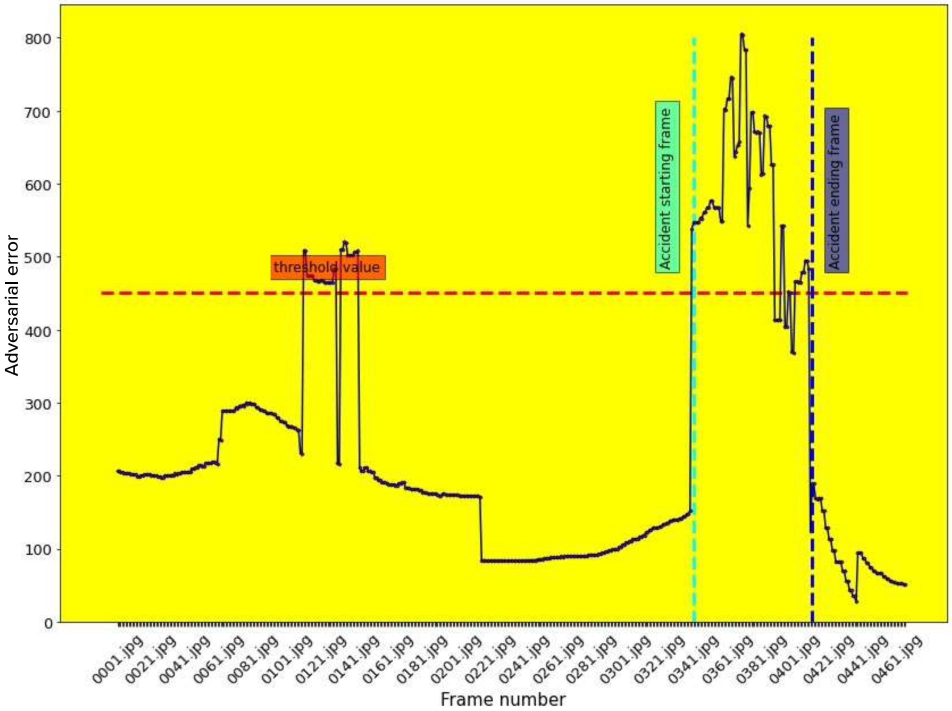

Figure 13: Adversarial error analysis

nications conference (GLOBECOM)], 1–6, IEEE (2018).

[6] Xue, R., Chen, J., and Fang, Y., "Real-time anomaly detection and feature analysis based on time series for surveillance video," in [*2020 5th International Conference on Universal Village (UV)*], 1–7, IEEE (2020).

[7] Nasaruddin, N., Muchtar, K., Afdhal, A., and Dwiyantoro, A. P. J., "Deep anomaly detection through visual attention in surveillance videos," *Journal of Big Data* **7**(1), 1–17 (2020).

[8] Sample, C. and Schaffer, K., "An overview of anomaly detection," *IT Professional* **15**(1), 8–11 (2013).

[9] Zenati, H., Foo, C. S., Lecouat, B., Manek, G., and Chandrasekhar, V. R., "Efficient gan-based anomaly detection," *arXiv preprint arXiv:1802.06222* (2018).

[10] Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., and Ahmad, F., "Network intrusion detection system: A systematic study of machine learning and deep learning approaches," *Transactions on Emerging Telecommunications Technologies* **32**(1), e4150 (2021).

[11] Van, N. T., Thinh, T. N., et al., "An anomaly-based network intrusion detection system using deep learning," in [*2017 international conference on system science and engineering (ICSSE)*], 210–214, IEEE (2017).

[12] Gurung, S., Ghose, M. K., and Subedi, A., "Deep learning approach on network intrusion detection system using nsl-kdd dataset," *International Journal of Computer Network and Information Security* **11**(3), 8–14 (2019).

[13] Fernandes, M., Corchado, J. M., and Marreiros, G., "Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: a systematic literature review," *Applied Intelligence* , 1–35 (2022).

[14] Maniraj, S., Saini, A., Ahmed, S., and Sarkar, S., "Credit card fraud detection using machine learning and data science," *International Journal of Engineering Research* **8**(9), 110–115 (2019).

[15] Bin Sulaiman, R., Schetinin, V., and Sant, P., "Review of machine learning approach on credit card fraud detection," *Human-Centric Intelligent Systems* , 1–14 (2022).

[16] Arpitha, V. and Pani, A., "Machine learning approaches for fault detection in semiconductor manufacturing process: A critical review of recent applications and future perspectives," *Chemical and Biochemical Engineering Quarterly* **36**(1), 1–16 (2022).

[17] Wang, X., Ding, H., Gu, X., Yuan, J., and Shen, Q., "Study of traffic incident detection with machine learning methods," in [*Education and Awareness of Sustainability: Proceedings of the 3rd Eurasian Conference on Educational Innovation 2020 (ECEI 2020)*], 725–728, World Scientific (2020).

[18] Abdelzad, V., Czarnecki, K., Salay, R., Denounden, T., Vernekar, S., and Phan, B., "Detecting out-of-distribution inputs in deep neural networks using an early-layer output," *arXiv preprint arXiv:1910.10307* (2019).

[19] Karras, T., Aila, T., Laine, S., and Lehtinen, J., "Progressive growing of gans for improved quality, stability, and variation," *arXiv preprint arXiv:1710.10196* (2017).

[20] Adam, A., Rivlin, E., Shimshoni, I., and Reinitz, D., "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE transactions on pattern analysis*

and machine intelligence **30**(3), 555–560 (2008).

[21] Kim, J. and Grauman, K., "Observe locally, infer globally: a space-time mrf for detecting abnormal activities with incremental updates," in [*2009 IEEE conference on computer vision and pattern recognition*], 2921–2928, IEEE (2009).

[22] Cong, Y., Yuan, J., and Liu, J., "Sparse reconstruction cost for abnormal event detection," in [*CVPR 2011*], 3449–3456, IEEE (2011).

[23] Tung, F., Zelek, J. S., and Clausi, D. A., "Goal-based trajectory analysis for unusual behaviour detection in intelligent surveillance," *Image and Vision Computing* **29**(4), 230–240 (2011).

[24] Reddy, V., Sanderson, C., and Lovell, B. C., "Improved anomaly detection in crowded scenes via cell-based analysis of foreground speed, size and texture," in [*CVPR 2011 WORKSHOPS*], 55–61, IEEE (2011).

[25] Liu, S. and Agaian, S. S., "Covid-19 face mask detection in a crowd using multi-model based on yolov3 and hand-crafted features," in [*Multimodal Image Exploitation and Learning 2021*], **11734**, 162–171, SPIE (2021).

[26] Liu, X., Wang, L., Wong, D. F., Ding, L., Chao, L. S., and Tu, Z., "Understanding and improving encoder layer fusion in sequence-to-sequence learning," *arXiv preprint arXiv:2012.14768* (2020).

[27] Lu, Y., Kumar, K. M., shahabeddin Nabavi, S., and Wang, Y., "Future frame prediction using convolutional vrnn for anomaly detection," in [*2019 16Th IEEE international conference on advanced video and signal based surveillance (AVSS)*], 1–8, IEEE (2019).

[28] Qiang, Y., Fei, S., Jiao, Y., and Li, L., "Anomaly detection of predicted frames based on u-net feature vector reconstruction," in [*Journal of Physics: Conference Series*], **1627**(1), 012014, IOP Publishing (2020).

[29] Graves, A., "Long short-term memory," *Supervised sequence labelling with recurrent neural networks* , 37–45 (2012).

[30] Yao, L. and Guan, Y., "An improved lstm structure for natural language processing," in [*2018 IEEE International Conference of Safety Produce Informatization (IICSPI)*], 565–569, IEEE (2018).

[31] Kowsher, M., Tahabilder, A., Sanjid, M. Z. I., Prottasha, N. J., Uddin, M. S., Hossain, M. A., and Jilani, M. A. K., "Lstm-ann & bilstm-ann: Hybrid deep learning models for enhanced classification accuracy," *Procedia Computer Science* **193**, 131–140 (2021).

[32] Di Mattia, F., Galeone, P., De Simoni, M., and Ghelfi, E., "A survey on gans for anomaly detection," *arXiv preprint arXiv:1906.11632* (2019).

[33] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A., "Generative adversarial networks: An overview," *IEEE signal processing magazine* **35**(1), 53–65 (2018).

Author Biography

Taraka Rama Krishna Kanth Kannuri is pursuing a Master's in Engineering and Sustainable Technology Management, focusing on the Mobility and Automotive industry at SRH Berlin University of Applied Sciences. He received his Bachelor's in Mechanical Engineering from Koneru Lakshmaiah University, Vijayawada, India, in 2018. His research interests include autonomous driving, computer vision, ethical decision, and deep learning.

Kirsnaragavan Arudpiragasam holds a Master's degree in Engineering and Sustainable Technology Management from SRH Berlin University of Applied Sciences, where he focused on the Mobility and Automotive Industry. He previously earned his Bachelor's degree in Manufacturing and Industrial Engineering in 2019. His research interests are computer vision, deep learning, multimodal learning, and ethical considerations related to autonomous driving.

Klaus Schwarz received his B.Sc. and M.Sc. in Computer Science from Brandenburg University of Applied Sciences (Germany) in 2017 and 2020, respectively. He is currently a Ph.D. student at the University of Granada, Spain. His research interests include IoT and smart home security, OSINT, mechatronics, additive manufacturing, embedded systems, artificial intelligence, and cloud security. As a faculty member, he is developing a graduate program in Applied Mechatronic Systems focusing on Embedded Systems at SRH Berlin University of Applied Sciences.

Reiner Creutzburg is a Retired Professor for Applied Computer Science at the Technische Hochschule Brandenburg in Brandenburg, Germany. Since 2019 he has been a Professor of IT Security at the SRH Berlin University of Applied Sciences, Berlin School of Technology. He has been a member of the IEEE and SPIE and chairman of the Multimedia on Mobile Devices (MOBMU) Conference at the Electronic Imaging conferences since 2005. In 2019, he was elected a member of the Leibniz Society of Sciences to Berlin e.V. His research interest is focused on Cybersecurity, Digital Forensics, Open Source Intelligence (OSINT), Multimedia Signal Processing, eLearning, Parallel Memory Architectures, and Modern Digital Media and Imaging Applications.