eISSN 2783-6851

VILNIUS TECH
Vilnius Gediminas Technical University

NEW TRENDS in COMPUTER SCIENCES

2024
Volume 2
Issue 1
Pages 1–18
https://doi.org/10.3846/ntcs.2024.20516

REVIEW AND EXPERIMENTAL COMPARISON OF GENERATIVE ADVERSARIAL NETWORKS FOR SYNTHETIC IMAGE GENERATION

Gabriela VDOVIAK ✉, Henrikas GIEDRA

Deparment of Electronic Systems, Faculty of Electronics, Vilnius Gediminas Technical University, Vilnius, Lithuania

Article History:
- received 3 December 2023
- accepted 26 April 2024

Abstract. The application of machine learning algorithms has become widespread particularly in fields such as medicine, business, and commerce. However, achieving accurate classification results with these algorithms often relies on large-scale training datasets, making data collection a lengthy and complex process. This paper reviews the current utilization of generative adversarial network (GAN) architectures and discusses recent scientific research on their practical applications. The study emphasizes the significance of addressing data scarcity in the process of training the machine learning algorithms and highlights the potential of advanced GAN architectures, in particular StyleGAN2-ADA, to mitigate this challenge. The findings contribute to ongoing efforts aimed at enhancing the efficiency and applicability of artificial intelligence across diverse domains by presenting a viable solution to the constraint of limited training data for image classification tasks.

Keywords: computer vision, convolutional neural networks, deep learning, generative adversarial networks, image classification, image synthesis.

✉Corresponding author. E-mail: ***gabriela.vdoviak@vilniustech.lt***

1. Introduction

The rapid growth of machine learning has had a significant impact on the modern technology market, where increasing hardware capabilities have led to more efficient machine learning algorithms. These tools of automation are substituting human labor in industries, medicine, and business. However, the accuracy of these models is greatly influenced by the quantity and quality of training data (Alzubaidi et al., 2021). The collection of data remains a complex process, taking considerable time, financial resources, and collaboration between specialists in various fields. Both the commercial and scientific sectors often face a lack of such data. To address this issue, image augmentation methods are often employed, but its impact is not always substantial (Y. Chen et al., 2022).

Today, one of the main challenges in the field of artificial intelligence is to train models using limited datasets (Ahmed et al., 2023). In many cases, it is impossible to create sufficient datasets, which could ensure the quality of the image classification systems. Recently, there has been notable interest in generative adversarial networks (GAN) as a promising solution for addressing data scarcity. Images created through generative adversarial network models showcase exceptional quality and distinctive features (Salimans et al., 2016). GAN technology stands out as a top method for crafting synthetic images, capable of generating new visuals from limited existing datasets. The implementation of this technology has the potential to improve the effectiveness of classification tasks. This approach has garnered considerable attention due to ongoing research and multiple modifications in recent years (Saxena & Cao, 2021).

2. Synthetic image generation

The process of image generation involves manipulating signals to extract desired information through various operations. Images can be subjected to a variety of filters and transformations to isolate specific values corresponding to the features of objects captured in them. Humans can effortlessly categorize real-world objects based on their features, but it is a challenging task for artificial intelligence. To overcome this, it is crucial to provide diverse variations of the same object to train precise classifiers. Generative adversarial networks serve this purpose by generating synthetic images that showcase potential variations of objects.

GAN is an unsupervised deep learning method comprised of two artificial convolutional neural networks: a generator and a discriminator. These components engage in continuous competition, with the generator learning from real and previously generated images, while the discriminator tries to distinguish between real and synthetic images. The discriminator's architecture consists of a fully connected neural network classifying images into two classes: real or synthetic. The generative part employs the inverse architecture, reconstructing an image from training data and a random noise vector. This progressive cycle continues until the discriminator accurately classifies the images, recognizing synthetic images as authentic (Goodfellow et al., 2020). The GAN architecture is illustrated in Figure 1.

However, another issue arises when evaluating the quality of the generated images. In the generation of images with GANs it can be difficult to assess the image quality due to the lack of a precise evaluation methodology (Iglesias et al., 2023). Common criteria include average log-likelihood, classification, and visual inspection, each with pros and cons (Borji, 2019). Qualitative assessment is based on human judgement, but this is not practical in large datasets, and it presents time and cost challenges. In addition, experts can comprehend images differently. Therefore, many researchers have improved the initial GAN architecture by incorporating various elements that enhance the generated image quality (Wang et al., 2021).

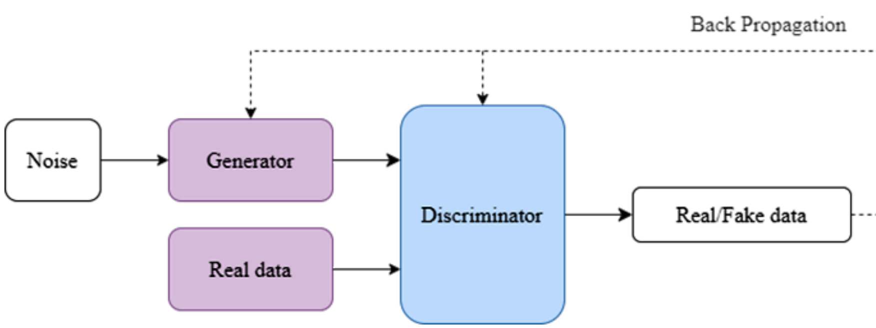

Figure 1. GAN architecture

3. An analytical review on deep learning methods for synthetic image generation

In recent years, many researchers have modified the conventional framework of generative adversarial networks by adapting various network architectures, loss functions, and evolutionary methods. As a result, the GAN architecture significantly improved and generated images became more realistic. Synthetic GAN generated images can be used to solve imbalanced datasets, data leakage, feature extraction, data shortage, inaccurate data labeling, and other challenges. This paper reviews the currently used GAN architectures and discusses scientific research on the practical use of GAN.

3.1. DCGAN network

The architecture of the Deep Convolutional Generative Adversarial Network (DCGAN) replaces the multilayered perception network with the deep convolutional artificial neural network, ensuring stable training of the generative component (Dash et al., 2023). The methodology is designed to project the input of the generator as a high-dimensional tensor and uses convolutional operations to generate the output image. In the generator, these convolutional layers manipulate the image, thus expanding and increasing its resolution (Radford et al., 2015).

This approach is suitable for applications with low-resolution color images. An illustration of the change in the dimension of the tensor (from 14×14×6 to 28×28×1) can be observed in Figure 2.

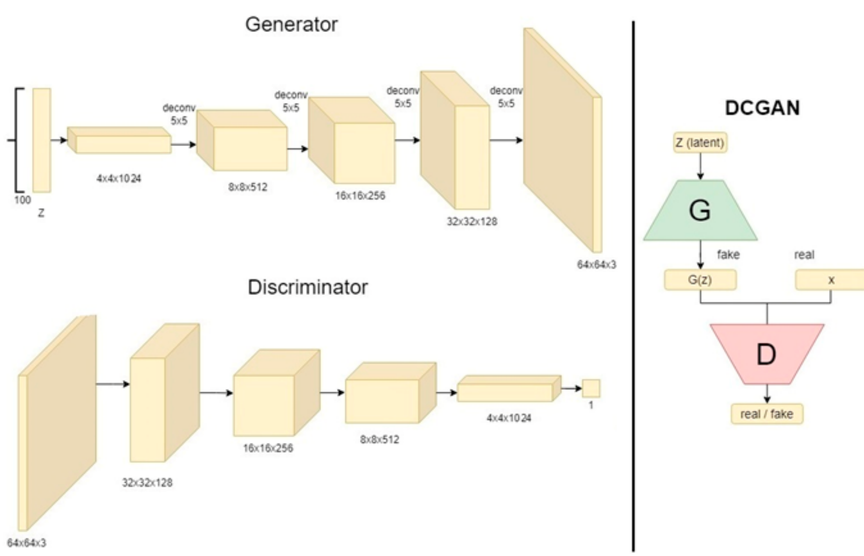

Figure 2. DCGAN model architecture

3.2. CGAN network

The Conditional Generative Adversarial Network (CGAN) is a widely used model in the realm of Generative Adversarial Networks, falling under the category of expansionary GAN architectures. Structured upon the standard GAN network architecture, CGAN incorporates an additional input layer in both its generator and discriminator, including conditional information such as class labels (Mirza & Osindero, 2014), as illustrated in Figure 3.

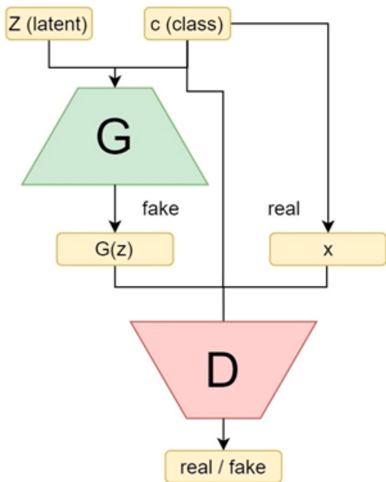

Figure 3. CGAN model architecture

The main objective of this method is to generate realistic images based on specific labels associated with each image in the dataset. The role of the discriminator goes beyond the distinction between real images and false ones; it also ensures that appropriate labels are assigned to the images. This extra information conditions the generation process, allowing for more controlled and targeted image generation (Mert, 2023).

3.3. InfoGAN network

Another widely used model is the Information Maximizing Generative Adversarial Network (InfoGAN), which falls into the category of expanded GAN architectures. InfoGAN addresses the unconstrained usage of noise vectors in the generative part by proposing a division into two parts: an uninterpretable noise source and a latent code (X. Chen et al., 2016). The structural design of this approach is illiustrated in Figure 4.

The InfoGAN network is a fully unsupervised method capable of learning representations of both interpretable and uninterpretable aspects within complex datasets. It is well suited for generating various 3D images, faces, and objects. Furthermore, InfoGAN training is relatively straightforward and requires minimal financial resources during the implementation process (Feng et al., 2023).

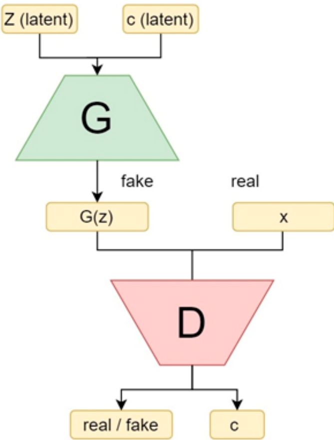

Figure 4. InfoGAN model architecture

3.4. StackGAN network

The accumulated generative adversarial network, StackGAN (Stacked Generative Adversarial Network), is designed to generate images from text employing hierarchically stacked conditional GAN models. However, generating realistic lifelike images through this method is highly complex. The structure proposed by Zhang et al. (2017), addresses the complexity of generating realistic images from text. This modified GAN employs hierarchically stacked conditional GAN models to overcome training instability and nonsensical results encountered in previous attempts. The StackGAN (Figure 5) divides the image generation process into two stages.

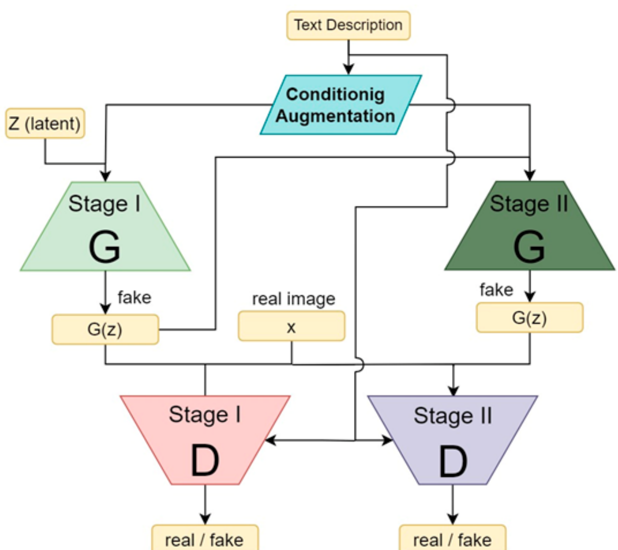

Figure 5. StackGAN model architecture

In the first stage, a low-resolution image is generated based on text criteria, and in the second stage, high-resolution images are produced using the original text and the low-resolution image from the first stage. This two-stage approach allows the second stage to refine and enhance details within the generated objects, utilizing the results of the first stage. Specifically, the second stage of StackGAN combines the low-resolution image and the original text description as conditioning input for the generator, enabling the network to capture more nuanced features and produce higher-resolution, more realistic images compared to using text alone as the sole input (Thamotharan et al., 2023).

3.5. Pix2Pix network

The Pix2Pix generative adversarial network is a method designed to train a deep convolutional neural network to perform image-to-image translation transformations. The generated output image is conditionally transformed based on the initial input image. Both the input and output images are fed into a discriminator that evaluates whether the resulting image is suitably transformed from the original input. Adversarial loss guides the training of the generator, ensuring credible output, and an L1 loss coefficient updates the generator based on the disparity between the synthetic and desired output images (Dash et al., 2023).

This versatile GAN model, illustrated in Figure 6, has successfully tackled tasks such as converting map images to satellite photos, grayscale images to color, and transforming sketches into product photographs using either U-Net or ResNet architecture (Henry et al., 2021).

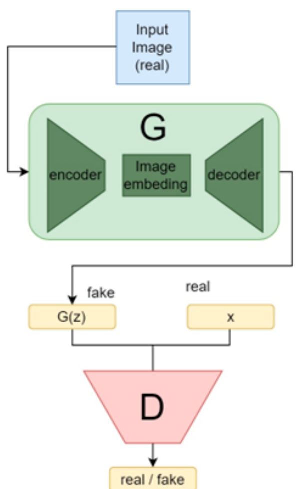

Figure 6. Pix2Pix model architecture

3.6. CycleGAN network

The CycleGAN network belongs to the advanced category of generative adversarial networks (GANs) and is widely used for image transformations. Unlike Pix2Pix, CycleGAN utilizes two unrelated datasets, addressing the difficulty and costs associated with assembling paired

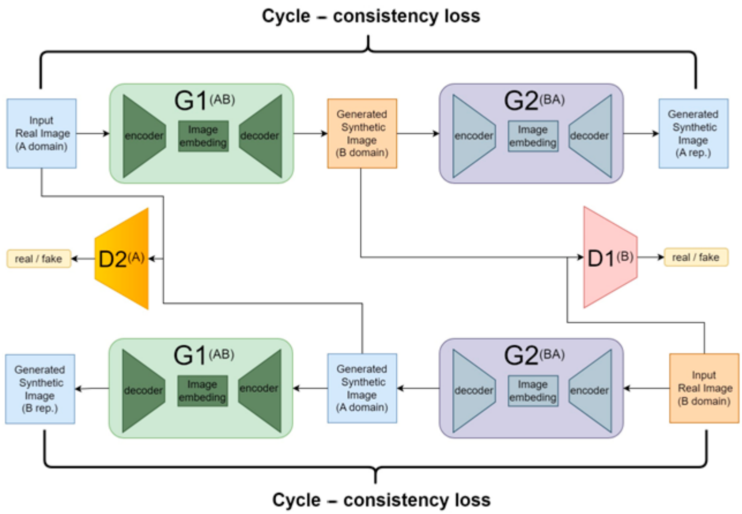

Figure 7. CycleGAN model architecture

training data (Son et al., 2023). For example, in tasks such as semantic image segmentation, only a few combined datasets currently exist, and even these datasets often lack sufficient data (Zhu et al., 2017).

The network comprises two generators and discriminators, each focused on converting images between domains, employing an encoder-decoder architecture with convolutional layers and skip connections. The key innovation is the introduction of cycle consistency loss (Zhu et al., 2017), ensuring that translated images maintain resemblance to the originals through forward and backward transformations. This model's applications include generating high-quality images without inherent correlations, simplifying dataset construction for GAN training, and excelling in medical imaging and photo quality enhancement. The architecture of this approach is shown in Figure 7.

Such network is well suited for medical imaging, the manipulation of objects within images, and the improvement of photo quality. However, challenges arise in modifying video geometry, and the generated images often closely resemble the originals, limiting diverse transformations (Son et al., 2023).

3.7. Progressive GAN network

The Progressive Growing Generative Adversarial Network described in Karras et al. (2017), proposes a concept of the progressive expansion of both the network generator and the discriminator.

Beginning with low-resolution images, the model undergoes incremental augmentation with new layers during subsequent training stages, enhancing synthesized image details. This synchronous growth of the generator and discriminator ensures seamless integration of newly generated layers, depicted in Figure 8 of the architecture. The approach strategically focuses on refining smaller details as additional network layers are added, leading to efficient and stable training with significant time savings compared to traditional methods (Pérez & Ventura, 2023).

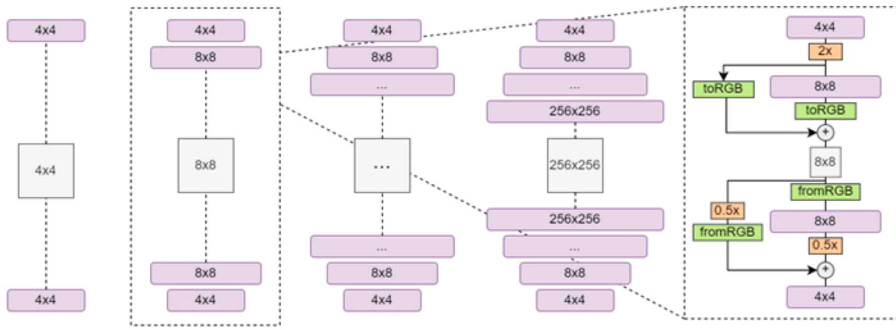

Figure 8. Progressive GAN model architecture

3.8. StyleGAN network

The Style-Based Generative Adversarial Network (StyleGAN) enhances control over image generation by modifying the generator without altering the discriminator. The network's generator embeds a latent algorithmic code along with a noise vector as input at various locations within the model, which as a result significantly influences the code's behavior (Iglesias et al., 2023). As described by Karras et al. (2019), the StyleGAN training process begins with a learned constant and progressively adjusts the images styles generated in each convolutional layer according to the transmitted latent code. This capability allows the algorithm to have direct control over the intensity of visual features in different zones as illustrated in Figure 9.

This model has demonstrated the ability to generate realistic images of faces with various accessories (such as eyeglasses, etc.). The realism achieved in the images generated by StyleGAN surpasses that of its predecessors and traditional GAN models. However, its training requires substantial computational resources.

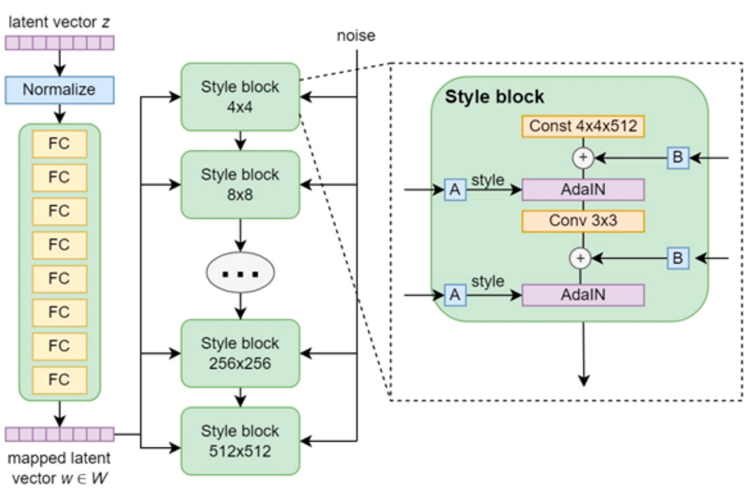

Figure 9. StyleGAN model architecture

3.9. Comparative analysis of Generative Adversarial Network architectures

The advancement of Generative Adversarial Networks (GANs) has led to the development of various architectural frameworks, each with a specific purpose to address challenges in image generation and manipulation. This section provides a comparative analysis of prominent GAN architectures in Table 1.

Table 1. Comparative analysis of Generative Adversarial Network architectures

	Input	Generator / Discriminator	Architecture
DCGAN	Only noise vector	Traditional*	Traditional
CGAN	Noise vector and class information	Traditional	Traditional
InfoGAN	Noise vector and latent code	Traditional	Traditional
StackGAN	Noise vector and embedded text description	Traditional	Two stages
Pix2Pix	Real image	Generator based on encoder-decoder architecture	Traditional
CycleGAN	Real image in Stage I and synthetic image in Stage2	Generator based on encoder-decoder architecture	Two stages
Progressive GAN	Only noise vector	Progressively growing generator and discriminator	Traditional
StyleGAN	Noise vector and algorithmic latent code-based style information	Progressively growing generator and discriminator	Has adaptive instance normalization for image style adjustment

Note: * – GAN network has only one generator and one discriminator. ** – GAN network has only one stage for image generation process.

A summary of the unique features, benefits and limitations of these architectures is presented in Table 2.

Table 2. Comparative analysis of Generative Adversarial Network unique features, advantages and limitations

	Unique features	Advantages	Limitations
DCGAN	Traditional GAN architecture based on deep convolution and serves as the foundation for the subsequent architectures.	Demonstrates stable convergence and produces coherent images.	Suitable only for low-resolution color image applications and has limited control over the attributes of generated images.
CGAN	Continuation of the DCGAN, incorporating additional feature (class information).	Enables precise control over the generated image content by conditioning on auxiliary information such as class labels or attributes.	Requires labeled data and additional conditioning variables, increasing complexity and computational requirements for training.

End of Table 2

	Unique features	Advantages	Limitations
InfoGAN	Continuation of the CGAN, incorporating one or more additional features (control variables) and mutual information from an auxiliary model.	Learns interpretable latent representations through unsupervised training, capturing key image features without labeled data.	Requires labeled datasets for training. May require careful tuning of hyperparameters for optimal performance. Limited interpretability of the learned latent codes.
StackGAN	Has a hierarchical stack of conditional GAN models.	Produces high-quality images with fine-grained details from textual descriptions.	Can result in training instability and nonsensical results. Two-stage process may increase computational requirements.
Pix2Pix	U-Net or ResNet generator architecture.	Effective for image-to-image translation tasks. Provides a direct mapping between input and output domains, resulting in precise image translations.	Requires paired training data for effective training. Sensitive to variations in input data and may struggle with diverse transformations.
CycleGAN	Utilizes cycle consistency loss to enforce unpaired image-to-image translation between two domains without requiring corresponding pairs during training.	Removes the need for paired training data, allowing for more flexible and diverse image translation tasks.	May not guarantee one-to-one mapping between input and output domains. Performance heavily depends on the quality and diversity of the training datasets, leading to potential mode collapse or suboptimal translations.
Progressive GAN	The training process is divided into multiple stages of different resolutions. Has a minibatch standard deviation layer. Employs pixel-wise feature vector normalization.	Produces high-quality images with fine details and textures, achieving state-of-the-art results in image synthesis.	Requires significant computational resources and prolonged training times due to the incremental growth of network complexity and resolution.
StyleGAN	Style-based modification of the generator. Includes mapping network to transform input latent code into intermediate style vectors. Employs stochastic noise injection at various stages of the generator.	Enables precise manipulation of image attributes and generation of diverse and high-quality images with realistic details.	Complex architecture and training procedure, demanding substantial computational resources and expertise in hyperparameter tuning.

4. Experimental comparison of Generative Adversarial Networks results

To address the issue of limited datasets, three generative adversarial network architectures were investigated: Conditional GAN (CGAN), Deep Convolutional GAN (DCGAN), and StyleGAN2-ADA. The first two models were chosen to explore the capability of fundamental GAN models in generating representative images of cable faults. These networks use less computational resources compared to more advanced ones. Other widely used models mentioned in the scientific literature review, such as StackGAN or Pix2Pix, designed for image generation from text description or image-to-image translation tasks, were considered unsuitable for addressing the specific problem of the cable dataset. Upon reviewing the latest innovative GAN architectures, StyleGAN2-ADA was selected for investigation due to its distinctive features and capabilities. This architecture is designed for generating high-resolution synthetic images with a minimum amount of data, addressing the challenges posed by the limited cable dataset.

4.1. Limited dataset

A limited dataset of industrial images was selected to conduct experimental research, specifically images depicting cable defects. The selection of this dataset was based on its insufficient size and the type of images, which allowed the dataset to be evaluated without the need for professional expertise. The MVTEC AD Dataset includes 5069 high-resolution images of objects from different classes. The dataset images are categorized into 15 different classes, each containing a specific number of images with and without defects. All images are of size 1024×1024 pixels. The available classes in the dataset include bottles, cables, pills, carpets, grids, nuts, leather, metal, capsules, screws, tiles, toothbrushes, transistors, wood, and zippers.

Among all possible classes, the cable class was chosen because of the greater variety of object defects within this class. The cable subset of the dataset consists of 316 colorful images, with 92 containing various defects and 224 without defects. The original images are illustrated in Figure 10, with images of cables without defects at the top and those with various damages at the bottom. Also, it is crucial to note that arrangement of cable strands is significant: a green strand must always be at the top, followed by blue and gray strands at the bottom (from left to right). Otherwise, the image belongs to the defect class. The dataset was partitioned into training and testing subsets, the latter consisting of 15% of the images from the original dataset.

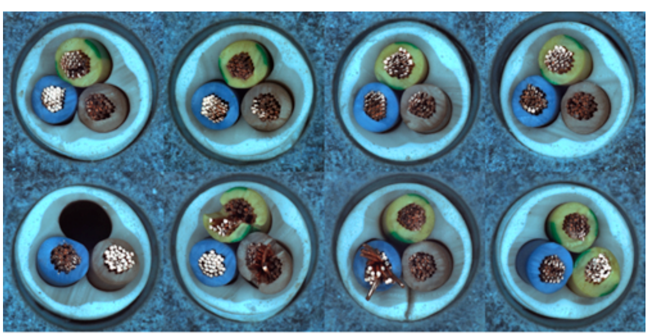

Figure 10. The MVTEC AD Dataset cable class images

4.2. Metrics for evaluating synthetic images

Two metrics are widely used to evaluate GAN-generated images: Fréchet Inception Distance (FID) and Kernel Inception Distance (KID). The first is based on the distance between the distributions of feature vectors obtained from real and synthetic images using the trained Inception-v3 network. The second metric is based on the maximum mean deviation between the feature vectors of the real and synthetic images. These vectors are determined similarly to FID, but additionally apply the gram activation matrix at each network layer. (Karras et al., 2020) conducted an additional study with small datasets and found that the KID metric is a more suitable measure for assessing GAN results with smaller datasets. This is because, unlike FID, the KID metric is independent of selection bias and better accounts for the distribution of image data. Due to this reason, only KID metric was calculated during the experiment.

4.3. CGAN and DCGAN networks experimental results

In the initial experiment, the CGAN and DCGAN models were investigated to expand the dataset. These models were chosen to explore whether traditional GAN models suffice for limited dataset expansion. KID metric was calculated for evaluation of the quality of synthetic images after training, as their calculation during training process greatly extends the network's training duration. The decision when to stop training was based on visual inspection of the generated images at intervals of 20 epochs. The DCGAN and CGAN architectures are designed to process relatively small image input, with the convolutional layer filters optimized to extract meaningful features from small image tensors. Typically, the image sizes processed by DCGAN and CGAN are 32×32, 64×64, and 128×128 pixels. However, these networks can be adapted to generate images with larger resolutions by adding more convolutional layers in both the discriminator and generator. In the research, images of size 256×256 pixels were used as inputs and outputs for both networks. The chosen image resolution was found to be suitable for a thorough examination of details in the generated synthetic images, ensuring that the images are large enough for an effective evaluation by the human eye. An increase in network layers also required more computational resources.

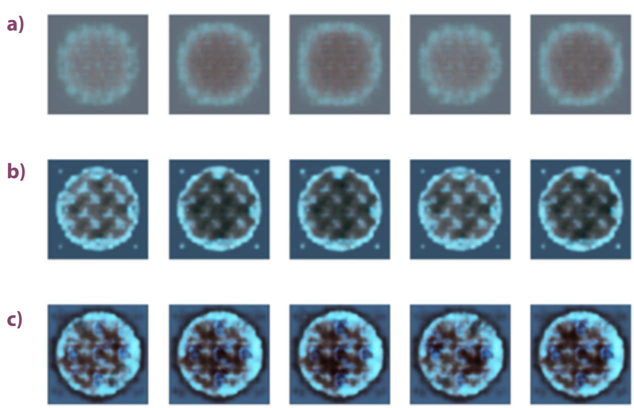

Figure 11. CGAN model generated synthetic images: a – 500 epochs; b – 800 epochs; c) – 1400 epochs

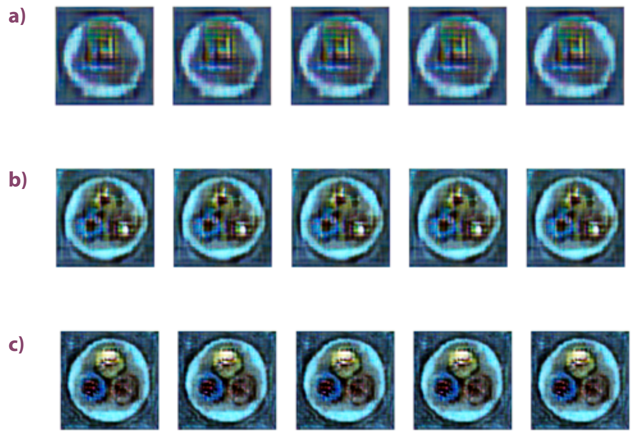

Figure 12. DCGAN model generated synthetic images:
a – 500 epochs; b – 800 epochs; c – 1400 epochs

The models were trained from scratch for approximately 6-8 hours. Images generated by the CGAN model at 500, 800, and 1400 epochs, with a fixed learning rate of $1{\cdot}10^{-5}$ are illustrated in Figure 11. Similarly, images generated by the DCGAN model using the same parameters are shown in Figure 12.

The experiment with both models was carried out by iteratively adjusting the learning rate from $1{\cdot}10^{-4}$ to $1{\cdot}10^{-6}$, however no significant results were obtained, and the quality of generated images remained similar. After evaluating KID metric (Table 3), it is evident that 1400 epochs are insufficient for both models to generate images of adequate quality, and the improvement in quality stops with continued training. It is hypothesized that such results are achieved due to limited dataset as several studies have found the CGAN and DCGAN models to excel with large-scale datasets such as ImageNet (Chakraborty et al., 2024). In such cases, the application of transfer learning technique, specifically fine-tuning, could be considered. However, since CGAN and DCGAN networks are not capable of generating high-resolution and high-quality images, the transfer learning technique was not further explored for these networks. It is crucial to ensure that various cable faults are clearly visible in the generated images. Therefore, it was concluded that the CGAN and DGAN models are not suitable for a limited dataset expansion.

Table 3. KID Metric Evaluation for CGAN and DCGAN

		500 epochs	800 epochs	1400 epochs
CGAN	good class	1.207	0.792	0.684
	defect class	0.996	0.964	0.628
DCGAN	good class	1.275	1.146	1.081
	defect class	0.922	1.033	0.813

4.4. StyleGAN2-ADA networks experimental results

During the second experiment, the investigation focused on StyleGAN2-ADA, which utilizes PyTorch framework. This model incorporates advanced architectural solutions, resulting in enhanced image quality, increased control over generated images, and improved network stability during training. The authors of StyleGAN2-ADA publicly released the implementation of their method on the Github platform, and this open-source software has been employed in subsequent experiments (Karras et al., 2020).

During the experiment, the decision was made to adapt the StyleGAN2-ADA model to inputs of 512×512 image resolution. This decision was made because:

- The 512×512 resolution allows for a robust visual assessment of image quality.
- Higher-resolution images convey a greater amount of useful information.
- The StyleGAN2 model is designed to generate high-resolution synthetic images, and processing at 512×512 input did not require high computational resources (within the Colab environment).

Under this condition, the data preparation and processing stage was performed. The implementation code of StyleGAN2-ADA provides a script tailored to automatically process the desired dataset and save it in a suitable format. Consequently, the prepared dataset is optimized to facilitate an efficient network training process.

StyleGAN2-ADA was chosen to employ transfer learning due to several key reasons:

- Training this model from scratch can take several days or even months. Transfer learning allows to lessen the amount of needed computer resources and achieve desired results faster.
- A pre-trained network has already learned to appropriately discern useful image features, enhancing model performance when trained on a limited-sized dataset. This proves particularly beneficial in refining the model's performance with entirely new data and reducing overfitting.
- Images from the trained network and the available dataset may be related, albeit visually different. In such cases, utilizing transfer learning allows the network to generalize better and learn more efficiently on a new dataset.

The authors of the StyleGAN2-ADA model trained the network on several large datasets and made their models weights publicly available. Among the options provided, the network trained on 512×512 resolution animal face images were selected (AFHQWild).

The selected network was trained to 512×512 animal face images (AFHQWild). It was observed that these images bear characteristics similar to the chosen cable dataset images, therefore, influencing the selection of network weights.

The network was trained with a "kimg" parameter set to 300, indicating the number of thousands of real images presented to the discriminator during network training. This parameter is frequently used in GAN models to track the network's learning progress and provide a reference point for different training stages. It aids in analyzing the model's evolution throughout the learning process, especially when comparing various experiments or adjusting training parameters. Using a single Nvidia Tesla V100 GPU, the network training took approximately 5–8 hours. The duration of model training depends on factors such as the size of the dataset, image resolution, GPU quantity, the desired image quality for generation, and the selection of hyperparameters.

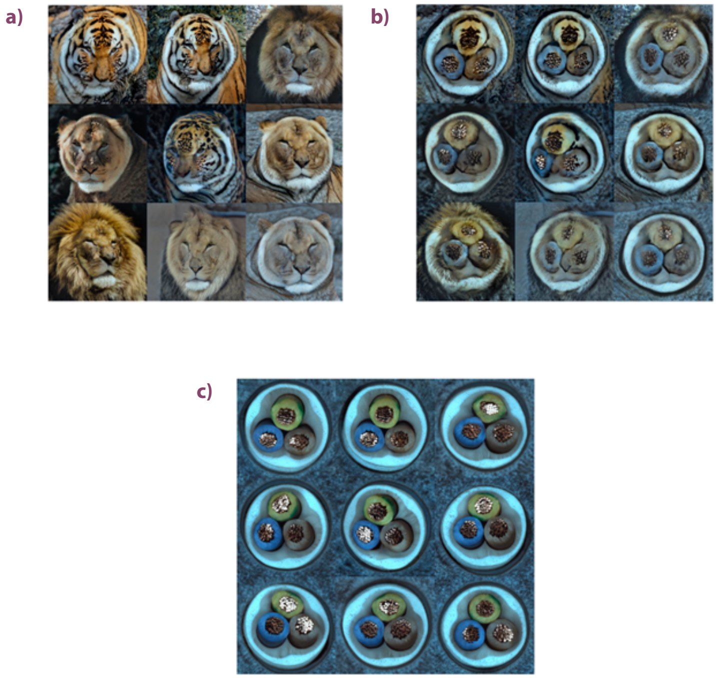

Figure 13. StyleGAN2-ADA samples during training: a – 16 kimg; b – 40 kimg; c – 136 kimg

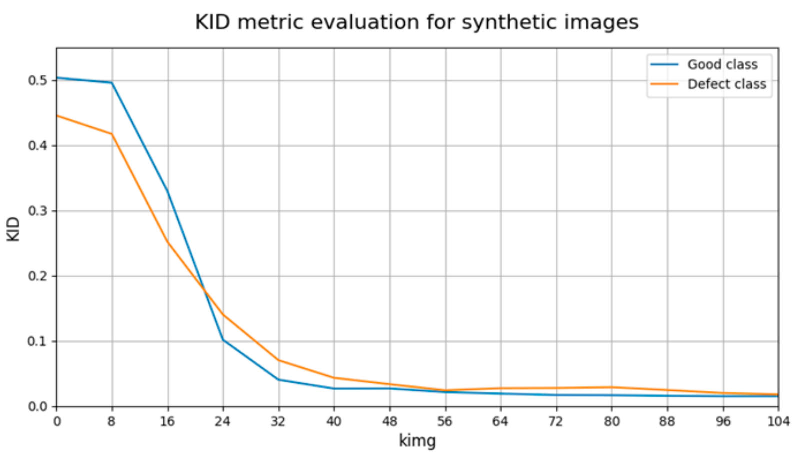

Figure 14. KID metric evaluation for synthetic images

During StyleGAN2-ADA training, examples of generated images with and without defects are depicted in Figure 13. The KID metric was not evaluated during the training process, as their calculation almost doubles the time of the training process. The metrics were evaluated separately after network training (Figure 14). From the obtained results, it is evident that employing transfer learning and a reduced volume of data prompts a faster convergence of the network. Nearing 100 kimg, the network generates images of sufficiently high quality, which is represented with low KID value of 0.02. The accelerated convergence can be attributed to the similarity in features present within the photographs of cables and animal faces. For these reasons, it was decided to stop the training at 104 kimg. If training continued, there was a risk that the synthetic images produced by the network would be identical to the images used for training.

5. Conclusions

The research aimed to address the challenge of limited dataset availability when training machine learning algorithms. In particular, it focused on the use of generative adversarial networks (GANs) to generate synthetic images of sufficient visual quality as a solution to this dataset problem. The study delved into the architectural advances of GANs, including models such as Deep Convolutional GAN, Conditional GAN, InfoGAN, StackGAN, Pix2Pix, CycleGAN, Progressive GAN, and StyleGAN. Each model showcased unique features and capabilities in generating synthetic images; however, not all methods were applicable in augmentation of limited datasets for CNN training.

Experimental comparisons were conducted using three GAN architectures: Conditional GAN (CGAN), Deep Convolutional GAN (DCGAN), and StyleGAN2-ADA. The investigation involved a limited dataset of cable defects and the results indicated that traditional GAN models such as CGAN and DCGAN were not suitable to effectively expand the dataset due to their limitations in generating high-resolution and high-quality images. Further research focused on the StyleGAN2-ADA model, leveraging transfer learning from a pre-trained network on animal face images. The experiments demonstrated that this approach led to faster convergence of the network, producing synthetic images of sufficiently high quality even with a limited amount of data. For generated image quality evaluation KID metric were calculated.

The findings contribute to ongoing efforts to improve the efficiency and applicability of artificial intelligence in diverse domains by providing a viable solution to the challenge of limited training data for image classification tasks. Future work could explore the use of GAN generated synthetic images to expand the training dataset of machine learning algorithms like Convolutional Neural Networks.

Author contributions

All authors contributed to this work equally.

Disclosure statement

The authors declare that they have no competing financial, professional, or personal interests that could influence the conduct or reporting of this research.

References

Ahmed, S. F., Bin Alam, M. S., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M., & Gandomi, A. H. (2023). Deep learning modelling techniques: Current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, *56*(11), 13521–13617. https://doi.org/10.1007/S10462-023-10466-8

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, *8*(1), Article 53. https://doi.org/10.1186/s40537-021-00444-8

Borji, A. (2019). Pros and cons of GAN evaluation measures. *Computer Vision and Image Understanding*, *179*, 41–65. https://doi.org/10.1016/j.cviu.2018.10.009

Chakraborty, T., Reddy, U. K. S., Naik, S. M., Panja, M., & Manvitha, B. (2024). Ten years of generative adversarial nets (GANs): A survey of the state-of-the-art. *Machine Learning: Science and Technology*, *5*(1), Article 011001. https://doi.org/10.1088/2632-2153/ad1f77

Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). *InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets*. arXiv. https://doi.org/10.48550/arXiv.1606.03657

Chen, Y., Yang, X.-H., Wei, Z., Heidari, A. A., Zheng, N., Li, Z., Chen, H., Hu, H., Zhou, Q., & Guan, Q. (2022). Generative adversarial networks in medical image augmentation: A review. *Computers in Biology and Medicine*, *144*, Article 105382. https://doi.org/10.1016/j.compbiomed.2022.105382

Dash, A., Ye, J., & Wang, G. (2023). A review of Generative Adversarial Networks (GANs) and its applications in a wide variety of disciplines: From medical to remote sensing. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3346273

Feng, Z., Daković, M., Ji, H., Zhou, X., Zhu, M., Cui, X., & Stanković, L. (2023). Interpretation of latent codes in InfoGAN with SAR images. *Remote Sensing*, *15*(5), Article 1254. https://doi.org/10.3390/rs15051254

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144. https://doi.org/10.1145/3422622

Henry, J., Natalie, T., & Madsen, D. (2021). *Pix2Pix GAN for image-to-image translation*. ResearchGate. https://doi.org/10.13140/RG.2.2.32286.66887

Iglesias, G., Talavera, E., & Díaz-Álvarez, A. (2023). A survey on GANs for computer vision: Recent research, analysis and taxonomy. *Computer Science Review*, *48*, Article 100553. https://doi.org/10.1016/j.cosrev.2023.100553

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). *Progressive growing of GANs for improved quality, stability, and variation*. ArXiv. https://doi.org/10.48550/arXiv.1710.10196

Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., & Aila, T. (2020). Training generative adversarial networks with limited data. *Advances in Neural Information Processing Systems*, *33*, 12104–12114.

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 4401–4410). IEEE. https://doi.org/10.1109/CVPR.2019.00453

Mert, A. (2023). Enhanced dataset synthesis using conditional generative adversarial networks. *Biomedical Engineering Letters*, *13*(1), 41–48. https://doi.org/10.1007/s13534-022-00251-x

Mirza, M., & Osindero, S. (2014). *Conditional generative adversarial nets*. ArXiv. https://doi.org/10.48550/arXiv.1411.1784

Pérez, E., & Ventura, S. (2023). Progressive growing of Generative Adversarial Networks for improving data augmentation and skin cancer diagnosis. *Artificial Intelligence in Medicine*, *141*, Article 102556. https://doi.org/10.1016/j.artmed.2023.102556

Radford, A., Metz, L., & Chintala, S. (2015). *Unsupervised representation learning with deep convolutional generative adversarial networks*. ArXiv. https://doi.org/10.48550/arXiv.1511.06434

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training GANs. *Advances in Neural Information Processing Systems*, 29.

Saxena, D., & Cao, J. (2021). Generative adversarial networks (GANs) challenges, solutions, and future directions. *ACM Computing Surveys* (*CSUR*), *54*(3), 1–42. https://doi.org/10.1145/3446374

Son, D. M., Kwon, H. J., & Lee, S. H. (2023). Enhanced night-to-day image conversion using CycleGAN-based base-detail paired training. *Mathematics*, *11*(14), Article 3102. https://doi.org/10.3390/math11143102

Thamotharan, B., Sriram, A. L., & Sundaravadivazhagan, B. (2023). A comparative study of GANs (Text to Image GANs). In C. Iwendi, Z. Boulouard, & N. Kryvinska (Eds.), *Lecture notes in networks and systems: Vol. 735. Proceedings of ICACTCE'23 – The International Conference on Advances in Communication Technology and Computer Engineering. ICACTCE 2023* (pp. 229–241). Springer. https://doi.org/10.1007/978-3-031-37164-6_16

Wang, Z., She, Q., & Ward, T. E. (2021). Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys* (*CSUR*), *54*(2), 1–38. https://doi.org/10.1145/3439723

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. N. (2017). Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 5907–5915). IEEE. https://doi.org/10.1109/ICCV.2017.629

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223–2232). IEEE. https://doi.org/10.1109/ICCV.2017.244

APPENDIX

Notations

Abbreviations

CNN – Convolutional Neural Network.

GAN – Generative Adversarial Network.

ADA – Adaptive Discriminator Augmentation.

GPU – Graphics Processing Unit.

FID – Fréchet Inception Distance.

KID – Kernel Inception Distance.