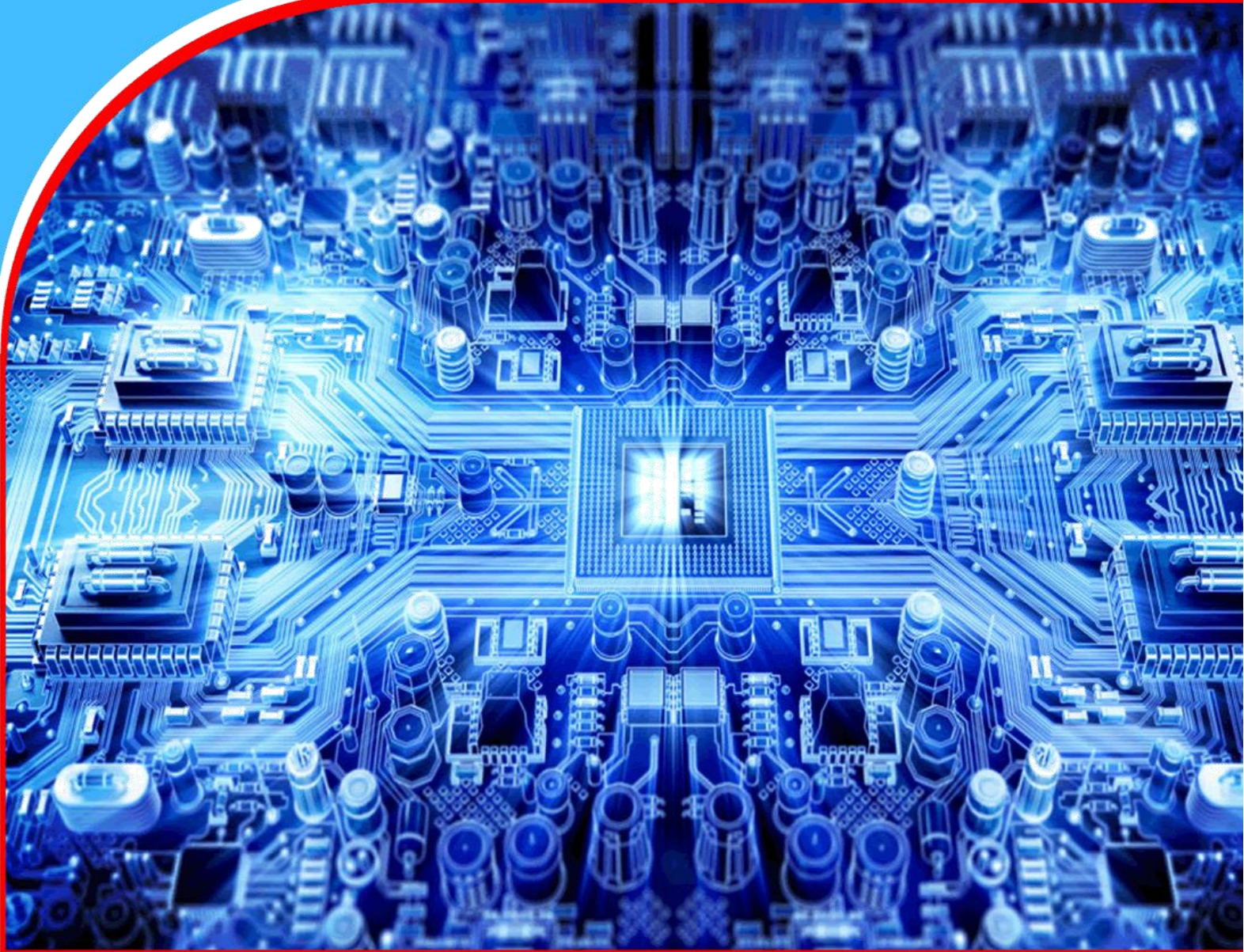


American Journal of Computing and Engineering (AJCE)




A State-of-the-Art Review on Image Synthesis with Generative Adversarial Networks

*Sai Mani Krishna Sistla, Suhas Jangoan & Ikram Ahamed
Mohamed*



A State-of-the-Art Review on Image Synthesis with Generative Adversarial Networks

 Sai Mani Krishna Sistla¹, Suhas Jangoan² & Ikram Ahamed Mohamed³
¹Soothsayer Analytics, USA
²Zendesk, USA
³Salesforce, USA



Article history

Submitted 08.01.2024 Revised Version Received 28.02.2024 Accepted 01.04.2024

Abstract

Purpose: The remarkable performance of Generative Adversarial Networks (GANs) in various applications has made them a popular subject in computer vision research, and they have also shown remarkable success in picture synthesis tasks.

Materials and Methods: Image processing, synthesis, generation, semantic editing, translation, super-resolution, inpainting, and cartoon creation are all areas covered in this article's presentation of the most recent GAN research. To demonstrate how they have improved the result, they analyze the methods used by these applications and describe them.

Findings: Insights into GAN research and a presentation of GAN-based applications in diverse contexts are the goals of this paper (Anon, 2022).

Implications to Theory, Practice and Policy: Following this, we will go over some of the difficulties encountered by GANs and provide solutions to these issues. We also discuss potential future areas of study for GANs, including video creation, 3D face reconstruction, and facial animation synthesis.

Keywords: *Generative Adversarial Networks, Image Synthesis, Image-To-Image Translation, Image Editing, Cartoon Generation*

1.0 INTRODUCTION

The term "artificial intelligence" (AI) has garnered a lot of attention in the media as well as on social media platforms. There has been tremendous advancement in image processing, particularly with the fast growth of deep learning. Researchers in the field of deep learning have recently taken an interest in generative models due to the massive volume of photographs used on social media. An increasing amount of research is focused on generative models because they show promise as unsupervised learning approaches that can effectively represent semantic information.

Out of all of them, the Variational Auto-encoder (VAE) fails to produce pictures with sufficient clarity. One model that has not seen much use yet is the Glow, which is based on flows. Impressive findings in image processing have generated increasing interest in Generative Adversarial Networks (GANs) within and outside of businesses and universities. Among the several current areas of study and applications using GANs are picture-to-image translation, art production, semantic segmentation, medical image processing, inpainting, and image creation.

In addition, GANs are widely used for face synthesis and editing, particularly for translating faces based on age and gender (Fan et al., 2020). When it comes to GANs, there are two current lines of inquiry:

- i. GANs in theory, with an emphasis on energy-based or information-theory models; this kind of study aims to address the open questions surrounding GAN training, including mode collapse, unstable training, and evaluation difficulty. In Section IX, we will quickly go over this issue area and the difficulties with GANs.
- ii. How GANs are used for different kinds of computer vision jobs. Researchers have shown that multiple GAN-variants increase GAN performance, even if there are still some open issues. For the most part, this paper is concerned with the second part of the ongoing GAN study.

Even though GANs have been the subject of a few studies so far, particularly in the realm of deep learning, GANs are advancing at a rapid pace. Research on GANs for picture synthesis in the last few years is the main topic of this article. It compares and analyzes different GAN-based applications based on their strengths and drawbacks. In addition, we review and summarize the approaches used by these apps to enhance the produced images.

We also go over some of the difficulties in training and evaluating GANs. Here are a few ways to train and evaluate GANs steadily. Video production, face animation synthesis, and three-dimensional face reconstruction are some of the probable areas for future study (Huang, Yu and Wang, 2018). Generative Adversarial Networks (GANs) train a discriminator and a generator neural network at the same time; Adversarial training is their basis. The role of the discriminator is to identify instances of synthetic data and their differences from real data, whereas the generator is responsible for creating samples of synthetic data that closely resemble real data. The discriminator provides the generator with input during training, which the generator utilizes to improve its ability to produce realistic samples. As a result, it is now better able to distinguish between real and fake data.

The generator's job is to give more compelling examples, whereas the discriminator's job is to accurately differentiate between real and fake data. A competitive environment is created by this dynamic interaction. Each node in a deep neural network contributes to the overall processing of

input data and the final classification decision. The generator takes in noise at random and produces data samples that are not real. Both the discriminator's and the generator's loss functions are optimized during GAN training. The discriminator's function seeks to minimize the classification error between real and synthetic data, while the generator's function maximizes the likelihood that its samples will be classified as genuine.

Mode collapse, where the generator produces a limited number of sample types, is one issue that hinders GANs' otherwise remarkable performance. We employ a number of methods, such as spectral normalization and minibatch discrimination, to make sure that the samples we produce are more diverse and reliable, and to stabilize the training process overall. By delving deeper into these core ideas, readers will be better equipped to understand GANs and all the various applications they have, ranging from data augmentation to image synthesis.

Body Generative Adversarial Networks

Image processing's immense potential contributes to the exceptional performance of GANs in image-related tasks. Their use is crucial in many fields, and they are often believed to be the best approach to picture production.

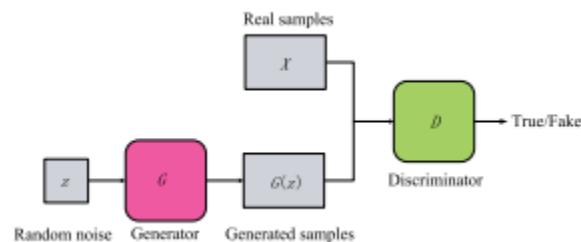


Figure 1: Generative Adversarial

Generator G's job is to trick discriminator D into thinking the samples it generates are genuine when, in fact, they are only random noise. Using the discriminator D, we can verify that the sample we generated using generator G is genuine. The discriminator and generator are always at war over who can fool it into believing an image is real or not. The objective of training GANs is to achieve the Nash equilibrium, if we see it as a two-player minimax game. To calculate the GAN loss function, use this formula (Mao et al., 2019):

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log (1 - D(G(z)))] \quad (1)$$

Where $P_{data}(x)$ denotes the true data distribution, $P_z(z)$ denote the noise distribution. Due to the special network structure and the generation performance of GANs. Extensive research has produced numerous applications based on GANs. as shown in Figure 2.

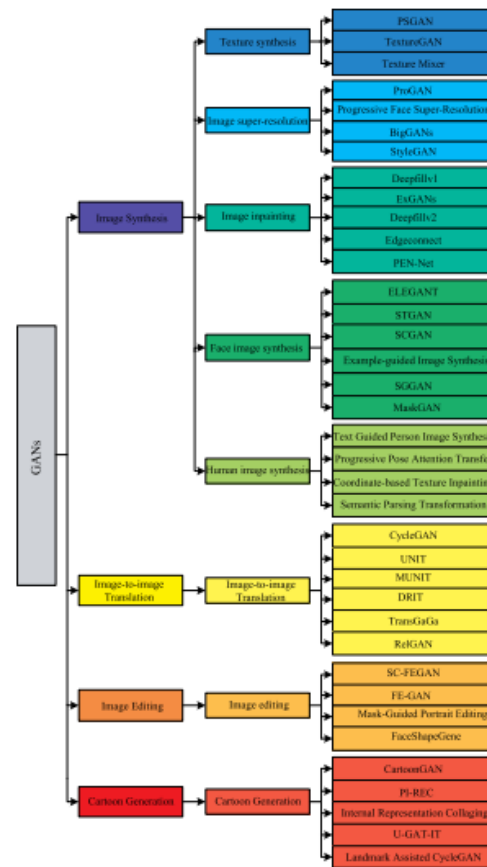


Figure 2: Taxonomy of GANS

Image Synthesis

Image synthesis has piqued many people's curiosity because of its widespread use in social media. In the field of image synthesis, GANs such as GauGAN have shown considerable promise. So far, several methods for image synthesis have been created.

A Synthesis of Textures

Both coarse-grained and fine-grained texture synthesis fall under the umbrella of image synthesis. In contrast to coarse-grained texture synthesis, which looks for similarities between input and output pictures, fine-grained texture synthesis attempts to find out if the synthetic texture is similar to the ground truth (Pierrick Bourgeat, 2022).

The PSGAN

A novel approach to texture creation, termed Periodic Spatial GAN (PSGAN) by Bergmann et al., is based on the Generative Adversarial Network. Pictured in Figure 3 is the PSGAN model.

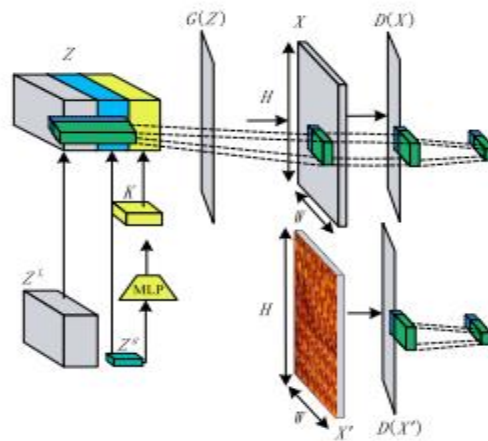


Figure 3: The PSGAN

The loss function of the PSGAN is defined as:

$$\begin{aligned} \min_G \max_D V(D, G) &= \frac{1}{LM} \sum_{\lambda=1}^L \sum_{\mu=1}^M E_{Z \sim P_Z(Z)} [\log(1 - D_{\lambda\mu}(G(Z)))] \\ &+ \frac{1}{LM} \sum_{\lambda=1}^L \sum_{\mu=1}^M E_{X' \sim P_{data}(X')} [\log D_{\lambda\mu}(X')] \end{aligned}$$

From complex information, including large photos, PSGAN can learn several textures. The method can do more than only learn periodic textures correctly; it can also generate new samples that seem to be between the textures in the original dataset and construct smooth interpolations between them in a structured noise space. Picture data sources and textures are no problem for PSGAN. This technique can create images of any size since it is incredibly scalable (Qiao et al., 2021).

Texture GAN

Xian et al. presented TextureGAN, a technique for synthesizing textures that integrates sketch, color, and texture to create realistically rendered pictures. You can see the training process in Figures 4 and 5.

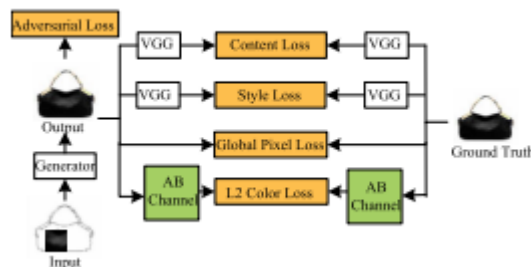


Figure 4: Texture GAN Pipeline

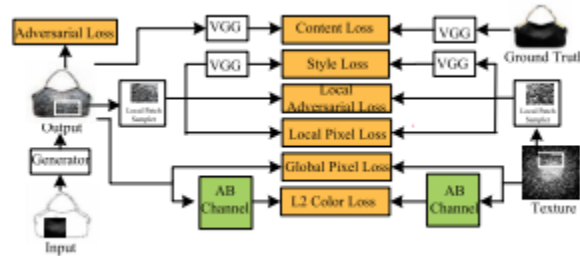


Figure 5: External Texture

One way to manipulate the texture of synthetic pictures is using TextureGAN, a technique for image synthesis. Users may modify the output texture by placing a texture patch anywhere on the drawing and adjusting its size. In addition to doing well in both sketch and texture-based picture synthesis, it can handle a wide variety of texture inputs and produce texture compositions that adhere to sketch outlines (Shamsolmoali et al., 2021).

The objective function of ground-truth pre-training defined as:

$$L = L_F + W_{ADV}L_{ADV} + W_S L_S + W_P L_P + W_C L_C \quad (2)$$

The objective function of external texture fine-tuning defined as:

$$L = L_F + W_{ADV}L_{ADV} + W_P L_P + W_C L_C + L_t \quad (3)$$

Texture Mixer

The technique integrates two distinct texture patterns into a seamless whole utilizing deep learning and GAN-enabled controlled interpolation. It suggests using a neural network that has been trained for both reconstruction and generation to project the sample's texture into a latent space. Then, to guarantee both realistic produced results and intuitive control, it projects linear interpolation onto the picture domain. It also performs well in texture generation across the board, including controllability, smoothness, and realism, and it outperforms a lot of baseline approaches (Anon, 2022).

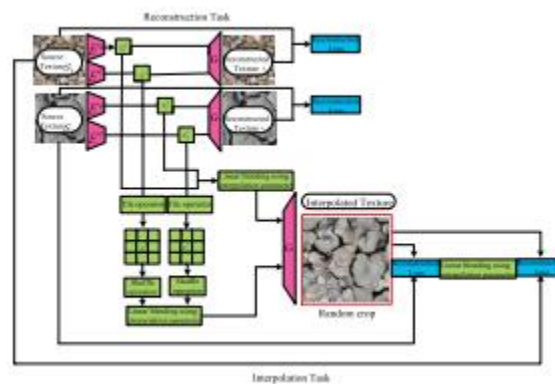


Figure 6: Texture Mixture

Other Methods

Markovian Generative Adversarial Networks (MGANs) are an effective approach for texture creation that Li and Wand presented. suggested, that spatial GAN (SGAN), works well for

synthesizing textures. A high-quality synthesis of texture pictures and the fusion of several source photos into complex textures are both possible with this technology.

By using interpolation, GAN-based texture synthesis may achieve a natural transition in texture creation while still producing realistic details. To make GANs adhere to certain rules, one might use interpolation or extrapolation. Incorporating constraints during GAN training and then using projection on the constraints space for extrapolation to enforce them after each step is a clever approach. Nevertheless, texture synthesis models might experience "mode dropping" and have trouble converging during training on occasion (Fan et al., 2020).

Image Super-Resolution

Even though it has always been difficult to produce big, high-quality pictures, the goal of the image-making model is to investigate techniques to create a desired picture. The capacity of GANs to generate visually pleasing and high-quality pictures is a significant benefit, and they have achieved remarkable progress in this area. A growing number of GAN-based models are emerging to generate pictures with more detail (Huang, Yu and Wang, 2018).

Pro GAN

The ProGAN technique for generating images was proposed by Karras et al. In Fig. 7, we can see the ProGAN architecture. This method involves beginning with a low resolution and gradually increasing the generator and discriminator by adding layers as training progresses. This creates the model and develops fine features. As a result, the training process may be both accelerated and greatly stabilized. The approach usually produces good-quality results when compared to previous GAN work, and the training is stable at high resolution. Having said that, this approach is not without its flaws. Consider how the limitations of the dataset affect semantic awareness and comprehension (Mao et al., 2019).

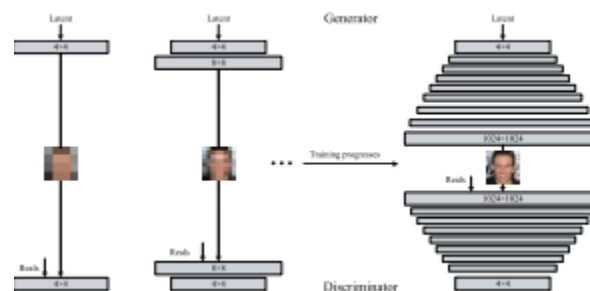


Figure 7: The Pro GAN Architecture

Progressive Face Super-Resolution

A new face super-resolution (SR) approach was suggested by Kim et al., and it can produce photorealistic face pictures with all the features preserved. Figure 8 depicts the network architecture. To achieve stable training, the authors use a progressive training method that divides the network into stages, with each step generating outputs of increasingly better resolution. The restoration of face characteristics is given more attention at each stage by using a revolutionary facial attention loss that multiplies pixel differences and heatmap values to restore them in more detail. To reduce training time and provide appropriate landmark heatmaps for face super-resolution (SR), they also suggested a condensed version of the face alignment network (FAN).

In addition, it can mimic human faces with astonishing realism by learning to restore facial characteristics and producing super-resolution photographs of the face. From a quantitative and qualitative perspective, the findings outperform the previous methodologies, particularly when it comes to perceptual quality (Pierrick Bourgeat, 2022).

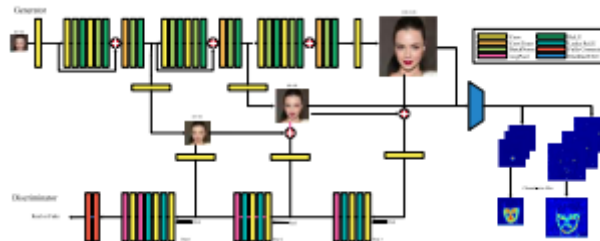


Figure 8: Network Architecture

Big GANs

The task of producing varied and high-resolution pictures from complicated datasets was made possible by models introduced by Brock et al., known as BigGANs. In Figure 9, we can see an example of a BigGAN network design.

Several high-resolution examples were extracted from the complicated dataset ImageNet using this approach. It can produce pictures of unparalleled quality and is the most comprehensive Generative Adversarial Network that has been taught so far. In terms of the final picture, it produces far more realistic results than previous approaches. To deal with the unique instability of such a scale, the authors minimized the latent space and applied orthogonal regularization to the generator; this allowed them to manage the diversity and quality of the produced pictures (Qiao et al., 2021).

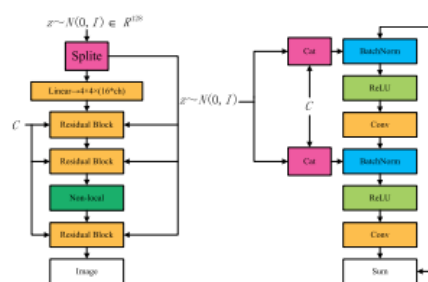


Figure 9: Big GANs

Image Inpainting

Deep learning technology has made significant advancements in the field of image painting throughout the past few years. Restoring and recreating pictures using background information is called image inpainting. It is anticipated that the produced pictures would seem very realistic and hard to differentiate from the actual reality. For picture inpainting to be of high quality, it is necessary for both the created content's semantics and its texture to be sensible. There has been encouraging progress in recent times towards using deep learning techniques, particularly GAN-based algorithms, for picture inpainting (Shamsolmoali et al., 2021).

Deepfillv1

Deepfillv1 incorporates deep learning algorithm solutions when it comes to the advantages of traditional methods. Improved picture quality is the result of further network enhancements that allow it to automatically repair images with numerous or large holes. The method enhances prediction accuracy and can potentially generate new picture structures using contextual picture data.

The authors experimented by interpreting multi-hole pictures using a feedforward and fully convolutional neural network. If this system can learn feature representations for explicit matching and focus on broad patches of background, it might potentially achieve better outcomes when inpainting images. Included in this framework for generative image inpainting, which moves from coarse to fine, is a unique contextual attention module (Anon, 2022).

Explicit GANs

When a person closes their eyes in a natural environment, they may be "painted" with high-quality, personalized results based on example data that represents the place. In the closed-to-open eye challenge, it may also characterize the item using a perceptual code and generate a customized, photorealistic picture based on perception and semantics. Conditional GANs, of which ExGANs are a subset, may boost the descriptive power of adversarial networks by injecting additional data at various places across the network. With reference pictures or perceptual codes as identifying information, it is a helpful approach for image production or inpainting that produces improved perceptual outcomes (Fan et al., 2020).

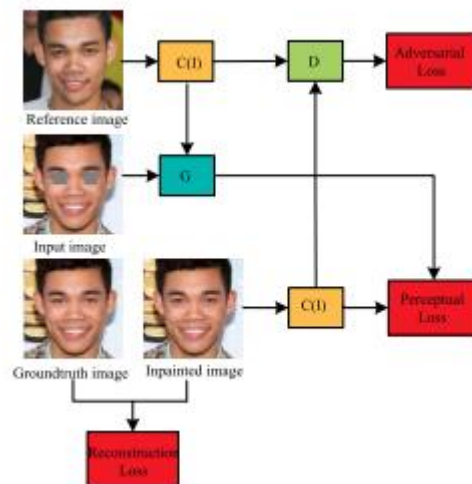


Figure 10: Explicit GANs

The Pen-Net

When part of a picture is missing, the PEN-Net may fill it up with plausible material using a technique that was suggested for high-quality image inpainting. Following a high-level semantic feature map and then applying the learned attention to a lower-level feature map, the authors suggest a pyramid-context encoder to progressively train the network area affinity. It has the potential to fix damaged photos by filling in the afflicted regions, leading to visually and

semantically convincing solutions. Another important part of the process is learning contextual semantics from full-resolution data and then using that knowledge to decode and reconstruct images.

By systematically moving from the most detailed to the most basic levels of a pyramid, we can guarantee that the produced material is visually and semantically consistent. Concurrently, a novel loss function was suggested by the authors to expedite training and provide more accurate results. When compared to the old technique, the network's ability to produce visually realistic and semantically sound pictures is far greater.

When compared to their more conventional predecessors, recent GAN-based image inpainting approaches provide results that are both realistic and semantically coherent. At present, some approaches use gated convolutions to recover pictures using free-form masks. These masks may either fill in gaps in images with non-standard forms or restore images with many holes. Nevertheless, the placement and magnitude of the masks impact the painting's quality (Huang, Yu and Wang, 2018).

Image-To-Image Translation

Recently, there has been significant advancement in the translation process from one image format to another. The objective of picture translation is to maintain the content of the original image while transforming its style or other attributes into another domain. Accomplishing this requires being an expert in a mapping that links several domains (Mao et al., 2019).

Image-To-Image Translation

Researchers in supervised and unsupervised learning have taken an interest in the use of generative adversarial networks for image-to-image translation. In contrast to image-to-image GANs, which can produce a wide variety of visuals from pictures, noise-to-image GANs can produce realistic images from samples of random noise. When tested on image-to-image translation tasks, many GAN variations demonstrated encouraging outcomes.

Cycle GAN

As a new approach, Cycle GAN is revolutionizing the area of unsupervised picture translation. Unsupervised photo translation using Cycle GAN has been the subject of several research projects. According to their proposal, cycle consistency loss can learn the mapping even in the absence of a training set of matched picture pairings. This approach works well for a wide variety of translation jobs that include changing colors and textures, transferring collection styles, transforming objects, and transferring seasons. On the other hand, when geometric adjustments are necessary, it falls short (Pierrick Bourgeat, 2022).

Unit

The technique relies on variational autoencoders and generative adversarial networks. It enforces a shared latent space by interacting with an adversarial training goal and a weight-sharing constraint, and it creates matching pictures in two domains. Additionally, it uses variational autoencoders to establish a relationship between the input pictures and the translated images in each domain. Among the many unsupervised image translation tasks that may be executed using the UNIT technique are the presentation of high-quality image translation results and the execution of tasks involving faces or street scenes. Having said that, this approach does have two drawbacks.

First, the translation model is unimodal since it assumes a Gaussian latent space. Conversely, training instability might be caused by the saddle point seeking difficulty (Shamsolmoali et al., 2021).

Munit

The source domains used by MUNIT are multimodal conditional distributions, which allow it to provide various outcomes. To create multimodal graphics, it trains two auto-encoders: one for the image's content and another for its style. By breaking down the picture into its parts, we may derive domain-invariant content code and domain-specific style code. To change the image's domain, the technique recombines the content code with a randomly selected style code from the destination domain's style space. Not only that, but the distribution of styles differs across the two domains, even when the content distribution is the same. For better style control, it could be helpful to compare the translated text with a large collection of high-quality photos (Anon, 2022).

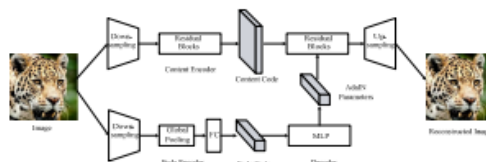


Figure 11: Munit

SC-FEGAN

For face editing, the SC-FEGAN technique requires a free-form mask, an image, and some color. This technique can repair huge parts with precise textures and mend regions of any form or size. Utilizing both human input and an end-to-end trainable convolutional network, it generates color- and shape-guided pictures. Instructing an extra style loss is another way the method might provide practical effects. By using the suggested network design and loss algorithms, it is capable of producing precise and high-quality outcomes (Fan et al., 2020).

GAN Fe-GAN

When FE-GAN edits fashion photos, it often uses drawings and color strokes. By combining free-form painting with sparse color strokes, it modifies fashion photos utilizing semantic structural information. Using a color-based, free-form parsing network, this technique guides the development of visual parsing that is eerily similar to human performance. Using a parsing-aware inpainting network inspired by the human parsing map, it generates semantically driven, very detailed textures. Finally, the inpainting network's decoder enhances the output pictures using a new attention normalization layer. Using a foreground-based partial convolutional encoder, the method can produce high-quality pictures with realistic details (Huang, Yu and Wang, 2018).

Fe-GAN

A fashion photo edited and modified with FE-GAN will have drawings and color strokes added to it. It uses information about semantic structures to add free-form doodling and sparse color strokes to fashion photos. The method's human-generated parsing is controlled by an unstructured parsing network that considers the drawing and color. Thanks to its parsing-aware inpainting network, it can follow the human parsing map to generate complicated textures. In addition, the inpainting

network's decoder incorporates a new attention normalization layer, which enhances the output pictures. Applying the approach to photos allows it to produce high-quality results with believable details, thanks to its foreground-based partial convolutional encoder (Mao et al., 2019).

Cartoon Generation

The engaging plot makes the show a hit with kids and teens. Researchers in the area of cartoon generation have also taken an interest in GANs, and they have suggested several novel and intriguing approaches to cartoon creation.

Cartoon GAN

Trained with both actual and cartoon pictures, Cartoon GAN is an easy-to-use generative adversarial network (GAN) approach to cartoon generation. The approach may aim to mimic the style of certain painters while using real-world photos as a foundation to generate high-quality cartoon drawings with smooth color shading and tidy borders. To address the apparent visual differences between cartoons and images, it suggests a semantic content loss, which is a sparse regularization of VGG network high-level feature maps. For Cartoon GAN to keep its evident edges, it proposes an adversarial loss that encourages edges. Adding an initiating step could also help the network get closer to the target manifold. Using this method, it is possible to transform high-resolution images of real-life items into cartoons (Pierrick Bourgeat, 2022).

U-GAT-IT

The unsupervised image-to-image translation method U-GAT-IT takes geometric changes between domains into account, allowing it to comprehend pictures with substantial or extensive form changes. The two domains, source, and target are separated by a single learnable normalization algorithm and an attention module. The auxiliary classifier builds an attention map to assist the attention module in focusing on the critical aspects of the model. Using the suggested Adaptive Layer-Instance Normalization (AdaLIN), the attention-guided model may additionally adjust the degree of shape and texture change on the fly. Additionally, it's a method to create more attractive anime faces by combining the attention module with AdaLIN. To generate cartoons using GANs, normalization is the primary method.

It can create high-quality cartoons from real-life images. However, it does sometimes create images that have artifacts. Furthermore, it may not provide an accurate representation of a particular person's look. Although networks are now the most used approach for generating pictures, there is currently no easy way to compare and assess the quality of images generated by GANs. Prior research on GAN-based picture creation relied heavily on subjective visual evaluations. Despite the difficulty in putting a numerical value on the quality of the produced pictures, research assessing GANs has started to surface.

Take, for instance, the most popular metrics for quantitatively assessing produced images: the Inception score (IS) and the Fréchet Inception Distance (FID). In addition, a method for scene-level visualization and understanding of GANs was suggested by Bau et al. To create a universally accepted human criterion for generative realism, Zhou et al. suggested Hype, or Human Eye Perceptual Evaluation. To keep tabs on the training process and identify failure modes such as unstable mode collapse and divergence, Grnarova et al. provide an assessment metric. Research on alternative assessment techniques is ongoing. Generative adversarial networks will continue to

see increased usage across a variety of domains as training and assessment techniques for GANs continue to advance and GANs themselves make significant strides (Mao et al., 2019).

Future

With the help of GANs, computer vision research has made great strides in recent years, and many new applications have emerged. The bulk of these applications pertain to image processing. Few studies have focused on video processing using GANs, despite the abundance of literature on video processing in general (e.g., video production, colorization, inpainting, motion transfer, and face animation synthesis). Furthermore, 3D colorization, 3D face reconstruction, 3D character animation, and 3D textured object production are just a few examples of the many applications where GANs have been used for 3D model synthesis and development, however, the outcomes have been less than satisfactory.

Even now, huge data sets are the foundation around which GANs are constructed. Less data use is going to be the norm in the future. Weakly supervised learning has been the subject of some study, but it is still in its infancy, and the outcomes are far from ideal. Among GANs' many promising applications in data augmentation is their capacity to produce photographs of excellent quality. This is particularly evident in data-poor domains like medical image analysis. The use of synthetic medical images produced by Frid-Adar et al. utilizing a GAN-based approach has the potential to enhance medical problem performance with minimum data. Han et al. proposed a GAN-based two-step data augmentation technique to decrease the number of annotated images required for medical imaging applications. Sandfort et al. created a GAN-based data augmentation method to improve the performance of medical imaging tasks.

Conditional Progressive GANs (CPGGANs) is a data augmentation approach introduced by Han et al. to lessen the burden on human medical specialists who annotate medical applications. In several biological datasets, Schlegl et al.'s Fast AnoGAN (f-AnoGAN) can identify aberrant images. A technique for data augmentation known as 3D Multi-Conditional GAN (MCGAN) was developed by Han et al. to tackle the issue of inadequate medical data. Very little research has focused on other equally important fields, such as those in the game and modular domains. Recent work by Lin et al. introduced Conditional Coordinate GAN (COCO-GAN), an innovative approach to picture component generation that achieves excellent generation quality by using spatial coordinates as the condition. For example, for huge field-of-view generation, this method may produce pictures that are bigger than any training sample. We conclude that these domains, in particular, have promise for further GAN study and implementation (Qiao et al., 2021).

Though Generative Adversarial Networks (GANs) show great potential in many domains, they are not without their limitations and challenges. Dataset bias is a major issue because it means that training data does not necessarily represent the distribution variety in the real world. The generation of synthetic data that fails to capture the complete spectrum of real-world variability could compromise the quality and diversity of the samples if this bias leads to its implementation. In addition, GANs may struggle to adjust to different domains, especially when asked to supply data for a domain that is significantly different from the training domain. The enormous amounts of labelled data and the painstaking fine-tuning of network designs that are usually needed to adapt GANs to new domains are some of the practical challenges that come up in real-world applications. Moreover, regardless of how visually appealing GAN-generated images may be, the problem of

their generalizability to unknown data remains. Created samples may have flaws that real data does not in critical applications where accuracy and reliability are paramount, such as autonomous driving or medical imaging. Efforts to address these challenges should focus on data augmentation strategies, domain adaptation tactics, and evaluating the transferability and resilience of GAN-generated samples in different contexts. If researchers can identify and address these limitations, they can work towards fully utilizing GANs and ensuring their appropriate and successful deployment in real-world contexts.

2.0 CONCLUSION AND RECOMMENDATIONS

Within the scope of this article, they went over the fundamentals of GANs and detailed their uses in picture synthesis. Along with the GAN applications, we also provide you with the benefits and downsides of each. Along with that, we summed up the techniques employed in GAN applications that enhanced the resulting photos' performance. Despite the growing maturity of GAN research, GANs continue to confront some obstacles, such as unstable training and difficult evaluation, which is why we presented new approaches to their training and evaluation. Future studies might focus on video creation, 3D face reconstruction, and facial animation synthesis, among other possible areas. There are still many undiscovered uses for GANs, and as more GAN-variants are developed, their performance will only get better. More intriguing GAN-based applications will likely emerge in the future.

Similarly, significant but understudied approaches exist in the modular and game domains. Conditional Coordinate GAN (COCO-GAN) is a new approach that Lin et al. presented; it generates high-quality pictures by sections using spatial coordinates as the condition. For big field-of-view creation in particular, this method may produce pictures that surpass the size of any training sample. We draw the conclusion that these domains, in particular, have promise for further GAN study and implementation.

REFERENCES

- Anon (2022). *IEEE Xplore Full-Text PDF*: [online] ieeexplore.ieee.org. Available at: <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10372211> [Accessed 1 Mar. 2024].
- Fan, J., Liu, T., Li, G., Chen, J., Shen, Y. and Du, X. (2020). *Relational Data Synthesis using Generative Adversarial Networks: A Design Space Exploration*. [online] arXiv.org. doi:<https://doi.org/10.48550/arXiv.2008.12763>.
- Huang, H., Yu, P.S. and Wang, C. (2018). An Introduction to Image Synthesis with Generative Adversarial Nets. *arXiv (Cornell University)*, 07(6). doi:<https://doi.org/10.48550/arxiv.1803.04469>.
- Mao, Q., Lee, H.-Y., Tseng, H.-Y., Ma, S. and Yang, M.-H. (2019). *Mode Seeking Generative Adversarial Networks for Diverse Image Synthesis*. [online] openaccess.thecvf.com. Available at: https://openaccess.thecvf.com/content_CVPR_2019/html/Mao_Mode_Seeking_Generative_Adversarial_Networks_for_Diverse_Image_Synthesis_CVPR_2019_paper.html [Accessed 1 Mar. 2024].
- Pierrick Bourgeat (2022). *Ea-GANs: Edge-Aware Generative Adversarial Networks for Cross-Modality MR Image Synthesis | IEEE Journals & Magazine | IEEE Xplore*. [online] ieeexplore.ieee.org. Available at: <https://ieeexplore.ieee.org/abstract/document/8629301/> [Accessed 1 Mar. 2024].
- Qiao, Y., Chen, Q., Deng, C., Ding, N., Qi, Y., Tan, M., Ren, X. and Wu, Q. (2021). R-GAN: Exploring Human-like Way for Reasonable Text-to-Image Synthesis via Generative Adversarial Networks. [https://dl.acm.org/doi/abs/10.1145/3474085.3475363\(08\)](https://dl.acm.org/doi/abs/10.1145/3474085.3475363(08)). doi:<https://doi.org/10.1145/3474085.3475363>.
- Shamsolmoali, P., Zareapoor, M., Granger, E., Zhou, H., Wang, R., Celebi, M.E. and Yang, J. (2021). Image synthesis with adversarial networks: A comprehensive survey and case studies. *Information Fusion*, [online] 72(09), pp.126–146. doi:<https://doi.org/10.1016/j.inffus.2021.02.014>.

License

Copyright (c) 2024 Sai Mani Krishna Sistla , Suhas Jangoan, Ikram Ahamed Mohamed



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/). Authors retain copyright and grant the journal right of first publication with the work simultaneously licensed under a [Creative Commons Attribution \(CC-BY\) 4.0 License](https://creativecommons.org/licenses/by/4.0/) that allows others to share the work with an acknowledgment of the work's authorship and initial publication in this journal.