https://doi.org/10.24867/GRID-2024-p36 Preliminary report

GENERATIVE ART METHODS FOR IMAGE CREATION

Ana Agić Cmrk , Nikolina Stanić Loknar , Lidija Mandić
University of Zagreb, Faculty of Graphic Arts, Zagreb, Croatia

Abstract: *In this paper, several generative art methods for image creation are discussed. Generative art is a form of image and video creation using artificial intelligence and machine learning. The methods explored include Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Denoising Diffusion Probabilistic Models (DDPMs). GANs involve two submodels: a generator, which generates image examples, and a discriminator, which classifies whether the generated examples are real or fake. VAEs consist of an encoder, which compresses input data into latent variables, a latent space representing the compressed data, and a decoder, which reconstructs the original data from the latent samples. Diffusion models generate image data by gradually adding noise to the data (diffusion) and then reversing the process (denoising) to produce new data. This paper also discusses the challenges, applications, ethical considerations, and future directions of generative art in the field of graphic design.*

Key words: generative art, GAN, VAE, diffusion models

1. INTRODUCTION

When speaking of generative art, there are several definitions that explain the creation of such art. One states that generative art is generated by a system that operates autonomously, and that artist may create the system itself, set parameters that influence image creation but ultimately the system creates the outcome (Tempel, 2017). Cao et al. in their paper define generative art as Artificial Intelligence Generated Content (AIGC), which refers to content generated by generative artificial intelligence (GAI) which can automate the creation of large amount of content in short period of time (Cao et al., 2023). Content can be textual - conversational (as ChatGPT or Microsoft Gemini), it can be images (Dalle -2, Midjourney, Microsoft Designer, Stable Diffusion and many others) and video creation (as Open AI Sora or Runway ML). These tools allow designers, artists and enthusiasts to create art without technical knowledge. Historically, generative art has existed for some time, and its origins trace back to the mid-20th century. There were several artists who contributed to computer generated art, like Vera Molnar, Harold Cohen, Herbert Franke, Manfred Mohr and others. Harold Cohen created AARON, computer software for independent art creation (Crysalis, 2021; Hencz, 2023). In more recent period, with the rise of machine learning generative art is on new heights, enabling creations of many different artistic styles which can be inspired by some already existing art or new, not seen before creations. This all could not be possible without advancements in artificial intelligence, machine learning models such as Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Denoising Diffusion Probabilistic Models (DDPMs) which are being discussed in section 2. of this paper. Besides these advancements in technology, ethical questions come into the picture. Questions concerning originality of the text or image, authorship and changes in the design process. The aim of this paper is to explore technical aspects of several generative art algorithms, with visual examples of today's possibilities, and to explore possible future directions of generative art.

2. GENERATIVE ART CREATION METHODS

2.1 Generative Adversarial Networks (GANs)

Generative Adversarial Network (GAN) is an algorithm that consist of two neural networks which work in opposition – a generator and a discriminator. The generator attempts to learn from real data in order to create new data, while discriminator attempts to determine whether the input is from the real data space or not (Cao et al., 2023). Figure 1 depicts a simple GAN algorithm.

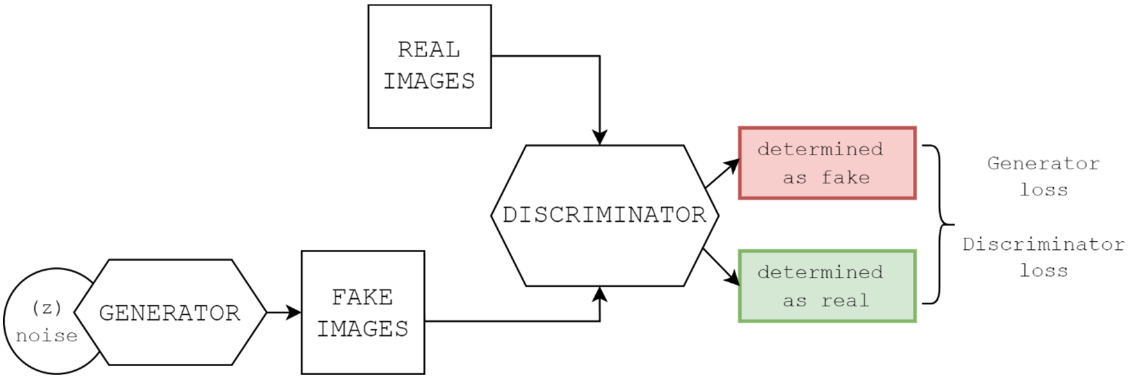

Figure 1: GAN network (adopted from (Wang et al., 2021).)

Both the generator and discriminator are trained simultaneously, where generator creates images and tries to fool the discriminator which tries to distinguish from real images and generated ones, and objective of GAN is to learn the generators distribution that approximates real data distribution (Wang et al., 2021). The generators loss function tries to measure and minimize the probability that the discriminator determines generated data as fake, and the discriminator loss function tries to measure and maximize correctly determined data as real. GANs require to find the Nash equilibrium, which is a state of end game between generator and discriminator, a place where they can't improve in learning. There are several GAN architectures and Das (2023) numbered six of them in his article (CycleGAN, StyleGAN, PixelRNN, text-2-image, DiscoGan and IsGan. He also states that all of these architectures are built on adversarial loss, and have generator and discriminator. Wang et al. (2021) also give an extensive taxonomy of recent GANs in their paper. Figure 2 depicts an image created from Artbreeder website containing "dark brown dog sitting on a bench in a park in front of tall green trees" prompt.

Figure 2: GAN image created from Artbreeder

2.2 Variational Autoencoders (VAEs)

Variational Autoencoders (VAEs) are a type of generator based on probability and include a latent space. Variational means that this technique uses Bayesian machine learning – variational Bayes (Foy, 2021). This feature makes them different from classical autoencoders which are not capable of producing new data, but they are similar because of encoder and decoder (Doersch, 2016). Autoencoders use a fixed (deterministic) latent code to map the input to, while Variational Autoencoders will replace this with a (Gaussian) distribution (Wenzel, 2022). VAEs work by compressing data into latent space (via encoder) where data is now presented as vectors and then the decoder reconstructs the data from the latent space and generates new data. Figure 3 depicts a simple VAE algorithm.

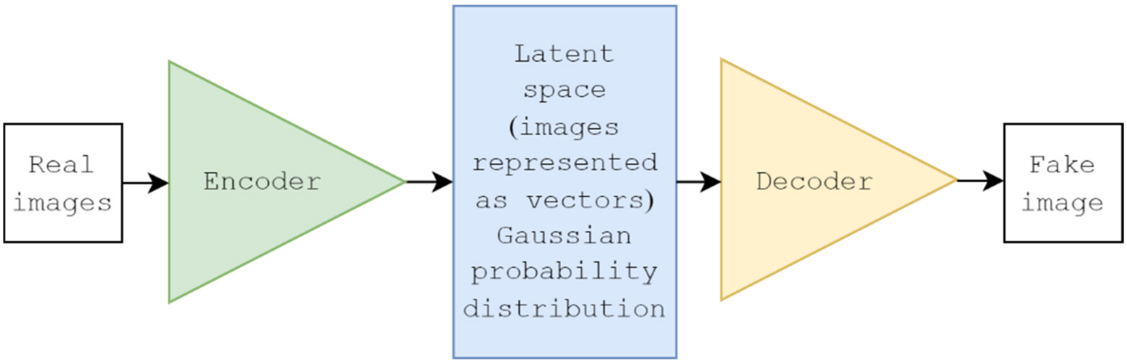

Figure 3: Simple VAE algorithm

Architecture of VAEs has three key components – encoder, latent space and decoder. At the end of encoder, we got samples of data, but presented in a probability distribution, allowing for meaningful interpolation between data samples. After this latent space, the samples of data (vectors) can pass through the decoder which are then passed through the neural network (Foy, 2021). Decoder converts data back into original form, reconstructing them from a lower-dimensional latent space into a higher-dimensional data domain. There can be many data samples in the latent space from which decoder can create new images. VAEs also have a reconstruction loss, which is similar to GAN reconstruction loss (Kullback-Leibler divergence) (Wenzel, 2022). This loss functions ensure that generated data is similar as possible to the original. Kingma & Welling in their paper list several applications of generative art, like training VAE on chemical structures with aim to search for new structures, then to generate natural language sentences, even to do simulations of observations of far galaxies (Kingma & Welling, 2019). Figure 4 depicts an image reconstruction using VAEs. Pros of VAEs are that they are stable and more reliable than GANs, cons are blurry images and not so realistic images.

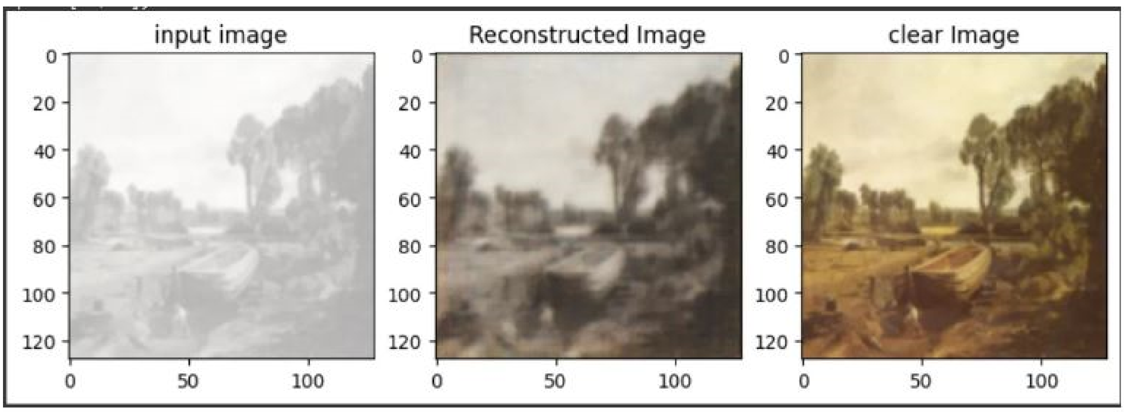

Figure 4: Image reconstruction – image curtesy of: https://pub.aimind.so/image-restoration-using-deep-learning-variational-autoencoders-8483135bb72d

2.3 Denoising Diffusion Probabilistic Models (DDPMs)

Diffusion models are a class of generative probabilistic models that produce images through a process of gradual noise injection into the original data until it turns into the known noise distribution, before reversing this process with noise reduction (Cao et al., 2015). DDPM models make use of two Markov chains - one that adds noise to the data and other which converts data from the noise (Yang et al., 2024). Figure 5 below depicts a Denoising Diffusion Model algorithm. Models are trained on large image datasets, the larger the dataset, the better generated outcome.

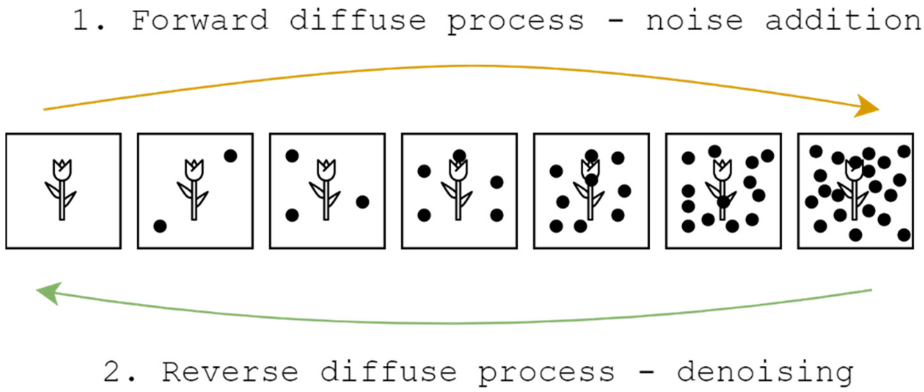

Figure 5: Denoising Diffusion Model algorithm

Ho et al. (2020) in their paper argue that diffusion models are straightforward to define and efficient to train, and show with their experiments that this model can produce high quality outputs (Ho et al., 2020). Besides image creation, diffusion models are now capable of video creations (GANs and VAEs are also capable) from text input (Zhou et al., 2024). Pros of stable diffusion models are high quality images and diverse outputs and cons are the time needed to sample an image, in comparison to other models, and artefacts which can be created during sampling of the image.
And to demonstrate visually stable diffusion, Figure 6 depicts an image created using said model and generating "dark brown dog sitting on a bench in a park in front of tall green trees" prompt.

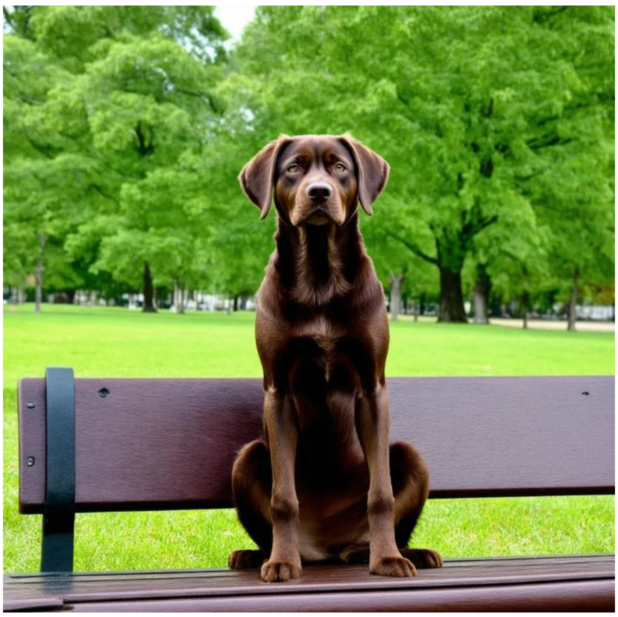

Figure 6: Image created with stable diffusion model

3. CONCLUSION

In this paper, some of the most prominent generative art models for image creation have been explained in terms of functioning and output generation. Explained were main differences between GANs – which employs generator and discriminator which adversarial positions enable image creation; VAEs – which uses two main components encoder and decoder, which in between them contain a latent space where samples of data in vector format are stored, and finally DDPMs – Diffusion models rendering two processes named forward diffuse process and reverse diffuse process where noise is added and removed in image creation process. Not all generative art models were mentioned in this paper, like Autoregressive models (PixelRNN), Neural Style Transfer, Transformers, Clip, Variational Diffusion Models, but will be in a sequential paper. Generative models have several practical applications in graphic design, digital art, gaming industry and video generation, and can accelerate processes which would usually take longer. Some downsides of this text/image/video production are beginning to unveil as inability to recognize real from artificially generated content, especially as deepfake videos where fake news can be easily placed in order to spread misinformation. Also, problems with generating delicate and illegal content can be encountered. Finally, it is unlikely that generative image creation will slow down in progress, and rather than avoiding and resisting its presence we can use it as a supplement and use it responsibly.

4. REFERENCES

Cao, H., Tan, C., Gao, Z., Xu, Y., Chen, G., Heng, P-A., & Li, S. Z. (2015) A Survey on Generative Diffusion Model. *Journal of Latex Class Files*. 14 (8), 2814-2830. Available from: 10.48550/arXiv.2209.02646

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S. & Sun, L. (2023) A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT. *The Journal of the Assosiation for Computing Machinery. 37 (111).* Available from: doi: 10.48550/arXiv.2303.04226

Crysalis. (2021) *A Whirlwind History of Generative Art: from Molnár to Hobbs*. Weblog. Available from: Https://Www.Crysalis.Art/Crysalis-Ai-Art-Research/a-Whirlwind-History-of-Generative-Art-from-Molnr-to-Hobbs [Accessed 15th August 2024].

Das, S. (2023) *6 GAN Architectures You Really Should Know*. Available from: doi: https://neptune.ai/blog/6-gan-architectures [Accessed 15th August 2024].

Doersch, C. (2016) *Tutorial on Variational Autoencoders.* Available from: doi: 10.48550/arXiv.1606.05908

Foy, P. (2021) *Generative Modeling: What is a Variational Autoencoder (VAE)?* Weblog. Available from: Https://Blog.Mlq.Ai/What-Is-a-Variational-Autoencoder/#introduction-to-Variational-Autoencoders [Accessed 15th August 2024].

Hencz, A. (2023) *What is Generative Art? From Seminal Experiences to New Frontiers*. Available from: Https://Magazine.Artland.Com/Generative-Art/ [Accessed 15th August 2024].

Ho, J., Jain, A. & Abbeel, P. (2020) Denoising Diffusion Probabilistic Models. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Available from: doi: 10.48550/arXiv.2006.11239

Kingma, D. P. & Welling, M. (2019) An Introduction to Variational Autoencoders. *Foundations and Trends in Machine Learning*. 12 (4), 307-392. Available from: doi: 10.1561/2200000056

Tempel, M. (2017) Generative art for all. *Journal of Innovation and Entrepreneurship*. *6* (1), 1–15. Available from: doi: 10.1186/s13731-017-0072-1

Wang, Z., She, Q. & Ward, T. E. (2021) Generative Adversarial Networks in Computer Vision. *ACM Computing Surveys*, *54*(2), 1–38. Available from: doi: 10.1145/3439723

Wenzel, M. (2022) *Generative Adversarial Networks and Other Generative Models*. Available from: doi: 10.48550/arXiv.2207.03887

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B. & Yang, M.-H. (2024) Diffusion Models: A Comprehensive Survey of Methods and Applications. *ACM Computing Surveys*. 56 (4), 1–39. Available from: doi: 10.1145/3626235

Zhou, P., Wang, L., Liu, Z., Hao, Y., Hui, P., Tarkoma, S. & Kangasharju, J. (2024) A Survey on Generative AI and LLM for Video Generation, Understanding, and Streaming. Available from: doi: 10.48550/arXiv.2404.16038

© 2024 Authors. Published by the University of Novi Sad, Faculty of Technical Sciences, Department of Graphic Engineering and Design. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license 3.0 Serbia (http://creativecommons.org/licenses/by/3.0/rs/).