

## IMAGE GENERATION TECHNIQUES USING GENERATIVE ADVERSARIAL NETWORKS

*Abstract:* GANs were first described in a year 2014, which is quite recently for algorithms. Although, during its time of existence, lots of various modifications and areas of possible usage were found. One of such area is image generation sphere, in which this algorithm is able to achieve results that, in some cases, do not differ from pictures drawn by a person or photographs of certain objects.

*Keywords:* Artificial intelligence, Neural networks, Generative adversarial networks, image generation, deep learning.

### Problem description

Machine data generation of the text and images, which will have any sense for a human being, is a very complex task. People tried to solve it for decades and created hundreds of different algorithms, which demonstrated different quality levels, but none of them can come even close to results, that will be indistinguishable from human work itself.

One of the first of such algorithms was GAN network, which appeared in the year 2014, which is quite a short term for an algorithm of such complexity. Moreover, it gets significant improvements over past eight years due to some changes in basic algorithm, increased quality of used datasets and increasing of computing power, which allows to use much higher volumes of high quality data. Currently, there are several main modifications of this network that are actively used and it's most likely that most web-resources that are generating images or apply certain styles to them are using these modifications.

### Usage of GANs and its modifications

A significant number of different modifications of Generative Adversarial Networks (GAN) exist. They can be used to generate various images in many different fields of human life [1]. These modifications include changes that have been applied to them to improve model for specific, more specialized tasks, such as image enhancement [2], generation of faces or their parts [3-4], creating license plates [5], learning drug elements [6] and much more.

Although all of them have diverse applications and are actively used for data acquisition, all of them are based on only a few changes to the model itself, based on which they can be divided into main categories based on what tasks they can perform. These categories include four main models:

CGAN, which utilizes an additional input to add a label to each input and generated

output [7]. This allows to get more deterministic results [8] and improves model, so it will be able to generate a large number of diverse, deterministic images by learning once.

Progressive GAN (PGAN) is a modification of the standard algorithm which creates high resolution images starting from the same images but with a lower resolution. It uses multiple networks, the first of which generates low-resolution images, and the subsequent ones improves its quality and resolution to get better results at the end. It allows to significantly improve the quality of an image and achieves higher resolution without requiring large computational power.

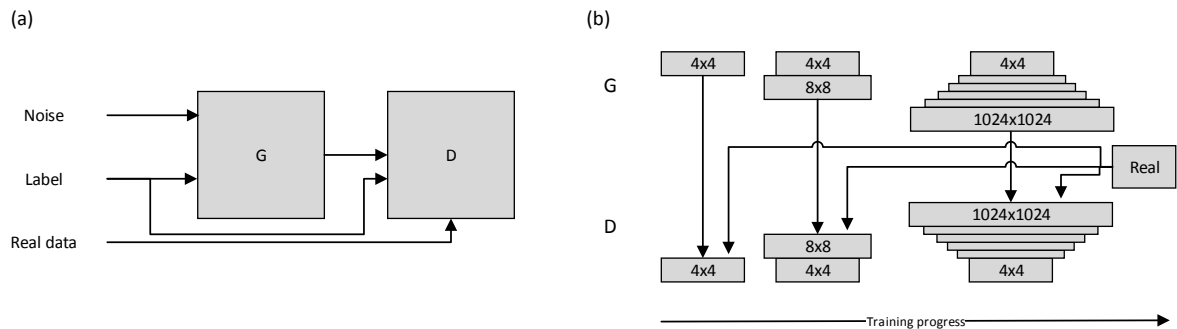


Figure 1. Working principles of CGAN (a) and PGAN (b)

StyleGAN is another type of GAN that generates images by transferring the style of one image onto another, allowing for images to be generated in a particular style (such as watercolor painting or only dark tones one). It takes two images as an input, one of which is the image to be altered, while the other one is a default image with the desired style, which should be applied to the first one.

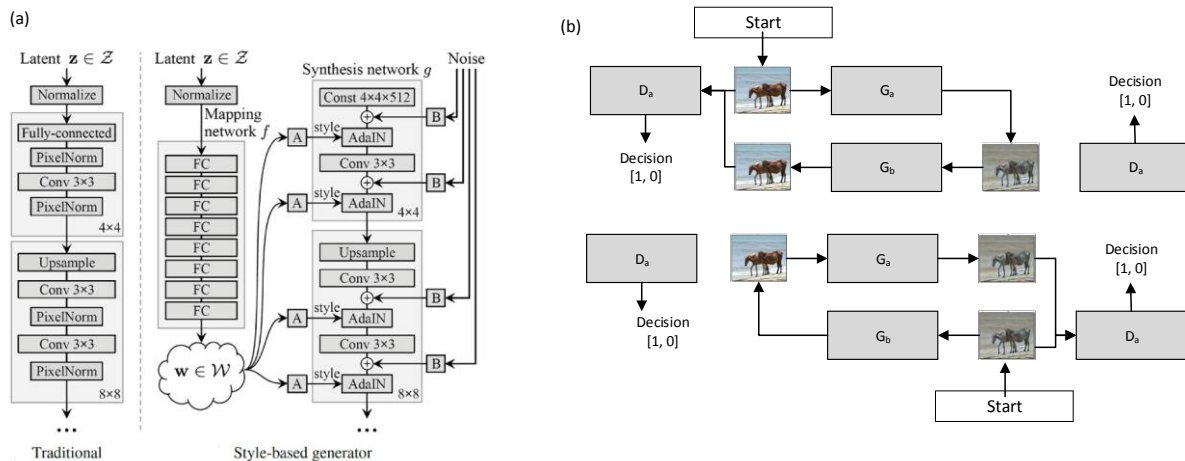


Figure 2. Working principles of StyleGAN (a) and CycleGAN(b)

CycleGAN is the latest of the described types of GANs, which can be used for image-to-image translation mechanisms. This algorithm provides the ability to add additional attributes to the image, such as certain objects like glasses, ties, change image parts, such as

hair or eyes color, clothing types and others. Thus, the algorithm finds and applies changes only to certain parts of the image, rather than working with it entirely. The main benefit on CycleGAN is the functionality to independently determine where to add a particular object, providing the ability to easily make quite complex changes.

### GAN principles of work

The architecture of the GAN model consists of two main parts: a generator, which is used to generate new example/instances from input data in the latent space, and a discriminator, which is trained to classify instances as real (it depends on the domain what means real) or fake.

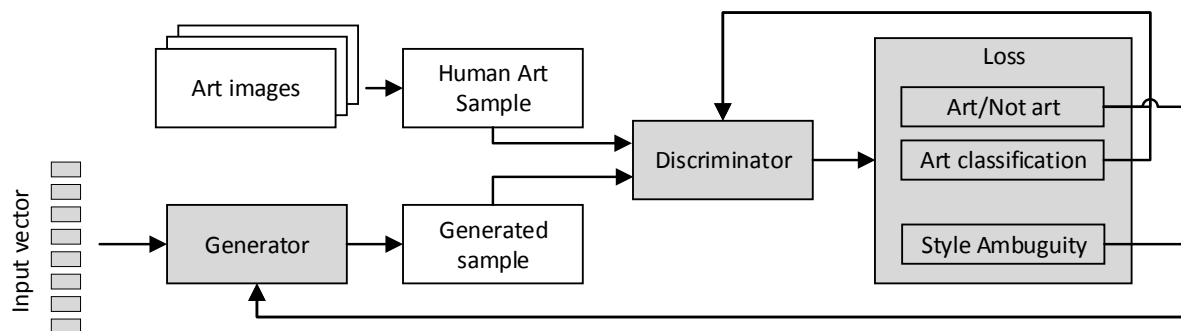


Figure 3. GAN architecture

The training of both models is happening by the predictions of the discriminator, which determines whether the data, presented to its input is a work of art or generated one, or classifies it according to certain parameters, such as providing a value from 0 to 1, where 1 is a real image [9]. The coefficients changes are applied based on the game theory algorithm, which is based on MinMax algorithm, in which elements are competing on minimizing their cost function and, on the same time, maximizing opponent's one.

The standard algorithm of such element as discriminator is a typical classification model, that decides whether the input image is real – 1 or generated – 0. Its main modification is Wasserstein GAN (WGAN) model, which is a regression model. In WGAN discriminator outputs value in a range from 0 to 1, depending on which both models are adjusted simultaneously. It allows for a significant reduction in the likelihood of the model converging.

Training stops when the discriminator can no longer determine whether the image in front of it is real or generated and always produces 50/50 results, which means that generator can already create images that are close to or identical to the original ones. However, it is not possible to start training having only one perfectly trained model of either the discriminator or the generator, as it will not provide the other model with the ability to understand which way its parameters should be changed. This way, learning will go in a random way and won't get any specific results.



One more modification of GAN is Layered Recursive GAN [12], which generates images not as a whole picture, but in parts, which often consists of a background and a foreground. This allows to generate one background image and then reuse it multiple times. To achieve this, the model uses recursion and then combines the layers to obtain a natural image. The model contains at least three generators that are partially trained separately – the foreground generator, the background generator and the generator, that combines these layers. The resulting model produces high-quality results and can be used to add multiple objects to one background.

Another modification of GAN is the asymmetric learning model, or CVAE-GAN [13], which adds an auto-encoder (AE) to the standard GAN model, which allows image generation in a specific category. Image generation occurs as a composition of the label and the latent attributes of the model. Therefore, by adding various category labels, different variations of images can be obtained and the output of the model can be controlled. The specific feature of this model is a usage of cross entropy loss function for the discriminator and mean discrepancy objective function for generator. Additionally, the encoder network was adapted to learn the relationship between the latent space and real images and pairwise feature matching was used to preserve the structure of the generated images.

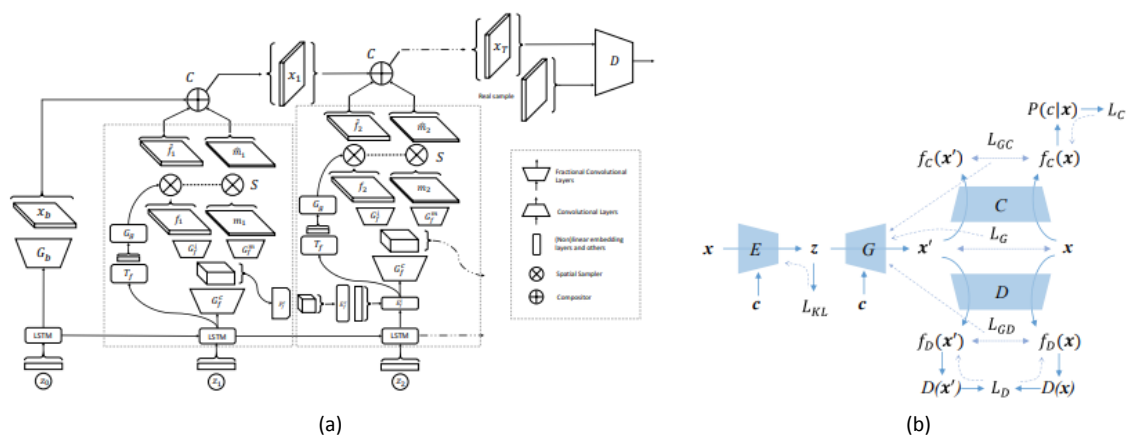


Figure 5. The architecture of Layered Recursive GAN (a) [12], and the architecture of CVAE-GAN [13]

The most modern evolution step of GAN is the Stable Diffusion network, which is based on a similar model training principle. By utilizing diffusion process in the generator model, this approach enables overcoming of model instability and collapsing. Furthermore, this algorithm allows to generate images, based on a text request, which significantly increases the output determinacy and allows using created and learned model in lots of different spheres.

## Conclusion

The article describes modifications of the image generation algorithms that use a GAN network at their basis. These modifications include networks that add labels during generation to obtain more deterministic output, transform images from low to high resolution, while rapidly increasing the quality of it, transfer styles or objects between different images. These models allow only a specific action to be performed, but more complex models and systems can be built on top of them.

Also, article describes the architecture of the network and some of its possible modifications. They introduce certain changes to the learning or generation process, which can improve such learning characteristics as. the quality of the output image, speed and stability of learning process and the amount of details on it.

## REFERENCES

1. Durgadevi M. Generative adversarial network (gan): a general review on different variants of gan and applications // 6th International Conference on Communication and Electronics Systems (ICCES). Coimbatore, 2021. C.1-8.
2. Photo-realistic single image super-resolution using a generative adversarial network / C. Ledig та ін. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Silver Spring, 2017. C.4681-4690.
3. Group consistent similarity learning via deep crf for person re-identification / D. Chen та ін. // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018. C.8649-8658.
4. Li C., Wand M. Precomputed real-time texture synthesis with markovian generative adversarial networks // European Conference on Computer Vision. Springer, 2016. C.702-716.
5. Byung-Gil H., Jong Taek L., Kil-Taek L. License Plate Image Generation using Generative Adversarial Networks for End-To-End License Plate Character Recognition from a Small Set of Real Images // Applied Sciences. 2020. № 10, 2780.
6. Hussain S., Anees A., Das A. High-content image generation for drug discovery using generative adversarial networks // Neural Networks. 2020. № 132. C.353-363.
7. Minaee, S., & Abdolrashidi, A. (2018). Iris-GAN: Learning to Generate Realistic Iris Images Using Convolutional GAN. arXiv:1812.04822.
8. Heim E. Constrained Generative Adversarial Networks for Interactive Image Generation // CVPR 2019. Long Beach, 2019.
9. Minaee S., Abdolrashidi A. Iris-GAN: Learning to Generate Realistic Iris Images Using Convolutional GAN // arXiv. 2018. № 1511.06434.
10. StyleSwin: Transformer-based GAN for High-resolution Image Generation / Z. Bowen та ін. // CVPR 2022. New Orleans, 2022.

11. CWGAN: Conditional Wasserstein Generative Adversarial Nets for Fault Data Generation / Y. Yu та ін. // 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). Dali, 2019.

12. Yang J., Kannan A., Batra D. LR-GAN: Layered recursive generative adversarial networks for image generation // ICLR 2017. Toulon, 2017.

13. Bao J. Fine-Grained Image Generation through Asymmetric Training // 2017 IEEE International Conference on Computer Vision (ICCV). Venice , 2017.