Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/52/20241111

Investigation of generative capacity related to DCGANs across varied discriminator architectures and parameter counts: A comparative study

Fan Yi

School of Computer Science, Fudan University, Shanghai, 200433, China

21307130047@m.fudan.edu.cn

Abstract. Generating lifelike images through generative models poses a significant challenge, where Generative Adversarial Networks (GANs), particularly Deep Convolutional GANs (DCGANs), are commonly employed for image synthesis. This study focuses on altering the DCGAN discriminator's structure and parameter count, investigating their effects on the characteristics of the resulting generated images. Assessment of these models is carried out using the Fréchet Inception Distance (FID) score, a metric that gauges the quality of generated image samples. The research specifically involves substituting some convolutional layers with fully-connected layers, and the ensuing outcomes are thoroughly compared to discern the impact of these structural changes. Furthermore, dropout was used to study the number of the parameters' influence. This study compared the FID score of the models when the probability is 0, 0.2, 0.4, 0.6 and 0.8. Experimental results showed that the DCGAN with the fully-connected layers' generated ability was stronger than the original one. Besides, when the probability of the dropout is 0.6, the images generated was the most realistic. Finally, the paper explained the possible reasons for the difference and proposed a better generative model based on DCGAN.

Keywords: Generative Adversarial Network, Deep Convolutional Generative Adversarial Network, Discriminator, Dropout.

1. Introduction

Image generation refers to the process of using computer algorithms and techniques to synthesize images. Generative Adversarial Networks (GANs) are a class of generative models which generate new data points based on the distribution of the training data, which can be used in both semi-supervised and unsupervised learning [1]. The network mainly consists of two neural networks, which are called the generator and discriminator. The generator takes random noise or other inputs and generates fake data that resemble real data, while the discriminator tries to distinguish between real and fake data. Through adversarial training, the network achieves the goal of generating convincing samples. Thereinto, the architecture configuration of GAN has a great influence on the performance of generated images, which necessitates a thorough discussion.

Image generation technology has been widely applied to numerous scenarios including design, artistic creation, virtual reality, medical imaging and special effects. Throughout the years, there are many models developed for the image generation. In [2], Kingma et al. introduced the concept of

Variational Autoencoders (VAEs) which was a strong framework for probabilistic encoding and decoding of image data. It combined the probabilistic graphical modeling with the autoencoder structure, mapping the input data into a latent space distribution and generating new data by sampling from this distribution. And in [3], the Google DeepMind PixelRNN team proposed Pixel Recurrent Neural Networks (PixelRNN), an autoregressive neural network model aiming for pixel-level data. The model used the convolutional neural network and the recurrent neural network to maintain the context and coherence of the generated data. The GANs was introduced in 2014 [4] to generate realistic data, including images, audios and more. Karras et al proposed a structure called StyleGAN [5] to generate image data with unprecedented control their visual appearance. The main innovation was that the model separated the latent space and the style space, allowing for more fine-grained control over the generated images. The study [6] introduced a training methodology for GAN known as Progressive GAN that involves increasing the resolution of generated model gradually during training. This progressive training process makes the training more stable and generator more robust, leading to high-resolution images. Deep Convolutional Generative Adversarial Network (DCGAN) demonstrated in [7] was a significant step forward in the field of image generation because it showed the potential of combining convolutional network and GAN to generate coherent and visually appealing images. However, there are few research diving in the relationship between the structure of the discriminator and the performance of the GAN model. Few focus on the influence of the number of the discriminator's parameters on the quality of the generated images.

To solve the limitation mentioned above, this research puts the emphasis on the structure of the discriminator, especially the number of the parameters, of the DCGAN. This study will show the changes of the quality of the images generated when replacing a convolutional layer with fully connected layers in the discriminator. Also, the research will discuss how the number of the parameters of the discriminator makes an influence on the performance of the GAN model and the quality of the new images. Fréchet Inception Distance (FID), which is used to evaluate the quality and diversity of the generated model, will be applied to judge the model's performance and decide the structure with the best performance.

2. Methods

2.1. Dataset description and preprocessing

In this study, the animation dataset extracted from the Danbooru 2018 dataset was used. The original dataset comprises 51, 224 animation images, each with dimensions of 96×96 pixels, and all images are in RGB color mode. Figure 1 is some visualization of the dataset this study used.

Figure 1. Visualization of the dataset (Photo/Picture credit: Original).

In terms of the preprocessing stage, it consists of three parts. The first is to resize the data. This study transformed the size of the input images from 96×96 pixels to 64×64 pixels. Subsequently, this study normalized the data to adjust it into a specific range to optimize the model training.

DOI: 10.54254/2755-2721/52/20241111

2.2. Proposed approach

The Generative Adversarial Network is an excellent model for image generation. The GAN typically consists of two neural networks, which are the discriminator and the generator [8, 9]. The generator takes a random noise as an input and outputs a synthetic data sample through a series of fully-connected layers. The generator is trained to generate samples that are similar to the training data. The discriminator also employs a deep feedforward neural network based on the fully-connected layers. It takes in a real sample from the training data and a fake sample from the generator and outputs a binary classification, identifying whether the image is real or fake. During the training, the generator and discriminator compete against each other and plays a minimax game. The discriminator tries to maximize the probability of correctly classifying reals vs fakes, while the generator tries to minimize the probability of the discriminator making the correct classification. The objective functions of the discriminator and the generator are as followed [4]:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \left[\log D\left(x^{(i)}\right) + \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right)\right] \tag{1}$$

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^{m} \log\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \tag{2}$$

where $\nabla_{\theta_d}$is the gradient with respect to discriminator parameters θ_d, $D(x)$represents Discriminator 's probability that x is a real sample, $G\left(z^{(i)}\right)$represent the i-th fake sample generated by generator G from random noise $z^{(i)}$ and m means the number of trainings samples.

This adversarial training pushes the generator to generate more realistic samples to fool the discriminator, thus accomplishing the goal. DCGAN shown in Figure 2 is a variant of the GAN, which replaces the fully-connected layers in the discriminator with the strided convolutional layers and the ones in the generator with the factioncallly-strided convolutional layers.

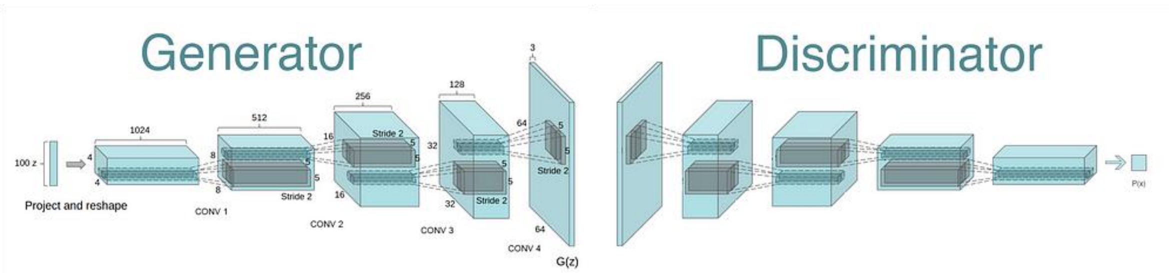

Figure 2. The basic structure of the DCGAN [7].

This study referred the DCGAN and make a little change to the discriminator. The generator consists of 5 transposed convolutional layers and each layer kernels with a size of 4×4 and the stride is 1 with no padding. A batch normalization after each transposed convolutional layers was adopted to help the training more convergent and make the model more robust. The input of the generation is a random noise tensor with dimensions of 1×1×100 and after 5 layers, it gives out a tensor with a height of 64, a width of 64 and a depth of 3. The discriminator composed 3 convolutional layers and 3 fully-connected layers in total. Each convolutional layer has 4×4 size kernels and the stride are 1 with no padding. The batch normalization was adopted the same way as the generator. Also, this study used the dropout method in the fully-connected layers. The discriminator's input is a 64×64 size RGB image, and the output is the probability of the image authenticity. If the probability is higher than 0.5, the image is predicted as real, and it is considered as fake image if the probability is under 0.5. The architecture is shown in Figure 3.

DOI: 10.54254/2755-2721/52/20241111

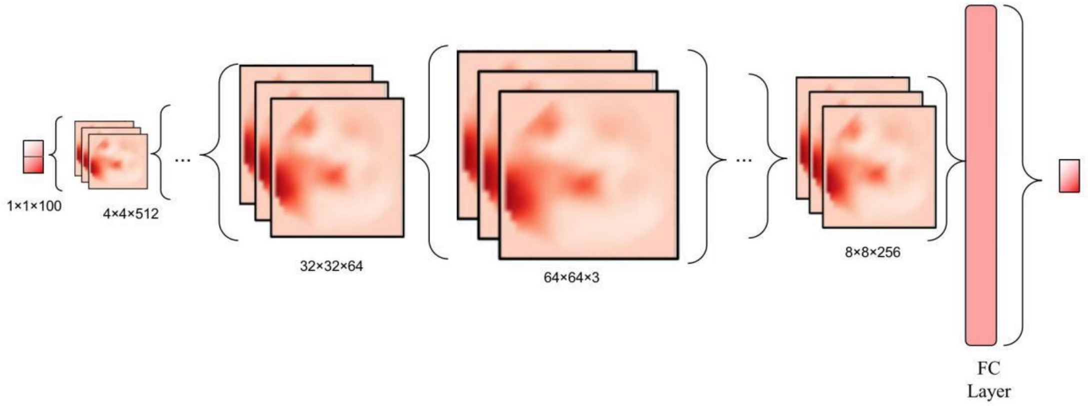

Figure 3. The structure of the proposed DCGAN (Photo/Picture credit: Original).

2.3. Implementation details
The size of the training data batch is 100. This study used the Adam optimizer to optimize the parameters of the generator and the discriminator. The learning rate is set to 0.015 and the decay rate for the first moment and the second moment is 0.5 and 0.999 respectively.

As the quality of the generated image is hard to evaluate through human eye, this study adopted the Fréchet Inception Distance (FID) as a metric to measure the difference of the quality of the images [10]. It facilitates in assessing the quality and diversity of generated images by comparing their feature statistics to those of real images. The lower the FID score is, the better of the quality of the generated samples, thus meaning the better performance of the GAN model. The calculation formula of FID is as followed:

$$\mathrm{FID} = ||\mu - \mu'||_2^2 + \mathrm{tr}\left(\Sigma + \Sigma' - 2\left(\Sigma^{1/2}\Sigma'\Sigma^{1/2}\right)^{1/2}\right) \tag{3}$$

where μ μ' Σ Σ' represent the mean the mean vector and covariance matrix of the real images and the generated images

To study the influence of the structure of the discriminator, especially the number of the parameters, of the DCGAN to the model performance, this study compared the FID score of the model with different dropout possibilities. The possibility was set as 0, 0.2, 0.4, 0.6 and 0.8 and the FID score was recorded.

3. Results and discussion
By replacing part of the convolutional layers in the discriminator, it is shown in the Figure 4 that DCGAN with the fully-connected layers can achieve the better performance than the original DCGAN model. The generator can generate high-quality image samples using less epochs than the original one. Furthermore, the best image sample generated has lower FID score than the best sample generated by the original DCGAN, meaning the image sample is more realistic. This may be due to the reason that fully-connected layer has a better ability of learning the features of the real images so the discriminator can better discriminator the real images from the fake ones, encouraging the generator to generate more realistic images to fool the discriminator, thus leading to better performance.

DOI: 10.54254/2755-2721/52/20241111

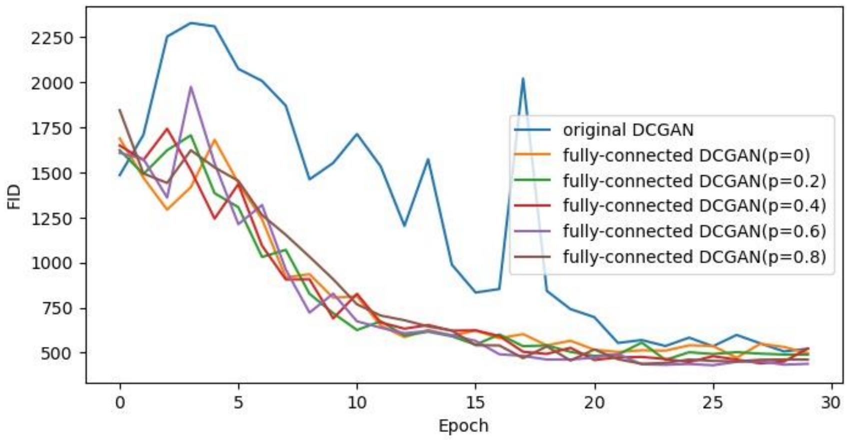

Figure 4. The changing curve of FID score based on different models with epoch (Photo/Picture credit: Original).

Table 1. The FID score of the original DCGAN and the proposed model with different dropout probability.

Model		Minimum FID score
DCGAN		506.56
	p = 0	417.58
	p = 0.2	481.77
Fully-connected DCGAN	**p = 0.4**	440.49
	p = 0.6	429.98
	p = 0.8	438.20

In the fully-connected DCGAN, the dropout probability, which represents the number of parameters in the discriminator, also has a influence in the performance of the model. The Table 1 shows that when the dropout probability equals 0.6, the images generated are the most realistic. The 0.8 and 0.4 dropout probability have roughly the same performance. This phenomenon could stem from the situation where a low dropout probability results in a high parameter count. This circumstance can potentially lead to overfitting. By decreasing the number of parameters, it is possible to enhance the discriminator's capability. From the experimental results, it is shown that the dropout probability is 0.6 leads to the best performance of the model. The final outcome of the generated images is shown in Figure 5.

DOI: 10.54254/2755-2721/52/20241111

Figure 5. The best images generated by the employed DCGAN with the dropout (p = 0.6) (Photo/Picture credit: Original).

4. Conclusion

This study focused on the influence of the structure and the number of the parameters of the discriminator on the images generated by the DCGAN model. By comparing the Frechet Inception Distance (FID) scores of various DCGAN models, this study aimed to uncover the factors contributing to the observed differences in generated image quality. The investigation involved training different DCGAN configurations and evaluating their FID scores based on the generated image samples. Notably, the findings highlight a significant improvement in model performance when substituting certain convolutional layers with fully-connected layers within the discriminator network. This structural adjustment demonstrated a profound effect on the quality of generated images. Furthermore, the study identified an optimal dropout probability of 0.6, leading to the most realistic images produced by the DCGAN model. This insight into dropout probability provides valuable guidance for achieving visually compelling results in image generation tasks. Looking ahead, the potential for further research lies in exploring alterations to the structure and the number of parameters within the generator network. Additionally, deepening the neural network architecture offers promising avenues for investigating its impact on the quality and realism of the images generated. These future endeavors could significantly contribute to the advancement of generative models and their practical applications.

References

[1] Creswell A et al 2018 *Generative Adversarial Networks: An Overview* (IEEE Signal Processing Magazine vol 35 no 1) pp 53-65

[2] Kingma D P and Max W 2013 *Auto-encoding variational bayes* (arXiv preprint) arXiv 1312.6114

[3] Van D O et al 2016 *Pixel recurrent neural networks* (International conference on machine learning) pp 1747-1756

[4] Goodfellow I et al 2014 *Generative adversarial nets* (Advances in neural information processing systems) 27

[5] Karras T Laine S Aila T 2019 *A style-based generator architecture for generative adversarial networks* (Proceedings of the IEEE/CVF conference on computer vision and pattern recognition) pp. 4401-4410

[6] Karras T Aila T Laine S et al 2017 *Progressive growing of gans for improved quality, stability, and variation* (arXiv preprint arXiv) 1710.10196

[7] Radford A Metz L and Chintala S 2015 *Unsupervised representation learning with deep convolutional generative adversarial networks* arXiv preprint arXiv:1511.06434

DOI: 10.54254/2755-2721/52/20241111

[8] Creswell A White T Dumoulin V et al 2018 *Generative adversarial networks: An overview IEEE signal processing magazine* 35(1): 53-65.

[9] Wang K Gou C Duan Y et al 2017 *Generative adversarial networks: introduction and outlook* IEEE/CAA Journal of Automatica Sinica 4(4): 588-598.

[10] Obukhov A Krasnyanskiy M 2020 *Quality assessment method for GAN based on modified metrics inception score and Fréchet inception distance* Software Engineering Perspectives in Intelligent Systems: Proceedings of 4th Computational Methods in Systems and Software 2020, Vol. 1 4. Springer International Publishing 102-114