Proceedings of the 4th International Conference on Signal Processing and Machine Learning
DOI: 10.54254/2755-2721/48/20241114

Exploring the influence of generator channel number on the quality of anime-style portrait generation based on DCGAN

Ye Huang

School of AI and Advanced Computing, Xi'an Jiao tong-Liverpool University (XJTLU), Suzhou, 215400, China

ye.huang21@student.xjtlu.edu.cn

Abstract. In the realm of contemporary image synthesis, this research delves into a crucial objective: exploring the connection between the quantity of generator channels and the production of anime-style portraits through Deep Convolutional Generative Adversarial Networks (DCGAN). Employing an extensive dataset of anime faces encompassing diverse artistic styles, this study systematically examines the nuanced interplay between architectural parameters and the fidelity and intricacy of the generated images. By employing the Frechet Inception Distance (FID) as a metric for image quality, this investigation contributes significantly to the field by enhancing the understanding of how the number of generator channels impacts the ultimate quality of anime-style portraits. The DCGAN framework, and in particular its variants, is the backbone of this investigation. The generator and discriminator components are involved in adversarial training, a competitive process that improves image quality through iterations. The findings reveal a non-linear relationship between the number of generator channels and image quality. While increasing the number of channels initially improves image quality and decreases the FID value, exceeding the optimal threshold leads to diminishing returns and image quality degradation. The intricate interplay between structure selection and image quality is further confirmed by the dynamics of the generator and discriminator loss functions. By elucidating the trade-off between complexity and image fidelity, this study contributes to the advancement of image synthesis techniques and encourages future exploration of architectural nuances in the field of artistic image generation.

Keywords: Anime-style Portrait Generation, GANs, Generator Architecture, Image Quality, Channel Number Optimization.

1. Introduction

In contemporary times, the application of deep learning techniques to image generation has made significant progress in various creative fields. One particularly appealing area is animation-style portraits, which have gained considerable attention due to their artistic and entertainment value [1]. Characterized by its emotional portrayal of characters, dynamic expression, and complex visual elements, animation-style art occupies a unique and precious place in the visual arts. The art form encompasses a wide range of genres from fantasy to science fiction, each with unique visual characteristics. Signature features of anime-style portraits include large, expressive eyes, elaborate hairstyles and rich color interplay [2, 3]. These visual elements not only capture emotion and storytelling but are an important part of narrative and character development in anime culture.

DOI: 10.54254/2755-2721/48/20241114

The introduction of Deep Convolutional Generative Adversarial Networks (DCGAN) has revolutionized image generation through the introduction of adversarial training and deep convolutional architectures. These networks are capable of creating complex images with extraordinary realism and complexity. DCGAN is comprised of a generator (which is assigned to synthesize images from random noise) and a discriminator (which is being used to differentiate between actual images and generative images) [4]. The generator and the discriminator engage in a competitive learning process that pushes each other to improve over time.

Despite advances in image synthesis using Generative Adversarial Network (GAN), there is a significant gap in generating independent animation-style portraits. Previous research has focused on wider applications of GAN, including style transfer techniques or enhancement of existing images with animation style attributes. However, the research community has yet to systematically explore the complex relationship between various generator architectures and the final quality of autonomously generated animation style portraits. For example, works by Andreini et al. and Mukherkjee et al. emphasize GAN for style transfer and customization of existing images to conform to anime aesthetics [5, 6]. However, these studies did not explore the creation of anime-style portraits from scratch. Similarly, studies by Casanova et al. and Smith et al. utilized conditional GAN for attribute control, but did not extensively address the architectural impact on overall image fidelity and stylization in the context of anime-style portraits [7, 8]. This identified research gap motivated the study to conduct an in-depth investigation to provide insights into the architectural components that influence the outcome of the generated animation style portraits.

To address this gap, this study explores various generator architectures within the DCGAN framework. By regularly varying the number of channels in the generators and comparing the generated results, this article aims to reveal the complex architectural attributes that play a key role in influencing image fidelity and stylization of animated stylized portraits [9,10]. This study builds on existing research on the impacts of building in the GAN presented by Radford et al, Huang et al, and Gao et al. [4, 9, 10], emphasizing the impact of design choices on image generation results. The current study hypothesizes that specific model structural changes affect the generation of animation-style portraits. Through this in-depth investigation, this study hopes to provide valuable insights into the field of animation style portrait generation.

The remaining part of this paper is structured in the following way: section 2 will provide a discussion of the proposed methodology of the study and the dataset. More specific structural and experimental details are given within Section 3. Section 4 furnishes a description of the experimental outcomes data and offers some possible reasons for these results. Section 5 discusses the conclusions of this study and future work.

2. Method

2.1. Dataset preparation

In this study, this study uses the anime faces dataset provided by Kaggle [11], which consists of a wide variety of anime character portraits covering a wide range of styles and expressions. Overall, the dataset contains approximately 60,000 images. The images range in size from 40x40 to 120x120, all in RGB format. A representative screenshot of this dataset is shown in Figure 1.

DOI: 10.54254/2755-2721/48/20241114

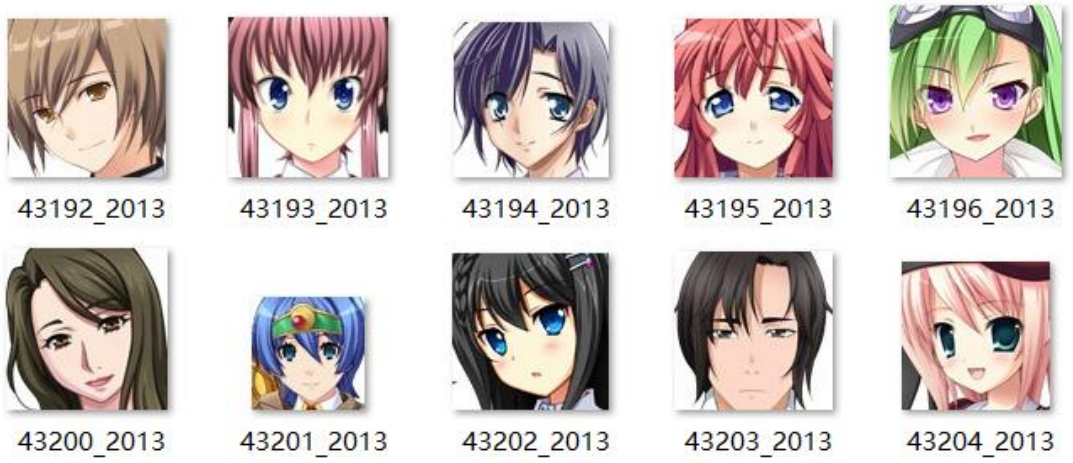

Figure 1. Representative samples from the dataset.

Since images vary in size, to prepare the training dataset, a series of preprocessing steps were applied to ensure consistency and data quality. Firstly, the images were resized to the same size of 64×64, which balances computational efficiency and image quality, thus ensuring that the images are compatible with the architecture of the DCGAN model. In addition, the pixel values are normalized to the range [-1,1] to ensure convergence and stability during the training process. The normalization process is comprised of subtraction of each pixel value by 0.5 and division by 0.5.

2.2. The proposed GAN

The core of this research method lies in the use of GANs, a well-known deep learning framework introduced by Goodfellow et al. in 2014. GAN is composed of two main elements: generators and discriminators. The generator is trained to synthesize real data instances, such as images, whereas the discriminator is tuned to discern between authentic and generated samples. Through adversarial training, the generator is taught to produce images that are increasingly convincing to the discriminator, resulting in high-quality images. To implement the GAN architecture, a DCGAN variant is used in this study. DCGAN extends the basic GAN framework by adding deep convolutional layers to improve the quality and stability of image generation. The DCGAN architecture used in this study is shown in Figure 2.

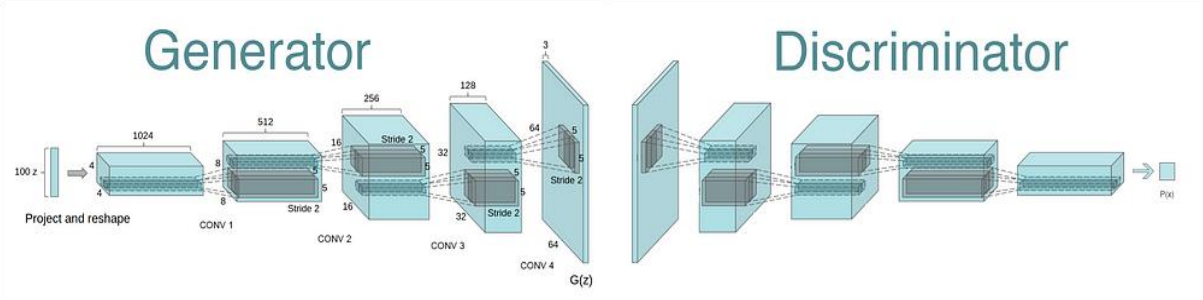

Figure 2. Architecture of the DCGAN used in the study [4].

As shown in Figure 2, a generator and a discriminator are incorporated in the DCGAN architecture used in this paper, and the Generator network consists of five transposed convolutional layers. Each layer progressively upsamples potential noise vectors into more complex representations. Batch normalization and ReLU activation functions follow each convolutional operation to promote training stability. The output of the generator is an RGB image of size 64x64 pixels. The discriminator consists

DOI: 10.54254/2755-2721/48/20241114

of a series of convolutional layers, each of them accompanied by a LeakyReLU activation function. Batch normalization is applied after each convolution to ensure balanced learning. The final layer employs a sigmoid activation function that produces a single probability value indicating the discriminator's confidence in the authenticity of the input image. A latent noise vector, of dimension 100, is used as the initial input to the generator. The impact of the generator channel configuration was explored, where the number of channels was varied to assess its effect on image quality. Specifically, experiments were conducted with 512, 1024 and 2048 generator channel configurations.

2.3. Implementation details

In terms of consistency, standardized implementation details were applied in all experiments: for the loss function, the Binary Cross Entropy (BCE) loss function was chosen for both generator and discriminator networks. This loss function measures the scatter between the true and generated distributions and leads to the convergence of the network. The batch size of 100 was determined to be the best compromise between computational efficiency and convergence stability. As for the optimizer, the Adam optimizer was used for both the generator and discriminator networks. The learning rate of the discriminator was set to 0.0004 while the learning rate of the generator was set to 0.0001, beta1 parameter was set to 0.5. In this study, the DCGAN model was trained for a total of 20 epochs. The entire experiment was executed using the PyTorch framework, leveraging the power of GPU acceleration. This optimization resulted in faster training times and allowed individuals to explore various configurations and options efficiently.

3. Results and discussion

3.1. The performance of the model

In this experiment, this study evaluated images generated under different numbers of generator channels, using Fréchet Inception Distance (FID) as an evaluation metric. The FID value aims to quantify the difference in the distribution of the generative image and the actual image in the feature space, and thus provides a measure of the generative images' quality. In the experiments, it is observed that the FID value shows a more obvious trend as the number of generator channels increases. The experimental results demonstrate that the FID value in the initial stage decreases gradually with the increase of the number of generator channels, however, when the count of channels hits a certain threshold, the FID value begins to increase, implying a decrease in the mass of the generative images. The experiments further analyze the variation of the loss function for both the generator and the discriminator. The experimental results show that as the count of channels of the generator increases, the loss value of the generator shows a trend of decreasing and then increasing. Meanwhile, the loss value of the discriminator shows the opposite trend, with the increase in the number of generator channels, the loss value for the discriminator increases at the beginning and eventually decreases.

Table 1. Performance of the generator for different number of channels.

Number of generator channels(epoch=20)	Generator loss (average)	Discriminator loss (average)	Average FID (Frechet Inception Distance)	Best FID
100,512,256,128,64,3	2.1	0.8	1334.65	442.74
100,1024,512,256,128,3	1.5	1.3	979.95	402.62
100,2048,1024,512,256,3	1.8	1.1	1563.05	790.02

DOI: 10.54254/2755-2721/48/20241114

Figure 3. The fake images generated by the generator (From left to right, correspond to the number of generator channels as 100,512,256,128,64,3; 100,1024,512,256,128,3; 100,2048,1024,512,256,3.).

3.2. Discussion

Table 1 illuminates that as the number of channels increases, the FID values show a tendency to decline and then enlarge, revealing a complex relationship between the mass of the generated images and the number of generator channels. The initial decrease in the FID value can be attributed to the fact that the increase in the number of channels improves the expressive ability of the model, which makes the generated images better approximate the real distribution. However, when the number of channels is too high, the model may become too complex and overfitting occurs, which has resulted in an increase in the FID value and a noticeable decrease in the quality of the generated image. The anomalous variations in the loss functions of the generator and the discriminator also reveal the effect of the number of generator channels on model training. The trend of the generator's loss value first decreasing and then increasing implies that a moderate number of channels can facilitate the generator to learn the data distribution, but too many channels may lead to overfitting of the model and reduce the generalization capability. The trend of the loss value of the discriminator is opposite to that of the generator, as the number of channels increases, the discriminating ability of the discriminator decreases, leading to a decrease in the loss value. It can also be seen from Figure 3 that the mass of the generative pictures first becomes better and then becomes worse as the number of generator channels becomes greater.

4. Conclusion

In conclusion, this study is an investigation of the effectiveness of different numbers of generator channels on the animation-style portrait quality generated using DCGAN and demonstrates the complex interaction between the number of generator channels and the quality of the generated animation style portraits. As expected, the results showed a non-linear relationship between the number of generator channels and image quality as measured by the FID score. This discovery aligns with our initial hypothesis, which posited an ideal range for the number of channels, beyond which we encounter diminishing returns and susceptibility to overfitting. The core methodology of this research revolves around a systematic series of experiments that unveil the influence of generator channel count on aspects such as image fidelity, loss function, and model complexity. These insights enhance our comprehension of structural parameters within DCGAN and their pivotal role in maintaining a balance between image quality and model generalization. The experimental outcomes shed light on the crucial equilibrium between increasing the channel count and the associated risk of overfitting, thereby providing valuable guidance for selecting optimal generator parameters to achieve peak performance. In addition, this study facilitates the advancement of image synthesis techniques and creates a foundation for further exploration of the nuances of generator architectures in the field of artistic image generation. However, the current study has some limitations, such as focusing only on channel numbers without considering other architectural elements. To address these limitations, future research should explore comprehensive and meticulous generator architectures and employ regularization techniques to enhance model generalization.

DOI: 10.54254/2755-2721/48/20241114

References

[1] Chen J Liu G and Chen X 2020 Int. symp. intell. comput. appl. (Singapore: Springer) pp 242–56
[2] Lin C Z Lindell D B Chan E R and Wetzstein G 2022 arXiv 1703.10593
[3] Paier W Hilsmann A and Eisert P 2020 IET Comput. Vis. 14 359–69
[4] Radford A Metz L and Chintala S 2015 arXiv 1511.06434
[5] Andreini P Bonechi S Bianchini M Mecocci A and Scarselli F 2020 Comput. methods programs biomed. 184 105268
[6] Mukherkjee D Saha P Kaplun D Sinitca A and Sarkar R 2022 Sci. Rep. 12 9141
[7] Casanova A Careil M Verbeek J Drozdzal M and Romero Soriano A 2021 Adv. Neural Inf. Process. Syst. 34 27517–29
[8] Smith K E and Smith A O 2020 arXiv 2006.16477
[9] Huang J Johanes M Kim F C Doumpioti C and Holz G C 2021 Technol. Archit. Des. 5 207–24
[10] Gao C Chen Y Liu S Tan Z and Yan S 2020 Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. pp 5680–9
[11] Kaggle 2022 Anime face dataset https://www.kaggle.com/datasets/splcher/animefacedataset