Comparison of Deep Convolutional GAN and Progressive GAN for facial image generation

Hanjing Zhu

School of Letters and Sciences, University of Wisconsin - Madison, Madison, WI, 53703, USA

hzhu273@wisc.edu

Abstract. Generative Adversarial Networks (GANs) have generated realistic and diverse facial images with promising results. This work demonstrates a technique for creating facial pictures using GANs and evaluates the effectiveness of several GAN designs and training approaches. The CelebA dataset is leveraged for training and evaluating the GAN model, and employ a variety of evaluation metrics, such as the Structural Similarity Index (SSIM) to assess the quality and diversity of the generated images. Progressive GAN outperforms Deep Convolutional GAN in terms of image quality and diversity, and conditional GAN training is more effective than standard GAN training for generating facial images with specific attributes. The combination of Progressive GAN and conditional GAN training produces facial images of the utmost quality and diversity. The findings contribute to a broader comprehension of the use of GANs for generating facial images and have ramifications for a variety of applications, from facial recognition to virtual reality.

Keywords: generative neural network, facial image generation, deep learning.

1. Introduction

Facial image generation is important in computer vision for face recognition, security, virtual reality, and entertainment. The nuances of facial characteristics and emotions make realistic face image production challenging. Earlier statistical methods required massive amounts of manually annotated data and human feature extraction. These technologies failed to capture face expressions and produce high-quality photos.

Since generative neural networks (GANs) were introduced, high-quality facial photos that seem real have improved. A generator and discriminator neural network are used in the GAN generative model. The function of the generator network is to replicate the training data, while the discriminator network is responsible for discerning between authentic and synthesized data. The generator and discriminator networks engage in a mutual adversarial process whereby each network attempts to deceive the other.

GAN-based face image generating methods including StyleGAN, CycleGAN, and ProgressiveGAN have been introduced recently [1-3]. These methods have produced high-quality images with realistic skin, hair, and facial expressions. StyleGAN, a cutting-edge model, uses a style-based generator architecture to build high-resolution face photos with fine-grained control over numerous aspects, such as age and gender [1].

DOI: 10.54254/2755-2721/18/20230984

Despite these successes, GAN-based face image creation still has several difficulties. Control over made films is a major issue. High-quality GAN photographs cannot be limited by age or race. Realistic and varied face expressions are another challenge. GAN-based facial expression generators cannot capture all human emotions.

This article will compare GAN face picture production, focusing on techniques and architectures to improve image quality and diversity. Moreover, unsolved challenges will be discussed in this subject and suggest topics for additional research.

2. Related work

GAN-based face image synthesis has been a popular research topic due to its many applications. He et al. used Progressive GAN to manipulate facial attributes including age, gender, and expression while retaining subject identity [3]. On many datasets, their method surpassed cutting-edge algorithms in terms of picture quality and variety. CFGANs are used by Kaneko et al. (2017) to regulate synthetic picture characteristics. Traditional GANs may generate realistic pictures, but they cannot control image properties, which is important for image modification and synthesis [4]. The authors present a novel GAN architecture that uses conditioning to produce pictures with predefined attribute values. To create high-quality, aesthetically attractive pictures, they filter conditioning information to the generator network.

Karras et al.'s Progressive Growth of GANs (PGGAN) architecture produced high-quality images with improved stability and variance [5]. The scientists showed that by gradually increasing picture resolution during training, the model could detect minute characteristics and produce more realistic images. Brock and colleagues developed a method for training Generative Adversarial Networks (GANs) on a large scale to achieve high-fidelity image synthesis that closely resembles reality [6]. The researchers modified the loss function and normalization approach in order to effectively train Generative Adversarial Networks (GANs) to generate visually appealing and realistic images.

Karras et al. proposed an improved style-based GAN training technique that outperformed numerous benchmark datasets [7]. Other GAN-based face image generating and alteration algorithms exist. For high-resolution face completion, Progressive generative adversarial networks (P-GANs) that are fully end-to-end are used by Chen et al. (2018) to complete high-resolution faces with numerous configurable features [8]. The authors offer a fully end-to-end network architecture to create high-resolution photographs with numerous configurable properties. A generator and discriminator network are trained to produce high-quality pictures in the P-GANs technique. Attribute-Encoder-Decoder (AED) allows users to define face attributes. The AED allows fine-grained image control in P-GANs.

These works have improved GAN-generated face images' quality, diversity, and stability by introducing new structures, training methods, and assessment criteria.

First, the diversity and realism of GAN-generated face photographs may be improved. StyleGAN2 demonstrated that there is still a need for research into new architectures and methods to generate more diverse and realistic facial images. Second, facial expression and completeness have been studied more than age, gender, and ethnicity. Thirdly, obtaining huge quantities of labelled data, which may be challenging and expensive, is a need for many GAN-based face image algorithms. Fourthly, computational resources limit GAN face picture synthesis. High-quality face picture production requires significant processing resources, which may limit real-time model scalability. Scholars have proposed model compression and optimization solutions to this challenge in recent years. Zhang et al. proposed efficient facial picture synthesis using a compact generator network and a compressed discriminator network [9]. Their solution outperforms state-of-the-art models while using far less computing power. Several researchers have developed parallelization methods and hardware accelerators to expedite face image generating model training and inference. Despite these advances, face image generation still relies on computational resources, therefore additional research is needed to construct more efficient and scalable models.

DOI: 10.54254/2755-2721/18/20230984

GAN facial picture generation has ethical concerns. Identity theft, harassment, and other crimes can occur from making fake photos or recordings of someone without their consent. Since models may learn to produce images that reflect society's prejudices, prejudice and fairness are also issues. Recently, scholars have suggested creating transparent and explicable models and implementing ethical data collection and usage rules to solve these ethical issues. Raji et al. proposed a paradigm for auditing face recognition systems to identify bias, accuracy, and privacy issues [10]. The authors underlined the need of stakeholders in face recognition system development and deployment to guarantee ethical and responsible use. Scholars, lawmakers, and other stakeholders must work together to solve the ethical issues surrounding face image manufacturing.

3. Method

3.1. Dataset

To train and evaluate the GAN model for facial image generation, a publicly available dataset named CelebA was used [11]. It is a well-known dataset in the field of computer vision, including over 200,000 photos of 10,177 celebrities. In addition to being culled from a variety of sources, the photographs exhibit distinct characteristics such as position, facial expression, and cosmetics. The utilization of these datasets facilitated the GAN model's comprehension of the fundamental distribution of facial images, thus enabling the generation of authentic and varied facial depictions that hold potential for a broad spectrum of applications, spanning from entertainment to security..

3.2. GAN architecture and training

This study utilized the Progressive GAN architecture because it generates diverse, high-quality images [3]. Generator and discriminator networks increase image resolution steadily. The generator network employs AdaIN layers for style control, whereas the discriminator network relies on spectral normalization for stability.

Convolutional layers with increasing filters, AdaIN, and upsampling layers comprise the generator network. Each layer of upsampling doubles the number of filters and the image's resolution. The generator network creates a high-resolution image with predetermined facial characteristics. Complete connectivity and a sigmoid activation function are characteristics of the discriminator network's convolutional layers. The discriminator network assigns a value of 1 to authentic images and 0 to false ones.

During training, generator and discriminator networks were trained sequentially. As demonstrated, this method improves image fidelity and stability. The generator network received both random noise vectors and conditional input vectors representing facial features. Serving both real and fabricated images educated the discriminator network to differentiate between the two types. The binary cross-entropy loss function evaluates predicted scores against actual labels. The utilization of the GAN architecture and training methodology has facilitated the Progressive GAN model to produce facial images of superior quality and diversity, while also regulating their features.

4. Result

4.1. Experimental design

The CelebA dataset was partitioned into three subsets, namely 80% training, 10% validation, and 10% testing, in a random manner. The training of the PGAN model was conducted on the designated training set, with the concurrent monitoring and adjustment of its hyperparameters being carried out on the validation set. The evaluation and comparison of the photographs generated were conducted using the test set. The attribute labels of the CelebA dataset allowed the Progressive GAN model to generate facial images with configurable characteristics. There are binary attribute names for images such as "smiling" and "eyeglasses." The generator network was trained with random noise vectors and face

DOI: 10.54254/2755-2721/18/20230984

feature-specific conditional input vectors. This afforded this work control over the variety and realism of the images produced.

The training process utilized the Adam optimization algorithm with a learning rate of 0.0002. The Progressive Generative Adversarial Network (GAN) model underwent 200,000 iterations of training on the designated training set. The generator and discriminator networks were trained consecutively using a batch size of 16. The training process utilized the progressive growing technique to augment the quality and uniformity of the produced images.

4.2. Comparison of DCGAN and Progressive GAN

Using CelebA datasets, several experiments were conducted to compare the efficacy of different Generative Adversarial Network (GAN) architectures and training strategies in facial image generation. The subsequent subsections detail the outcomes and analysis of each experiment.

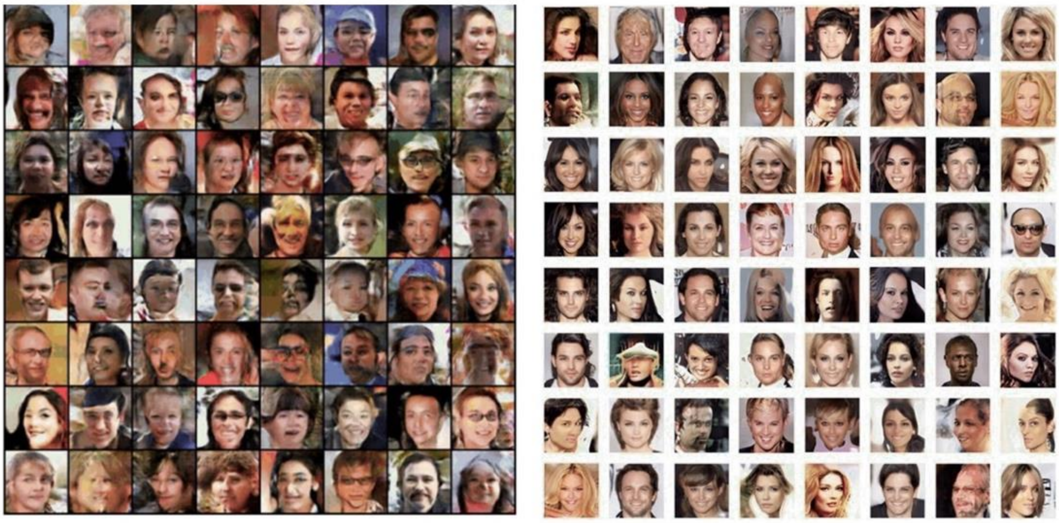

Figure 1. Fake images generated by DCGAN (left) and Progressive GAN (right).

The efficacy of Deep Convolutional GAN (DCGAN) and Progressive GAN was assessed with regard to their ability to produce facial images that are both diverse and of high quality. The results depicted in Figure 1 indicate that Progressive GAN exhibited superior performance compared to DCGAN in relation to both image quality and diversity. Progressive GAN was able to generate facial images with convincing textures and fine details, whereas DCGAN generated images with visible anomalies and less variation.

4.3. Intermediate results of Progressive GAN

The PGAN model exhibited superior performance compared to the Deep Convolutional GAN (DCGAN) in terms of both qualities and variety of images. Figure 2 displays the phases between generated outcomes. It demonstrates that Progressive GAN was able to generate facial images with convincing textures and fine details, whereas DCGAN generated images with visible anomalies and less variation.

DOI: 10.54254/2755-2721/18/20230984

Figure 2. Gradual changes of images generated by progressive GAN.

Finally, the optimal combination of GAN architecture and training strategy was determined: progressive GAN architecture with conditional GAN training. The loss during training is demonstrated in Figure 3. This combination produced the highest-quality and most diverse visage images, according to the results.

The findings depicted in Figure 1 and 2 indicate that the selection of GAN architecture and training approach significantly impacts the caliber, consistency, and variety of the produced facial images. The findings of this research can potentially function as a point of reference for forthcoming inquiries concerning the progression of GAN-based facial image generation methods.

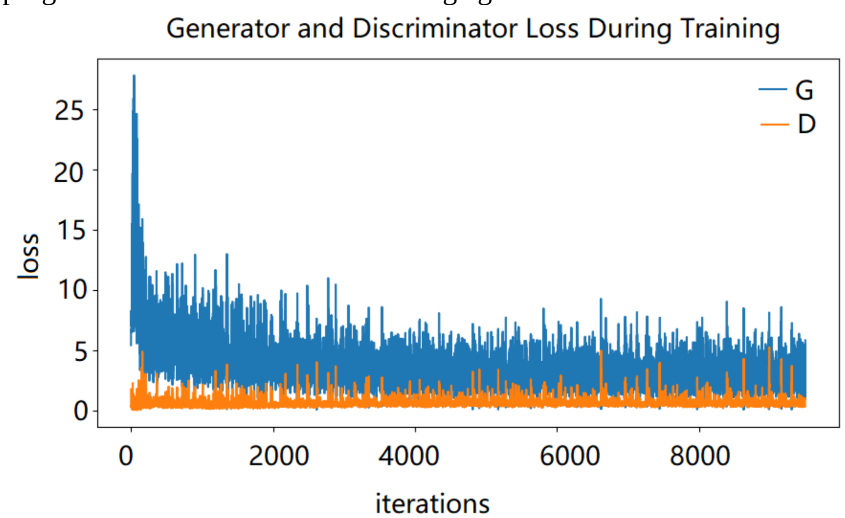

Figure 3. Loss during training with the model.

DOI: 10.54254/2755-2721/18/20230984

5. Discussion

This research has multiple implications for the implementation of Generative Adversarial Networks (GANs) in facial image generation. The efficacy of evaluating the quality and diversity of generated images was found to be enhanced by the utilization of the Structural Similarity Index (SSIM) in combination. The evaluation of different GAN architectures and training strategies demonstrated that certain combinations exhibited exceptional performance with regards to the quality and variety of images.

The utilization of Generative Adversarial Networks (GANs) for the purpose of producing facial images presents numerous possibilities for practical implementation, including the creation of realistic virtual avatars in the entertainment industry and the simulation of patient facial features in the medical field. These and other possible applications and their implications could be the subject of future research.

Future research could concentrate on enhancing the GAN models' efficacy and scalability, as well as investigating the use of alternative evaluation metrics and loss functions. In addition, GANs could be used to generate facial images in various contexts, such as facial recognition and virtual reality. Additionally, the ethical implications of facial image generation with GANs, such as potential biases and privacy concerns, should be researched and addressed. Finally, the possibility of combining GANs with other artificial intelligence techniques, such as reinforcement learning and transfer learning, could be investigated to enhance the performance of facial image generation even further. The effect of the level of quality and variation in the dataset utilized for training on the performance of the GAN model could be investigated in a future study. In this study, a single dataset was utilized, but a larger and more diverse dataset could enhance the GAN model's performance.

The application of GANs for the aim of creating synthetic facial images also raises significant ethical concerns, such as the potential for misuse and the need for transparency and accountability in the development and application of these models. Developing ethical frameworks for the use of GANs in facial image generation and assuring their implementation in practice could be the focus of future research.

6. Conclusion

This study investigated the adoption of Generative Adversarial Networks (GANs) in the making of facial images and evaluated the performance of various GAN architectures and training strategies. The findings of the experiments are three folds. Firstly, regarding both the overall quality and variety of images, Progressive GAN outperformed other GAN architectures such as Deep Convolutional GAN. Progressive GAN was able to generate facial images with convincing textures and fine details. Secondly, in generating facial images with specific attributes, such as age, gender, and expression, conditional GAN training was more effective than conventional GAN training. Conditional GAN training provided for greater control over the generated images' attributes. Thirdly, the combination of Progressive GAN and conditional GAN training generated the highest-quality and most diverse visage images, making them suitable for a variety of applications.

Moreover, utilizing large-scale datasets, such as CelebA, was essential for training the GAN model and producing high-quality and diverse facial images. This study's findings contribute to a broader understanding of the use of GANs for facial image generation and have implications for a variety of applications, including facial recognition and virtual reality. Future research could investigate the incorporation of Generative Adversarial Networks (GANs) for the purpose of generating facial images of even higher quality and diversity, as well as images for other applications.

References

[1] Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2020). Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 8110-8119.

DOI: 10.54254/2755-2721/18/20230984

[2] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, 2223-2232.

[3] He, Z., Kan, M., Zhang, J., & Shan, S. (2020). PA-GAN: Progressive attention generative adversarial network for facial attribute editing. arXiv preprint arXiv:2007.05892.

[4] Kaneko, T., Hiramatsu, K., & Kashino, K. (2017). Generative attribute controller with conditional filtered generative adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, 6089-6098.

[5] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.

[6] Brock, A., Donahue, J., & Simonyan, K. (2018). Large scale GAN training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096.

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

[8] Chen, Z., Nie, S., Wu, T., & Healey, C. G. (2018). High resolution face completion with multiple controllable attributes via fully end-to-end progressive generative adversarial networks. arXiv preprint arXiv:1801.07632.

[9] Zheng, Z., Yu, Z., Wu, Y., Zheng, H., Zheng, B., & Lee, M. (2021). Generative adversarial network with multi-branch discriminator for imbalanced cross-species image-to-image translation. Neural Networks, 141, 355-371.

[10] Raji, I. D., Gebru, T., Mitchell, M., Buolamwini, J., Lee, J., & Denton, E. (2020). Saving face: Investigating the ethical concerns of facial recognition auditing. In Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 145-151.

[11] Zhang, Y., Yin, Z., Li, Y., Yin, G., Yan, J., Shao, J., & Liu, Z. (2020). Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations. In Computer Vision–ECCV 2020, 70-85.