A Brief Analysis of GAN Variants on Image Classification and Generation

Mingjing Zhu1*

1 Highschool Affiliated to Shanghai Jiaotong University, Shanghai, China
*Corresponding author’s e-mail: mjzhu@cumt.edu.cn

Abstract. Recently, the Generative Adversarial Network (GAN) has attracted much attention in a large range of applications, such as image classification, plausible image generation and image completion. Lots of GAN variants have been proposed. This paper focused on two popular GAN variants, including GAN and Auxiliary Classifier Generative Adversarial Network (ACGAN) and made a comparison between them. The experiment on CIFAR-10 and anima dataset shows that ACGAN can perform better than original one with better accuracy. Furthermore, we find ACGAN has a worse inception score (IS), which indicates that ACGAN is not capable of generating realistic pictures.

1. Introduction

Generative Adversarial Network (GAN) attracts much attention in recent years [1,2] and has been applied to various domains, including computer vision [3,4,5,6], natural language processing [7,8,9] and temporal-based tasks [10,11]. GAN is effective in handling sharp estimated density functions, desired sample generation and deterministic bias elimination. Thus, GAN has been widely used in vision tasks, such as image generation [12,13,14], image-to-image translation and image classification. GAN is a combination of two subnetworks, namely generator and discriminator. The generator will learn from the input data and generate a fake image. The images will go to the discriminator combine with some real images. Discriminator tries to distinguish which image is real and which is fake. They act like two players that try to beat the other. The generator tries to fool the discriminator and the discriminator tries not to be fooled.

Two main research topics have been popular to improve the accuracy of GAN: (1) training improvement and (2) the deployment of GANs for real-world applications. Many works are proposed to improve the accuracy for corresponding tasks and the visual quality of generated images. Semi-supervised GAN (SGAN) is designed in the context of simi-supervised learning, a promising research area between supervised learning and unsupervised learning. SGAN does not need numerous labeled data, thereby achieving substantial results. Bidirectional GAN (BiGAN) aims to solve the problem that the original GAN network cannot project the data back to the latent space by adding an encoder. The results of MINIST dataset validate the effectiveness of BiGAN.

With the above observation, we aim to analyse various types of GAN variants, including typical GAN and ACGAN networks, in both image generation and classification tasks. We measure the results using the classification accuracy for image classification. And as for image generation, we employ the inception score to indicate the quality of generated images. We conduct several experiments on CIFAR-10 and anima datasets and conclude that ACGAN performs much better than the original GAN in handling image classification tasks. However, ACGAN is inferior to GAN with a worse inception score, which indicates the ACGAN still has room for improvement of the quality of generated images.
Our work can give guidance on selecting models in both image classification and generation. Furthermore, in the discussion section, we discuss the trade-off between the classification accuracy and generation quality, which can help researchers select proper algorithms in different tasks. The following sections are summarized as follows. In section 2, we detail our selected methods, a typical GAN and ACGAN, as well as dataset description and implementation details.

2. Method

2.1. Generation Adversarial Network (GAN)

Figure 1 demonstrates the architecture of a typical GAN network [16]. The framework contains two components, the discriminator D and the generator G. The discriminator D distinguishes the real samples and the fake samples generated by the following generator G. The generator G aims to create images to fool the discriminator with the input distribution Nz. The basic formula to train GAN is:

\[
\min_{G} \max_{D} V(D, G) = \mathbb{E}_{x \sim P_{data}} \log D(x) + \mathbb{E}_{z \sim P_{z}} \log (1 - D(G(z)))
\]  (1)

G is Generator, D is Discriminator, P data(x) is distribution of original data, P(z)(z) is input noise variables. The adversarial modeling framework (min, max) is used with multiple components.

![Figure 1. The framework of the original GAN network](image)

2.2. Auxiliary Classifier Generative Adversarial Network (ACGAN)

ACGAN [17], as known as auxiliary classifier generative adversarial networks, was proposed in 2017. Each generated sample in ACGAN has a corresponding class label Ly in addition to the noise Nz. The framework of ACGAN is shown in Figure 2.

The loss function of ACGAN can be expressed as follow:

\[
L_s = \mathbb{E}[\log P(S = \text{real}|X_{\text{real}})] + \mathbb{E}[\log P(S = \text{fake}|X_{\text{fake}})]
\]  (2)

\[
L_c = \mathbb{E}[\log P(C = c|X_{\text{real}})] + \mathbb{E}[\log P(C = c|X_{\text{fake}})]
\]  (3)

Ls is the log-likelihood of the correct source, which is for the noise. LC is the log-likelihood of the correct class, which is for Ly. D trained to maximize Ls+Lc, whereas G trained to maximize Lc-Ls. The proposed ACGAN achieves improvement on the visual quality of the generated images and has a higher model diversity.

We note that the main difference between GAN and ACGAN is in the discriminator part. ACGAN is an extension of the conditional GAN that changes the discriminator to predict the class label of a given image rather than receive it as input.

2.3. Dataset Description

**CIFAR-10 Dataset.** CIFAR-10 dataset is one of the most popular datasets used for training machine learning and deep learning-based methods, which contains 60000 image samples whose size is 32 * 32. All the images are colorized images with 3 channels and they are categorized into 10 different classes.
The classes represent airplane, car, bird, cat, deer, dog, frog, horse, ship and truck. The samples are processed and balanced. Each of the categories contains the same number of images. An example of the CIFAR-10 dataset is shown in Figure 3.

**Figure 3.** The example of the CIFAR-10 dataset. 10 categories are contained in the dataset.

Anima Dataset. In this section, we describe our dataset. We employ an anima dataset containing multiples 64*64 animation people to train our network. The dataset is from Mr. Li, which is publicly accessed. The dataset contains various female animated characters, with lots of styles and appearances. Furthermore, a few male individuals are also included in this dataset. An example of the amima dataset is shown in Figure 4.

**Figure 4.** The example of the anima dataset.
Female faces of various types and appearances are contained in the dataset.
2.4. Implementation Details
In our experiment, we utilize the same setting for training origin GAN and ACGAN. The discriminator trains 5 times in each epoch, whereas the generator trains 1 time in each epoch to ensure that the discriminator can justify the images correctly. The learning rate is 0.0001. Lambda is set to 5, which is a value that shows the relationship between classifications and generation. We test the model when training around 250 epochs, where both models have already been convergent. We also find that the ACGAN has a fast convergence speed than the original GAN. Both of our methods are implemented using Python with Pytorch deep learning platform.

3. Experiments and Discussions

3.1. Comparison Results on Image Classification
We conduct an analysis of image classification on the anima dataset. The results of GAN and ACGAN in sorting various types of images are depicted in Figure 5 and Figure 6, respectively. The characters in each image have different color in hair and eyes, which needs to be correctly categorized. The results are satisfying with few images are mistaken, which validates the effectiveness of ACGAN in image classification.

![Figure 5. The experimental results on classifying anima dataset using a typical GAN](image-url)
3.2. Comparison Results on Image Generation

This section shows the results on image generation mission using both original GAN and ACGAN algorithms. The analysis is conducted in the CIFAR-10 dataset, and the inception score is employed to measure the quality of the generated images. The comparison result is shown in Table 1. Though ACGAN has a superior performance in sorting different types of characters in the anima dataset, the generated images are not as good quality as the original GAN. The inception score of ACGAN is 5.2, while it of typical GAN is 5.7. Furthermore, we show the images generated by GAN and ACGAN in Figure 6 and Figure 7, respectively. Both quantitative and qualitative results show the typical GAN outperforms ACGAN in image generation with a margin.

Table 1. The Inception Score of GAN and ACGAN in the CIFAR10 Dataset

| Methods | Inception Score |
|---------|-----------------|
| GAN     | 5.7             |
| ACGAN   | 5.2             |
3.3. Discussion

The relationship between classification accuracy and image quality. In the original paper, ACGAN is described as “more discriminable” and “performs a naive resize operation”. In our experiment, ACGAN does perform great on clarity. In the CIFAR-10 dataset, we can also see that ACGAN is more stable than simple GAN. Pictures generate by ACGAN have similar quality. It can also generate more realistic outputs within fewer epochs. The accuracy of the generated image various from different datasets, but we can say that ACGAN is more accurate than GAN.

The IS value of the CIFAR-10 dataset may indicate that GAN and ACGAN cannot generate various of very realistic images because both of their values are very low. The IS value of ACGAN is even lower.
than GAN though it can generate more accurate images. This might because ACGAN will sacrifice some of the ability of generating images to classify them.

The paper also mentions some downsides of ACGAN, which is its visual discriminability of the 128 × 128 resolution model needs to be improved. This can also be seemed even in the CIFAR-10 dataset.

**The analysis of lambda value.** When we change the lambda value, the accuracy of classifying the outputs and the quality of outputs changes. As the classification increase, the quality decreases. This might because ACGAN needs to separate its computing power into two sections. In order to find a balance, the reality of the image and the accuracy of classification might need to be sacrificed.

4. **Conclusion**
Generative Adversarial Network has been more and more popular in various research areas. This paper analyses two typical GAN variants, including original GAN and ACGAN, in handling image generation and classification tasks. Compared with the original one, ACGAN achieves a better quality of generated images and effectively classifies which one is a fake image. Furthermore, we apply the inception score (IS) to evaluate the quality of generated images by those two GAN based algorithms. Though the accuracy is satisfying in the image classification tasks, the quality of generated image output by ACGAN needs to be improved. A new improved approach based on ACGAN is required to solve this problem. Our work can give a brief introduction on selecting a different type of GAN based algorithms and make a balance in classification accuracy and generated image quality.

**References**

[1] Jabbar, Xi L., Omar B. (2020) A Survey on Generative Adversarial Networks: Variants, Applications and Training. In Arxiv.

[2] Wang Z., Qi S., Ward T. E. (2020) Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy. In Arxiv.

[3] Dziugaite G. K., Roy D. M., and Ghahramani Z. (2020) “Training generative neural networks via maximum mean discrepancy optimization,” In Arxiv.

[4] Ma L., Jia X., Sun Q., Schiele B., Tuytelaars T., and Van Gool L. (2017) Pose guided person image generation,” in NIPS.

[5] Vondrick, Pirsiavash H., and Torralba A. (2016) Generating videos with scene dynamics,” In NIPS.

[6] Yang, Lu X., Lin Z., Shechtman E., Wang O., and Li H. (2017) High-resolution image inpainting using multi-scale neural patch synthesis, In CVPR.

[7] Yang Z., Hu J., Salakhutdinov R., and Cohen W. (2017) Semi-supervised QA with generative domain-adaptive nets, In ACL.

[8] Dai Z., Yang Z., Yang F., Cohen W. W., and Salakhutdinov R. R. (2017) Good semi-supervised learning that requires a bad GAN, In NIPS.

[9] Jetchev N., Bergmann U., and Vollgraf R. (2016) Texture synthesis with spatial generative adversarial networks, In Arxiv.

[10] Donahue, McAuley J., and Puckette M. (2018) Synthesizing audio with generative adversarial networks, In Arxiv.

[11] Hartmann K. G., Schirrmeister R. T., and Ball T. (2018) EEG-GAN: Generative adversarial networks for electroencephalographic brain signals, In Arxiv.

[12] Karras T., Laine S., and Aila T. (2018) A style-based generator architecture for generative adversarial networks, In Arxiv.

[13] Wang T.-C., Liu M.-Y., Zhu J.-Y., Tao A., Kautz J., and Catanzaro B. (2018) High-resolution image synthesis and semantic manipulation with conditional GANs, in CVPR.

[14] Poole B., Alemi A. A., Sohl-Dickstein J., and Angelova A. (2016) Improved generator objectives for GANs, In Arxiv.