Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/27/20230189

Combine model fine-tuning freezing layers and adaptive filter modulation to implement transfer learning for GANs

Chen Yang

School of Computer Science & Informatics, Cardiff University, Cardiff, CF24 3AA, United Kingdom,

rickyyang1997@163.com

Abstract. Generative Adversarial Network (GAN) requires more resources to train than other deep learning models and its loss function converges more slowly. For this reason, scholars at home and abroad have proposed a GANS algorithm based on transfer learning, which is applied to fewer samples, thus improving the training effect of GANS algorithm. In this paper, we provide a new way to perform the transfer of genetic algorithms and combine the two ways. On this basis, we will compare and analyze a variety of transfer learning algorithms to verify the feasibility and effectiveness of the joint application.

1. Introduction

Generative Adversarial Network (GAN) is a breakthrough model that has been widely used since its appearance, such as: data augmentation, style transfer. The GAN is a combination of two networks (a discriminator and a generator) that are optimised by conflict each other during training [1]. However, with the widespread use of GAN, a number of problems have been identified by users, such as the instability of GAN training and difficulty in convergence, pattern collapse and blurred images generated by training on smaller datasets [2]. Therefore, in order to solve these problems, GAN has been continuously improved, like DCGAN (improve the network structure),WGAN(improve the loss function) and ProGAN(improve the training method). These improvements allow GAN to generate higher-quality images, and higher-pixel images (1024*1024) can also be generated.

However, after these improvements were proposed, the training cost of GANs was greatly increased. Obviously, this situation is not conducive to researchers who do not have powerful GPU resources and scientific research institutions or companies will waste more resources if they want to train a GAN that generates high-resolution images. In order to make the loss function of GAN converge quickly and use the minimized data set for training, some researchers try to use transfer learning technology in the training process of GAN [3]. In machine vision, when the source and object regions are similar, the transition learning method is often used to improve the training effect of the model. In addition, transfer learning can also increase the usage rate of pre-trained models, which means it can allow pre-trained models to be used in more tasks [4]. Transfer learning can maintain the commonness of the two fields, so when it is trained, it can be relearned, thus reducing the resource consumption of the system.

One of the most common migration methods is to introduce samples from existing object regions into existing model adjustments. For GANs, more transfer learning techniques need to be explored, such as: freezing layers, adaptive filter modulation [5]. Furthermore, combining these techniques will

DOI: 10.54254/2755-2721/27/20230189

maximise the advantages of these technologies. To demonstrate this conclusion, this project intends to combine the two methods to verify the theory.

The following is an outline of this paper. The Section 2 will introduce the related work, such as GAN and transfer learning. The Section 3 is methodology which describes in detail the methods I have used in this work. The Section 4 will show the experimental results. Finally, I will present my conclusions and the problems I found in the experiments in Section 5.

2. Related work

This part will introduce the related work of this project which can help readers understand this paper more efficiently. First of all, I will explain the related knowledge of GAN and show some common improvements to GAN. Then some transfer learning methods will be introduced.

2.1. Generative adversarial network (GAN)

In the research of machine learning and deep neural network, the effective training of the model often depends on a large amount of data. However, in some cases, such as some rare medical images, it is difficult to collect large amounts of information. To extend the original dataset, Goodfellow proposed a revolutionary model in 2014 which is called Generative Adversarial Network (GAN) [1]. GAN network consists of two neural networks, which are generator and discriminator. During the training process, these two networks are optimized by confronting each other [1]. For example, if the loss function of the producer decreases, the loss function of the discriminator will also increase. Finally, after weighing with the discriminator, a similar image is produced. However, a number of problems were found in the training phase of the GAN, including pattern collapse (the generated images were very similar), gradient disappearance and training instability. Because to these problems, there is a high probability that the generator will not produce high quality images. In addition, GAN is almost impossible to generate high-resolution images. In general, GAN was not suitable for most scenarios when it was first invented.

If you want to deal with the above problems well, GAN has been improved much times since it was proposed, such as: DCGAN, WGAN, ProGAN, BigGAN, StyleGAN. Alec,R. and Luke,M. proposed DCGAN in 2016 that improves the network structure of GAN. The author's main idea is to replace the linear layer in GAN with convolutional layesr/deconvolutional layers and use batchnormal for the discriminator and generator [6]. Compared with GAN, the network structure of DCGAN is more stable during training. In addition, WGAN improves the loss function of GAN it solves pattern collapse and gradient disappearance [2].

2.2. Transfer Learning

Transfer learning refers to a new method that applies new, new and meaningful methods and methods acquired by one discipline to other disciplines or problems. For example, when we train a classifier for an animal face, the classifier can distinguish the faces of cartoon characters. Because cartoon faces and animal faces have similar features, transfer learning reuses these features to improve training efficiency and increase training speed. In this method, the first step is to preview a sample called the source domain. Then this pre-trained model can be used in a data set similar to the source domain, which is called the target domain [4]. One of the most common methods is to introduce the data fields to be processed into existing data sets. This method improves the existing prediction algorithm so that it can work normally in the target area [3]. In addition, some researchers have found that the layers far away from the loss function in the neural network contain common features and freezing these layers can further increase the training speed [7].

3. Methodology

This paper proposes a combination of three transfer learning methods for GAN, including: Model Fine-tuning, Freeze Layers and Adaptive Filter Modulation. First I will train a pre-trained model in the source domain and use the improved GAN in order to make it more stable. Then, I will compare the results of using these transfer learning techniques individually and using them in combination.

3.1. Pre-trained model

As described in the Section 2, GAN has many problems like: mode collapse, gradient disappearance, unstable training. Therefore, I refer to the network structure of DCGAN in building the generator and discriminator.

In order to avoid mode collapse and gradient disappearance, I used WGAN's loss function and used the RMSprop optimizer [2]. The loss function of WGAN is based on the Wasserstein distance (Earth-Mover distance) and it is shown in the Formula 1:

$$\mathrm{E}_{\mathrm{x}\sim\mathrm{P}_\mathrm{g}}[\mathrm{f}_\mathrm{w}(\mathrm{x})] - \mathrm{E}_{\mathrm{x}\sim\mathrm{P}_\mathrm{r}}[\mathrm{f}_\mathrm{w}(\mathrm{x})]$$

Formula 1. The loss function of WGAN [2]

Compared with KL dispersion and JS dispersion, Wasserstein spacing has greater advantages, because Wasserstein spacing can reflect the relationship between the two distributions.

3.2. The combination of transfer learning

The core of this paper is the combination of transfer learning techniques. First, the following part will introduce them separately. Then I will analyze the advantages of using them in combination based on their characteristics.

1) Model fine-tuning: After a deep learning model is trained on a certain dataset, some of the features that the layers in this model will learn are available in other datasets. Therefore, the parameters of the pre-trained model only need to be fine-tuned when it is trained on other data sets.In addition, model fine-tuning is also applicable to GAN, even if the distance between the source and target domains is large [5].

2) Freezing layers: The principle of freezing layers is to fix the layers which contain common information in the neural network [7]. For example, researchers often freeze convolutional layers and optimise fully connected layers when solving image classification problems.

3) Adaptive Filter Modulation: This technique was improved from the freeze layers by adding two learnable parameters to the fixed convolutional layer and the authors gave the a diagram which will be shown on the Figure 1. Compared with model fine-tuning, this method can effectively avoid overfitting.

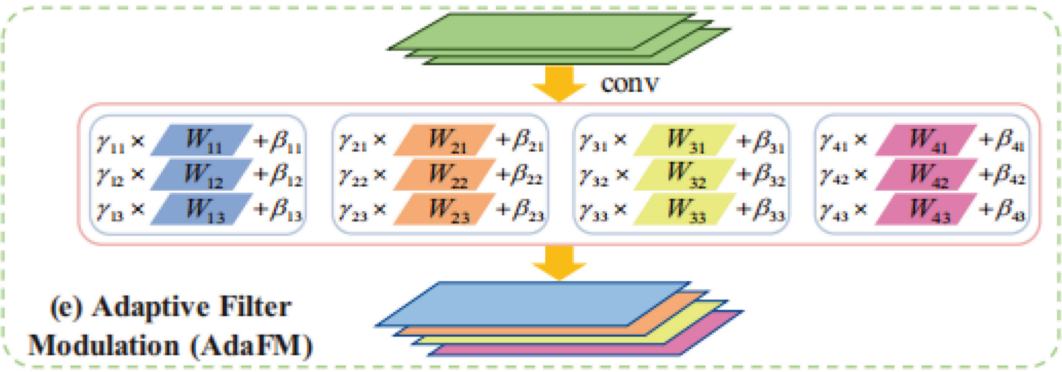


Figure 1. Adaptive Filter Modulation [8]

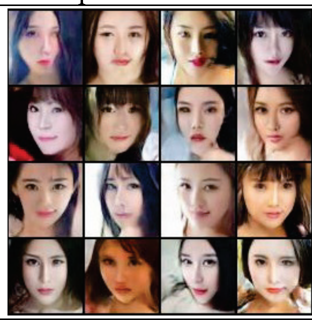
The purpose of this method is to maintain the commonness of the existing dataset in the source domain and to make the existing dataset have strong adaptive ability. First, the model fine-tuning directly optimizes all parameters so I will use it in special parts of the network. Then, the frozen layers can be used in the public part of the network. Finally, I think that some of the layers are between the special and public parts, which will be referred to as the **"middle part" i**n this paper. I will use adaptive filter modulation for these layers.

4. Experimental results

4.1. Train a pre-trained model in the source domain

This topic intends to pre-train the Asian face database and transplant it on the animated face database (the database comes from) [9]. First, I build the generator (Deconvolution Layer + BatchNorm2d + ReLU/Tanh activation function) and the discriminator (Convolution Layer + LeakyReLU activation function + Linear layer) based on DCGAN [6]. After that, WGAN's loss function is used in the training process and after each optimization the discriminator parameters are clipped to between -0.01, 0.01 [2]. Under the condition of 0.00005, the generator and discriminator are optimized by RMS algorithm. Table 1 shows the images generated from a trained schema and a picture obtained in a database.

Table 1. Comparison of training set and images generated by GAN.

Images in the dataset	Images generated by the pre-trained model
	

4.2. Implementing transfer learning on a limited dataset

The Section3 analyzes three transfer learning techniques and they need to be used in different parts of the network. After comparing different combinations, I found the best combination scheme that can generate higher quality images. For the pre-trained model used in this project, the optimal transfer learning combination scheme is shown by the Figure 2.

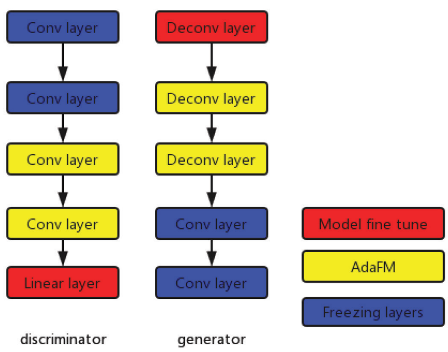


Figure 2. Combination scheme of transfer learning technology.

4.3. Comparison with other transfer learning techniques

In this subject, I use FID technology to evaluate the image quality produced by gallium nitride. In some examples, the training group of the resulting model will be different from the original version 3 [10]. Moreover, the FID is calculated using both generated and real data, which is more reasonable than IS [11]. The Table 2 compare the use of transfer learning methods in combination with using transfer learning methods alone.

DOI: 10.54254/2755-2721/27/20230189

Table 2. Results and FID of different transfer learning techniques.

	Model fine-tuning	Freezing layers	AdaFM	Combine them
Images				
FID	185	226	189	**157**

5. Conclusion

In this paper, I use a combination of three transfer learning techniques that allow the generator to generate better images. In addition, it demonstrates that there is great potential for fusion transfer learning techniques. In many scenarios, using a specific combination of transfer learning can make transfer learning more successful than using a particular technique alone. In general, this paper demonstrates that combining transfer learning techniques can significantly improve training efficiency and reduce the training cost of GANs.

However, only a limited number of GANs were used in this project and more transfer learning techniques need to be explored. Therefore, some work needs to be done in the future, such as using more GANs as pre-trained models and combining more transfer learning techniques.

References

[1] Goodfellow,I.J. et al. (2014) Generative adversarial nets. Advances in neural information processing systems, 27:2-9.
[2] Arjovsky,M., Chintala,S., Bottou,L., (2017) Wasserstein GAN. arXiv preprint. arXiv:1701.07875.
[3] Wang,Y.X. et al. (2018) Transferring GANs: generating images from limited data. In: ECCV, pp. 2-17.
[4] Zhao,W. (2017) Research on the deep learning of the small sample data based on transfer learning. In: AIP Conference Proceedings.
[5] Fr´egier, Y., Gouray, J.B. (2019) Mind2Mind : transfer learning for GANs. arXiv preprint arXiv:1906.11613v1.
[6] Radford, A., Metz, L., Chintala, S. (2016) Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. arXiv preprint arXiv:1511.06434
[7] KAIST, S.M., POSTECH, M.C., KAIST, J.S. (2020) Freeze the Discriminator: a Simple Baseline for Fine-Tuning GANs. arXiv preprint arXiv:2002.10964v2.
[8] Zhao, M.Z., Cong, Y.L., Carin, L. (2020) On Leveraging Pretrained GANs for Generation with Limited Data. arXiv preprint arXiv:2002.11810v3.
[9] Gwylab. (2019) Download and introduction of the data set. http://www.seeprettyface.com/mydataset.html.
[10] Maximilian, S. (2020) pytorch-fid: FID Score for PyTorch.Version 0.1.1. https://github.com/mseitzer/pytorch-fid.
[11] Xu, Q.T. et al. (2018) An empirical study on evaluation metrics of generative adversarial networks. arXiv preprint arXiv:1806.07755v2.