The research process of generative adversarial networks

Wenjie Luo1*, Pei Wang2, Jiahui Wang1 and Wei An1
1School of Electronic Science, National University of Defense Technology, Changsha, China
275831 forces, Guangzhou, China

*Corresponding author e-mail: 838244306@qq.com

Abstract. This paper reviews the research progress of the generative adversarial nets (GAN). Firstly, we introduce the principle and structure of GAN. Secondly, we explain its training method and the classification and typical types of GAN, such as fully connected GAN and conditional GAN. Finally, we give a brief introduction of GAN in the intelligent life, and show the outstanding significance in the field of image processing.

1 Introduction
Generative Adversarial Nets (GAN) [1], as one of the generative approach, is the most popular neural networks. It can be used in a variety of applications, including image synthesis, semantic image editing, image style conversion, image super-resolution and automatic driving. Different from other network structures, such as forward propagation neural network, convolutional neural network, recurrent neural network, GAN consists of generative network G and discriminant network D. Generating network and discriminating network is a confrontational relationship. Generating network can generate result which is expected from a set of random number, and the result can match as possible as the real data. Discriminant network can be trained to distinguish whether the result is from training data or generated by the generative network. After many rounds of iteration, the generator can generate a very realistic result, so that the discriminator can’t judge the faked result. The structure of GAN is shown in Fig.1

![Figure 1. The structure of Generative Adversarial Nets](image-url)

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
2 The training procedure of GAN

The optimization procedure is a minimax two-player problem, fixed one model during training and updated the parameters of another model. The training goal of GAN is to obtain the parameters that maximize the discriminator classification accuracy and to obtain the generator parameters that deceive discriminator as much as possible. The cost training process can be evaluated by cost function \( V(G,D) \), and the training process can be expressed as:

\[
\max_D \min_G V(G,D)
\]

where

\[
V(G,D) = \mathbb{E}_{x \sim p_{data(x)}} \log D(h) + \mathbb{E}_{z \sim p_{noise(z)}} \log(1 - D(z))
\]

where \( h \) represents the true training samples and \( z \) represents the generating data.

During the training processing, the parameters of one model are updated while the parameters of the other are stay the same.

3 The classification of GAN

3.1 Fully connected based GAN

In fully connected network based GAN, both the structure of generative network and discriminant network are fully connected network. Due to the limitation of the numbers of parameters, they can be only applied on simply dataset, such as MNIST, CIFAR-10, etc.

The original GAN has difficulty in training, and the loss function cannot guide the training process of GAN. Gulrajani [2] proposed Wasserstein GAN (WGAN), which theoretically analysed the reasons for the difficulty of traditional GAN training, and designed the Wasserstein distance, the definition is as follows:

\[
W(P_r, P_g) = \inf_{\gamma \in \Pi(P_r, P_g)} \mathbb{E}_{(x,y) \sim \gamma} \left[ \|x - y\| \right]
\]

then Formula 1 is converted to

\[
W(P_r, P_g) = \frac{1}{K} \sup_{\|f\| \leq K} \mathbb{E}_{x \sim P_r} - \mathbb{E}_{x \sim P_g} \left[ f(x) \right]
\]

The distance \( W(P_r, P_g) \) is smooth and can be optimized using the gradient descent method.

![Figure 2. The conditional variables of CGAN](image-url)
3.2 Conditional GAN
In the case of a large number of pixels for a picture, the training method is free and the result is hard to predict. Mirza et al. proposed the conditional generative adversarial network (CGAN) [3]. It introduced conditional variables in both the generation model G and the discriminant model D. As shown in Fig 2, the conditional variable can be category label or other data, it can guide the generator to generate new data. The loss function of the conditional adversarial network is as follows:

\[ V(G, D) = E_{h,l\sim p_{data}(h,l)} \log D(h,l) + E_{z\sim p_{data}(z)} D(G(z,l),l) \log(l - D(G(z,l),l)) \]  

(4)

where \( l \) is the conditional variables.

3.3 Convolutional based GAN
Since the convolutional neural network has great advantages on image processing, the GAN network also uses the structure of the convolutional network, such as LAPGAN [4], DCGAN.

Different from other image generation algorithms, LAPGAN can generate high resolution images by Laplace pyramid, which can generate a series of high resolution images from low resolution images.

\[ L_i = G_i - \text{UP}(G_{i+1}) \otimes G_{5x5} \]  

(5)

where \( G_i \) represent the image of the i-th layer. \( \text{UP}() \) represents the upsampling operation, symbol \( \otimes \) represents the convolutional operation.

Similar to LAPGAN, SimGAN [5] can make the image more realistic by unlabelled real images.

In the convolutional image processing method, the Laplacian pyramid needs to be generated by high resolution image. Denton used CGAN to generate Laplacian pyramid.

Radford et al. proposed a deep convolutional generative adversarial networks (DCGAN), and use the convolutional network for unsupervised training, the learning effect of the generated network is greatly improved.

3.4 other GAN
Chen et al. proposed InfoGAN [6], which adds part of the judgement of mutual information. Mutual information can be used to measure the degree of information in the random variable X containing the random variable Y. While judging the picture generated by generator, the value of the mutual information needs to be calculated.

4 The application of GAN

4.1 Image generation
Ledig [7] et al. proposed to use GAN to transform a low resolution image into an image with rich detail. Phillip et al. proposed pix2pix GAN to complete image translation tasks. The input and output of pix2pix GAN are both images and the generated images is matched to the input image. However, the network needs to input images that are paired with each other. Such training data is really seldom. Cycle GAN [8] just need to input two different styles of pictures to train the network to generate different styles of pictures. It is of great significance in image style transition.

4.2 In automatic drive
Santana et al. use GAN to generate the image that is consistent with the actual traffic scene [9].

5 Conclusion
This paper reviews the research progress of the generative adversarial nets (GAN). After the GAN was proposed, it was immediately focused by the researchers. GAN is a generative model, it learns to estimate the potential distribution of new data. This ability to generate new samples makes it has great application in the field of image processing, automatic driving and so on.


References

[1] Goodfellow, I., et al. Generative adversarial nets. in Advances in neural information processing systems. 2014.

[2] Gulrajani, I., et al. Improved training of wasserstein gans. in Advances in Neural Information Processing Systems. 2017.

[3] Mirza, M. and S.J.a.p.a. Osindero, Conditional generative adversarial nets. 2014.

[4] Denton, E.L., S. Chintala, and R. Fergus. Deep generative image models using laplacian pyramid of adversarial networks. in Advances in neural information processing systems. 2015.

[5] Shrivastava, A., et al. Learning from Simulated and Unsupervised Images through Adversarial Training. in CVPR. 2017.

[6] Chen, X., et al. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. in Advances in neural information processing systems. 2016.

[7] Ledig, C., et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. in CVPR. 2017.

[8] Zhu, J.-Y., et al., Unpaired image-to-image translation using cycle-consistent adversarial networks. 2017.

[9] Santana, E. and G. Hotz, Learning a driving simulator. arXiv preprint arXiv:1608.01230, 2016.