Image Transformation Based on Generative Adversarial Networks

Chen Jie
School of Computer Science and Engineering
Xi’an Technological University
Xi’an, China
e-mail: 1181814961@qq.com

Zhao Li
School of Computer Science and Engineering
Xi’an Technological University
Xi’an, China
e-mail: 332099732@qq.com

***Abstract*—Image transformation is a hot topic in the field of computer vision. Its purpose is to use the existing similar or semblable images to learn the mapping between the input image and the output image. Generative Adversarial Networks model is a powerful generative model with the idea of Zero-sum game theory, Co-training the two through the confrontation learning method of the generator and the discriminator, so as to estimate the potential distribution of data samples and generate new data samples. This paper is based on the generative adversarial networks and the existing network, the image transformation is realized by combining the two network structures of DCGAN and Cycle GAN. Experimental results show that this method not only effectively solves the problem that paired images are not easy to obtain, but also fully demonstrates the superiority of the generated adversarial networks in image transformation.**

Keywords-Image Transformation; GAN; Deep Learning; CNN; Adversarial Training

In recent years, with the rapid development of artificial intelligence, machine learning methods represented by statistical machine learning and deep learning are one of the main research directions [1].Among them, the model of deep learning can be divided into discriminant model and generative model. Because of the invention of Logistic Regression, Support Vector Machine, Conditional Random Field and other algorithms, the discriminant model has developed rapidly. However, the development of generative model is slow because of the difficulty of generative model modeling. In recent years, until the invention of the most successful generation model--Generative adversarial networks model, this field has been revitalized. Since its introduction, generative adversarial networks has been receiving great attention from the academic and industrial circles. With the rapid development of GAN in theory and model, it has been applied more and more in-depth applications in computer vision, natural language processing, human-computer interaction and other fields, and continues to extend to other fields.

I. THE BACKGROUND OF THE GAN

GAN is developed based on the depth generative model. This section first briefly introduces the basic ideas and development history of deep learning, artificial intelligence and depth generative model so as to better understand the research progress and application field of GAN.

A. Deep learning and artificial intelligence

In recent years, with the improvement of computing power and the dramatic increase of data volume in various industries, Artificial intelligence has achieved rapid development, which makes researchers pay more attention to artificial intelligence and the public's expectation of artificial intelligence unprecedented improvement. Deep learning is a method of realizing artificial intelligence. Artificial Intelligence is a broad concept, the purpose of Artificial Intelligence is to make computers think like human beings. Machine Learning is a branch of Artificial Intelligence, which focuses on how computers simulate or implement human Learning behaviors to acquire new knowledge or skills so that they can continuously improve their performance. Deep Learning is a method of machine learning, which attempts to abstract data at a high level using multiple processing layers consisting of complex structures or multiple non-linear transformations. Compared with general machine learning methods, the main difference of deep learning is that it does not rely on manual feature extraction. Deep learning uses Multilayer neural network to characterize data learning. A neural network is a group of algorithms that roughly imitate the structure design of human brain to recognize patterns. The patterns it can recognize are numerical, so all real-world data such as images, sounds, text and time sequences must be converted to numerical values. The neural network interprets the

DOI: 10.21307/ijanmc-2019-053

sensor data through the machine perception system and can mark or cluster the original input data.

In terms of model, algorithm and hardware facilities, deep learning has changed the difficulties of traditional neural network optimization, limited application, slow calculation and low recognition, making the influence of deep learning expanding. At present, deep learning has become a mainstream method in artificial intelligence research. The breakthrough of deep learning in supervised learning tasks, especially in the field of image, is particularly noticeable. Compared with traditional neural network method, deep learning has made breakthroughs in the following aspects: Convolutional neural network and Recurrent neural network are used in this paper. These new network structures greatly enhance the modeling ability of the neural network. New activation functions, regularization methods and optimization algorithms such as Rectified Linear Unit (ReLU), Dropout and Adam are used. These new training techniques effectively improve the convergence speed of the neural network and make large-scale training of the neural network possible. New computing devices such as Graph-processing Unit (GPU), distributed system and so on are used. These devices make the training time of the neural network greatly shorten, so it has the possibility of actual deployment [2] [3].

B. *Depth generative model*

Generation method and discrimination method are two important branches of supervised learning method in machine learning. The generative method has great research value, involving the assumption of distribution of explicit or implicit variables of data, learning the fitting and training of distribution parameters or models containing distribution, and generative new samples according to the learned distribution or model. Generative model is a model acquired by generative method learning, which occupies an important position.

In the early research of deep learning, people have made a series of attempts in order to achieve good results in generative models. For example, for maximum likelihood estimation of real samples, parameter updates come directly from data samples, which limit the learning of generative models. Because the target function is difficult to be solved, the generative model obtained by the approximation method can only approach the lower bound of the target function in the learning process and is not a direct approach to the target function. Markov chain method can be used for both the training of generating model and the generative of new samples, but the computational complexity of markov chain is high [4].Generally, the generated model established does not directly estimate or fit the distribution, but obtains the sampled data from the distribution that has never been explicitly assumed, and modifies the model through these data [5].In this way, the generated model lacks interpretability, and there will be over-fitting phenomenon, which cannot be well generalized to generate diverse samples. To solve this problem, the researcher has proposed a method named Stochastic back-propagation [6].By adding additional random noise z independent of the model, the deterministic neural network f(x) can be transformed into a random f(x,z),and trained by back propagation method. This method can improve the diversity of the generated model output samples.

II. INTRODUCTION OF GAN

A. *The basic idea*

Generative adversarial networks is a generative model proposed by Goodfellow etal in 2014.The generative adversarial networks is composed of generator and discriminator. The purpose of a generator is to estimate the distribution of a data sample from a given noise and generate synthetic data. The purpose of a discriminator is to distinguish the input data from the data generated by the generator or the real data. The relationship between generator and discriminator is adversarial. The idea of confrontation comes from the zero-sum game in game theory. In an equal game, the two players of the game use each other's strategies to change their own confrontation strategies, so as to achieve the goal of winning [7].Extended to the generating confrontation network, that is, the generator and discriminator are both players in the game, and the optimization goal is to reach the Nash equilibrium [8]. The generator tries to produce more real data. Accordingly, the discriminator tries to more accurately distinguish between real data and generated data. As a result, the two networks make progress in the confrontation, and continue to fight after the progress. The data obtained from the generated network will be more and more perfect and approximate to the real data, so that the desired data can be generated.

B. *The training principle*

Let z be random noise and x be real data. The generator generating the adversarial network is G and the discriminator is D. The training process of GAN is as follows: sampling random variable z from a probability distribution PZ(gaussian distribution or normal distribution) as the input of generator G, random noise z passes through the nonlinear mapping function of generator G and the output is as close as possible to the generated data G(z) of real data distribution Pdata. The discriminator D takes G(z) or x as input, by calculating the probability that it belongs to the real data, it can judge whether the input data comes from the real data or from the data generated by the generator. Generator G and discriminator D generally adopt highly nonlinear and differentiable deep neural network structure, such as multi-layer perceptron [9].Therefore, end-to-end learning strategies can be used for training. The process of GAN is shown in Figure 1.

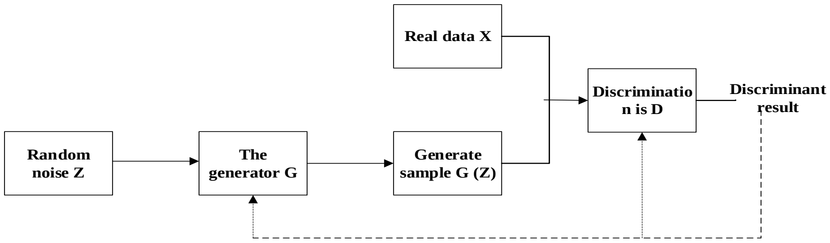

Figure 1. Data flow diagram of GAN

In the training of G and D, we adopt the strategy of antagonistic learning to make their training objectives opposite. The objective of training discriminator D is to maximize the logarithmic likelihood function, judge G(z) as generated data and judge x as real data. E() is the calculation of expected value, x is sampled from the real data distribution Pdata(x), z is sampled from the prior distribution Pz(z).In contrast, the goal of G minimize the logarithmic likelihood function. In order to learn the distribution of data x, the generator constructs a mapping space G(z;θ g) from the prior noise distribution Pz(z), which Pg the distribution of G(z) Pg to approximate the distribution Pdata of real data x. The performance of discriminator D and generator G is continuously improved by iterating the training process and updating the parameters of discriminator D and generator G alternately. When the equilibrium state is reached, it is considered that the generator has learned the distribution space of the real data x. At this time, G(z) and x have no difference in the distribution, and the discriminator D cannot make a correct judgment on the data source.

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)] + E_{z\sim P_z(z)}[\log(1-D(G(z)))] \tag{1}$$

The logD(x) of the first term represents the judgment of real data by discriminator, and the second log(1- D(G(z)) represents the judgment on the synthetic data. Through such a max-min game, G and D are respectively optimized alternately to train the required generated network and discriminant network, until the Nash equilibrium point is reached.

III. IMAGE TRANSFORMATION BASED ON GENERATIVE ADVERSARIAL NETWORKS

A. Training process based on generative adversarial networks image transformation

As a medium of transmit information, images can be expressed in many ways, such as grayscale, color, sketch, gradient, etc. Image transformation is to map a known picture into another required picture, that is, pixel to pixel mapping, such as a known sketch, and then generate a color photo. Over the years, these problems have often been solved in a specific way, but there is no general method. After the emergence of GAN and CGAN, these tasks can use exactly the same network structure and objective function, the image transformation task can be realized only by changing different training data sets, that is, the variant based on CGAN – CycleGAN[10].

CycleGAN is an image translation method without paired data. Let X, Y be two types of images , PX and PY be the mutual mapping between the two types of images. Cycle GAN is composed of two pairs of generators and discriminators, GX→Y, DY and GY→X, DX. Based on wgan, class X images are transformed into class Y images, and a GAN discriminator loss function can be constructed, The expression is as follows:

$$L_Y(D_Y, G_{X\to Y}) = E_{y\sim P_Y}[\log D_Y(y)] + E_{X\sim P_X}[1-\log D_Y(G_{X\to Y}(x))] \tag{2}$$

Similarly, class Y images are transformed into class X images, and a GAN discriminant loss function is constructed. The expression is as follows:

$$L_{\mathrm{X}}(D_X, G_{Y\to X}) = E_{x\sim P_X}[\log D_X(x)] + E_{y\sim P_Y}[1 - D_X(G_{Y\to X}(y))] \tag{3}$$

Besides, a relatively novel idea in CycleGAN is cycle-consistent. The basic idea is that two types of images, after two corresponding mappings and return to the original image. Therefore cyclic consistency can be written as

$$L_{cyc}(G_{X\to Y}, G_{Y\to X}) = E_{x\sim P_x}(\| \mathrm{x} - \mathrm{G}_{\mathrm{Y\to X}}(\mathrm{G}_{\mathrm{X\to Y}}(\mathrm{x})) \|_2) + E_{y\sim P_Y}(\| y - G_{X\to Y}(G_{Y\to X}(x)) \|_2) \tag{4}$$

Therefore, the total loss function consists of three parts: the loss function of class X images into class Y images, the loss function of class Y images into class X images, and the loss function of cyclic consistency.

$$\min_{\mathrm{G}_{\mathrm{X\to Y}}, G_{Y\to X}} \max_{D_X, D_Y} L_{CycleGAN} = L_Y + L_X + \lambda_c L_{cycle} \tag{5}$$

The □□ is constant.

CycleGAN successfully solved the problem of lack of data in the field of image translation. It can transform images without pairing data sets. This paper is implemented based on the combination of DCGAN[11] and CycleGAN, The training parameters recommended by DCGAN are used for training, Adam optimization algorithm is used for training [12],the learning rate lr is 0.0002. The first convolutional layer in the network is not normalized. The output of other convolution layers is normalized in batches. DCGAN suggests not to use normalization in the first convolutional layer. If this regulation is not followed, the range of approximate fuzzy images generated will be -0.7 to 0.7 instead of -1.0 to 1.0.Batch-normalization is an important technology in deep learning. It not only makes it easier to train deeper networks and accelerate convergence, but also has some regularization effect so as to reduce the dependence between each layer and improve the independence between each layer [13] [14],it also prevents model over-fitting. The Batch-normalization layer is used to replace the Instance normalization layer of the CycleGAN textual model. This is because Instance normalization calculates the mean and variance independently for each sample and for each channel, using the same statistics for training and testing. Normalization of each sample into a single style has replaced the Instance normalization operation in order to increase the randomness of style.

The network structure of the generator is shown in Table 1. It consists of two modules: convolution and deconvolution. The convolution operation uses a convolution kernel of 32, 64, 128 to perform feature extraction on the image. Similarly, the deconvolution operation uses a convolution kernel of 128, 64, 32, 3 for image synthesis. The size of the convolution kernel is 7×7 and 3×3. Instead of using the pooling layer in the generator network, the convolution is used for up-sampling and down-sampling. The first three layers of the generator use dropout, and each layer randomly removes 50% of the nodes to prevent over-fitting. In addition to the output layer to ensure that the output image range is within [0, 255], the activation function uses the relu function. When training the generator network, the input image is pre-processed and scaled to a size of 256×256. Because the image conversion network is a complete convolutional network and there is no pooled sampling layer, so in the generator network, the down-sampling is achieved using a convolution kernel with a moving step size of 2. In deconvolution operation, the moving step of convolution core is 1/2 to realize up-sampling, so that the size of output and input is the same.

The network structure of the discriminator is as shown in Table 2. The convolution operation is performed using a 4×4 convolution kernel. The number of convolution kernels is 64, 128, 256, 512, respectively. The discriminator uses the LeakyReLU activation function and the leak slope of the model is set to 0.2.

TABLE I. NETWORK STRUCTURE OF THE GENERATOR

operation	Convolution kernel size	pace	Convolution kernel number	Batch normalization	activation function
convolution	7×7	1	32	No	Relu
convolution	3×3	2	64	Yes	Relu
convolution	3×3	2	128	Yes	Relu
convolution	3×3	2	128	Yes	Relu
convolution	3×3	2	128	Yes	Relu
convolution	3×3	2	128	Yes	Relu
convolution	3×3	2	128	Yes	Relu
convolution	3×3	2	128	Yes	Relu
deconvolution	3×3	1/2	128	Yes	Relu
deconvolution	3×3	1/2	128	Yes	Relu
deconvolution	3×3	1/2	128	Yes	Relu
deconvolution	3×3	1/2	128	Yes	Relu
deconvolution	3×3	1/2	128	Yes	Relu
deconvolution	3×3	1/2	64	Yes	Relu
deconvolution	3×3	1/2	32	Yes	Relu
deconvolution	7×7	1	3	Yes	Relu

TABLE II. NETWORK STRUCTURE OF THE DISCRIMINATOR

operation	Convolution kernel size	pace	Convolution kernel number	Batch normalization	activation function
convolution	4×4	2	64	No	Leaky Relu
convolution	4×4	2	128	Yes	Leaky Relu
convolution	4×4	2	256	Yes	Leaky Relu
convolution	4×4	2	512	Yes	Leaky Relu

B. Experiment and result analysis

In order to verify the effectiveness of the algorithm proposed in this paper, Python language and TensorFlow deep learning framework are used in the Ubuntu platform. The images are downloaded from ImageNet database by using the keywords winter and summer. ImageNet is a database created by Stanford computer scientists to simulate human recognition systems. Summer→Winter of the public data set ImageNet was used for the experiment. The image size is 256 × 256 pixels, Summer's image has 1540, Winter's image has 1200. The ratio of 8:2 is used for training and testing. The results of the transformation are shown in Figure 2.The left column is the original image and the right column is the generated image.

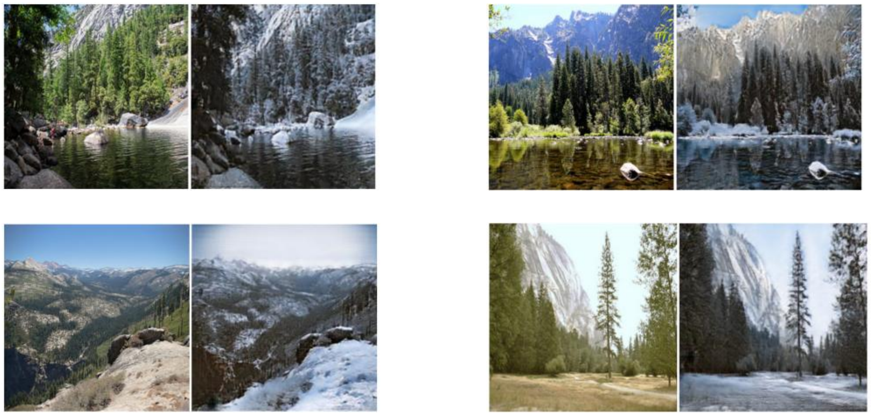

Figure 2. Experimental results

As can be seen from the experimental results, under the condition that the image structure remains unchanged, a summer picture is transformed into a winter picture and the transformed image has no color unbalanced distribution phenomenon, which proves the effectiveness of the algorithm in this paper.

IV. SUMMARY AND PROSPECT

This paper combined DCGAN and CycleGAN network structure optimization, using this model structure to achieve image transformation. A pairless image can be built to realize image transformation. The next step is to delve into the details of the image transformation, the convergence speed, the discriminator and the generator need a metric to tell the model when it can be optimal. In addition, GAN as a depth model, because of its powerful generative ability is also a good model to solve the natural language processing(NLP). How to apply GAN in NLP field is also the next step to solve.

REFERENCES

[1] Yilun Lin, Xingyuan Dai,li Li,et al.The new frontier of artificial intelligence research:generative adversarial networks[J]. IEEE/CAA Journal of Automatica Sinica (JAS) ,2018,44(05):775-792.

[2] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

[3] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2414-2423.

[4] Kunfeng Wang, Chao Gou, Yanjie Duan, et al.Research progress and prospect of Generative adversarial networks[J]IEEE/CAA Journal of Automatica Sinica (JAS),2017,43(03):321-332.

[5] Bengio Y, Laufer E, Alain G, et al. Deep generative stochastic networks trainable by backprop[C]//International Conference on Machine Learning. 2014: 226-234.

[6] Opper M, Archambeau C.The variational Gaussian approximation revisited[J]. Neural computation, 2009, 21(3): 786-792.

[7] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.

[8] Ratliff L J, Burden S A, Sastry S S. Characterization and computation of local nash equilibria in continuous games[C]//2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE, 2013: 917-924.

[9] Xianlun Tang, Yiming Du, Yuwei Liu et al.An image recognition method based on conditional depth convolution Generative adversarial networks[J]IEEE/CAA Journal of Automatica Sinica (JAS) ,2018,44(05):855-864.

[10] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2223-2232.

[11] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.

[12] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[13] Salimans T, Goodfellow I, Zaremba W, et al. Improved techniques for training gans[C]//Advances in neural information processing systems. 2016: 2234-2242.

[14] Gulrajani I, Ahmed F, Arjovsky M, et al. Improved training of wasserstein gans[C]//Advances in Neural Information Processing Systems. 2017: 5767-5777