Research on the Application of an Image Translation Technology Based on GAN

Zhiliang Zhang¹⁺, Tao Yang¹ and Ximing Pei¹
¹Experimental Training Center, Dalian Neusoft University of Information, Dalian, Liaoning, China, 116023

*Corresponding author e-mail: zhangzhiliang@neusoft.edu.cn

Abstract. Image translation is a field with broad prospects. Recently, diverse research methods have been developed and applied in this field, among which GAN is the most potential technology. GAN belongs to artificial neural network, which not only has the advantages of artificial neural network, but also has its own unique characteristics. Although it still has some shortcomings, with the continuous improvement of researchers, it has been widely used as an effective algorithm in the field of image translation.

Keywords: Image Translation, GAN, ANN

1. The present situation of image translation technology

Image translation is to convert the image representation of an object into another image representation of that object, which means that engineers or scientists should find a function that can map the image in one domain to another, which can be applied to many practical problems, such as style transfer, attribute transfer, improving image resolution and so on [1]. Image translation tasks are widely used in many visual fields, such as image stylization, super-resolution image generation, colour filling, four seasons transformation and so on.

Compared with the previous image processing technology, the current digital image translation technology has great advantages [2]:

1) Good reproducibility

Digital image processing technology will not lead to a degradation of image quality because of the transformation operation in this picture, and the digital image can always retain the reappearance of the picture.

2) High precision

Digital image processing technology calculates and processes image data by compiling various algorithms by computer, which ensures the accuracy of calculation [3].

3) Wide range of applications

Digital image processing technology can accurately process electron microscope images, aerial photographs, and remote-sensing images, which cannot be achieved by general image processing technology.

4) High flexibility
Image translation by network technology can complete both linear and complex nonlinear image processing.

Image translation technology was first applied to aerospace image processing, and then it was also applied to the field of medicine. After entering the 21st century, the high development of computer technology has created a broader environment for the application and development of image translation technology. At present, image processing technology is widely used in China, such as microbial science, construction industry and so on[4].

2. Application of neural network in the field of image translation

With the development of technology, image translation has become an indispensable and powerful tool in people’s life. Traditional methods cannot fit the demands, scientists began to find novel and better effective methods with high efficiency, in which using neural networks in the field of image translation seems to be the most promising direction.

Artificial neural network (ANN) is not a very strict concept. Researchers tend to call those algorithms or structural models which have 1) a large number of simple computing units, 2) extensive connections between units and, 3) the strength of connections that can be adjusted according to input and output data as a kind of artificial neural network. Different neural network models are formed by the combination of different neuron types, network structures and learning algorithms[5]. It is only more than half a century since the birth of ANN. Because of its advantages of distributed information storage and parallel processing, ANN has been more and more widely used in many fields, including pattern recognition, intelligent control and system modeling[6]. In particular, the multi-layer feedforward network (BP), based on error back propagation algorithm can approach arbitrary continuous functions with any accuracy. Therefore, it is widely used in many fields, such as nonlinear modeling.

Initially, ANN is used as a kind of clustering technology in image translation[7]. With the further study of ANN theory, the features of neural network have been thoroughly realized and fully applied in lots of aspects of image translation, including handwritten character recognition, face detection, and so on.

At present, BP is the most widely used algorithm, and it has many advantages:

1) the network achieves the aim of mapping from input to output, which is agree with solving problems with complex mechanisms [8].
2) the network can voluntarily extract reasonable solving rules by examples, that is, it has the ability of self-study.
3) it has a strong ability to popularize and generalize.

At the same time, this algorithm also has its own shortcomings:
1) BP uses gradient descent method, and the speed of the algorithm is slow. However, the objective function to be optimized is complex, so it will reduce the image processing effect.
2) the complexity of the optimized objective function leads to a flat area when the output of the neuron is close to 0 or 1, which makes the process slow.
3) the traditional searching method is unable to calculate the step size of each iteration, leading to a result that the update rules of the step size should be uploaded to the network beforehand, which also reduces the efficiency.

3. The advantages and applications of GAN in the field of image translation

Based on the defects of some traditional artificial neural networks, the researchers developed a new algorithm, which is called Generative Adversarial Nets (GAN), and applied it in the process of image translation.

3.1. The principle of GAN

GAN consists of a generator and a discriminator. The goal of generator is generating a real sample which can deceive discriminator, while that of the discriminator is distinguishing between the real and
generated sample. Under this kind of antagonistic game, the performance of the generator and discriminator can be continuously improved. After reaching the Nash equilibrium, the generator can realize the output of the false and the real.

![Figure 1. Basic structure of GAN.](image)

GAN can take any distribution as input $Z$, and take $Z \sim N(0 \sim 1)$ in the experiment. The parameter of generator $G$ is $\theta$. The input $Z$ gets $G(z; \theta)$ under the generator, and the outlet is able to be regarded as the sample taken from $G(z; \theta) P_g$. The data distribution of training sample $x$ is $P$, and the training goal of generating model $G$ is to make $P_g$ seem to be more similar with $P$. Discriminator $D$ is to discriminate the true and false of the generated and the real sample. The generator attempts to generate real data to cheat the discriminator, while discriminator tries to discern the difference of real and synthetic data. The building block of original GAN is an entirely connected layer. Later, some researchers proposed to use convolution neural network to realize more perfect outcome. Since then, the convolution layer has become the core component of many GAN models.[9]

3.2. Application of GAN in the field of image translation

Since the birth of GAN in 2014, a large number of articles around GAN have been published in major journals and conferences in order to improve the quality of GAN generation and expand the scope of application of GAN in image generation. The specific methods are as follows:

1) Pixel2Pixel

It uses conditional generation countermeasure network (CGAN) structure. Compared with the crude generation confrontation network, CGAN adds condition $y$ to the input of the generator and discriminator respectively. That $y$ can be any type of data (a category label, or other types of data, etc.). The purpose of this activity is conditionally monitoring the data gained by the generator, making the pattern in which the generator generates results is not completely free and unsupervised.

![Figure 2. Basic structure of CGAN](image)
2) CycleGAN
The structure is shown in the figure, where Gline F is two different generators, and Dx and Dy are two different discriminators. Generators G and F can generate images with the same distribution as the target domain.

![Figure 3. Basic flow structure of CycleGAN](image)

3) StarGAN
Both Pixel2Pixel and CycleGAN can only complete one-to-one image translation tasks. If the research needs to be converted between multiple fields, then a model needs to be retrained between each two fields to solve the problem. For K areas, you need to train k (k-1) generators. However, this method is relatively inefficient and needs to be further improved.

3.3. The defects and improvement methods of GAN
Although GAN is very effective, its training process is very unstable and there is a problem of mode collapse. The discriminator does not need to consider the type of sample generated, but only focuses on determining whether each sample is real, which makes it sufficient for the generator to generate only a few high-quality images to fool the discriminator.

For example, if a data set contains digital images from 0 to 9, in extreme cases, the generator only needs to learn to generate one of the ten numbers perfectly to completely deceive the discriminator, and then the generator stops trying to generate the other nine digits so that the other nine digits cannot be implemented. This is also called an in-class mode crash. At present, many methods have been proposed to solve the problem of model collapse.

1) MiniBatch characteristics.
The principle of this technique is to make the discriminator compare small batches of real samples and small batches of generated samples. In this way, the discriminator can learn to determine whether the generated sample is too similar to some other generated samples by measuring the distance of the sample in the potential space. Although this method works well, its performance depends largely on the characteristics used in distance calculation.

2) Add an encoder to convert the samples in the data space back to the potential space.
3) Wasserstein distance is used to judge the similarity between real and learning distribution. Although it avoids pattern collapse in theory, the convergence time of the model is longer than that of the previous GAN.
4) Use gradient penalty instead of weight reduction. This usually greatly avoids schema collapse, and this approach can be easily applied to other GAN models.

4. Conclusion
Cases of GAN application in image translation can be found everywhere. GAN not only has the advantages of self-organization, non-linearity and self-learning ability of artificial neural network, but also has its own unique characteristics, which makes it come to the forefront in many fields. With the development of ANN theory and practice, GAN will be more functional in the field of image translation in the near future.

Acknowledgments
This work was financially supported by Natural Science Foundation of Liaoning Province (2019-MS-013).
References

[1] J. Uthayakumar, Mohamed Elhoseny, K. Shankar, Highly Reliable and Low Complexity Image Compression Scheme using Neighborhood Correlation Sequence Algorithm in WSN, IEEE Transactions on Reliability, In Press, 2020.

[2] Sun Yulan; Research on the Application and Development of Digital Image Processing Technology [J]; Computer Knowledge and Technology; 2014.

[3] E. Laxmi Lydia, Joshua Samuel Raj, R. Pandi Selvam, Mohamed Elhoseny, K. Shankar, Application of discrete transforms with selective coefficients for blind image watermarking. Transactions on Emerging Telecommunications Technologies, 2019, In press.

[4] Liu Zhonghe, Wang Ruixue, Wang Fengde, Ma Changqing, Liu Xianxi; Present situation and Prospect of Digital Image processing Technology [J]; Computer Era; 2005.

[5] Chen Hua; Introduction of the image Recognition technology and methodology in neural network [J]; Journal of Jining Teachers’ College; 2006.

[6] Chen Jingshui, Li Yeqin, Liu Yu, Cai Xiansheng; Application of artificial neural network in image processing [J]; China Medical Equipment; 2010.

[7] Xu Feng, Lu Jiangang, Sun Youxian; Application of neural network in image processing [J]; Information and Control; 2003.

[8] Mohamed Elhoseny, Bian Gui-Bin, S.K. Lakshmanaprabu, K.Shankar, Amit Kumar Singh, Wu Wanqing, Effective Features to Classify Ovarian Cancer Data in Internet of Medical Things, Computer Networks, Volume 159, 4 August 2019, Pages 147-156

[9] Wu Meiyin, Chen Li, Tian Jing; Video image distortion detection and classification based on CNN [J]; Application Research of Computers; 2016.