IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Generating Calligraphy Works Based on DiscoGAN

To cite this article: Yuqing Zhao and Guangyuan Fu 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **565** 012007

IOP Conf. Series: Materials Science and Engineering **565** (2019) 012007 doi:10.1088/1757-899X/565/1/012007

Generating Calligraphy Works Based on DiscoGAN

Yuqing Zhao[1*] and Guangyuan Fu[1]

[1]Xi'an research institute of high-tech, Shaanxi,710025, China

[*]Email: zoe_rabbi@126.com

Abstract. Copying the works of a good calligrapher is a necessary step for many beginners. However, due to the large number of Chinese characters, it is a challenge to automatically generate Chinese characters in the style of calligraphers. With the development of deep learning, the recognition of Chinese characters, style migration and other research have achieved remarkable results. However, there are few studies on the generation of calligraphy works, especially on the lack of paired training data. In this work, we propose a method for generating calligraphy works. This method is generally applicable to the generation of calligraphy works of various styles. And the validity of our method is verified by experiments on data sets.

1. Introduction

Calligraphy is a unique form of artistic expression of the beauty of Chinese characters in China and surrounding countries and regions deeply influenced by Chinese culture. Among them, Chinese character calligraphy is an original expressive art of the Han nationality, which is known as: wordless poetry, invisible dance, pictures without pictures, silent music. Like other literary theories, calligraphy theory includes both the technical theory of calligraphy itself and its aesthetic theory. From ancient times to the present, a group of outstanding calligraphers have emerged, among them Xizhi Wang (303-361,Eastern Jin Dynasty), Xun Ouyang (557-641,Tang Dynasty), Zhenqing Yan (709-784,Tang Dynasty), Gongquan Liu (778-865, Tang Dynasty), Mi Fu (1051-1107,Northern Song Dynasty) and so on have left a famous masterpiece. Long time ago, many famous works lost. The existing part also cannot cover all the commonly used words. As a result, more and more people are beginning to try to imitate more famous masterpieces in modern ways.

In this work, we formulate the problem of generating calligraphy works as a problem of learning the mapping from unpaired printed fonts to calligraphers' fonts. Our main contributions are:

- Based on DiscoGAN [1], the method of generating calligraphy works is proposed.
- Identity consistency loss is proposed to adapt to the specific application environment.
- The effectiveness and versatility of the proposed method are proved by experiments.

2. Related work

2.1 Reorganizing Chinese characters by decomposing strokes

Many of the previous ideas [2-4] were expressed as the basis of Chinese characters through the level of simple strokes. Decompose existing Chinese characters into stroke trees and establish an accurate correspondence between strokes. Combine new Chinese characters by re-typing. However, this method is difficult to guide the synthesis process with high regular statistics.

2.2 Neural style transfer

A major advance in deep learning is the neural style transfer, which was proposed by Leon Gatys et al in 2015 [5]. Style transfer refers to applying the style of the reference image to the target image while retaining the content of the target image. The key concept of implementing style transfer is consistent with the core idea of the deep learning algorithm: define a loss function to specify the desired goal and then minimize the loss.

$$\begin{aligned} \text{loss} &= distance(style(reference\_image) - style(generated\_image)) \\ &+ distance(content(original\_image) - content(generated\_image)) \end{aligned} \tag{1}$$

Where distance is a norm function; Content is a function that takes an image and computes its representation; Style is a function that inputs an image and computes the representation of its style. Minimizing this loss will make the style of the generated image close to the reference image, and the content of the generated image also close to the target image, so as to achieve the style transfer. Although style transfer is applied to many images, the generation of calligraphy works involved in this paper is not entirely a matter of neural style transfer. In particular, the particularity of Chinese characters, such as: stroke position and Angle slightly change will lead to the change of different characters.

2.3 Generative adversarial networks (GANs)

Generative adversarial networks (GANs) [6] consists of a generator G and a discriminator D, which compete in a two-player minimax game: the generator attempts to generate realistic images to deceive the discriminator, while the discriminator attempts to distinguish the synthetic images from the real data. Formally, D and G are two players of min-max problem to find Nash equilibrium:

$$\min_G \max_D \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}\left[\log\left(1 - D\big(G(z)\big)\right)\right] \tag{2}$$

Goodfellow et al. (2014) proved that when $p_g = p_{data}$, this minimax game has the global optimal value, and under mild conditions (for example, G and D have sufficient capacity), p_g converges to p_{data}.

2.4 Matching image conversion

Pix2Pix [7] adopts the idea of cGAN [8] and takes the input image as the condition to Generative adversarial networks. In the design of the network structure, Pix2Pix basically refers to the structure of DCGAN [9], using convolution layer, batch normalization and ReLU activation function. Pix2Pix learns how to convert one type of image into another. Zi2Zi [10] uses GAN to convert Chinese characters between fonts in an end-to-end manner. As shown in Figure 1(a), Zi2Zi USES pairs of Chinese characters in source and target fonts as training data. Zi2Zi's network structure is based on Pix2Pix, adding a variety of font categories embedded. This enables Zi2Zi to use a training model to convert characters into several different fonts. However, since it is impractical to obtain a set of large number of paired training samples for the generation of personalized handwritten Chinese characters, Zi2Zi is not applicable to our problem.

IOP Conf. Series: Materials Science and Engineering **565** (2019) 012007 doi:10.1088/1757-899X/565/1/012007

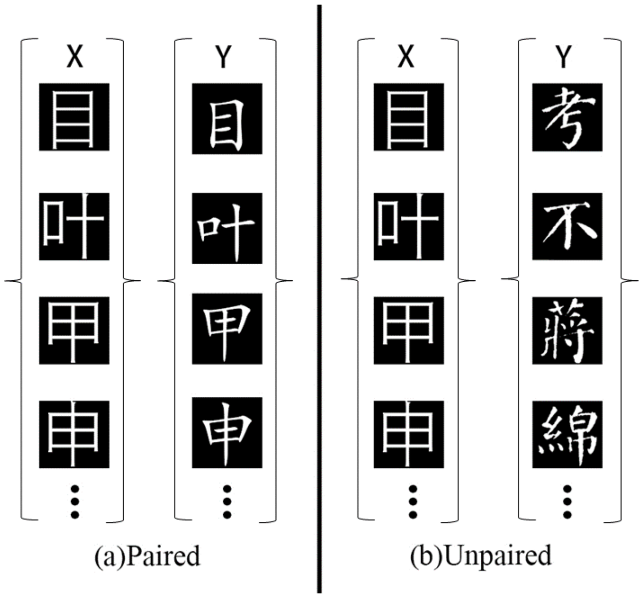

Figure 1. Paired and unpaired train data. (a) Paired training set contains training examples $\{x_i, y_i\}_{i=1}^{N}$, where there exists correspondence between x_i and y_i. (b) Unpaired training set contains training examples, where the source set X and the target set Y have no paired matching information.

3. Our method

Pix2Pix handles image conversions that match data sets very well, but in many cases, there is almost no data set that is absolutely matched, and collection is difficult. However, the amount of data that is not matched in the two fields is still very large. As shown in Figure 1(b), we can get the calligrapher's font from some inscriptions. If you can achieve image conversion through unmatched data, this work becomes very meaningful. In 2017, DiscoGAN and CycleGAN [11] proposed an image conversion scheme for solving unmatched data sets. The schematic diagram of the DiscoGAN structural framework is shown in Figure 2. Essentially two different GANs are coupled together, and each GAN maps one domain to the corresponding domain.

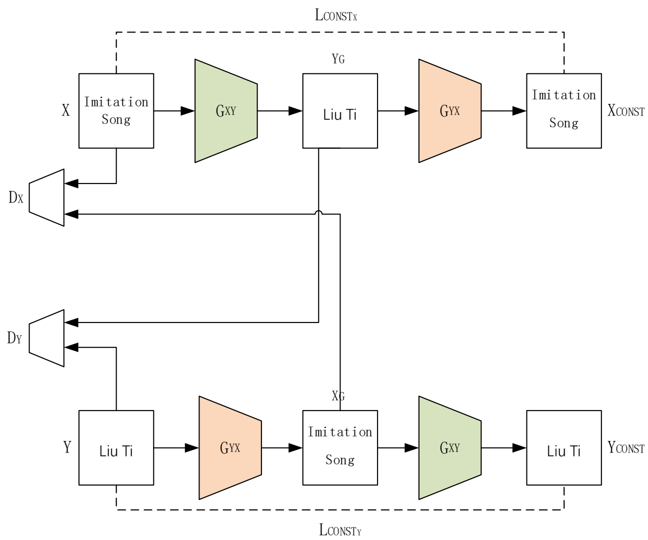

Figure 2. DiscoGAN model diagram.

In section 3.1, we introduce the network architecture adopted in this paper, and in section 3.2, we describe several loss functions adopted in this paper.

3.1 Structural framework

Figure 3 shows the network structure of the generator, which is composed of three parts: coding layer, conversion layer and decoding layer.

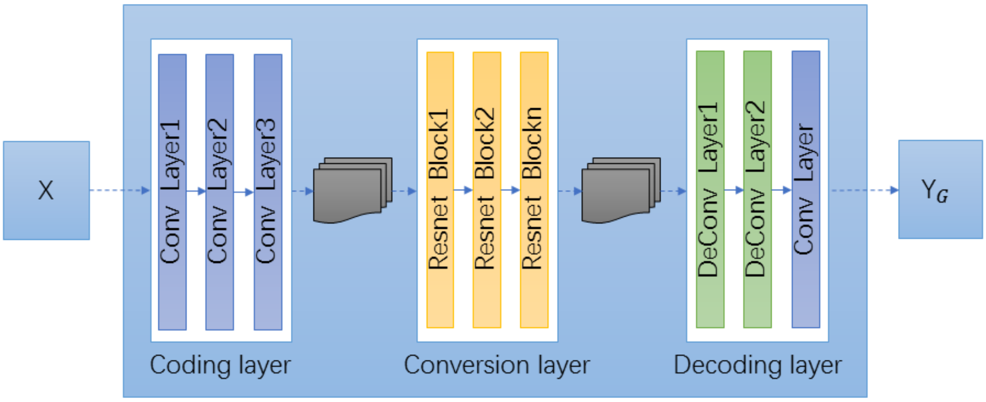

Figure 3. The generator network structure of DiscoGAN.

3.2 Loss function

The conversion between non-matching pairs of image domains is not as simple as two GANs.

$$L_G = L_{G_{XY}} + L_{G_{YX}} = L_{GAN_Y} + L_{CONST_X} + L_{GAN_X} + L_{CONST_Y} \tag{3}$$

$$L_D = L_{D_X} + L_{D_Y} \tag{4}$$

Where, L_{GAN_Y} is the loss of generation of Y domain in the X domain, and L_{CONST_X} is the reconstruction loss of X.

If the generator G_{X2Y} that converts X to Y expects the discriminator D_Y to discriminate the generated image as true, then G_{X2Y} preferably does not extract any information related to X, but independently learns and generates data from Y. At this time, although we got a realistic Y-domain image, this image has nothing to do with the X-domain. Therefore, a concept of reconstruction loss is introduced in DiscoGAN. In Figure 1, we combine the two sets of generated confrontation networks. Image X generates YG through G_{X2Y}, and Y_G is restored to X_{const} through G_{Y2X}. Obviously, the more consistent X and X_{const}, the closer our entire network is to our intended goal. Therefore, the distance between X and X_{const} is called the reconstruction loss:

$$\begin{aligned} L_{CONST} = (G_{X2Y}, G_{Y2X}) = \mathbb{E}_{x\sim p_{data}}(x)[\|G_{Y2X}(G_{X2Y}(x) - x)\|_1] \\ + \mathbb{E}_{y\sim p_{data}}(y)[\|G_{X2Y}(G_{Y2X}(y) - y)\|_1] \end{aligned} \tag{5}$$

4. Experiments

In this section, we will perform experimental verification on the dataset. In the experimental set, the original domain and the target domain have both common attributes and visible differences.

Song style is a kind of Chinese character printing font, which imitates the font carved in the book of song dynasty. The song dynasty typeface is people imitates the song dynasty typeface the structure, the brushwork meaning, in order to adapt to the printing, need to evolve, is very different with the writing style of the brush. The refers to the general name of the calligraphy works of Liu Gongquan, one of the last great calligraphers of the Tang Dynasty and one of the four masters of the book. The Liu Ti is balanced and thin, and the paintings are cool and pretty, the bones are strong, and the knots are tight. Compared with Yan Zhenqing's calligraphy work, it is slightly even and thin.

The training data consists of 5,000 imitations from the computer font library and 1,302 face characters (including duplicates) from the mysterious tower monument. All words are pre-processed to 256 x 256. After 200 rounds of training with our method, the results shown in Figure 4 were obtained, including many simplified characters that were not in the training set, which better preserved the style of the willow.

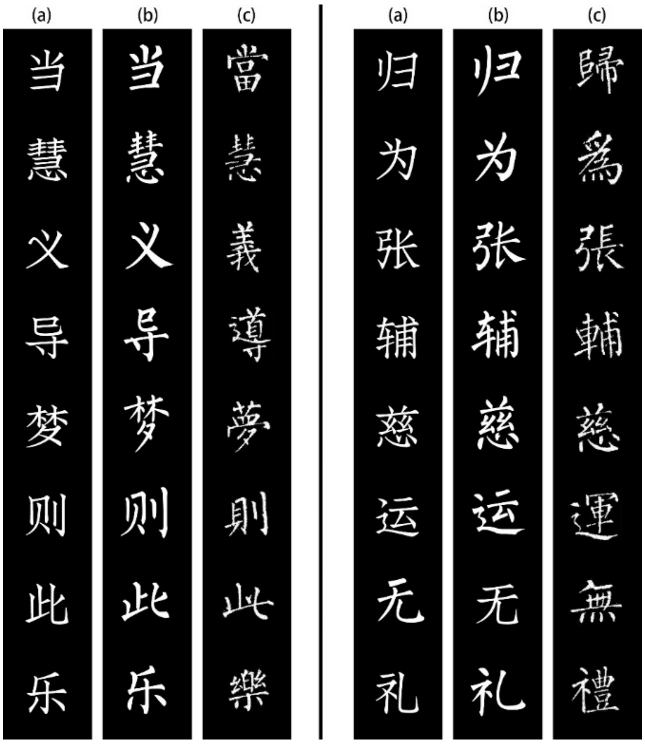

Figure 4. Partially generated word set (a) imitation Song source characters. (b) willow characters generated using DiscoGAN. (c) willow source characters.

5. Conclusion

In this work, we describe the problem of generating a style calligraphic work as a mapping problem from the existing print font to the calligrapher style. We use DiscoGAN to solve this problem. In order to make the model more suitable for our problem, the loss of identity consistency is guaranteed in the loss function, and the validity of the method is verified by experiments on the dataset. In the next work, we try to propose an evaluation function to quantify the effect of font conversion; on the other hand, we plan to study more ways of art style transfer.

References

[1] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]. Proceedings of the 34th International Conference on Machine Learning-Volume 70, 2017: 1857-1865.
[2] Xu S, Lau F C, Cheung W K, et al. Automatic generation of artistic Chinese calligraphy[J], 2005, 20(3): 32-39.
[3] Xu S, Jin T, Jiang H, et al. Automatic generation of personal chinese handwriting by capturing the characteristics of personal handwriting[C]. Twenty-First IAAI Conference, 2009.
[4] Zong A, Zhu Y. Strokebank: Automating personalized chinese handwriting generation[C]. Twenty-Sixth IAAI Conference, 2014.
[5] Gatys L A, Ecker A S, Bethge M J a P A. A neural algorithm of artistic style[J], 2015.

IOP Conf. Series: Materials Science and Engineering **565** (2019) 012007 doi:10.1088/1757-899X/565/1/012007

[6] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. Advances in neural information processing systems, 2014: 2672-2680.
[7] Isola P, Zhu J-Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 1125-1134.
[8] Mirza M, Osindero S J a P A. Conditional generative adversarial nets[J], 2014.
[9] Radford A, Metz L, Chintala S J a P A. Unsupervised representation learning with deep convolutional generative adversarial networks[J], 2015.
[10] Tian Y. zi2zi: Master chinese calligraphy with conditional adversarial networks, 2017.
[11] 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.