Advances in Intelligent Systems Research, volume 147

International Conference on Network, Communication, Computer Engineering (NCCE 2018)

Font Style Conversion Based on Deep Learning

Da Lv [a)], Yijun Liu

School of Computer, Guangdong University of Technology, Guangzhou 510006, China

[a)] Corresponding author: 502472551@qq.com

Abstract. In view of the cost of traditional design of new fonts, a method of combining deep learning for font style conversion is proposed. By using a U-Net type network structure combining the training method of generative adversarial network, supervised learning part of fonts, the font style conversion ability of the font generator network is constantly enhanced so that the fonts can be converted to another style through the generator network. The experimental results show that this method convert font structure generated clear and smooth, less noise, and the original character of the same font highly consistent in size, weight, style etc.

Key words: Deep learning; style conversion; generative adversarial network; structure generated clear.

INTRODUCTION

With the development of new technology, deep learning has launched a new wave in academia. Machine learning, including computer vision, has been rapidly occupied by deep learning, and deep learning is also showing great splendor in every application. The following two methods are proposed in the aspect of font style conversion. A method is to use recurrent neural network such as [1-2], create a point by way of using specific generation methods, until the end of production so far, the method can generally clear font structure, but generally cannot generate word heavy, difficult to express the font style; the other is a font style with image convolution neural network extraction and migration literature such as [3], but the font generation method of fuzzy structure, often incomplete. Recently, deep learning has also made an important breakthrough. The generation of network Generative Adversarial Networks[4] (GAN) and later people put forward a variety of generation countermeasures network [5-6].So this paper presents a conversion method based on deep learning font style, the computer supervised learning part Chinese characters style makes the computer has the ability to convert other fonts out another style font, using the training methods of the generation of confrontation, the method of converting font structure generated clear and less noise. In the human visual and character in the font is consistent.

FONT STYLE CONVERSION

This paper puts forward the method of font style conversion, Convert the font A into a font like the font B.The font A in the generator output font C, through the network of C and D discriminator font font to distinguish between the target B style is the same, in order to deceive the D generator G discriminator will learn to adjust B parameters to the font style, the generated C fonts as possible and font style similar to B. D will continue to adjust the parameters of the discriminator to enhance the ability to distinguish between C and B font font. The font C generated by the final generator G is difficult to distinguish in style similar to the font B.

loss function

The font is the font style conversion process by neural network style transfer [7-8], our goal is to train a G generator, making a style font into another style, in order to generate the corresponding font style, the number of loss function we design consists of two parts, the content of loss and loss of style.

Content loss

The general time by mapping one image to another image of the same sample size, the use of a point estimate of the quality evaluation standard index always points estimate and the true value of the parameters. The function of the distance function, the most commonly used is the square of the distance between the mean square error (MSE), in this paper, we map a style font A to another style font, C also uses MSE loss function as a content loss function, so as to ensure that the generated font C is consistent with the structure of font A, and is reliable and reliable:

$$\text{context\_loss} = \frac{1}{2n}\sum_{i=1}^{n}(A_i - C_i)^2 \tag{1}$$

Style loss

In order to convert the font A into the font C that is the same style as the font B.The probability distribution of the set of A data sets is defined as pA, and the probability distribution of the font B data set is pB. A network D(b)represents the input a character image of B, and the probability of return to B to pB, when the image of the font to as pB when D(b) returns a value of nearly 1, if pB to close to 0 of the value is not to return. Therefore, the goal of D is to maximize D (b) for font b~p_B that comes to font set B, and to minimize D(b)for font B that does not come to font set B, while G aims to generate font fonts that can deceive the font from the font A. Thus, the generator needs to maximize D(G(a)), or the same minimization of 1-D (G(a)), while D, on the contrary, needs to minimize D(G(a)), or the same maximization 1-D(G). The whole process is done under the following minimax game:

$$\min_G \max_D V(D,G) = E_{b\sim p_B}[\log D(b)] + E_{a\sim p_A}\left[\log\left(1 - D\left(G(a)\right)\right)\right] \tag{2}$$

Among them: B and a represent the font B and the font A, respectively.

So in order to make the output of the font style, the style of the loss function with network confrontation with the use of a standard:

$$\text{style\_loss} = \sum_{i=1}^{n}\log\left(1 - D\left(G(A_i)\right)\right) \tag{3}$$

Therefore, the total loss of (1) (3) is as follows:

$$\text{total\_loss} = \alpha * \text{context\_loss} + \beta * \text{style\_loss} \tag{4}$$

Among them, α, β are super parameters.

Network Structure

The generator network structure is adopted in this paper, pix2pix[9] and U-Net[10] from the structure consists of two parts encode and decode, are using the 5×5 convolution kernel, using convolution extraction of the input image until the 2×2 ×1024 space, and then the recovery of 64 × 64 ×1 using a series of deconvolution, convolution and skip connection, and on each layer with Batch Normalization[11] structure is shown in Figure 1 (a), network structure parameters are shown in table 1. The generator has reached the purpose of network transformation to a certain extent, but the background is often fuzzy noise multiple strokes is not complete and other issues, so we use generative against network thought using a discriminator to generate the results of optimization, make the image clearer, more realistic and less noise. A network structure is: input a 64 x 64 x 1 pictures, are using the 5×5 size into several layers of laminated volume convolution kernel, after each layer convolution image size decrease and increase in the number of

half channel is two times the previous layer, the output layer is a two classification, as shown in Figure 1 (b) shows. The specific parameters are set in Table 2.

TABLE 1. The parameters of the generator network

Encode	Decode
Conv1:$64^2\times1$ →$32^2\times64$ Conv2:$32^2\times64$→$16^2\times128$ Conv3:$16^2\times128$→$8^2\times256$ Conv4:$8^2\times256$→$4^2\times512$ Conv5:$4^2\times512$ →$2^2\times1024$	Deconv1:$2^2\times1024$→$4^2\times512$ Skip1:$4^2\times(512+512)$→$4^2\times1024$ Deconv2:$4^2\times1024$→$8^2\times256$ Skip2:$8^2\times(256+256)$→$8^2\times512$ Deconv3:$8^2\times512$→$16^2\times128$ Skip3:$16^2\times(128+128)$→$16^2\times256$ Deconv4:$16^2\times256$→$32^2\times64$ Skip4:$32^2\times(64+64)$→$32^2\times128$ Deconv5:$32^2\times128$ →$64^2\times1$

TABLE 2. The parameters of the discriminator network

Conv1: $64^2\times1$ →$32^2\times64$
Conv2: $32^2\times64$→$16^2\times128$
Conv3: $16^2\times128$→$8^2\times256$
Conv3: $8^2\times256$→$4^2\times512$
Full: $4^2\times512$ →8192

In the experiment, we used Adam [12] to do gradient optimization and adjusted Adam's beta 1, learning rate and training times. Finally, we tested on many different font sets, and achieved good results.

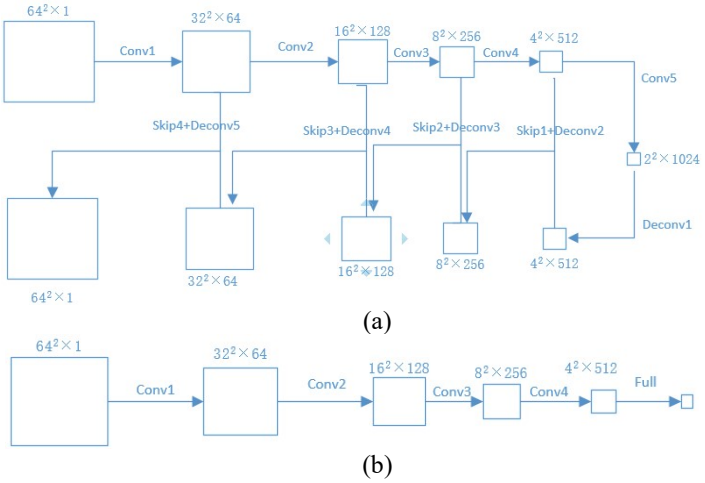

FIGURE 1. (a) Generator network structure (b) Discriminator network structure

EXPERIMENTS

In order to validate the experiment, using Simsun, Fangsong,Xinhua clerical scripts, Zhun yuan, official script and so on ,for font conversion experiment.

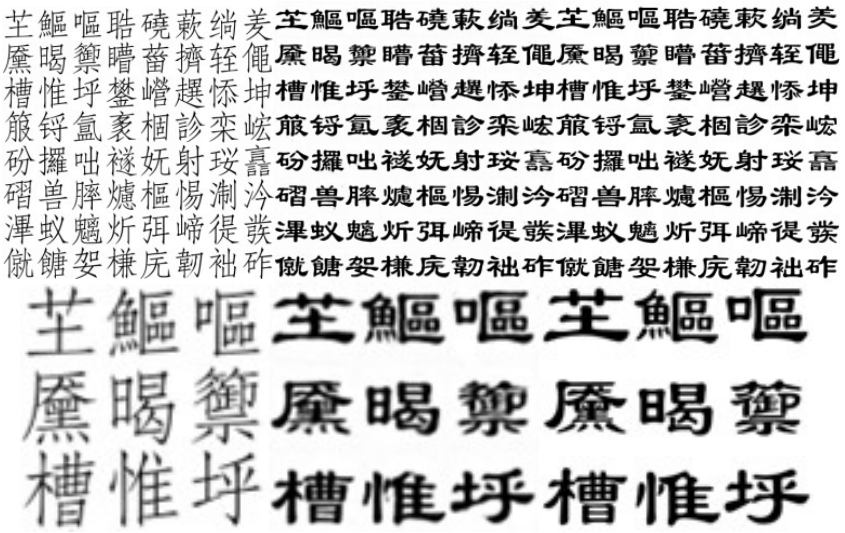

FIGURE 2. style conversion results. Upper left: source font; middle and upper: generated fonts; right up: real target font; Lower left: magnified source font; middle and down: magnified generation font; right down: magnified real target font;

As shown in Figure 2, we use this algorithm convert fang song to clerical script. Watch the upper row of pictures from source font and generated fonts to real target, we find that the generated font's structure is clear, at first glance, not observed differences, the lower the generated font magnifier can send images generated with the real image of the target structure is basically the same, the details generated is clear and consistent, less noise, heavy style also expressed words very clearly, more real, some of the tiny details such as point ,pie and na, also can generate delicate strokes.

CONCLUSION

There is a possibility that deep learning tends to be perfect, and it is no doubt that it has a wider future than the artificial algorithm. In this paper, we propose a new method of font style conversion. This method uses U-Net network to preserve font structure through content loss and generates font to be changed to another style by generating network loss. The design method of generating fonts regardless of the size or weight of the word font style with font size and style of word manual design corresponding to the basic agreement, but also looks clear, truly accords with human visual perception, the application of this method can greatly improve the efficiency in the design of a font, can save manpower, material and time.

ACKNOWLEDGEMENTS

The paper is supported by funds from Department of Science and Technology at Guangdong Province and Guangzhou City (No. 201604010051, 2015B090901060, 2016B090903001, 2016B090904001, 2016B090918126，2016KZ010101)

REFERENCES

1. Xu SJin T, Jiang H, et al. Automatic G-eneration of Personal Chinese Handwrit-ing by Capturing the Characteristics ofPersonal Handwriting[C]//IAAI. 2009.
2. Zhang X Y,Yin F,Zhang Y M,et al.Drawingand Recognizing Chinese Characters withRecurrent Neural Network[J].IEEE Trans-actions on Pattern Analysis & Machine Intelligence,2016,PP(99):1-1.
3. Baluja S. Learning typographic style: from discrimination to synthesis[J]. M-achine Vision and Applications, 2017, 28(5-6): 551-568.
4. Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in neural information processing systems. 2014: 2672-2680.

5. Radford A, Metz L, Chintala S. Unsuper-vised representation learning with dee-p convolutional generative adversarial networks[J]. arXiv preprint arXiv:1511.06434, 2015.
6. Denton E L, Chintala S, Fergus R. Deep generative image models using a laplac-ian pyramid of adversarial networks[C]//Advances in neural information proces-sing systems. 2015: 1486-1494.
7. Gatys L A, Ecker A S, Bethge M. A neur-al algorithm of artistic style[J]. arXiv preprint arXiv:1508.06576, 2015.
8. Johnson J, Alahi A, Fei-Fei L. Percept-ual losses for real-time style transferand super-resolution[C]//European Conf-erence on Computer Vision. Springer, C-ham, 2016: 694-711.
9. Isola P, Zhu J Y, Zhou T, et al. Image-to-Image Translation with Conditional Adversarial Networks[J]. arXiv preprint arXiv:1611.07004, 2016.
10. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.
11. Ioffe S. Batch renormalization: Towards reducing minibatch dependence in batch-normalized models[C]//Advances in Neu-ral Information Processing Systems. 2017: 1942-1950.
12. Kingma D P, Ba J. Adam: A method for s-tochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.