Image Transformation Method to Refine Handwritten Characters Using Generative Adversarial Networks

Takuya Takahashi^a, Hideaki Orii^b, Hideaki Kawano^a

^a Department of Electrical Engineering and Electronics, Faculty of Engineering, Kyushu Institute of Technology 1-1 Sensui-cho, Tobata-ku, Kitakyushu, Fukuoka, 804-8550, JAPAN

^b Department of Electrical Engineering, Faculty of Engineering, Fukuoka University 8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, JAPAN

*Corresponding Author: takahashi.takuya361 @ mail.kyutech.jp, oriih @ fukuoka-u.ac.jp

Abstract

It is easier for hand-written characters to convey the individuality of the writer than characters printed by a printer or the like. Even when digitization progresses, it is used when you want to communicate important information with your own thought to the other party. For this reason, knowledge, experience, and technology are important to write well-written characters. In this paper, we propose an image processing method that corrects handwritten characters that contain imperfections in local shapes such as balance of character, 'tome', 'hane', 'harai', etc. To well-balanced character form. Specifically, using a Generative Adversarial Network (GAN), simultaneously learn a generator that generates a beautiful character image from a poorly written character image and a discriminator that identifies a poorly written character image and a well-written character image.

Keywords: GAN, DCGAN.

1. Introduction

It is easier for hand-written characters to convey the individuality of the writer than characters printed by a printer or the like. Even when digitization progresses, it is used when you want to communicate important information with your own thought to the other party. On the other hand, in Japanese, it is possible to express characters by mixing a wide variety of characters such as hiragana, katakana, kanji, numbers, alphabets and the like. When describing it by handwriting, it is necessary to write those characters one by one in a well-balanced manner. For this reason, knowledge, experience, and technology are required to write well-written characters. The shape balance has local shapes

Fig. 1.1. Example of 'tome', 'hane', 'harai'.

such as 'tome', 'hane', 'harai' and so on. In this paper, we propose an image processing method to correct handwritten characters (poorly written characters) including those deficiencies into characters with good appearance (well-written characters). Specifically, using a Generative Adversarial Network (GAN), simultaneously learn a generator that generates a beautiful character image from a poorly written character image and a discriminator that identifies a poorly written character image and a beautiful character image. model used this time uses CNN to generate an image, and generates an image with good resolution. In Generator, dimension reduction is performed using an automatic encoder to extract local feature quantities. Also, average pooling is used for the output layer.

2. Related work

2.1 GAN

Generative Adversarial Network (GAN) is a wonderful generation network proposed in 2015(3). GAN is a kind of generative model, and there is a generator and a discriminator at the time of learning. In learning in the case of learning, in the learning of the discriminator, learning is performed as a binary classification model that determines

Fig. 2.1. Overview of GAN

whether the output of the generator obtained from the input image or actual data. In learning of the generator, learning is performed so that an output obtained from the latent variable causes the discriminator at that time to distinguish it from actual data. As a result, a generator is learned which outputs data such that the discriminator determines the generated data as actual data. By alternately repeating these learnings, the discriminator model and generator model can finally become accurate the model. Here, let $p_{data}(x)$ be the probability distribution of real data x , and $p_{data}(z)$ be the probability distribution of images input to the generator. The final goal is to have $p_{data}(x) = p_{data}(z)$ It is. The generator learns that the output value $D(G(z))$ of the discriminator gives a large value when it inputs its output $G(z)$ into the discriminator. The discriminator performs learning so as to increase the output value $D(x)$ for the actual data x and reduce the output value $D(G(z))$ for the generated data. These are expressed as the maximization problem of the value function $V(D, G)$ as follows.

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (2.1)$$

2.2 DCGAN

DCGAN (Deep Convolutional GAN) is a GAN model using CNN proposed by Alec Radford et al. In 2015. CNN can efficiently generate high-resolution images. DCGAN not only incorporates CNN, There are various ideas. In CNN, downsampling is usually performed by maximum pooling. But in the discriminator model, it is replaced by the convolution layer of stride 2. Pooling is not used in the generator model. In the discriminator model, not all connection layers are used, but instead, global average pooling is used. This slows down the convergence and has

the effect of preventing overlearning. ⁽⁴⁾ In the paper by Alec Radford The image of the bedroom was learned by DCGAN to construct a generation model. It produces an image that is visually indistinguishable from the visually generated data.

3. Proposed methods

Figures 3.1 and 3.2 show the generator and discriminator models. As shown in the figure, the neural network used in the proposed method is a generator network that converts hand-written lower-hand characters into well-written characters. And a discriminator that identifies the converted image and the well-written characters image obtained by the generator. It consists of two parts of the network. In network learning, the process of learning the generator so that the discriminator parameters are fixed and the discriminator outputs beauty character judgment, and the generator parameters are fixed and the input to the discriminator is a well-written characters or converted image The process of learning the discriminator is alternately performed so that it can be determined correctly. Since the discriminator learns quickly, the converted image of the generator outputs well-written characters of different characters. Dimensional compression was performed using an auto encoder. So, in this research, we reduced the convolution layer of discriminators and let the generator learn first. This will allow the generator to generate an image that will mislead the discriminator well if the network is trained well. That is, the generator after learning can convert the poorly written character images into a well-written character images.

Fig. 3.1. Detailed structure of Generator.

Fig. 3.2. Example of experimental image.

4. Experiments

4.1. Experiments settings

In the experiment, images for learning were created using eight variations handwritten (eight characters) for 80 first-grade kanji characters in Japan and pictograms

Fig. 4.1. Example of experimental image.

obtained from a commercial dictionary of well-written character images (poorly written character: $80 \times 80 = 640$ patterns, well-written character: 80 patterns). The image was rotated, scaled, and misaligned to increase the number of well-written characters and poorly characters to 10000 each. Figure 4.1 shows an example of an experimental image.

4.2. Experiments results

Figures 4.2 and 4.3 show some beautified characters by correcting some types of poorly written characters. Fig. 4.4 shows a correction and beautification of multiple inferior characters of the same type. The generator's input is poorly written character images and the generator's output is well-written characters images.

In Fig. 4.2, 'hane', 'harai' and 'tome' are corrected and beautified. In Fig. 4.3, it is understood that correction is made to each lower-case letter one by one, and it is not just the replacement of the beautiful letters. In Figure 4.3, the characters of the same type are corrected differently. You can see that we are not replacing characters.

Fig. 4.2. Images focusing on 'hane', 'harai' and 'tome'. The first line is a poorly written character and the second line is a well-written character

Fig. 4.3. Corrected image keeping position.

The first line is a poorly written character and the second line is a well-written character

Fig. 4.4. Comparison image of the same character. The first line is a poorly written character and the second line is a well-written character

5. Conclusions

In this research, we proposed an image processing method to correct poor characters into well-written characters using GAN. We reduced the layer of CNN in discriminator and slowed learning so we could see that the input and output were beautified with the same kind of characters. From the experimental results, it was found that the local parts of 'hane', 'harai' and 'tome' were corrected. In addition to replacing the characters, it turned out that the way of correction was applied to each character.

References

- (1) Hiroshi Fukai, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi : “Attention Branch Network : Learning of Attention Mechanism for Visual Explanation”, CVPR , 2018

- (2) Anders Boesen Lindbo Larsen , Søren Kaae Sønderby , Hugo Larochelle and Saburo Sangyo : “Autoencoding beyond pixels using a learn Ole Winther ed similarity metric”, Machine Learning,, 2015
- (3) Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio : “Generative Adversarial Nets”, NIPS, 2014
- (4) Alec Radford, Luke Metz and Soumith Chintala : “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, ICLA, 2015