Automatic Chinese Font Generation System Reflecting Emotions Based on Generative Adversarial Network

Lu Chen 1,*, Feifei Lee 1,*, Hanqing Chen 1, Wei Yao 1, Jiawei Cai 1 and Qiu Chen 2,*

1 School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China; 182560443@st.usst.edu.cn (L.C.); 183790736@st.usst.edu.cn (H.C.); 182560447@st.usst.edu.cn (W.Y.); 182560444@st.usst.edu.cn (J.C.)
2 Major of Electrical Engineering and Electronics, Graduate School of Engineering, Kogakuin University, Tokyo 163-8677, Japan
* Correspondence: feifeilee@usst.edu.cn (F.L.); chen@cc.kogakuin.ac.jp (Q.C.)

Received: 2 August 2020; Accepted: 26 August 2020; Published: 28 August 2020

Abstract: Manual font design is difficult and requires professional knowledge and skills to perform. Therefore, how to automatically generate the required fonts is a very challenging research task. On the other hand, there are few people who have studied the relationship between fonts and emotions, and common fonts generally cannot reflect emotional information. This paper proposes an Emotional Guidance GAN: an automatic Chinese font generation framework based on Generative Adversarial Network (GAN), which enables the generated fonts to reflect human emotional information. First, an elaborated questionnaire system was developed from Tencent company, which aims to quantitatively figure out the relationship between fonts and emotions. A visual expression recognition part is designed based on the trained model to provide a font generation module with conditional information. Moreover, the Emotional Guidance GAN (EG-GAN) with EM Distance and Gradient Penalty, as well as classification strategies, is proposed to generate new fonts with combined multiple styles that infer by an expression recognition module. The results of the evaluation experiments and the resolution of the synthesized font characters show the credibility of our model.

Keywords: generative adversarial network (GAN); font generation; expression recognition; questionnaire system; emotional guidance

1. Introduction

Recently with the development of advertising and multiple media technology rushing into our daily lives, text and characters are generally taken, in the area of publishing-and-printing and screen display, to show some emotions. The reason why most of these application scenarios express emotions using fonts that are appropriate for the characters is because characters alone cannot express enough emotions to users, especially for hieroglyphs such as Chinese and Japanese. Therefore, during the process of font design, designers tend to incorporate emotional information in them. [1,2] show that characters with specific styles can effectively exert the expression function of emotions, and at the same time, it can be a designed font that can connect characters with people’s emotions and resonate with the audience. What is more, modern Chinese character font expressions need to adapt to the different visual pursuits of modern people according to different purposes, environments, and artistic styles. It is very practical to study the relationship between fonts and emotions. Although the fonts’ emotion analysis has been well studied for analyzing the expression of the fonts’ characteristic, which is deemed necessary to study the emotion that the font shows, quantitative research on the emotional intention is rarely utilized. Considering the discussion above, we propose a questionnaire system to combine both
qualitative and quantitative aspects to exactly analyze the specific font attached to the corresponding facial expressions.

Different from western characters, common Chinese characters include more than 3500 characters, and each character has various fonts with different thickness, length, and straightness, as well as partial strokes. The design of Chinese fonts is very time-consuming and requires professional skill, artistic sense, and expertise in calligraphy. Compared to Latin alphabet font generation [3], as proposed by Hideaki Hayashi, the task of Chinese character generation is more complicated and difficult. Most previous works have focused on local strokes of characters rather than the whole. In contrast, Zi2Zi [4] is directly originated from the Pix2Pix [5] model, which uses paired images as training data to generate fonts automatically. However, the Zi2Zi model is hard to train due to mode collapse and instability. Since the performance of Generator and Discriminator is difficult to balance without knowing the metrics of the model training, some generated font images are somewhat blurred. Therefore, we propose the use of Wasserstein distance [6], which can reflect the training performance by measuring the difference between generated images and real images. We also employ Gradient Penalty [7] to enhance the stability of the model and improve the quality of the generated images.

Much research has been done to find what to recommend and how to choose appropriate fonts for users. Wada et al. [8] propose a system for automatically generating fonts reflecting Kansei. However, the proposed system operates based on the genetic algorithm that directly operates the parameters related to Kansei engineering many times to creates various kinds of fonts, which is complicated and time-consuming. Therefore, we propose a recommended automatic font generation system, which allows ordinary users who do not need professional font design capabilities to change the font to convey certain emotions. In order to obtain accurate style results, we incorporate a classification loss in the model. The novelty of this research is to generate a new font that more appropriately reflects the emotion contained in the input image.

Our main contributions of this paper can be summarized as follows:

1. We propose and design a questionnaire system to quantitatively and qualitatively study the relationship between fonts and facial expressions. Data analysis shows that the system has high credibility, and the result provides a dataset for further research.
2. In our model, we propose an Emotional Guidance GAN (EG-GAN) algorithm; employing an emotion-guided operation on the font generation module, the automatic Chinese font generation system is able to generate new style of Chinese fonts with corresponding emotions.
3. We incorporate EM Distance, Gradient Penalty, and a classification strategy to enable the font generation module to generate high-quality font images and make sure that each single font has a consistent style.
4. We conduct various experimental strategies on various Chinese font datasets. The experimental results are utilized as the basis for other questionnaires we propose to analyze, and it shows that generated fonts are credible with specific emotions.

The rest of this paper is shown as follows. Section 2 discusses about the review of related literatures, and a proposed questionnaire is elaborated in Section 3. A detailed model is presented in Section 4. Experimental results and analyses are described in Section 5, and the conclusion is drawn in Section 6.

2. Related Works

2.1. Font Emotion Research

The imagery and characteristics of Writing Characters have been keenly studied for decades. Most of this research focus on the evolution [9] and recognition [10], as well as the construction of Chinese characters [11]. Recently, some studies have demonstrated that fonts can reflect relative emotions [1,2,12,13]. Accordingly, these studies show that fonts can express some extra emotions.
besides characters’ meaning, [12] shows that a character’s features are based on the method that proposed which emotion is affected by the elements of character such as Genre, Serif, Tool kind, and Aspect ratio, which includes thickness, length, and straightness as well as the partial strokes. Wada et al. [8] propose a system to automatically generate font reflecting Kansei. In this paper, we focus on exploring which fonts can reflect facial emotions by using the questionnaire method to estimate the correlation between fonts and emotions. However, with regard to the deficiencies of these proposed methods of analyzing fonts’ emotion. Promotions are incorporated into our work to enhance the reliability of relationship between alternative fonts and emotions.

2.2. Generative Adversarial Network (GAN)

Previous studies [14,15] have introduced a generative model based on Generative Adversarial Network (GAN), which includes two parts as follows: Generator G to generate a distribution from a random noise, and the discriminator D is adversarially trained to distinguish the credible degree. However, traditional GAN tends to have a phenomenon of model collapse and gradient disappearance, which is elaborated in [6]. Various variants to improve the quality of GAN have been done in different tasks. For instance, the results of Least squares GANs (LS-GANs) [16], Wasserstein-GANs (WGANs) [6], and WGAN with gradient penalty (WGAN-GP) [7] show that using some useful tricks can be helpful to generate high-quality images.

2.3. Automatic Font Generation

Various research has been carried out in previous studies on automatic font generation. Different from Western letters generative tasks [3,17], Chinese characters generation tasks with the problems time-consuming and tough. Former studies are usually based on the hierarchical representation of strokes to represent Chinese characters [18,19]. In recent years, with the wide application of deeply generative model, the Generative Adversarial Network (GAN) [15] has been one of the very important models that have been applied for font generative tasks [4,20–22]. Zi2Zi [4], one of the automatic font generation methods, is based on paired images font generation, which considers each character as a whole and learns to transform between fonts. There are many studies like Combined cCVAE and cGAN [20] that carry out automatic font generation tasks.

2.4. Style Embedding Generation

The task of generating a specific style of fonts is studied by different means. Style migration based on the Convolutional Neural Network (CNN) [23] is employed to create a font with artistic style [24]. The task of handwritten font generation [20,21] mainly focuses on the characteristic of the font, which attracts the feature of resource input to generate a font with a specific style. Besides, some style learning tasks on font generation based on GAN use a style embedding strategy to generate new fonts [3,17,22,25]. However, since these generated new fonts are various and it is hard to guide the process of image synthesis, our proposed algorithm is useful and meaningful.

3. Questionnaire for the Relationship between Facial Expressions and Fonts

Common facial expressions are divided into eight categories [26] including ‘anger,’ ‘contempt,’ ‘disgust,’ ‘fear,’ ‘happiness,’ ‘neutral,’ ‘sadness,’ and ‘surprise.’ As illustrated in Section 1, the Chinese fonts can accurately reflect emotions. In order to obtain a better comprehending of people’s emotional tendency towards fonts, we began to quantitatively and qualitatively research the relationship between facial expressions and fonts. The questionnaires are set based on the relationship between facial expressions and fonts.
3.1. Questionnaire Design

We selected 110 Chinese fonts in the font collection, and these different fonts composed of Simplified Chinese are designed by artists. Thereafter, we chose 30 fonts that could best express a certain degree of emotional meaning. The fonts described by artists included classic Chinese fonts, such as Songti style, Kaiti style, and Lishu style. After that, we created a survey based on five calligraphy experts with more than ten years of experience and 25 common respondents, asking them to select 10 fonts that could reflect emotions from all the given options. We collected the survey results of 30 respondents (consist of five calligraphy experts and 25 common respondents). For considering the importance of representative samples, we adopted the following weight assignment: for the choices of the calligraphy experts and the common respondents, the weights are set to 1.5 and 1, respectively, which are empirically determined. Finally, we obtained the 10 most selected fonts as shown in Figure 1 for the next questionnaire.

![Figure 1. Options for questionnaire determined by five calligraphy experts and 25 common respondents.](image)

The eight facial expressions used in this questionnaire are posted on the Tencent platform according to the order of ‘anger,’ ‘contempt,’ ‘disgust,’ ‘fear,’ ‘happiness,’ ‘neutral,’ ‘sadness,’ and ‘surprise.’ A total of 452 male and female respondents who usually have the opportunity to see different fonts in their daily lives were tasked to pick up one of the ten fonts. The content of the questions and the corresponding example are shown in Table 1 and Figure 2, respectively.
Table 1. Questions 1–4 aim to acquire basic information of respondents, which is useful to improve the credibility of questionnaire. Questions 5–12 are main content of questionnaire.

| Question Sequence | Question Content |
|-------------------|------------------|
| Question 1        | What’s your name? |
| Question 2        | What’s your gender? |
| Question 3        | How about your experience on calligraphy? |
| Question 4        | Talking about fonts reflecting emotions |
| Question 5        | Choose a font best corresponds to “anger”. |
| Question 6        | Choose a font best corresponds to “contempt”. |
| Question 7        | Choose a font best corresponds to “disgust”. |
| Question 8        | Choose a font best corresponds to “fear”. |
| Question 9        | Choose a font best corresponds to “happiness”. |
| Question 10       | Choose a font best corresponds to “neutral”. |
| Question 11       | Choose a font best corresponds to “sadness”. |
| Question 12       | Choose a font best corresponds to “surprise”. |

Choose a font best corresponds to “anger”.

![Image of a person expressing anger]

○ Font 1 ○ Font 2 ○ Font 3 ○ Font 4 ○ Font 5
○ Font 6 ○ Font 7 ○ Font 8 ○ Font 9 ○ Font 10

Figure 2. A typical example of the questionnaire, which asking respondents to choose a font that corresponds to the emotion shown in the facial image (The emotion “anger” is shown here).

3.2. Questionnaire Results

Figure 3 shows the results of the questionnaire for emotions, and we can see that the results below show it is distinguishable for this questionnaire. It is found that font 3 is the most selected with 43.6% having chosen it, followed by font 7, which has been chosen by 19.9% of the participants for the emotion of anger. For the emotion of contempt, font 7 is the most selected with 45.4% having chosen it, followed by font 4, which has been chosen by 23.3% of the participants. For the emotion of happiness, font 6 is the most selected with 56.2% having chosen it, followed by font 10 with up to 13.5%. For the emotion of neutral, font 5 is the most selected with 76.1% having chosen it, followed by font 6, which has been chosen by 12.8% of the participants. For the emotion of surprise, font 8 is the most selected with 30.1% having chosen it, followed by font 1, which has been chosen by 13.5% of respondents. These results are relatively well distinguished. However, the choice of font for disgust, sadness, and fear is relatively unclear because the results’ difference in the questionnaire is small. Therefore, we choose font 9 because it is the highest option for all three emotions.
Figure 3. The main results of the questionnaire. For each emotion, the first two probabilities are marked, respectively.

According to the results of the questionnaire, we get some correspondence as shown in Table 2. For example, we find “Anger” best corresponds to font 3, as well as “Contempt” to font 7, etc.

Table 2. Quantitative correspondence between fonts and emotions.

| Emotion               | Font  |
|-----------------------|-------|
| Anger                 | Font 3|
| Contempt              | Font 7|
| Happiness             | Font 6|
| Neutral               | Font 5|
| Disgust, Fear, Sadness| Font 9|
| Surprise              | Font 8|

4. Architecture

In this task, we propose an Emotional Guidance GAN (EG-GAN) algorithm based on Zi2Zi merged with Earth Mover (EM) distance and the Gradient Penalty, as well as classification loss, which learns to map from random noise vector z, and we observe source font image x combined with two conditional singles that contain style information s and classification information f to y, \( G(x, s, f, z) \rightarrow y \).

The generator G is trained to produce domain images that cannot differentiate between the generated image and the real image by the corresponding trained discriminator, D, which is trained to recognize the generator’s output as well as possible. A bunch of paired images that are determined by the results of the questionnaire are fed to the font generation module. The training procedure of the font generation module is shown in Figure 4. Finally, with a guided information \( c(c_1', c_2') \) that contains \( c_1' \) and \( c_2' \) controlled by inputting the facial expression image, the generated Chinese fonts are taken to another questionnaire and other metrics to evaluate the creditable degree. The whole process is diagrammed in Figure 5.
surprise’ are calculated by the expression recognition module. Considering that the Fer2013 [29]
font generation.

As is shown in Figure 6, a state-of-the-art model illustrated in [27,28] is adapted to recognize
emotions, which takes facial image as an input and calculates the first two probabilities and makes a
process of data adjustment to guide new font generation with a specific style. The probabilities of
seven kinds of expressions including ‘anger,’ ‘disgust,’ ‘fear,’ ‘happiness,’ ‘neutral,’ ‘sadness,’ and
‘surprise’ are calculated by the expression recognition module. Considering that the Fer2013 [29]
expressions dataset is manually annotated and have certain errors, to make sure the generated fonts
(consists of source font images x and target domain images y that are ground-truth).

4.1. Network Architectures

4.1.1. Facial Information Extraction Module

As is shown in Figure 6, a state-of-the-art model illustrated in [27,28] is adapted to recognize
emotions, which takes facial image as an input and calculates the first two probabilities and makes a
process of data adjustment to guide new font generation with a specific style. The probabilities of
seven kinds of expressions including ‘anger,’ ‘disgust,’ ‘fear,’ ‘happiness,’ ‘neutral,’ ‘sadness,’ and
‘surprise’ are calculated by the expression recognition module. Considering that the Fer2013 [29]
expressions dataset is manually annotated and have certain errors, to make sure the generated fonts of
Chinese character exactly reflects emotions, we select the first two probabilities \( c_1 \) and \( c_2 \) of the expression
recognition module’s results to make sure that the two data \( c_1, c_2 \) are adjusted into the standard data
\( c'_1 = c_1 / (c_1 + c_2) \) and \( c'_2 = c_2 / (c_1 + c_2) \) usage of regularization as combined style labels for the process of
font generation.
4.1.2. Font Generation Module

As illustrated in Section 1 of our paper, the Zi2Zi model faces some problems as a Chinese character generation model. On the one hand, the model is difficult to train, easy to collapse, and often unstable. On the other hand, generating images leads to relative low resolution. Considering those problems mentioned above, we adopt some strategies to solve these issues. We use EM Distance [6]:

$$W(P_r, P_g) = \inf_\omega \Pi (P_r, P_g) \mathbb{E} (p_1, p_2) \omega [||p_1 - p_2||],$$  

which is continuous and differentiable, and which could solve the phenomenon of gradient disappearance. Where $\Pi(P_r, P_g)$ is the set of all possible joint distributions combined by $P_r$ and $P_g$, the $P_r$ is the distribution of real paired images (consisting of source font images and real target domain images), and the $P_g$ is the distribution of fake paired images (consisting of source font images and generated images). The marginal distributions of $\omega(p_1, p_2)$ are, respectively, $P_r$ and $P_g$. Equation (1) can be interpreted as the minimum cost of transforming the distribution $P_r$ into the distribution $P_g$.

Meanwhile, as the EM Distance continues to shrink the closer the two distributions are, this can provide indicators to guide the process of model training and visible supervision over the process of training and judging the metrics model of effectiveness, and this can further the enhancement towards the revolution of generating images. Furthermore, Gradient Penalty [7] is employed to enhance the quality of generation images.

$$L_{gradient\_penalty} = E_{x\sim P_r} [\|\nabla_x D(x)\|_2 - 1]^2, $$  

where $x$ is uniformly sampled from the straight line between pairs of points sampled from $P_r$ and $P_g$ [7]. We use L1 distance rather than L2, as L1 encourages less blurring:

$$L_{L1}(G) = E_{x,y,s,z} [\|y - G(x, s, z)\|_1].$$  

We define the generator loss:

$$L_{gen\_loss} = -E_{(x,z)\sim P_g} [D(G(x, z))],$$  

to encourage $G$ to fool the $D$ by the way of generating high-quality images as much as possible. The discriminator loss is

$$L_{disc\_loss} = E_{(x,z)\sim P_g} [D(G(x, z))] - E_{(x,y)\sim P_r} [D(x, y)].$$  

We use constant loss [30] to assume that the generated image and real image should reside in the same space and close to each other.

$$L_{const\_loss} = E_{(x,s,z)\sim P_g}[G_{encoder}(x, z)] - E_{(x,y)\sim P_r}[G_{encoder}(x, y)].$$
We use category loss [31] to solve the one-to-many function [32] by concatenating a non-trainable Gaussian noise as style information, and embedding $s$ to the character embedding to generate a target character. We incorporate the style information $s$ in the constant loss function and the category loss function.

$$L_{\text{category\_loss}} = L_{\text{real\_category\_loss}} + L_{\text{fake\_category\_loss}},$$

$$L_{\text{real\_category\_loss}} = -E_{(x,y) \sim p_t}(t \cdot \log \mathcal{D}(x,s,y) + (1-t) \cdot \log(1 - \mathcal{D}(x,s,y))),$$

$$L_{\text{fake\_category\_loss}} = -E_{(x,s,z) \sim p_g}(t \cdot \log \mathcal{D}(G(x,s,z)) + (1-t) \cdot \log(1 - \mathcal{D}(G(x,s,z)))),$$

where $t$ denotes target domain labels. Although those loss functions can be used in our model to generate high-quality images, some fonts generation will still be blurred. Therefore, we incorporate a classification strategy [22] to make the discriminator easier to distinguish the style of the generated characters.

$$L_{d\_cls} = -E_{(x,y,t') \sim p_t}(\log \mathcal{D}(t'|y)).$$

The $\mathcal{D}(t'|y)$ is a probability distribution on target domain labels computed by the Discriminator.

$$L_{g\_cls} = -E_{(x,t,z,f) \sim p_g}(\log \mathcal{D}(G(x,t,z,f))),$$

where $y$, $t$, and $f$ denote ground-truth images, ground-truth labels, and classification information, respectively.

Combining all the loss functions, the final objection function is

$$\min_{G} \max_{D} \left( L_G + L_D + \lambda L_{\text{gradient\_penalty}} \right),$$

and

$$L_G = L_{\text{gen\_loss}} + \phi L_{\text{const\_loss}} + \alpha L_{\text{L1}} + \beta L_{\text{fake\_category\_loss}} + L_{g\_cls},$$

$$L_D = L_{\text{disc\_loss}} + \gamma L_{\text{category\_loss}} + L_{d\_cls},$$

where $\lambda, \alpha, \beta, \gamma,$ and $\phi$ are hyper-parameters. The architecture of font generation module is shown in Figure 7.

**Figure 7.** The architecture of font generation. The features of the characters in the source font strive to have low-level information via the encoder. Then the abstracted information combines with style information $s$, classification information $f$, and guided information $c(c_1', c_2')$, the output of which can be regarded as the characteristics extracted in the combined target domains that depend on the guided information $c(c_1', c_2')$. The skip connections between encoder and decoder aim to share the low-level information. Finally, the generated images are decoded.
5. Experimental Results and Analyses

5.1. Comparison Experiments and Results

In Section 4, we illustrate our approach and the detailed model of generating fonts to reflect emotions. We divide the task into three parts. The first one is the expression recognition module, which is up to the state-of-the-art commonly used in expressions recognition on FER2013 [29]. In this part, the probabilities of seven kinds of expressions including ‘anger,’ ‘disgust,’ ‘fear,’ ‘happiness,’ ‘neutral,’ ‘sadness,’ and ‘surprise’ are calculated by the expression recognition module. Then the first two probabilities are adjusted into the standard data elaborated in Section 4.1.1. Intuitively, we can see the results of recognition and the generated font with corresponding emotions shown in Figure 8. The upper part of Figure 8 shows the results of expression recognition and the corresponding information after data processing. The bottom part of Figure 8 shows the characters of the source font and the generated characters mixed with two styles, respectively.

![Figure 8](image)

**Figure 8.** The results of expression recognition and the generated font with corresponding emotions. When inputting a facial image with specific emotions, a new font is generated by combining the first two probabilities of the emotions.

The font generation stage is performed into two parts: (1) the training process, and (2) the testing process. In the process of training, we adopt a strategy to split training into a two-step process. First, 27 fonts with 1000 characters are randomly sampled and are fed to the model described in Section 4.1.2; then, we freeze the parameters of the encoder. After that, we choose to fine-tune [4] to the six fonts with 3000 characters that are acquired from the questionnaire. By this way, the encoder learns the characters’ structure information with 27,000 characters in the first step, and the dedicated decoder can better focus on the characteristics of the target domain. Those loss functions elaborated in Section 4 are designed to make the proposed model generate high-quality images. We also confirm that the
two-timescale update rule (TTUR) [33] is effective, and we advocate using it specifically to address slow learning in regularized discriminators. We set the $\alpha = 100$, $\beta = 0.5$, $\phi = 15$, and $\gamma = 1$ during the first step, and $\alpha = 300$, $\beta = 0.5$, $\phi = 150$, and $\gamma = 0.9$ during the second step. We employ the style information $s$ and classification information $f$ in these two steps of the training process. In the process of the test, the guided signal $c$ is embedded into the generator $G$ to synthesize the specific style of font. The guided signal $c$ is adjusted from first two probabilities that are recognized in the process of expression recognition. The source font is Heiti style.

In our experiments, the EG-GAN model and EG-GAN $^1$ model are employed, respectively. “EG-GAN $^1$” represents the EG-GAN model without adopting the strategy of Gradient Penalty. Intuitively, compared with the Zi2Zi model, ours can successfully generate multiple style fonts with higher quality. The comparison results are shown in Figure 9.

Moreover, some evaluation metrics, structural similarity index (SSIM) [34,35], and peak signal-to-noise ratio (PSNR) [35], to show the quality of synthesized images, are employed in our experiment. SSIM is used to evaluate the aspects of luminance, contrast, and structure between two images. The higher the SSIM score, the clearer the description of image distortion. PSNR is a ratio between the real images and the reconstructed images to measure the quality of the images. We calculate the SSIM and the PSNR for grayscale images (8 bits). Input a reference image $u$ and a test image $v$, both of size $M \times N$, the PANR is defined by:

$$PSNR(u,v) = 10 \log \left( \frac{255^2}{MSE(u,v)} \right),$$

where

$$MSE(u,v) = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (u_{ij} - v_{ij})^2.$$  

The $MSE$ represents the cumulative squared error. The SSIM is defined as:

$$SSIM(u,v) = \text{lum}(u,v) \text{con}(u,v) \text{str}(u,v),$$

where

$$\begin{align*}
\text{lum}(u,v) &= \frac{2\mu_u \mu_v + \gamma_1}{\mu_u^2 + \mu_v^2 + \gamma_1}, \\
\text{con}(u,v) &= \frac{2\sigma_u \sigma_v + \gamma_2}{\sigma_u^2 + \sigma_v^2 + \gamma_2}, \\
\text{str}(u,v) &= \frac{\sigma_{uv}^2 + \gamma_3}{\sigma_u^2 \sigma_v^2 + \gamma_3}.
\end{align*}$$

The $\text{lum}(u,v)$ is the luminance function which calculates the proximity between $\mu_u$ and $\mu_v$. Here the $\mu_u$ and the $\mu_v$ are the mean luminance of input two images. The $\text{con}(u,v)$ is the contrast function which calculates the contrast of the input two images. Here the $\sigma_u$ and the $\sigma_v$ are the standard deviation of the input two images. The $\text{str}(u,v)$ is the structure function which calculates the correlation

---

**Figure 9.** The comparison results of fonts generated by various models. The images of the first row are the ground truth, and the second raw are generated by the Zi2Zi model. Images of the third and fourth rows are generated by EG-GAN $^1$ and the EG-GAN model, respectively.
of the input two images. Here the $\sigma_{uv}$ is the covariance between two images. The positive constants $\cap_1$, $\cap_2$ and $\cap_3$ are set to make sure that a non-null denominator.

A combination of the obtained six fonts image and every single font image is employed in our experiment. Table 3 shows the comparison results of the SSIM and PSNR metrics between Zi2Zi and our model in a combination of the obtained six fonts image. Table 4 shows the comparison results of the SSIM metric between Zi2Zi and our model in every single font image, and Table 5 shows the comparison results of the PSNR metric between Zi2Zi and our model in every single font image.

Table 3. Comparison of objective image quality metrics (SSIM, PSNR) of the proposed model.

| Model   | SSIM     | PSNR     |
|---------|----------|----------|
| Zi2Zi   | 0.8854   | 14.7226  |
| EG-GAN 1 | 0.8865   | 14.7501  |
| EG-GAN  | 0.8953   | 15.3035  |

1 Denotes EG-GAN without Gradient Penalty.

Table 4. Comparison of SSIM metrics for each single font.

| Model   | Font 3   | Font 5   | Font 6   | Font 7   | Font 8   | Font 9   |
|---------|----------|----------|----------|----------|----------|----------|
| Zi2Zi   | 0.8670   | 0.9006   | 0.9049   | 0.8820   | 0.8794   | 0.8809   |
| EG-GAN 1 | 0.8614   | 0.9036   | 0.9158   | 0.8732   | 0.8831   | 0.8835   |
| EG-GAN  | 0.8690   | 0.9082   | 0.9092   | 0.8997   | 0.8940   | 0.8915   |

1 Denotes EG-GAN without Gradient Penalty.

Table 5. Comparison of PSNR metrics for each single font.

| Model   | Font 3   | Font 5   | Font 6   | Font 7   | Font 8   | Font 9   |
|---------|----------|----------|----------|----------|----------|----------|
| Zi2Zi   | 14.2367  | 16.9022  | 16.3358  | 13.6608  | 13.7157  | 14.2021  |
| EG-GAN 1 | 14.1828  | 17.1461  | 17.1391  | 13.1789  | 14.0231  | 14.4951  |
| EG-GAN  | 14.6932  | 17.3627  | 16.4064  | 14.8043  | 14.8671  | 15.0289  |

1 Denotes EG-GAN without Gradient Penalty.

In order to confirm that employing the EM Distance can improve the convergence rate of the model, the sampled comparison results from training process of Zi2Zi and EG-GAN 1 are shown in Figure 10. We can see that our generated images have a higher quality and convergence rate in the same steps.

Meanwhile, we compare the details of EG-GAN $^{1,2}$ and EG-GAN $^{2}$ with EG-GAN $^{1}$ and EG-GAN, respectively. “EG-GAN $^{1,2}$” represents the EG-GAN model both without employing Gradient Penalty strategy and classification loss. “EG-GAN $^{2}$” represents the EG-GAN model without employing classification loss. We notice that the classification strategy is useful, and the comparison results are shown in Figure 11.
Meanwhile, we compare the details of EG-GAN 1,2 and EG-GAN 2 with EG-GAN 1 and EG-GAN, respectively. “EG-GAN 1,2” represents the EG-GAN model both without employing Gradient Penalty strategy and classification loss. “EG-GAN 2” represents the EG-GAN model without employing classification loss. We notice that the classification strategy is useful, and the comparison results are shown in Figure 11.

Figure 11. The comparison results of fonts generated by our models. The images of the first and second rows are the characters of Font 9, and the third and fourth rows are of Font 6. The models used to generate the images of the first to fourth rows are EG-GAN 1,2, EG-GAN 1, EG-GAN 2 and EG-GAN, respectively.

An evaluation experiment is conducted to verify the effectiveness of the proposed system. Specifically, the purpose is to verify whether the generated fonts can reflect the corresponding emotions. The experiment is in the form of a questionnaire posted on the Tencent platform that collects responses from 200 respondents for the generated six fonts shown in Figure 12, respectively. Respondents are asked to choose one font from options corresponding emotion. The results of this questionnaire present in Figure 13. According to the results of the evaluative questionnaire shown in Table 2, we get some
correspondence, which are as follows: (1) Font 6 reflects on “Happiness”; (2) Font 3 reflects on “Anger”; (3) Font 8 reflects on “Surprise”; (4) Font 7 reflects on “Contempt”; (5) Font 5 reflects on “Neutral”; and (6) Font 9 reflects on “Disgust,” “Fear,” and “Sadness.” All these questions are answered more than 50%, as shown in Figure 13, which confirms that our algorithm is credible.

![Fonts generated by our proposed model for reliability analysis.](Image)

**Figure 12.** Fonts generated by our proposed model for reliability analysis.

**Figure 13.** Evaluation results of fonts generated by our proposed model.

5.2. Discussion

The performance of the Generator and Discriminator is the key issue for ordinary GAN during the training process. However, it is difficult to balance the performance without knowing the metrics of the model training. In addition, in order to evaluate the difference between the real image and the generated image, the KL divergence and the JS divergence [6,7] for the Discriminator are used, but there is a problem of gradient vanishing during the initial training. The experiment results shown in Figures 9 and 10 intuitively confirm the effectiveness of incorporating the strategies of EM Distance and Gradient Penalty. Meanwhile, SSIM and PSNR metrics are employed in our experiments to imply lower numerical differences compared with the baseline. Generally, the Zi2Zi model can achieve the function of one-to-many. However, since the performance of the classification function proposed previously only focuses on the individual information, some generated fonts are somewhat blurred. Therefore, we incorporate conditional information to make sure that each single font has a consistent style. In order to verify the effectiveness of the classification strategy, the comparison results are shown in Figure 11. Meanwhile, these generated fonts shown in Figure 12 are used as questions for the questionnaire, and results with high reliability are shown in Figure 13. These comparison experiments mentioned above verify the effectiveness of our proposed font generation system.
6. Conclusions

In this paper, we have proposed an automatic Chinese font generation algorithm reflecting emotions based on Generative Adversarial Network (GAN). A questionnaire is designed to study the relationship between fonts and emotions. According to the results of the questionnaire, the fonts and emotions are associated with each other. The Emotional Guidance GAN (EG-GAN) model combines the guided signal recognized by the expression recognition module with the font generation module to synthesize new fonts that reflect corresponding emotions. We incorporate EM Distance, Gradient Penalty, and TTUR, as well as classification loss tricks, to improve the resolution of the generated images and to reasonably demonstrate the generated fonts with quantitative styles by using the weight of the adjusted labels from the results of the expression recognition module. Some comparison experiments are employed to confirm the effect of our models. In addition to these, an evaluation experiment is carried out further proving the credibility of this task. Moreover, we can also get some inspirations to generate fonts with various styles contained in different scenes.

Author Contributions: Methodology, L.C. and F.L.; software, L.C.; data curation, H.C.; validation, W.Y. and J.C.; writing—original draft preparation, L.C.; writing—review and editing, F.L. and Q.C.; supervision, F.L. and Q.C.; funding acquisition, F.L. and Q.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by The Programme for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, and also partially supported by JSPS KAKENHI Grant Number 15K00159.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Choi, S.; Aizawa, K. Emotype: Expressing emotions by changing typeface in mobile messenger texting. Multimed. Tools Appl. 2019, 78, 14155–14172. [CrossRef]
2. Amare, N.; Manning, A. Seeing typeface personality: Emotional responses to form as tone. In Proceedings of the 2012 IEEE International Professional Communication Conference, Orlando, FL, USA, 8–10 October 2012; pp. 1–9.
3. Hayashi, H.; Abe, K.; Uchida, S. GlyphGAN: Style-consistent font generation based on generative adversarial networks. Knowl. Based Syst. 2019, 186, 104927. [CrossRef]
4. Tian, Y.C. zi2zi: Master Chinese Calligraphy with Conditional Adversarial Networks. Available online: https://kaonashi-tyc.github.io/2017/04/06/zi2zi.html (accessed on 30 July 2020).
5. Isola, P.; Zhu, J.Y.; Zhou, T.H.; Efros, A.A. Image-to-image translation with adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
6. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 214–223.
7. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved training of wasserstein GANs. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach, CA, USA, 4–9 December 2017; pp. 5769–5779.
8. Wada, A.; Hagiwara, M. Japanese font automatic creating system reflecting user’s kansei. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Washington, DC, USA, 8 October 2003; pp. 3804–3809.
9. Dai, R.W.; Liu, C.L.; Xiao, B.H. Chinese character recognition: History, status and prospects. Front. Comput. Sci. China 2007, 1, 126–136. [CrossRef]
10. Liu, C.L.; Jaeger, S.; Nakagawa, M. Online recognition of chinese characters: The state-of-the-art. IEEE Trans. Pattern Anal. Mach. Intell. 2004, 26, 198–213. [PubMed]
11. Lan, Y.J.; Sung, Y.T.; Wu, C.Y.; Wang, R.L.; Chang, K.E. A cognitive-interactive approach to chinese characters learning: System design and development. In Learning by Playing. Game-Based Education System Design and Development. Edutainment 2009; Chang, M., Kuo, R.K., Chen, G.D., Hirose, M., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2009; pp. 559–564.
12. Chuang, H.C.; Ma, M.Y.; Feng, Y.C. The features of chinese typeface and its emotion. In Proceedings of the International Conference on Kansei Engineering and Emotion Research, Paris, French, 2–4 March 2010.

13. Shen, H.Y. The Image and the meaning of the chinese character for ‘enlightenment’. *J. Anal. Psychol.* 2019, 64, 32–42. [CrossRef] [PubMed]

14. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* 2014, arXiv:1411.1784.

15. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the Advances in Neural Information Processing Systems 27(NIPS), Montreal, QC, Canada, 8–13 December 2014; pp. 2672–2680.

16. Mao, X.D.; Li, Q.; Xie, H.R.; Lau, R.Y.K.; Wang, Z.; Smolley, S.P. Least squares generative adversarial networks. *arXiv* 2016, arXiv:1611.04076.

17. Yuan, D.J.; Feng, H.X.; Liu, T.L. Research on new font generation system based on generative adversarial network. In Proceedings of the 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Hohhot, China, 24–26 October 2019; pp. 18–21.

18. Zong, A.; Zhu, Y. Strokebank: Automating personalized chinese handwriting generation. In Proceedings of the 28th AAAI Conference on Artificial Intelligence, Quebec City, QC, Canada, 29–31 July 2014; pp. 3024–3029.

19. Pan, W.Q.; Lian, Z.H.; Sun, R.J.; Tang, Y.M.; Xiao, J.G. Flexifont: A flexible system to generate personal font libraries. In Proceedings of the 2014 ACM Symposium on Document Engineering, Fort Collins, CO, USA, 16–19 September 2014; pp. 17–20.

20. Kong, W.R.; Xu, B.C. Handwritten chinese character generation via conditional neural generative models. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–7 December 2017.

21. Chang, B.; Zhang, Q.; Pan, S.Y.; Meng, L.L. Generating handwritten chinese characters using CycleGAN. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 12–15 March 2018; pp. 199–207.

22. Chen, J.F.; Ji, Y.L.; Chen, H.; Xu, X. Learning one-to-many stylised chinese character transformation and generation by generative adversarial networks. *IET Image Process.* 2019, 13, 2680–2686. [CrossRef]

23. Balouchian, P.; Foroosh, H. Context-sensitive single-modality image emotion analysis: A unified architecture from dataset construction to CNN classification. In Proceedings of the 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 1932–1936.

24. GitHub. Available online: https://github.com/yuweiming70/Style_Migration_For_Artistic_Font_With_CNN (accessed on 29 July 2020).

25. Azadi, S.; Fisher, M.; Kim, V.; Wang, Z.W.; Shechtman, E.; Darrell, T. Multi-content gan for few-shot font style transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7564–7573.

26. Langner, O.; Dotsch, R.; Bijlstra, G.; Wigboldus, D.H.J.; Hawk, S.T.; Knippenberg, A.V. Presentation and validation of the radboud faces database. *Cogn. Emot.* 2010, 24, 1377–1388. [CrossRef]

27. GitHub. Available online: https://github.com/WuJie1010/Facial-Expression-Recognition.Pytorch (accessed on 29 July 2020).

28. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.

29. Goodfellow, I.J.; Erhan, D.; Carrier, P.L.; Courville, A.; Mirza, M.; Hamner, B.; Cukierski, W.; Tang, Y.; Thaler, D.; Lee, D.-H.; et al. Challenges in representation learning: A report on three machine learning contests. *Neural Netw.* 2015, 64, 59–63. [CrossRef] [PubMed]

30. Taigman, Y.; Polyak, A.; Wolf, L. Unsupervised cross-domain image generation. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2017.

31. Odena, A.; Olah, C.; Shlens, J. Conditional image synthesis with auxiliary classifier GANs. In Proceedings of the 34 International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 2642–2651.

32. Johnson, M.; Schuster, M.; Le, Q.V.; Krikun, M.; Wu, Y.; Chen, Z.; Thorat, N.; Viégas, F.; Wattenberg, M.; Corrado, G.; et al. Google’s multilingual neural machine translation system: Enabling zero-shot translation. *TAI.* 2017, 5, 339–351. [CrossRef]
33. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6629–6640.

34. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* 2004, 13, 600–612. [CrossRef] [PubMed]

35. Horé, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).