Embedding Secret Message in Chinese Characters via Glyph Perturbation and Style Transfer

Yao, Ye; Wang, Chen; Wang, Hui; Wang, Ke; Ren, Yizhi; Meng, Weizhi

Published in:
IEEE Transactions on Information Forensics and Security

Link to article, DOI:
10.1109/TIFS.2024.3377903

Publication date:
2024

Document Version
Peer reviewed version

Embedding Secret Message in Chinese Characters via Glyph Perturbation and Style Transfer

Ye Yao, Chen Wang, Hui Wang, Ke Wang, Yizhi Ren, Weizhi Meng

***Abstract*—Glyph perturbation adjusts the characters' structures and strokes to make the original characters change subtly, which cannot be detected by the naked eye. These generated variants with different glyph perturbation can represent different status of secret messages, which can be used to embed information in Chinese text documents. However, Chinese characters have characteristics in large numbers, complex structures, and diverse fonts, which limit the generation of glyph perturbation and make the design of Chinese characters time-consuming and laborious. Many font style transfer methods for Chinese characters have been proposed to improve the efficiency of Chinese character generation based on deep learning. At present, there are few studies on efficient font style transfer for glyph perturbation of Chinese characters. In this paper, a stylized glyph perturbation method based on style extractor and attention augmented convolution is proposed. It adopts a multi-head attention mechanism to enhance convolution in the font transfer, which concatenates the convolution feature maps and the self-attention activation maps to weaken the limitations of ordinary convolution in processing images. The extracted style features are sent into the decoder of the font transfer network so as to improve the stylized ability. Particularly, the impact of style extractor and attention augmented convolution on the glyph perturbation generation is addressed. The extraction accuracy and embedding capacity are tested in our experiments. The embedding capacity of secret message can achieve around 1.8 bit/character.**

***Index Terms*—Chinese character generation, glyph perturbation, information hiding, font style transfer, attention augmented convolution**

I. Introduction

GLYPH perturbation adjusts the structures and strokes of characters to make the original characters change subtly, which cannot be detected by the naked eye. These modified characters with different glyph perturbation can represent different status of secret messages. It can be used for information hiding in text documents.

For example, considering the Chinese character '讯' as shown in Fig. 1, the movements (up & down or left & right) of two selected strokes – marked in red, can produce four glyph variants accordingly. Specifically, both the horizontal turning with a rising stroke in the left part and the horizontal stroke in the right part of the character '讯' in FangSong font style are moved up and down, respectively. Therefore, different stroke-moving combinations can result in four glyph variants of '讯', totally. Each glyph variant represents a different state of secret message in two-binary digits.

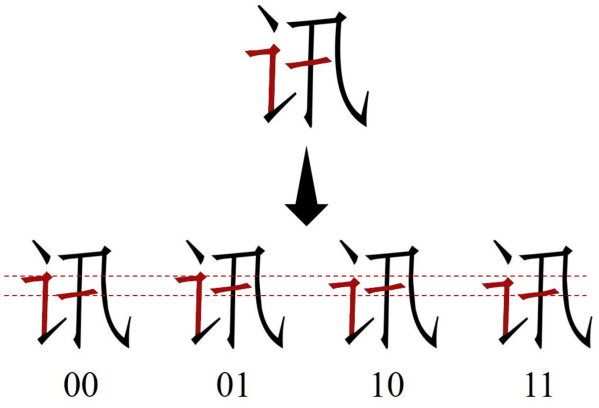


Fig. 1. Glyph perturbation of Chinese character '讯'. Secret message is embedded by moving the red strokes up and down to produce four glyph variants of Chinese character '讯' that represent different state of secret message in two binary digits.

However, the production of glyph variants by glyph perturbation is a laborious and time-consuming task. The structure of Chinese characters is more complex, and the number of Chinese characters is much larger than that of some other languages, such as English, French and German. For example, the standard coding for Chinese characters named GB2312-80 includes 6,763 commonly-used characters [1]. Moreover, there are a wide variety of font types in the font library, including different typography designed by professional designers. Therefore, it becomes an urgent task to deal with the problem caused by the complexity of Chinese character generation with glyph perturbation.

A reasonable approach to tackle this issue is to combine the glyph perturbation and style transfer of Chinese characters together, and to change the stroke position of Chinese characters based on deep learning. Many font style transfer methods for Chinese characters [2]–[4] have been proposed to improve the efficiency of Chinese character generation. However, there are few studies on font style transfer for glyph perturbation [5], [6] of Chinese characters at present. Compared with the existing font style transfer methods, glyph perturbation has to pay more attention to the position of strokes while transferring the font style and generating glyph variants from source to target fonts.

In this paper, we propose a font style transfer method for

Y. Yao, C. Wang, H. Wang, K. Wang and Y. Ren are with the School of Cyberspace, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China. W. Meng is with the Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark. (*Corresponding author: Yizhi Ren* and *Weizhi Meng*)

This work was funded in part by the Humanities and Social Sciences Foundation of Ministry of Education of China under Award Number 23YJA870013, and the Zhejiang Provincial Natural Science Foundation of China under Award Number LY24F020017.

Manuscript received June 5, 2023; revised November 10, 2023.

generating glyph variants of Chinese characters. It contributes to the generation of Chinese characters under the above mentioned difficulties. The Chinese characters in text documents can be replaced by generated glyph variants to convey secret message.

The main contributions of this paper are listed as follows:

1) We integrated the glyph perturbation into the field of font-style transfer to generate stylized glyph variants of Chinese characters. The generated Chinese glyph variants could be applied into secret message embedding and extraction. It could have a wide application prospects for data hiding in text documents.
2) The multi-head attention mechanism is adopted in the encoder of the font transfer network, which enhances the learning ability of the network on the skeleton of Chinese characters and avoids the disadvantages of ordinary convolution that fails to capture the internal correlation of different regions in image processing.
3) The style extractor network is responsible for sampling style features from Chinese characters with target font and splicing them into deconvolution layers of the font transfer network. It strengthens the learning ability of style features in the generated images.
4) To identify the position changes of strokes for Chinese characters, perturbation loss and patch-pixel loss are proposed and applied to comparing extracted features and counting white pixels in each fixed-size patch with the corresponding position.
5) The secret message embedding and extracting processes using the generated Chinese characters with glyph perturbation are described in details. Our proposed method is verified through experimental study on embedding capacity and extracting accuracy, respectively. The embedding capacity of secret message can achieve around 1.8 bit/character in our experiments.

The rest of the paper is organized as follows. Section II introduces the existing methods of Chinese font transfer and glyph perturbation of characters. A font style transfer method is proposed for glyph perturbation based on style extractor and attention augmented convolution in Section III. Section IV experimentally validates and discusses the effectiveness of style features and attention augmented convolution. Section V introduces the application of glyph perturbation for Chinese characters in text documents and Section VI summarizes the work and future directions.

II. Related Work

A. Chinese Font Transfer

The font transfer methods for Chinese characters based on deep learning are classified into stroke generation and Chinese character generation according to whether the characters are split or not. The former mainly generates desired strokes in target fonts, which constitute complete characters through a set of predefined rules [7]–[11]. The latter generates images of Chinese characters by extracting and learning font features.

Auto-encoder has the ability to extract features of the input. Its encoder extracts the necessary features that can represent the input to restore the real images. Therefore, many researchers have utilized it to design networks for Chinese character generation. Xiao et al. [12] introduced the font label with one-hot encoding into U-Net [13] for controlling font categories, and adopted the loss of mean absolute error to enhance the sharpness and clarity of generated images. This method accomplished one-to-many font transfer of Chinese characters. Aiming at the limitation of label controlling style features, they designed a transfer network for artificially controlling style and content features. Variational auto-encoder (VAE) [14] was used to extract style features and fuse them with font labels as content features. It is possible to transfer the font with small samples by distributing the encoder, but the generated images of Chinese characters are not as clear as those generated by the auto-encoder. SA-VAE [15] defined an informational rule to supplement the structural details of Chinese characters, which encodes each character into 133-bit coding based on its structures and strokes. Using the dependence between content and style of images, EMD [16], [17] extracted common style and content features from a set of style-reference images (with different contents but the same style) and content-reference images (with different styles but the same content). The extracted features are fused by a bilinear model to generate Chinese characters with specified style and content.

Recent developments in generative adversarial networks (GAN) have led to a renewed interest in the style transfer, and many researchers have introduced the idea of style transfer to the font transfer of Chinese characters. According to the way of learning style features of Chinese characters, these methods can be divided into three categories: self-learning style features, external style features, and extracted style features.

1) Self-Learning Style Features: Rewrite2 [18] is inspired by Rewrite, GAN, DCGAN and conditional GAN. It adopts adversarial loss as the objective optimization function. Despite some noise interference, these generated images of Chinese characters are recognizable. Based on a style transfer network for images named Pix2Pix [19], Unet-GAN [20] extended its network to 16 layers by increasing the number of convolution layers. It converted the font from printing to handwriting while retaining the structure details of Chinese characters. PEGAN [21] added cascaded refinement connection into the encoder of its generator and used the pre-trained VGG19 to calculate the perception loss for network optimization. HAN [22] described the global skeleton and local stroke details of Chinese characters through feature maps in lower and higher layers, generated corresponding images of middle layers, and sent the generated images to its discriminator. The loss generated by the middle layer images can improve the generator's ability to match the real images. Inspired by the self-attention mechanism that performs well in image generation tasks [23], SAFont [24] used self-attention blocks to calculate the features changes of Chinese characters before and after transferring. It added edge loss to the optimization function, which makes the strokes of Chinese characters clearer.

From the perspective of strokes and radicals, Lu et al. [25] selected the Chinese characters containing the maximum

number of radicals as the training set. The encoder is designed to extract the characters of source and target fonts separately, and a complete Chinese characters library with target fonts can be generated by learning small samples of Chinese characters with target fonts. StrokeGAN [26] introduced the concept of stroke encoding and defined the stroke-encoding reconstruction loss to preserve the stroke details of Chinese characters.

2) External Style Features: Zi2Zi [27] connected the font label to the embedding layer of auto-encoder based on Pix2Pix [19] and added judgment for font category. It is suitable for font transfer tasks of Chinese and Korean characters. CalliGAN [28] took dictionary sequences and features extracted by encoders as the content of Chinese characters, and concatenated them with one-hot vectors representing target fonts. Chen et al. [29] used a style specifying mechanism to combine the images with one-hot vectors of font labels, which can generate a variety of Chinese characters in different fonts. Losses of font category and semantic consistency were added to constrain the network optimization. SSNet [30] extracted and restored stroke features (Horizontal, Vertical, Left-falling, Right-falling, Tuning) of Chinese characters by structural module. It initialized the style features with random Gaussian noise and optimized the network with dual-masked Hausdorff distance to improve the quality of generated images. Gao et al. [31] proposed a font transfer network with three stages named ENet-TNet-RNet, which extracted the skeleton by using a set of mask matrices and font labels, then converted it into the skeleton with target font, and rendered stroke details to generate Chinese characters finally.

3) Extracted Style Features: AEGN [32] is a calligraphy font transfer method composed of two auto-encoder networks. The supervision network provided a transfer network with strokes details of target characters. In the transfer network, the residuals module connected the encoder and decoder to learn subtle differences in the spatial structures between the source and target. DCFont [33] used VGG16 to extract high-level features of images, which were fused with content features extracted from the font transfer network and sent into the decoder. CocoAAN [34] maps Chinese characters with source and target fonts to content and style features respectively through alternate optimization strategy. It adds a fully connected layer after the first three convolution layers of discriminators to improve the generalization ability of Chinese characters with new fonts.

FontGAN [2] integrated the stylization and de-stylization of Chinese characters into the framework. The font consistency module and content prior module were introduced to solve the problem of strokes missing in the process of de-stylization. TET-GAN [35] extended artistic Chinese characters to the field of font transfer and designed a stylized and de-stylized network under the framework of auto-encoder. To separate style and content features of Chinese characters, AGIS-NET [36] extracted common style features from a dataset of reference images with consistent style, fused content and style features, and used cooperated decoders to generate shape images and texture images at the same time.

For unpaired datasets of Chinese characters, Xiao et al. [37] proposed a multiple mapping model of font transfer for Chinese characters. It normalized the style features of Chinese characters in order to generate Chinese characters with multiple fonts. In addition, the Kullback-Leibler Divergence loss was added to make the style features extracted by the style encoder consistent with Gaussian distribution, which is suitable for the font transfer tasks of printing and handwriting. DG-Font [38] employed deformable convolution to learn content features of Chinese characters and introduced a feature deformation skip connection (FDSC) to predict displacement maps mixing the low-level feature maps in the content encoder and decoder. These mixed features are sent into the content decoder together with style features extracted by the style encoder, which generate the Chinese characters with target font.

B. Glyph Perturbation of characters

At present, text documents are widely used in the paperless office. Secret information can be embedded in the text by adding digital watermarks, which can be used for copyright protection and source tracing. The document watermarking methods based on glyph perturbation of characters adjust characters size, shape, brightness, and other characteristics to embed messages in text documents. Several researchers have proposed information hiding methods based on deep learning for the glyph perturbation of characters.

To embed additional information in text documents, FontCode [5] focused on assigning each English letter an integer and embedding the integer by perturbing the glyph of each letter to a precomputed codebook. When extracting information from text documents, it requires a high resolution of document characters to achieve reliable information retrieval. The embedded information might lose in the case of occlusion, contamination, destruction, etc. Due to the limitations of generating a font library, FontCode is unsuitable for Chinese characters with large dataset and complex structures.

To solve the problem of automatic watermarking and generation of Chinese characters, Sun et al. [39] embedded the watermark into images of Chinese characters, generated watermarked Chinese characters, and trained a neural network for extraction. It had strong robustness in digital transmission and print-scan scenarios, which alleviated the distortion in transmission. It is less sophisticated in embedding and extracting watermarks from photo scenes and suffers greatly from environmental instability.

III. Proposed Method

Because the structure of Chinese characters is more complex [6], Chinese characters are more suitable for glyph perturbation than English characters to embed secret messages. However, the large number of Chinese characters with various fonts increases the difficulty in implementation of glyph perturbation. Different from previous style transfer works and information hiding methods, we integrated the glyph perturbation into the field of font-style transfer to generate stylized glyph variants of Chinese characters. Our proposed method aims to learn style changes of strokes and generate glyph variants of Chinese characters in an expected font.

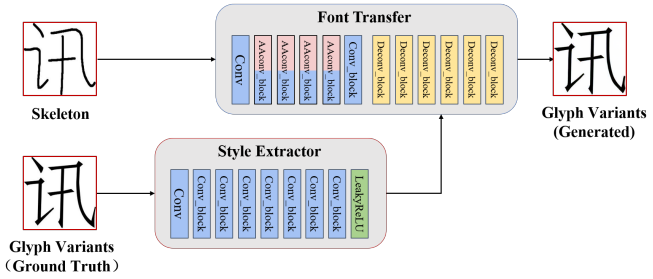


Fig. 2. Overview. (down) Our style extractor network takes the input images of Chinese characters and extracts the styles of strokes as the representation of style features. (up) In addition to learning features of Chinese characters, extracted style features are supplemented to a font transfer network. After training, the skeleton is converted by the font transfer network into stylized glyph variants of Chinese characters.

Fig. 2 presents an overall framework of our proposed stylized network, which can be divided into two stages: the extraction of style features and the stylization of glyph perturbation for Chinese characters.

A. Network Architecture

1) Attention Augmented Convolution: In the field of computer vision, the Convolutional Neural Network (CNN) performs well, especially in image classification. It is characterized by locality and translation equivariance through the receptive field and weight sharing. However, the receptive field only focuses on the region with a fixed size of the convolution kernel in processing images. The inherent locality of CNN makes it unable to obtain global information and ignores the relationship between each region of images.

Considering that Chinese characters are mainly composed of one or more radicals arranged and combined in two-dimensional space, global information is essential for glyph perturbation of Chinese characters. To make up this defect of CNN, the attention mechanism, as its name implies, is to sift through a large amount of information and focus on a small amount of important information. The greater the weight, the more important the information. Self-attention mechanism is a variant of attention mechanism, which reduces the dependence on external information, captures the internal correlation of data or features, and has the advantage of low computational complexity [40]. Unlike convolution operation in CNN, it generates a weighted average of the values calculated from the hidden cells, whose weight is dynamically generated by the similarity function between the hidden cells. Self-attention mechanism shows great potential in capturing long-range dependencies by calculating the interaction between words in texts. The calculation formula of the self-attention mechanism [23] is as follow:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

Take the translation from Chinese into English as an example, query vector Q represents the English characters, key vector K represents the Chinese characters, and value vector V represents the content of Chinese characters. In general, the dimensions of Q and K are equal, denoted by d_k. The weight is obtained through the similarity between Chinese and English characters, normalization, and an activation function named SoftMax.

Self-attention mechanism can be used as a layer in the neural network, which either replace the convolutional layer [23] or alternate with the convolutional layer [40]. When using separately, it generally requires position encoding to revise since the location of the input is ignored [23]. Multi-head attention mechanism (MHA) utilizes multiple queries to select multiple heads of information from the input in parallel, and each head focuses on different region, enabling it to capture different interactive messages in multiple feature subspaces.

Similar to convolution, attention augmented convolution has the translation equivariance of ordinary convolution, which is also applicable to images of different spatial dimensions. Thus, attention augmented convolution is conducive to improving learning ability in image classification and target detection tasks. The formulas of calculating attention convolution are as follows:

$$Attention_h = softmax(\frac{(ImgW_q)(ImgW_k)^T}{\sqrt{d_k^h}})(ImgW_v) \tag{2}$$

$$MHA(Img) = W_0[Attention_1 \oplus \cdots Attention_{N_h}] \tag{3}$$

$$AAConv(Img) = MHA(Img) \oplus conv(Img) \tag{4}$$

Here the input 'Img' is set as (B, C, H, W), where 'B' represents the batch size meaning the number of images accepted at one time, 'C' represents the number of channels of the inputs, 'H' and 'W' represent the height and width of the input images respectively. Suppose only one image is sent in at a time, the image is flat into a matrix denoted by $Img \in \mathbb{R}^{HW \times C}$, N_h represents the number of attention heads, d_k represents the dimension of Queries and Keys, and d_v represents the dimension of Values. In Eq. 3, for a certain self-attention mechanism h, W_q, $W_k \in \mathbb{R}^{C \times d_k^h}$, and $W_v \in \mathbb{R}^{C \times d_v^h}$ represent linear transformation coefficients from input images to Queries, Keys and Values. It is also called random initialization mapping matrix and can be expressed as: $Q = ImgW_q$; $K = ImgW_k$; $V = ImgW_v$. $W_0 \in \mathbb{R}^{d_v \times d_v}$ is an output mapping matrix associated with multi-head attention. After multi-head attention processing, the size of the input image is modified into (B, d_v, H, W). It is connected with the convolution feature maps on the dimension of C to obtain the final result.

In this paper, we adopted a multi-head attention mechanism to enhance convolution without increasing parameters. It connects the convolutional feature maps with self-attention activation maps to enhance convolution [41]. The translation equivariance can be kept while incorporating with relative position information, which makes the attention augmented convolution applicable for image processing.

2) Style Extractor Network: As shown in Fig. 3, the style extractor network consists of the convolution layer, seven LeakyReLU-Conv-BatchNormal blocks, and the LeakyReLU activation function. Each convolution layer adopts a convolution kernel with the size of 4×4 to conduct down-sampling in

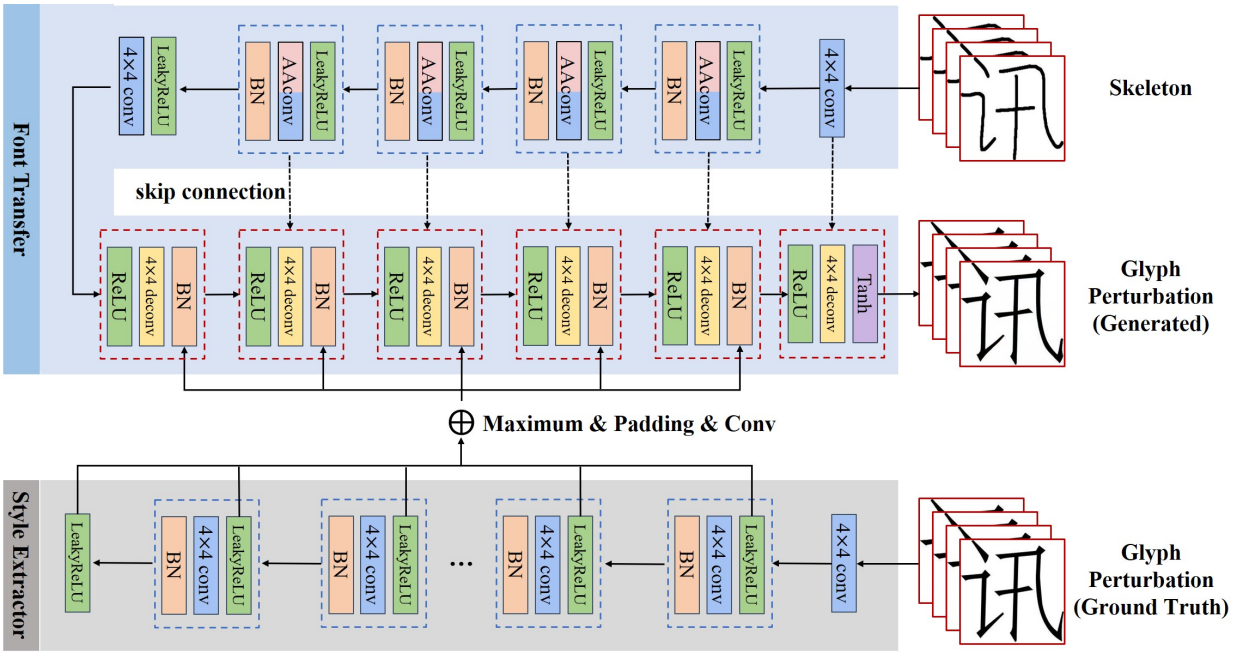


Fig. 3. Network architecture. The skeleton and stylized images are input into the network, which are both containing a single Chinese character. It consists of a font transfer network (up) and style extractor network (down). It generates stylized glyph variants of Chinese characters, in which each character is remolded by moving two selected strokes.

the stride of 2. The features output by each activation function layer are loaded line by line and selected the maximum to represent the features of the corresponding layer, and then connected and sent to the font transfer network.

3) Font Transfer Network: As shown in Fig. 3, the font transfer network consists of an encoder and a decoder. The former is composed of two convolution layers, four LeakyReLU-AAconv-BatchNormal blocks, and a LeakyReLU activation function. The latter is composed of six ReLU-Deconv-BatchNormal blocks but the last one replaces the normalization with the Tanh activation function.

In the encoder part, the convolution layer adopts a convolution kernel with the size of 4×4 and a stride of 2 to encode the input Chinese character image. Each LeakyReLU-AAconV-BatchNormal block uses multi-head attention to enhance convolution, as shown in Fig. 4. We concatenated the features of the self-attention mechanism with the features of the ordinary convolution and sent them to the next layer for normalization. Among them, the convolution kernel of 4×4 is used to sample three tensors named Q, K, and V with a stride of 2. Set the number of Head N_h to 4, the depth of Values D_v to 4, and the depth D_k of Queries and Keys to 40 in the following experiments.

In the decoder part, the encoded features of Chinese characters are upsampled by deconvolution layers with the kernel of 4×4 and stride of 2. In addition, the above style extractor network provides the style information to the decoder to generate Chinese characters with the stroke style of specified font. These style features are first filled to be consistent with the size of decoded features corresponding, and then fused with the decoded Chinese character features. It uses padding and a 1×1 convolution kernel to flexibly control the size

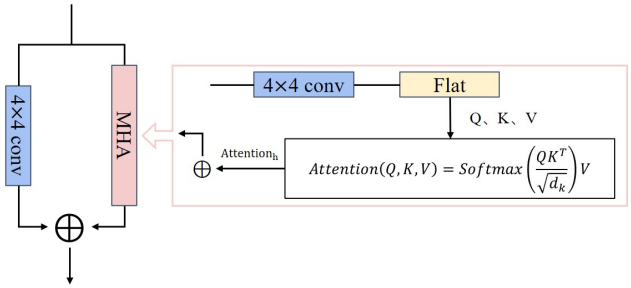


Fig. 4. The illustration of attention augmented convolution. The feature map of convolution, which emphasizes locality, is connected with the multi-head attention.

of style features and make it consistent with the features in corresponding deconvolution layers.

We use the discriminator of PatchGAN [19] to distinguish the authenticity of the generated and ground truth of Chinese characters. As shown in Fig. 5, the discriminator consists of two convolution layers, a LeakyReLU activation function layer, and three Conv-BatchNormal-LeakyReLU blocks. All convolution layers adopt the convolution kernel of 4×4. Except for the last two convolutional, whose stride is 1, the remaining convolutional layer is set to be 2.

In adversarial training, the style extractor network and font transfer network aim to learn content and style features of Chinese characters from the real samples and generate fake images to trick the discriminator into making wrong decisions. The purpose of the discriminator is to identify as accurately as possible whether the input data comes from the real samples or the font transfer network. They are constantly optimized during training to improve their generative and discriminant

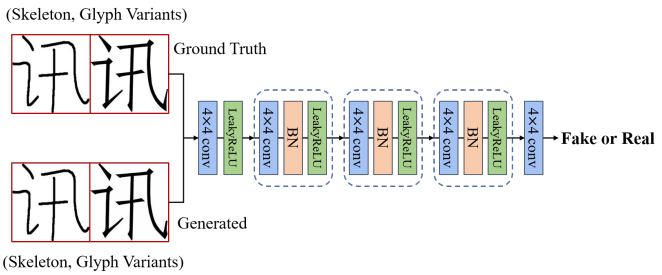


Fig. 5. The illustration of the discriminator. It determines the authenticity of the paired images, whether it is from real samples or artificial generation.

abilities. Through the game confrontation, the discriminator can hardly distinguish the real from the generated images.

B. Loss Function

The objective function of our model is presented as follow:

$$L_{total} = \lambda_{adv} L_{adv} + \lambda_1 L_1 + \lambda_p L_{perturb} + \lambda_{patch} L_{patch} \quad (5)$$

which consists of adversarial loss, L_1 loss, perturbation loss and patch-pixel loss with λ_{adv}, λ_1, λ_p and λ_{patch} control the weight of them.

1) Adversarial loss: Our proposed method consists of a style extractor network, font transfer network and discriminator, which are trained in adversarial learning. The main purpose of the style extractor network and font transfer network is to make the discriminator unable to correctly distinguish the generated images from the real, that is, to misjudge the authenticity of images. The discriminator is to judge the authenticity of the inputs. The images generated by the font transfer network mislead the discriminator that constitutes the loss function:

$$L_{adv} = \mathbb{E}_t \left[log \left(D \left(t \right) \right) \right] + \mathbb{E}_s \left[1 - log(D \left(G \left(s \right) \right)) \right] \quad (6)$$

where *s* and *t* denote real images of source and target font correspondingly, and *G(s)* denotes generated images of Chinese characters with target font.

2) L_1 loss: The generated images of Chinese characters are obtained by the font transfer network from skeleton images. Thus, the pixel-level loss between the real and generated images of target font is defined as follow:

$$L_1 = \mathbb{E}_{(s,t)} \left[\| \, t - G \left(s \right) \, \|_1 \right] \quad (7)$$

where *t* denotes ground truth images and *G(s)* denotes generated images with target font.

3) Perturbation Loss: Fig. 6 provides four glyph perturbations of character '法' in source and target fonts, respectively. They are paired according to the position changes of strokes using green or yellow lines. From green lines in Fig. 6, the character '法' is perturbed by moving the first horizontal stroke up and down. Also, it is apparent from yellow lines that the middle dot in the left part is moved up and down alternately.

$$L_{perturb\_g} = \sum_{img=s,G(s)} \mathbb{E}[\|img_1 - img_4\|_2] - \mathbb{E}[\|img_2 - img_3\|_2] \quad (8)$$

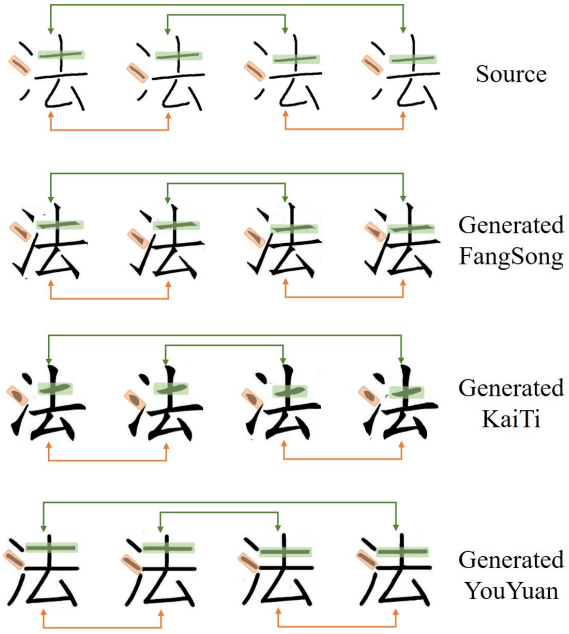


Fig. 6. Illustration of perturbation loss. They are paired according to the position changes of strokes using green or yellow lines. The green lines is perturbed by moving the first horizontal stroke up and down. Also, The yellow lines that the middle dot in the left part is moved up and down alternately.

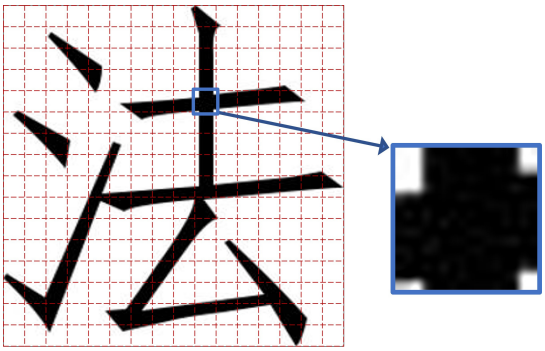


Fig. 7. Illustration of patch-pixel loss. The Chinese character '法' is split into small patches with the size 16×16 by red dotted lines, and one of the patches is explicitly shown on the right.

$$L_{perturb\_y} = \sum_{img=s,G(s)} \mathbb{E}[\|img_1 - img_2\|_2] - \mathbb{E}[\|img_3 - img_4\|_2] \quad (9)$$

$$L_{perturb} = L_{perturb\_g} + L_{perturb\_y} \quad (10)$$

Where $L_{perturb\_g}$ and $L_{perturb\_y}$ are perturbation loss labeled green or yellow correspondingly, img encompasses source and generated images of Chinese characters, and $\{img_1, \ldots, img_4\}$ denote four corresponding glyph perturbations for each character.

4) Patch-pixel Loss: To learn the spatial information of strokes better, we divided the real and generated images of the target font into the small patch with the size 16×16 and calculated the number of white pixels in each patch to form two one-dimensional vectors. For example, the Chinese

character '法' is split into patches by red dotted lines in Fig. 7, and one of the patches is explicitly shown on the right. The fitting degree of two vectors constitutes patch-pixel loss:

$$L_{patch} = H(W_t, W_{G(s)}) = -\sum_{k=1}^{K} w_t^k log(w_{G(s)}^k) \quad (11)$$

Where $W_t = \{w_t^1, \ldots, w_t^k, \ldots, w_t^K\}$ and $W_{G(s)} = \left\{w_{G(s)}^1, \ldots, w_{G(s)}^k, \ldots, w_{G(s)}^K\right\}$ denote white pixel vectors of real and generated images of target font, w_t^k and $w_{G(s)}^k$ mean the count of white pixels in the k-th patch, and K means the count of patches in each image.

C. Network Training

The proposed method consists of a stylized network and discriminator in adversarial learning. The former serves as a generator whose purpose is to learn the target styles and generate stylized glyph variants of Chinese characters. Instead, the latter distinguishes between generated and ground truth of Chinese characters. The skeleton images and ground truth are paired named (s, t) and sent into the stylized network to learn the target styles. It generates Chinese characters $G(s)$ corresponding to the skeleton and judges the authenticity of the input images by the discriminator. Adam optimizer and loss functions are used to optimize parameters, as shown in Algorithm 1. In training, the paired data are sent to different networks separately, in which the skeleton images are sent into the font transfer network, and the Chinese characters with target font are sent into the style extractor network. Under the set iterations, the parameters of the stylized network are fixed first, and training samples with batch size are sent in one time to calculate the gradient according to loss functions and update the parameters of the discriminator. Then the parameters of the discriminator are fixed, and the stylized network parameters are updated.

Algorithm 1: Network training using Adam optimizer

Data: training dataset $D = (\boldsymbol{s}^{(n)}, \boldsymbol{t}^{(n)})_{n=1}^{N}$, batch size B, loss function L, iterations I.
Result: network parameters $\hat{\theta}_G$ and $\hat{\theta}_D$ for stylized network and discriminator

repeat
  for $i \leftarrow 1$ ***to*** I **do**
    for $b \leftarrow 1$ ***to*** B **do**
      $//discriminator$
      $\theta_D \leftarrow Adam(\nabla\theta_D - \mu\frac{\partial L_{adv}}{\partial \theta_D}, \theta_D)$;
      $//stylized\ network$
      $\theta_G \leftarrow Adam(\nabla\theta_G - \mu\frac{\partial L_{total}}{\partial \theta_G}, \theta_G)$;
    end
  end
  $\hat{\theta}_G \leftarrow \theta_G$;
  $\hat{\theta}_D \leftarrow \theta_D$;
until *Reaches the maximum of iterations*;

IV. Experiments

A. Experimental Settings

We built an image dataset for glyph perturbation of Chinese characters in our experiment to verify our method's availability, which contains 995 Chinese characters and 3980 glyph variants in each font. It can be expanded by adding new characters and fonts in the data preparation. In addition, the experiments adopt skeleton images [36] as the source font and select three commonly used fonts of Chinese characters as the target: FangSong, KaiTi, and YouYuan. In the experiments, we take these glyph variants produced by 795 Chinese characters as a training dataset, and the remaining constituted test dataset.

All images of Chinese characters are with size 256×256. The initial learning rate of Adam optimizer is set to be 0.0002, and its β_1 is set to be 0.5. Weights in the loss functions are set as λ_{adv}=1.0, λ_1=100.0, λ_p=15.0 and λ_{patch}=1.0. For Chinese characters in each font, we use a batch size of 4 with 200 training epochs.

B. Competitors

Different from the font style transfer [42] methods, glyph perturbation has to pay more attention to the position of strokes while transferring the font style and generating glyph variants. At present, there are few studies on glyph perturbation for Chinese characters. Xiao et al. [5] proposed an information embedding technique for English documents by perturbing the glyphs of English characters while preserving the text content. Due to the limitations of generating a font library, It is unsuitable for Chinese characters with large dataset and complex structures.

Many font style transfer methods for Chinese characters [2]–[4] have been proposed to improve the efficiency of Chinese character generation. We compared our proposed method with five font transfer methods for Chinese characters with available source codes.

1) Rewrite2: Rewrite2 [18] is inspired by Rewrite, GAN, DCGAN and conditional GAN. It adopts adversarial loss as the objective optimization function. The generated images of Chinese characters are available to be recognized, despite there was some noise interference on them.

2) Zi2Zi: Zi2Zi [27] is a supervised font generation method based on Pix2Pix [19], it achieves font generation and uses Gaussian Noise as category embedding to achieve multi-style transfer. In addition to L_1 loss and adversarial loss, the constant loss between generated and real images is added to optimize the network. As for the printed characters with thick strokes and clear structures, the quality of generated images is higher.

3) TET-GAN: TET-GAN [35] extracts and combines the content features and style features of artistic characters, completing the task of stylization and de-stylization of Chinese characters. It provides skeleton images with obvious content features for unsupervised learning of font transfer.

4) DG-Font: DG-Font [38] is the state-of-the-art unsupervised font generation method. It introduces deformable convolution to enhance the learning ability of content features. Meanwhile, feature deformation skip connection performs

TABLE I
QUANTITATIVE EVALUATION COMPARISON WITH THE CLASSIC METHODS (BEST RESULT IN BOLD).

Method	Skeletonr⇒FangSong			Skeleton⇒KaiTi			Skeleton⇒YouYuan		
	RMSE↓	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑
Rewrite2 [18]	0.2865	11.1144	0.7876	0.3475	9.3540	0.7320	0.3841	8.4571	0.7239
Zi2Zi [27]	0.1699	15.5200	0.8425	0.2157	13.4430	0.7921	0.2209	13.2676	0.7836
TET-GAN [35]	0.1301	17.8151	0.8896	0.1151	18.8896	0.8993	0.2264	13.1226	0.8115
DG-Font [38]	0.3482	9.2977	0.6257	0.3806	8.5124	0.6012	0.4070	7.9318	0.5592
Glyph-Font [6]	0.1380	17.5177	0.9072	0.1373	18.2590	0.8989	0.1391	13.1183	0.7931
Proposed	**0.1121**	**19.2494**	**0.9107**	**0.1117**	**19.2039**	**0.9053**	**0.2223**	**13.4039**	**0.8047**

spatial transformation for low-level feature maps to retain strokes and radicals of Chinese characters, which effectively improves the integrity of generated characters.

5) Glyph-Font: Glyph-Font [6] is based on parallel auto-encoder networks for font transfer of Chinese characters, which focuses on the position of strokes while transferring the font in adversarial training. It simultaneously generates four glyph variants of each character in target fonts. A difference discriminator is trained to measure the difference between the real and generated images. In addition, the perturbation loss and patch-pixel loss are defined to distinguish position changes of strokes and amend incorrectly generated pixels in generated images.

C. Evaluation Metrics

Common indexes of quantitative evaluation in images include MSE, RMSE, PSNR, and SSIM [43]. From the perspective of pixels, MSE measures the pixel error of the corresponding positions of two images. RMSE is the square root of MSE. PSNR is used to evaluate the quality of image compression objectively. SSIM measures the difference between two images based on brightness, contrast, and structure.

The performance of methods is measured by comparing the quality of the generated images. Based on the fact that the images of Chinese characters are composed of black and white pixels, we selected RMSE, PSNR and SSIM [43] as evaluation indexes. It is considered that the images of Chinese characters generated by the font transfer network are more realistic, when the value of RMSE is smaller, the value of PSNR is larger, and the value of SSIM is closer to 1.

D. Experimental Results

1) Qualitative Results: In experiments, we selected three fonts frequently used in daily life: FangSong, KaiTi, and YouYuan. Fig. 8 show the Chinese characters with FangSong, KaiTi, and YouYuan generated by our proposed and five comparison methods mentioned above. From the generated results, the strokes generated by Rewrite2 were not smooth and complete with poor identifiability. DG-Font generates distorted strokes of glyph variants in KaiTi and YouYuan. In

Fig. 8. An example of Chinese characters generated by different methods.

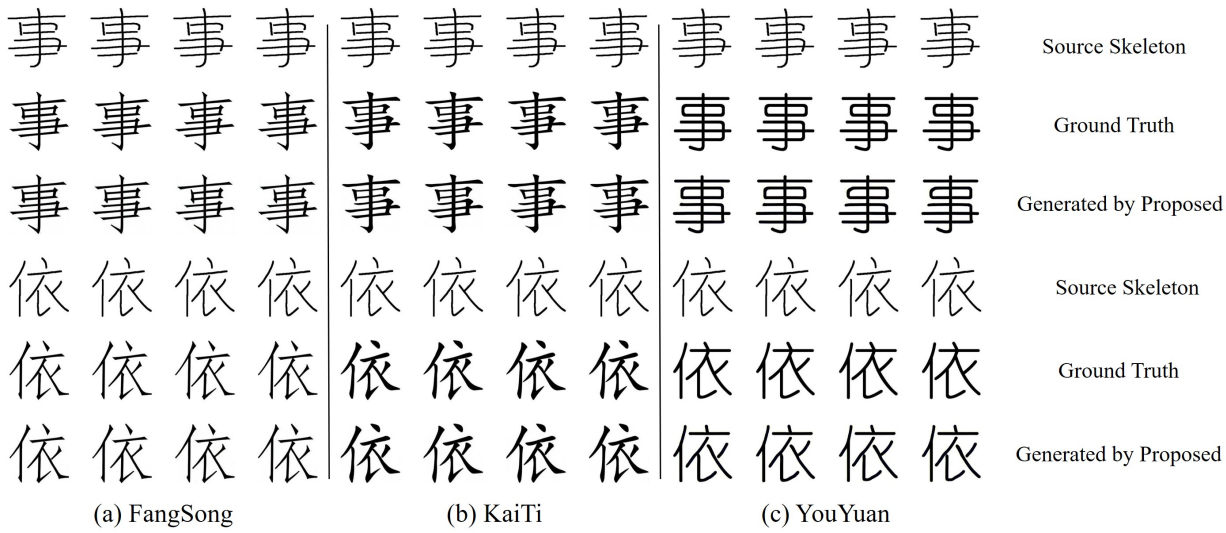


Fig. 9. Detailed comparison of generated characters with target fonts (such as FangSong, KaiTi, and YouYuan).

comparison, the Chinese characters generated by Zi2Zi, TET-GAN, Glyph-Font, and the proposed method are clear, and the style of target font is obvious. The stroke details of the method proposed in this paper are closer to the real images of Chinese characters. In addition, stylized glyph variants of Chinese characters with a target font named YouYuan are generated by our method. Part of the generated glyph variants of Chinese characters '事' '依' with three target fonts are shown in Fig. 9. The generated Chinese characters have clear strokes, complete structures, high authenticity, and obvious style of the target fonts.

2) Quantitative Results: The quantitative metrics of evaluation mentioned above measure the quality of generated images more directly. The smaller RMSE, the larger PSNR, and the closer SSIM to 1 mean the higher quality of generated Chinese characters. In order to compare the images generated by different methods, we uniformly set the image resolution and executed grayscale processing. Then, RMSE, PSNR, and SSIM of evaluation indexes are calculated. The comparison of quantitative results between the proposed method and five existed methods is shown in Table I. It shows that compared with the other methods, the proposed are superior in the RMSE, PSNR, and SSIM, demonstrating the better visual quality of the proposed scheme. Overall, the performances of the other five methods are relatively poor, while our method can generate realistic images that are difficult to be distinguished from the ground truth. It is more suitable for the fonts with clear transitions and forceful strokes relative to the font like YouYuan with smooth and delicate strokes.

E. Ablation Study

To verify the role of extracted style features and attention augmented convolution play, four cases are chosen for experimental validation with or without style extractor and attention augmented convolution, respectively.

1) Influence of extracted style features: The style features extracted by the style extractor network supply information about the target font. Our experiments compare two situations in which the style features are directly sent into the middle layer of the auto-encoder or deconvolution layers of the decoder in the font transfer network. The fourth and sixth lines of Fig. 10 respectively show the generation results of Chinese characters in different feeding ways of extracted style features. In terms of the generated results, the Chinese characters are more complete by connecting the style features with the outputs of each deconvolution layer in the decoder and sending them into the following normalization layer.

2) Influence of attention augmented convolution: In order to verify the effect of attention augmented convolution, attention augmented convolution is replaced by ordinary convolution to retrain the network in the encoder of the font transfer network. In the case that the extracted style features are sent into the middle layer of the auto-encoder, the Chinese characters generated by the font transfer network with or without attention augmented convolution correspond to the third and fourth lines of Fig. 10, respectively. In the case that the extracted style features are sent into each deconvolution layer of the decoder, the generated Chinese characters with or without attention augmented convolution correspond to the fifth and sixth lines of Fig. 10, respectively. Comparatively, attention augmented convolution further improves the completeness of strokes and the visual quality of generated results.

To more intuitively reflect the promotion effect of attention augmented convolution and style features on the stylized network of glyph perturbation, the comparison between the method proposed in this chapter and the quantitative results of the other three cases is shown in Table II. In the task of stylizing glyph variants of Chinese characters in FangSong and KaiTi, the method proposed in this chapter is superior to other comparative cases in objective evaluation indexes, which is consistent with the above qualitative results. By contrast, it can be seen that the observed meliorate in generated characters

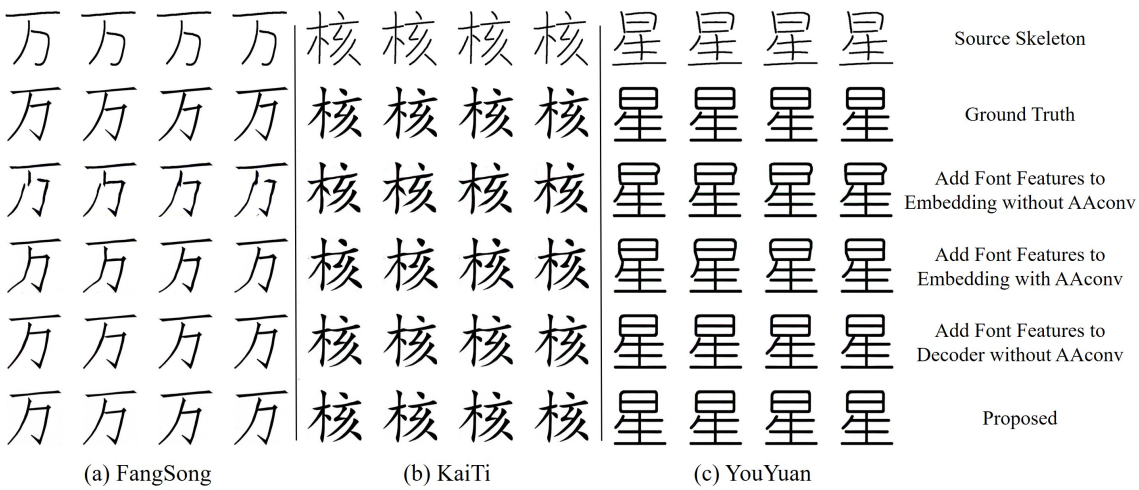


Fig. 10. The influence of extracted style features and attention augmented convolution on the font transfer of Chinese characters.

TABLE II
QUANTITATIVE EVALUATION COMPARISON WITH THE INFLUENCE OF EXTRACTED STYLE FEATURES AND ATTENTION AUGMENTED CONVOLUTION (BEST RESULT IN BOLD).

Method	Skeletonr⇒FangSong			Skeleton⇒KaiTi			Skeleton⇒YouYuan		
	RMSE↓	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑	RMSE↓	PSNR↑	SSIM↑
Add style features to embedding without AAconv	0.2937	10.7733	0.7193	0.3049	10.4280	0.7135	0.3280	9.8380	0.6961
Add style features to embedding with AAconv	0.2038	13.9869	0.8183	0.2255	13.0884	0.7897	0.2696	11.6676	0.7520
Add style features to decoder without AAconv	0.1298	17.9288	0.8853	0.1308	17.8014	0.8808	0.2238	13.3294	0.8042
Proposed	**0.1121**	**19.2494**	**0.9107**	**0.1117**	**19.2039**	**0.9053**	**0.2223**	**13.4039**	**0.8047**

could be attributed to attention augmented convolution and style features. It further verifies the improvement of visual quality in the qualitative results, which is effective for generating Chinese characters with high integrity.

V. APPLICATIONS

Our proposed method integrates glyph perturbation into the field of font style transfer to embed secret message in Chinese characters. These stylized glyph variants of Chinese characters can be used in text documents for hiding messages with wide application prospects. In this section, the secret message embedding and extracting processes using the generated Chinese characters with glyph perturbation are described in details. Our proposed method has been verified through experimental studies on embedding capacity and extracting accuracy, respectively.

A. Embedding Secret Message

This section describes the process of embedding the secret information into general text document by using the perturbed and deformed characters generated by our proposed method. This will be followed by performance analysis on the embedding capacity for secret information embedding.

1) The Embedding Process: In the general electronic text files, the most printed characters are in standard vector fonts. This paper proposes a method of generating perturbed characters for standard Chinese characters with glyph perturbation adjustment. The secret message embedding is realized by replacing the standard character with the generated perturbed character in the electronic files. The standard Chinese character set is denoted as $\hat{S}$. In our experimental design, only the commonly-used characters from $\hat{S}$ are selected as training dataset for generating glyph variants in this paper. The selected common Chinese character set is denoted as S, where $S \in \hat{S}$. For each standard Chinese character S_i in S, there are four corresponding perturbed characters in the glyph variants set of G, which are marked as G_i^{00}, G_i^{01}, G_i^{10} and G_i^{11}, respectively. The superscripts of G_i with value of 00, 01, 10 and 11 are the 2-bit binary numbers of the four perturbed characters.

Here the secret message to be embedded is denoted as M consisting of binary numbers. The process of embedding a secret message M in a general electronic file is described as follows.

1) For each Chinese character in the electronic file, the first step is to judge whether the character belongs to the commonly-used Chinese character set S. If it belongs to S, the four corresponding deformed characters in G can be used to carry secret information. The character is marked as S_i. If it does not belong to S, the character will be ignored, and the judgement loop keeps on going to process for the next character in the electronic file.
2) After the judgement processing, for each marked character S_i, 2-bit binary number is read out from the secret message M for embedding. According to the value of 2-bit binary number, one perturbed character is selected from the four glyph perturbed characters G_i^{00}, G_i^{01}, G_i^{10} and G_i^{11}, and is denoted as G_i^{sec}. For example, if the 2-bit secret message to be embedded is 10, then the G_i^{sec}

is recorded as G_i^{10}.

3) The character S_i in electronic file is replaced by the perturbed character G_i^{sec} to complete the 2-bit secret message embedding.
4) Repeat steps 1 to 3 to complete embedding the remaining part of secret message.

How to replace the standard Chinese characters in the document with the perturbed ones generated by the proposed method in this paper is a question belongs to the research scope of coding implementation and engineering application, which involves the technology of electronic document format parsing and the encapsulation of font library. Therefore, the related technologies on solving the above-mentioned question is not discussed in this paper.

2) The Embedding Capacity: In our experiments, the glyph perturbation is only applied to the selected commonly-used Chinese characters, which are part of the standard Chinese character set $\hat{S}$. Therefore, the average capacity for secret message embedding is less than 2 bits per character. However, since the selected Chinese characters have a higher usage frequency and covers most commonly-used characters in general document, the embedding capacity will be close to 2 bits per character, theoretically.

There are about 3500 primary commonly-used Chinese characters listed in the Table of General Standard Chinese Characters. According to the character frequency list of informative texts in Modern Chinese [44], the cumulative frequency of the first 1000 high-frequency used Chinese characters is around 91.48%. In other words, for most Chinese text document, the first 1000 high-frequency Chinese characters can cover more than 90% of the characters in the document. Here in our experiments, the first 995 high-frequency Chinese characters are selected to generate the glyph variants. Since the 995 characters selected in this paper are all belongs to the first 1000 high-frequency Chinese characters, the average embedding capacity of secret information is around $2\ bits \times\ 90\% \ =\ 1.8\ bit/character$.

B. Extracting Secret Message

In this section, the algorithm for secret message extracting is described in details, including the image preprocess after secret message embedding with or without print-scan scenarios, the glyph variants match process, and the experimental result analysis of extracting accuracy.

1) Image Preprocess: To extract the secret message from the electronic document with secret information hiding, we only need to convert the disturbed characters from the electronic document into the digital images, and matching them with the glyph variants which are generated by our designed glyph perturbation algorithm. However, it is commonly occurred that the electronic documents with secret information hiding are printed as paper documents in the process of transmission. The distortion caused by the printing, scanning or photographing will bring the interference to the matching algorithm between the distorted character images and the glyph perturbed character images.

According to the different application scenarios for message extraction, different image preprocessing algorithms are designed to obtain the deformed character images for extracting the secret information hiding in the document.

1) print-scan processing: The printed document is scanned to obtain a digital image by a scanner. The scanned image is relatively flat without obviously geometrical distortion. The character image patches are cut from the scanned image which will be used for message extraction by matching the glyph variants. Generally speaking, the binary image can be clearly obtained from the character images which are binarized using OTSU method. If the stroke disconnection or incoherence occurred in the binary image, the image expansion and corrosion operations will be applied into the binary image.
2) print-photo processing: The printed document is photographed to obtain a digital image by a phone. The photographed image is relatively not flat and always with more obviously geometrical distortion and distortion caused by uneven illumination. Therefore, the photographed image needs to processed by basic image operations such as perspective transformation and sharpening before binarization. Then, the OTSU binarization is carried out to obtain a clear binarized image. After that, the character image patches with clear stroke structure can be cut down from the binary image which will be used for message extraction by matching the glyph variants.

2) Glyph Variants Match: The algorithm of glyph variants matching is the key part in the process of secret information extraction. The input of the algorithm is the binary image after preprocessing, and the output is the secret message hiding in the glyph perturbed character images. The main flow of the perturbed character matching algorithm is as follows.

1) The pre-processed binarized image is divided into independent image patches in accordance with the character as the unit. All the character image patches are collected into a character image set, denoted as $\hat{C}$.
2) Determine whether each Chinese character in $\hat{C}$ in turn belongs to the common Chinese character set S. If it does, it means the character would carry the secret message, and would have corresponding glyph variant character in the set of G. Then the character is added into the set of C. If it does not belong to S, the character will be ignored, and the judgement loop keeps on going to process for the next character in $\hat{C}$.
3) Take out each character image patch C_i in the set of C and its four corresponding glyph variants in the set of G, denoted as T_i^{00} ,T_i^{01}, T_i^{10} and T_i^{11}. Scale the five character-images into the same size $M \times N$, where M and N are integral multiples of 16. Here the values of M and N are both set to be 256 in this paper.
4) The character image C_i and the its four corresponding standard glyph variants image T_i^{00}, T_i^{01}, T_i^{10} and T_i^{11} are divided into 16×16 blocks. The number of character strokes' pixels in each block is counted. The feature

TABLE III
THE EXTRACTION ACCURACY COMPARISON WITH THE DIFFERENT FONT TYPES AND DIFFERENT FONT SIZES.

	Font Type	Font Size (24×18 in A4)	Font Size (32×25 in A4)	Font Size (36×33 in A4)
print-scan	FangSong	98.84	98.43	97.64
	KaiTi	98.05	97.67	97.13
	YouYuan	98.39	97.04	96.35
print-photo	FangSong	94.30	95.43	91.88
	KaiTi	94.52	93.39	88.16
	YouYuan	93.83	91.37	91.31

vector of each character image is a sequence of the number of character strokes' pixels calculated from each 16×16 block. Then the feature vector V_i of the character image C_i and the feature vector V_i^{00}, V_i^{01}, V_i^{10} and V_i^{11} of the glyph variants T_i^{00}, T_i^{01}, T_i^{10} and T_i^{11} are obtained.

5) Normalize the feature vector V_i and the feature vectors V_i^{00}, V_i^{01}, V_i^{10} and V_i^{11}, respectively.
6) Calculate the Euclidean distance between the feature vector V_i and the its four corresponding feature vectors V_i^{00}, V_i^{01}, V_i^{10} and V_i^{11}, respectively. And regard it as the similarity degree.
7) Among the four calculated similarity degrees, take the feature vector V_i^{sec} with the most matching similarity. And the 2 bits secret message can be extracted from the value of sec. For example, if the Euclidean distance between V_i and the feature vector V_i^{10} is the smallest, the character image C_i is considered to carry the 2-bit secret message of 10.
8) Repeat steps 3 to 7 to complete matching of the remaining part of character images and the secret message extraction. The whole secret message would be obtained by combining all extracted secret message bits together.

3) Extraction Accuracy: Aiming at the scenario that the digital documents containing glyph variants characters will be printed into paper documents, our work designs several experiments to extract the secret message from the digital images which are captured from the paper documents by scanner and mobile phone, respectively. In the experiment, the accuracy of extraction was calculated according to different font type and different font size. The experimental results are shown in Table III. Here the Font type includes FangSong, KaiTi and YouYuan. The Font size includes three kinds of settings. In the experiment, all the Chinese characters are printed on a A4 paper. The font size is regarded as $m \times n$ where m is the number of characters per line and n is the total number of lines per page.

The extraction accuracy of secret message depends on both the image preprocessing algorithm and the glyph variants based Chinese character matching algorithm. It can be seen from Table III that the smaller the Font size is, the more characters are printed per page, and the lower the extraction accuracy becomes. It indicates that the strokeś glyph perturbation is not easy to recognize when the character image becomes small. Meanwhile, the extraction accuracy in print-scan scenario is higher than that of print-photograph scenario. This is because the print-photograph attack would cause more distortion in the acquired image. Before secret message extraction, the character image needs to be corrected by perspective transformation, which would introduce additional distortion.

C. Typical application scenarios

In this section, two typical application scenarios of the results of glyph perturbation for Chinese characters in text documents are introduced.

1) Information hiding in fixed-layout electronic files: Currently there is a popular electronic file format on the Internet that is independent of the application, operating system and hardware, such as PDF and OFD files. These files named fixed-layout files commonly have a standardized format. Utilizing the glyph variants of Chinese characters generated by our method to replace the standard characters in the fixed-layout files, invisible secret message can be embedded into the electronic documents. It is allowed to embed the attribution information of the electronic documents, the identity information of the user, etc. Embedding information in fixed-layout files invisibly makes the confidentiality responsibilities of files clarified. No matter whether these fixed-layout files are distributed or printed [45], the secret information within is not easily damaged, which can be applied in copyright protection and tampering detection.

2) Information hiding in streaming screen text: When users browse unformatted text through screen terminals, such as word files, web pages, and other streaming files, the forensics for leakage caused by taking photos of the screen is difficult. Despite many research aiming to prevent screen-shooting [46] by overlaying an visible image watermark on the screen, the visible watermark embedded greatly affects visual performance. Glyph variants of Chinese characters assigned unique identity are generated by our method, which can be installed on the screen terminal. If this screen terminal is photographed, it is possible to extract unique identity information from glyph variants in the leaked photograph for forensics.

D. Limitations and Discussion

In this paper, we only present the secret message embedding and extraction process under basic print-scan and print-photograph attacks. The analysis of the character recognition accuracy following multiple print-scan attacks or multiple photocopy attacks is not included in this study. We intend to investigate different image processing techniques to minimize interference in image recognition and improve character matching accuracy in future work.

Applying our method to other complex languages for information hiding in text documents is a highly meaningful research endeavor. The languages whose characters have similar stroke structure features to Chinese ones include Korean and Japanese. Currently, we lack training datasets of generated character variants with different glyph perturbation for these new languages. Designing and creating these datasets is a very time-consuming task. We will focus on this research direction and attempt to achieve cross-language transfer generation in future work.

VI. Conclusion

Our work addresses font transfer for glyph perturbation of Chinese characters designed with style extractor and attention augmented convolution. It is based on auto-encoder and divided into two stages: style extractor and font transfer. The style extractor network is responsible for sampling style features from Chinese characters with target font and splicing them into deconvolution layers of the font transfer network. It strengthens the learning ability of style features in the generated images. The multi-head attention mechanism is adopted in the encoder of the font transfer network, which enhances the learning ability of the network on the skeleton of Chinese characters and avoids the disadvantages of ordinary convolution (e.g., fails to capture the internal correlation of different regions in image processing).

The experimental results show that excellent agreement can be reached between the generated and ground truth images for Chinese characters, resulting in 0.1121, 19.2494, and 0.9107 for RMSE, PSNR, and SSIM in the font transfer with FangSong, intended to measure the extent to which results are generated. The font transfer with KaiTi and YouYuan achieved 0.1117, 19.2039, 0.9053 and 0.2223, 13.4039, 0.8047 for RMSE, PSNR, and SSIM respectively. Compared with the other methods, the Chinese characters generated by our approach are more recognizable and authentic with clear strokes, complete structures, and obvious styles. It is found that our method can be suitable for the regular and clear Chinese characters with common fonts, while further optimization and exploration are necessary for font transfer with more various fonts.

Specific results are derived for style extractor and attention augmented convolution, as they are defined for improving the quality of generated characters. Each of these cases is chosen for experimental validation with or without attention augmented convolution and different feeding ways of extracted style features. The results indicate that one cannot accomplish the same effect on generated characters without style extractor and attention augmented convolution. For the Chinese characters not appearing in the training, the stylized network can still showcase a strong generalization ability, but the generated quality requires further improvement.

References

[1] X. Zhu, Y. Wu, and X. Ding, "Recognition of 6763 printed chinese characters," *Journal of Ching Hua University*, vol. 27, no. 1, pp. 39 – 49, 1987.

[2] X. Liu, G. Meng, J. Chang, R. Hu, S. Xiang, and C. Pan, "Decoupled representation learning for character glyph synthesis," *IEEE Transactions on Multimedia*, vol. 24, pp. 1787 – 1799, 2022.

[3] W. Liu, F. Liu, F. Ding, Q. He, and Z. Yi, "Xmp-font: Self-supervised cross-modality pre-training for few-shot font generation," in *Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR*, 2022, pp. 7895 – 7904.

[4] Y. Liu and Z. Lian, "Fonttransformer: Few-shot high-resolution chinese glyph image synthesis via stacked transformers," *Pattern Recognition*, September 2023.

[5] C. Xiao, C. Zhang, and C. Zheng, "FontCode: Embedding information in text documents using glyph perturbation," *ACM Transactions on Graphics*, vol. 37, no. 2, 2018.

[6] C. Wang, Y. Zhu, Z. Shen, D. Wang, G. Wu, and Y. Yao, "Font transfer based on parallel auto-encoder for glyph perturbation via strokes moving," in *21st International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP)*, 2021, pp. 586–602.

[7] Z. Lian, B. Zhao, and J. Xiao, "Automatic generation of large-scale handwriting fonts via style learning," in *SIGGRAPH ASIA 2016 Technical Briefs*, 2016.

[8] Z. Lian, B. Zhao, X. Chen, and J. Xiao, "EasyFont: A style learning-based system to easily build your large-scale handwriting fonts," *ACM Transactions on Graphics*, vol. 38, no. 1, 2018.

[9] X.-Y. Zhang, F. Yin, Y.-M. Zhang, C.-L. Liu, and Y. Bengio, "Drawing and recognizing chinese characters with recurrent neural network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 849–862, 2018.

[10] X. Lin, J. Li, H. Zeng, and R. Ji, "Font generation based on least squares conditional generative adversarial nets," *Multimedia Tools and Applications*, vol. 78, no. 1, p. 783–797, 2019.

[11] Y. Huang, M. He, L. Jin, and Y. Wang, "RD-GAN: Few/Zero-shot Chinese character style transfer via radical decomposition and rendering," in *European Conference on Computer Vision*, 2020, pp. 156–172.

[12] F. Xiao, B. Huang, and X. Wu, "Automatic generation of chinese handwriting via fonts style representation learning," *arXiv preprint*, 2020. [Online]. Available: https://arxiv.org/abs/2004.03339

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[14] Z.-L. Yang, S.-Y. Zhang, Y.-T. Hu, Z.-W. Hu, and Y.-F. Huang, "VAE-Stega: Linguistic steganography based on variational auto-encoder," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 880 – 895, 2021.

[15] D. Sun, T. Ren, C. Li, H. Su, and J. Zhu, "Learning to write stylized chinese characters by reading a handful of examples," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, 7 2018, pp. 920–927.

[16] Y. Zhang, Y. Zhang, and W. Cai, "Separating style and content for generalized style transfer," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR*, 2018, pp. 8447–8455.

[17] Y. Zhang, Y. Zhang, and W. Cai, "A unified framework for generalizable style transfer: style and content separation," *IEEE Transactions on Image Processing*, vol. 29, pp. 4085–4098, 2020.

[18] Chang, Bo and Zhang, Qiong, "Rewrite2: A GAN based Chinese font transfer algorithm," 2017. [Online]. Available: https://github.com/changebo/Rewrite2/

[19] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5967–5976.

[20] J. Chang and Y. Gu, "Chinese typography transfer," *arXiv preprint*, 2017. [Online]. Available: https://arxiv.org/abs/1707.04904

[21] D. Sun, Q. Zhang, and J. Yang, "Pyramid embedded generative adversarial network for automated font generation," in *24th International Conference on Pattern Recognition (ICPR)*, 2018, pp. 976–981.

[22] J. Chang, Y. Gu, Y. Zhang, and Y. Wang, "Chinese handwriting imitation with hierarchical generative adversarial network," in *British Machine Vision Conference*, 2018, pp. 1–12.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS*, Long Beach, California, USA, 2017.

[24] C. Ren, S. Lyu, H. Zhan, and Y. Lu, "SAFont: Automatic font synthesis using self-Attention mechanisms," *Australian Journal of Intelligent Information Processing Systems*, vol. 16, no. 2, pp. 19–25, 2019.

[25] S.-Y. Lu and T.-R. Hsiang, "Generating chinese typographic and handwriting fonts from a small font sample set," in *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, 2018.

[26] J. Zeng, Q. Chen, Y. Liu, M. Wang, and Y. Yao, "StrokeGAN: Reducing Mode Collapse in Chinese Font Generation via Stroke Encoding," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, pp. 3270–3277.

[27] Tian, Yuchen and chongzhe, "zi2zi: Master chinese calligraphy with conditional adversarial networks," 2019. [Online]. Available: https://github.com/kaonashi-tyc/zi2zi/

[28] S.-J. Wu, C.-Y. Yang, and J. Yung-jen Hsu, "CalliGAN: Style and Structure-aware Chinese Calligraphy Character Generator," *arXiv preprint*, 2020. [Online]. Available: https://arxiv.org/abs/2005.12500

[29] J. Chen, X. Xu, Y. Ji, and H. Chen, "Learning to create multi-stylized chinese character fonts by generative adversarial networks," in *Proceedings of the ACM Turing Celebration Conference*, Chengdu, China, 2019.

[30] J. Zhang, D. Chen, G. Han, G. Li, J. He, Z. Liu, and Z. Ruan, "SSNet: Structure-semantic net for Chinese typography generation based on image translation," *Neurocomputing*, vol. 371, pp. 15–26, 2020.

[31] Y. Gao and J. Wu, "GAN-based unpaired chinese character image translation via skeleton transformation and stroke rendering," in *34th AAAI Conference on Artificial Intelligence, AAAI*, 2020, pp. 646 – 653.

[32] P. Lyu, X. Bai, C. Yao, Z. Zhu, T. Huang, and W. Liu, "Auto-encoder guided GAN for Chinese calligraphy synthesis," in *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, vol. 01, 2017, pp. 1095–1100.

[33] Y. Jiang, Z. Lian, Y. Tang, and J. Xiao, "DCFont: An end-to-end deep Chinese font generation system," in *SIGGRAPH Asia 2017 Technical Briefs*, 2017.

[34] Z. Zheng and F. Zhang, "Coconditional autoencoding adversarial networks for chinese font feature learning," *arXiv preprint*, 2018. [Online]. Available: http://arxiv.org/abs/1812.04451

[35] S. Yang, J. Liu, W. Wang, and Z. Guo, "TET-GAN: Text effects transfer via stylization and destylization," *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 1238–1245, 2019.

[36] Y. Gao, Y. Guo, Z. Lian, Y. Tang, and J. Xiao, "Artistic glyph image synthesis via one-stage few-shot learning," *ACM Transactions on Graphics*, vol. 38, no. 6, 2019.

[37] F. Xiao, J. Zhang, B. Huang, and X. Wu, "Multiform fonts-to-fonts translation via style and content disentangled representations of chinese character," *arXiv preprint*, 2020. [Online]. Available: https://arxiv.org/abs/2004.03338

[38] Y. Xie, X. Chen, L. Sun, and Y. Lu, "DG-Font: Deformable Generative Networks for Unsupervised Font Generation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 5126–5136.

[39] S. Sun, W. Zhang, H. Fang, and N. Yu, "Automatic generation of robust chinese document watermarking fonts," *Journal of Image and Graphics*, vol. 27, no. 1, pp. 262–276, 2022.

[40] T. Shen, J. Jiang, T. Zhou, S. Pan, G. Long, and C. Zhang, "DiSAN: Directional self-attention network for RNN/CNN-free language understanding," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 01, pp. 5446–5455, Apr. 2018.

[41] I. Bello, B. Zoph, Q. Le, A. Vaswani, and J. Shlens, "Attention augmented convolutional networks," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 3285–3294.

[42] J. Cho, K. Lee, and J. Y. Choi, "Font representation learning via paired-glyph matching," in *British Machine Vision Conference (BMVC)*, London, UK, 2022.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, pp. 600–612, 2004.

[44] J. Da, "Chinese text computing, character frequency list of informative texts in modern chinese," 2004. [Online]. Available: https://lingua.mtsu.edu/chinese-computing/statistics/index.html

[45] S. Joshi and N. Khanna, "Source printer classification using printer specific local texture descriptor," *IEEE Transactions on Information Forensics and Security*, pp. 160 – 171, 2020.

[46] H. Fang, W. Zhang, H. Zhou, H. Cui, and N. Yu, "Screen-shooting resilient watermarking," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 6, pp. 1403–1418, 2019.

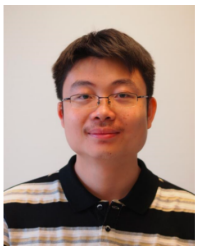
Ye Yao received the M.S. degree in computer science and the Ph.D. degree in communication and information systems from Wuhan University, Wuhan, China, in 2005 and 2008, respectively. He was a Visiting Scholar with New Jersey Institute of Technology, Newark, NJ, USA, from December 2016 to December 2017. He is currently an Associate Professor with the School of Cyberspace, Hangzhou Dianzi University, Hangzhou. His research interests include multimedia forensics and information security.

Chen Wang received the M.S. degree from Hangzhou Dianzi University, Hangzhou, Zhejiang, China, in 2022. She is currently a teacher with the School of Digital and Information Technology, Hangzhou Xiaoshan Technician College. Her current research interests include information security, video forensics, deep learning, and computer vision.

Hui Wang received the MSc degree in Security Technologies and Applications and Ph.D. degree from Computing Department, University of Surrey, Guildford, Surrey, UK in 2010 and 2014, respectively. She is currently a lecture with the School of Cyberspace, Hangzhou Dianzi University, Hangzhou. Her research interests include multimedia forensics, digital watermarking technologies and information security.

Ke Wang received the B.S. degree from Southwest Minzu University, Chengdu, Sichuan, China, in 2022. He is currently pursuing the master's degree with the School of Cyberspace, Hangzhou Dianzi University. His current research interests include multimedia forensics and information hiding.

Yizhi Ren received the Ph.D. degree in Computer software and theory from Dalian University of Technology, China in 2011. From 2008 to 2010, he was a research fellow at Kyushu University, Japan. He is currently a full professor with School of Cyberspace, Hangzhou Dianzi University, China. His research interests include data security, privacy preserving, and trust management. He has published over 70 research papers in refereed journals and conference proceedings. He served as the local chairs/PC members of more than 20 International conferences, such as, ICCCN 2018, ICC 2018, and DSC 2019.



Weizhi Meng received the Ph.D. degree in computer science from the City University of Hong Kong (CityU), Hong Kong SAR. He is currently an Associate Professor with the Department of Applied Mathematics and Computer Science, Technical University of Denmark (DTU), Denmark. Prior to joining DTU, he worked as Research Scientist with the Institute for Infocomm Research, Singapore. He won the Outstanding Academic Performance Award during his doctoral study. His primary research interests are cyber security and intelligent technology in security including intrusion detection, smartphone security, biometric authentication, HCI security, cloud security, trust management, blockchain in security, cyber-physical system security, and IoT security. He is a recipient of the Hong Kong Institution of Engineers (HKIE) Outstanding Paper Award for Young Engineers/Researchers in 2014 and 2017. He received the IEEE MGA Young Professionals Achievement Award in 2020 for his contributions to leading activities in Denmark and Region 8.