Heritage Science

RESEARCH Open Access

EA-GAN: restoration of text in ancient Chinese books based on an example attention generative adversarial network

Zheng Wenjun, Su Benpeng, Feng Ruiqi, Peng Xihua and Chen Shanxiong*

Abstract

Ancient Chinese books are of great significance to historical research and cultural inheritance. Unfortunately, many of these books have been damaged and corroded in the process of long-term transmission. The restoration by digital preservation of ancient books is a new method of conservation. Traditional character restoration methods ensure the visual consistency of character images through character features and the pixels around the damaged area. However, reconstructing characters often causes errors, especially when there is large damage in critical locations. Inspired by human's imitation writing behavior, a two-branch structure character restoration network EA-GAN (Example Attention Generative Adversarial Network) is proposed, which is based on a generative adversarial network and fuses reference examples. By referring to the features of the example character, the damaged character can be restored accurately even when the damaged area is large. EA-GAN first uses two branches to extract the features of the damaged and example characters. Then, the damaged character is restored according to neighborhood information and features of the example character in different scales during the up-sampling stage. To solve problems when the example and damaged character features are not aligned and the convolution receptive field is too small, an Example Attention block is proposed to assist in restoration. Qualitative and quantitative analysis experiments are carried out on a self-built dataset MSACCSD and real scene pictures. Compared with current inpainting networks, EA-GAN can get the correct text structure through the guidance of the additional example in the Example Attention block. The peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) value increased by 9.82% and 1.82% respectively. The learned perceptual image patch similarity (LPIPS) value calculated by Visual Geometry Group (VGG) network and AlexNet decreased by 35.04% and 16.36% respectively. Our method obtained better results than the current inpainting methods. It also has a good restoration effect in the face of untrained characters, which is helpful for the digital preservation of ancient Chinese books.

Keywords Digitization of ancient books, Text restoration, Font restoration, Ancient Chinese book text dataset

Introduction

Ancient books record the development history of human society and civilization. They help people to study historical facts and inherit culture. As the non-renewable wealth of human civilization, the conservation of ancient books is essential. With the development of digitalization, turning ancient books into data stored in disks and other carriers has become an emerging means of ancient book conservation [1]. The digitized images of ancient books both improve the efficiency of dissemination and protect the original books from unnecessary damage. However, during the long-term transmission of ancient books, many characters have been damaged and corroded.

*Correspondence:
Chen Shanxiong
csxpml@163.com
College of Computer and Information Science, Southwest University, Chongqing 400715, China

Restoration of damaged characters is a prerequisite for enabling these books to be studied and disseminated.

Chinese characters are hieroglyphics. In traditional text restoration work, researchers use contextual and perceptual information, i.e., using the pixels around the damaged character images and the characteristic elements in the complete standard characters to complete the deduction, and then use image processing tools to complete manual restoration [2]. This approach is inefficient and consumes a lot of time.

To improve the efficiency of text restoration work, researchers are eager to use image technology to complete automatically. In 2008, Zhang Wei et al. proposed a method for restoring the lettering of bamboo slips using Canny edge operator [3]. In 2014, Zhang Na et al. proposed a text restoration and recognition method based on horizontal and vertical projection [4]. Some researchers migrated the image inpainting algorithm to the text restoration field and also achieved good results. In 2020, Ge Song et al. proposed a method to restore modern Chinese handwritten characters based on self-attention and adversarial classification loss, which could restore partially occluded handwritten Chinese characters [5]. In 2021, Duan Ying et al. proposed a partial convolution-based text image irregular interference restoration algorithm for different ancient damaged text images [6]. In 2022, Chen Shanxiong et al. proposed a dual discriminator-based method for restoring Yi handwritten characters, which can effectively recover Yi characters [7]. Su Benpeng et al. constructed a multi-stage restoration network combining shape restoration network and texture restoration network, which realized the authentic restoration of ancient characters [8]. However, these methods are affected by image inpainting and mainly pursue the visual consistency of the restored damaged regions. They have little restriction on the freedom degree of the structure, and the obtained results have many structural errors. Character images are not simply equivalent to traditional images. Firstly, it requires the correct topology restoration of text strokes, rather than just visual consistency. Secondly, when critical information exists in a large damaged area, it is difficult to accomplish correct restoration by relying only on the edge information present. Meanwhile, there are many types of characters in Chinese ancient books, especially the same characters have different writing forms in different eras. The methods proposed before can only restore the character categories that exist in the training dataset by learning from the training data. In contrast, facing with character categories that do not exist in the training dataset, they can only give the most similar but structurally incorrect results based on the previously learned features. Therefore, to improve the restoration accuracy of the model, the dataset must be very adequate. However, in real scenarios, Chinese characters exist in different writing forms in different historical eras and regions, so it is very difficult to construct a dataset containing sufficient categories.

Observing the writing behavior of humans, it can be divided into two categories. One type is that knowledgeable people can write correct words based on the memory in their minds. The other is that people can imitate writing based on the given text samples, and in this way they can write previously unseen words. Inspired by this kind of human's imitation writing behavior, this paper proposes an ancient Chinese books text restoration model based on examples. This model adds an additional reference example in the restoration process to guide the recovery of missing regions of damaged Chinese text. Firstly, two down-sampling branches are used to extract the features of the damaged text and example text respectively, and the features of the two branches are added up and sent to the bottleneck layer for feature transformation. In the up-sampling process, example text features are used to guide the restoration of the damaged text. In this way, even if the damaged text lacks critical information, the text restoration can be completed based on the structural information of the reference example and the existing residual edge information of the damaged text, which improves the restoration accuracy. The previous text restoration methods can only restore the text categories existing in the training dataset by training the network to restore based on the residual edge information. In this paper, the network can simulate the human imitation behavior, and by giving suitable reference examples, it can restore the text categories not existing in the training set, which reduces the difficulty of dataset construction. At the same time, this paper proposes Example Attention block to solve two major problems in the restoration process: the damaged text features are not aligned with the reference example features, and the global features of the reference example cannot be observed due to the restricted receptive fields. The main contributions of this paper are as follows:

A) A Chinese text restoration network based on reference examples is proposed, which can efficiently and accurately restore the structure of damaged characters.
B) An Example Attention block is proposed to solve the problems of unaligned features and restricted receptive fields.
C) The text restoration effect of the model is explored. The experimental results show that the model proposed in this paper can learn an imitation writing behavior and can restore the Chinese character categories that do not appear in the training dataset.

Related work

Image inpainting methods based on deep learning

Deep learning is widely used in natural language processing [46], image classification [45], image inpainting and other fields. The current research of image inpainting mainly focuses on regular mask [9] and irregular mask [10] inpainting, denoising [11], defogging [12], old photo coloring[13], target removal [14]and other tasks. In recent years, many deep learning-based image inpainting algorithms have been proposed. Among these methods, D. Pathak et al. first proposed to use a trained deep learning network to restore missing pixels in the damaged images in 2016 [9]. This method can restore the missing regions to some extent, but it cannot keep the local consistency between the newly restored regions and the residual regions. To ensure the consistency of the image texture, in 2017, Iizuka et al. proposed to use both global and local discriminators to evaluate the adversarial loss [15]. The global discriminator evaluates whether the restored image has overall semantic consistency, while the local discriminator constrains the restoration of the damaged regions to enhance local consistency. However, this method is restricted by the convolution receptive field and can only reason and restore based on the information of the surrounding regions. These previous methods are only for regular masks, and do the same convolution for mask and non-mask regions. To better deal with irregular masks, partial convolution was proposed [10], which only performs convolution in the valid pixel region of the image and updates the mask automatically. But in this convolution mode, the mask uses hard update, i.e., the update rule is fixed, and all channels use the same mask. In 2019, Yu Jiahui et al. proposed gated convolution [16]. Unlike partial convolution, gated convolution provides a learnable dynamic feature selection mechanism for each spatial location and each channel in all layers, i.e., using soft mask. Although previous methods have been able to solve the texture inconsistency problem, it is still very difficult to restore the central area when the area of missing pixels is too large. In 2020, Li et al. designed a Recurrent Feature Reasoning network (RFR-net) [17]. The RFR module circularly reasoned the hole boundaries of the convolutional feature map, and then used the obtained results as clues for further reasoning. It strengthened the constraint on the hole center during the reasoning process, and finally obtained clear restoration results.

Compared with traditional diffusion-based [18, 19], and patch-based [20–22] image inpainting algorithms, the above algorithms can use deep networks such as GAN to learn shallow features and deep semantic features of non-damaged image regions driven by a large amount of data. Then, by using these features, the damaged regions are restored adaptively. The restored image structure and texture are consistent at the deep semantic level, and have a more realistic restoration effect [23]. Therefore, the text restoration model in this paper is also designed based on GAN network structure.

Visual attention

Inspired by the fact that the human visual processing system selectively focuses on certain parts of an image, in 2017, researchers introduced the attention mechanism into computer vision [24]. Its basic idea is to make the model automatically learn to assign different weights to each part of the input features and extract the important key information, which makes the model make more accurate judgments. Based on the research of residual networks, Residual Network, Wang et al. proposed Residual Attention Network using style attention module [25]. It can refine the network feature map to improve the accuracy and robustness of the network. To model the weight relationship among each channel in feature maps, Squeeze-and-Excitation Networks (SENet) was proposed in 2017 [26]. The SE module uses the global average pooling to calculate the channel attention, which can significantly improve the classification accuracy of the network. However, the SENet does not focus on spatial attention. Combining spatial and channel attention, Sanghyun et al. proposed Convolutional Block Attention Module (CBAM) [27]. CBAM uses two modules, channel attention and spatial attention, to obtain the corresponding attention maps, and then multiplies the attention maps with the input feature maps to achieve adaptive feature refinement. The spatial attention in CBAM calculates global maximum pooling and average pooling of feature maps in the channel direction, and then performs convolution to obtain the attention map. However, this method is limited by the receptive field of the convolution kernel and can only focus on local regions. Of course, using a fully connected layer can get global information, but it will lead to excessive parameters. To solve this problem, Wang Xiaolong et al. proposed non-local neural network [28], which obtains a position weight feature map by calculating feature weighted sum of all positions in the input feature map, and then normalizes it with the original feature map. This module can effectively learn global features without limiting the size of the input feature maps. After Transformer's brilliant performance, Alexey et al. applied standard Transformer to image tasks [29]. To solve the model input inconsistency problem, they split images into patches and used their linear embedding sequences as the input of Transformer. The deep features of images are extracted by Transformer structure, and the inductive bias of images is removed by multi-head self-attention mechanism, which has higher accuracy and better performance than the traditional Convolutional Neural Network (CNN) network.

In this paper, inspired by visual attention, the Example Attention is proposed to obtain the global features of the reference example, and fuse the example features with the damaged text features to guide the text restoration process.

Priori guided image inpainting

Existing learning-based image inpainting methods have achieved good restoration results. Some researchers believe that introducing appropriate priori knowledge guidance in the restoration stage can adapt to more scenarios. Initially, researchers introduced structural priors to guide image inpainting, including edges [30, 31], contours [32] and segmentation mapping [33, 34]. The experiments proved that the structure priori can effectively improve the quality of the inpainting results. However, when the damaged image has complex semantics or a large missing area, the inpainting tasks still face greater challenges. To obtain more priori knowledge, reference images with similar texture and structure are introduced in some inpainting tasks. The realistic and abundant texture features of the reference image are used to guide the restoration of the missing details, and more accurate and realistic inpainting results can be achieved for tasks such as image compression [35], image super resolution [36–38], etc. In addition, Li Xiaoming et al. achieved realistic inpainting of low-resolution face images by using the same subject but not identical images with higher resolution as reference images [39]. In 2022, Lu Wanglong et al. proposed EXE-GAN [40]. Under the guidance of face example images with rich texture and semantic information, EXE-GAN can effectively generate more realistic face restoration results. And Liu Taorong et al. encoded texture and structural features of input images and reference images and performing multi-scale fusion [41], which achieved more realistic detailed restoration of natural scene images.

Inspired by the methods above, this paper proposes EA-GAN based on the generative restoration model, which restores the damaged text under the guidance of example text. Through the two-branch input structure and Example Attention block, EA-GAN can use the residual feature information of the damaged text and the structure and detail features of the example text to reason the missing part, which achieve better text restoration effects.

Method

Overall network architecture design

To solve the damaged text restoration problem, this paper proposes a GAN-based restoration model EA-GAN. The whole model structure follows the GAN-based image inpainting model (shown in Fig. 1), which is composed of generator G (yellow box on the left of Fig. 1) and discriminator D (yellow box on the right of Fig. 1). Different from the traditional GAN model, the generator G of EA-GAN adopts a two-branch input structure (branch E1 and E2), which inputs both damaged text and example text images to guide the restoration. This structure learns a rule of imitation, making the restoration result details more realistic. And when a text category that does not exist in the training dataset is used as the network input, the correct

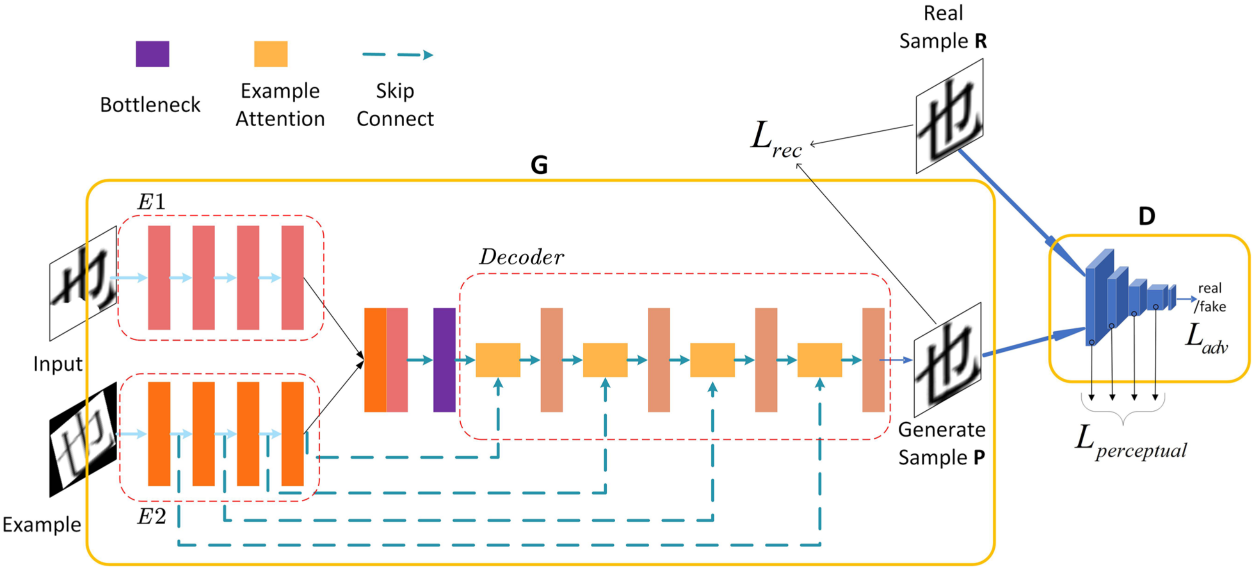


Fig. 1 Overall structure diagram of restoration network EA-GAN

restored text can still be obtained if a suitable example is given. The discriminator D of EA-GAN is used to distinguish generated samples from real samples during training. The structure of generator G and discriminator D in the network are described in detail next.

We first introduce the structure of the generator G. The generator consists of two encoder branches E1 and E2, a bottle-neck layer and a decoder. Both encoder branches are composed of four convolution layers, which are used to encode the features of the input damaged text image and example text image respectively. The encoder branch E1 encodes the damaged image into contextual features by convolution, and then sends them to the decoder through the bottle-neck layer for text restoration. The encoder branch E2 encodes the input example image into feature maps of different scales and dimensions, and then provides multi-scale information reference for damaged text restoration by skip connection layer. The bottle-neck consists of four residual blocks, which are used for feature transformation and can effectively remove high-frequency noise. The Decoder (*Decoder* in Fig. 1) consists of a deconvolution layer and an Example Attention block. The Example Attention block guides the restoration process with the help of example features transmitted by the skip connection layer. This block is mainly used to learn the contextual features in the example feature maps, avoiding the problems of insufficient receptive fields and unaligned features caused by traditional convolution.

Next, we introduce the discriminator D, which consists of five convolutional layers. It is a PatchGAN structure discriminator with Markov property. The traditional discriminator uses a fully connected layer at the last layer and outputs a probability between 0 and 1. The last layer of the PatchGAN structure discriminator outputs an N × N matrix, where each value (in the range 0 to 1) represents the true/false discriminant result for every local patch in the image. In this way, the discriminator can improve the recognition ability of local regions of the image. At the same time, the features obtained from each convolution layer in the discriminator D are also extracted to calculate the perceptual loss between the generated sample P and the real sample R.

Example attention block design

The example-based feature fusion has been applied in the field of super-resolution, mainly by aligning the key points of low-resolution images and high-resolution images, and then fusing the features by convolution. However, this is not applicable to the network in this paper. Firstly, this network uses features of different dimensions and scales, and there is no algorithm for aligning key points of such features at present. Secondly, convolution is limited by receptive fields. We consider that text restoration relies on the spatial layout of the whole word, which means that not only the similar regions in the example features need to be learned, but also the context needs to be referred. To solve this problem, an Example Attention block is designed in this paper, as shown in Fig. 2. In Fig. 2, x and y respectively represent the damaged text image feature map, and the example text feature map obtained by skip connection layer in the Decoder. The whole block takes x and y as input. First linear transformation is performed

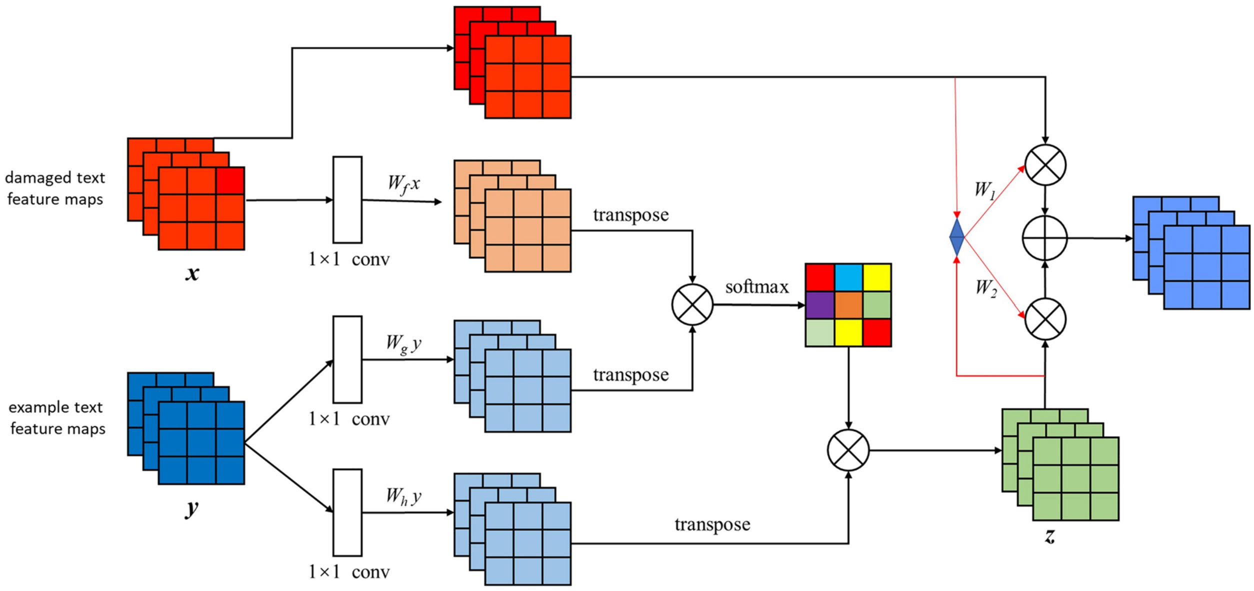


Fig. 2 Structure diagram of an example attention block

on the damaged text image feature map x and example image feature map y. Then the features of damaged text is used as query features and matrix multiplied with each pixel in the example image feature to calculate the correlation, thus obtaining a correlation feature map. The calculation formula is shown in Eq. (1):

$$f\left(x_i, y_j\right) = \left(W_f x_i\right)^T \left(W_g y_j\right) \tag{1}$$

In Eq. (1), W_f and W_g respectively represent linear transformation. x_i and y_j respectively represent the i and j column vectors of matrixes obtained from damaged text feature x and example feature y after linear transformation and transpose. By normalizing and summing the correlation feature map with the example features, the feature map z can be obtained, which is the features of damaged text feature x under the influence of sample feature y. The calculation formula is shown in Eq. (2):

$$z_i = \frac{f\left(x_i, y_j\right)}{\sum_{j=1}^{n} f\left(x_i, y_j\right)} \left(W_h y_j\right) \tag{2}$$

In Eq. (2), z_i is the i row in the feature map $z.W_h$ is the weight coefficient. The whole calculation can realize the query of every pixel in the example feature map and learn the global feature influence. Meanwhile, even if the feature map of the input example text is not aligned with the feature map of the damaged text, the information region with high correlation can be found in the example feature map. Finally, the damaged text feature x and the feature map z obtained by Eq. (2) are fused with different weights to obtain the next stage feature f. The calculation formula is shown in Eq. (3):

$$f = W_1 x + W_2 z \tag{3}$$

In Eq. (3),f is next stage feature map and W_1 and W_2 represent the weights. In the traditional feature fusion methods, different features are usually fused by using hyperparameters. However, the weight coefficient calculation in this paper refers to the spatial attention module in CBAM [27]. CBAM generates spatial attention map by modeling the spatial relationship between features, which emphasizes the attention to the more informative part and suppresses the less important feature information. The main idea of spatial attention is: first apply the average pooling and maximum pooling operations along the channel direction so that the information of each channel can be aggregated. Then concatenate them to produce efficient feature descriptors to highlight informative regions. Next, a spatial attention map is generated by using convolutional layers to encode the emphasized or suppressed parts. Finally, the spatial attention features are generated by activation function. The calculation formula is shown in Eq. (4):

$$\begin{aligned} W_i = \sigma(F_i^{7\times7}([&AvgPool([x;z]); \\ &MaxPool([x;z])])) \end{aligned} \tag{4}$$

In Eq. (4), $F_i^{7\times7}$ is the 7×7 convolution. σ is the activation function, and the one used in this paper is the Relu activation function.$AvgPool()$ is the average pooling operation, and $MaxPool()$ is the maximum pooling operation. The calculated spatial attention feature is used as the weight coefficient to multiply with the input feature, which can achieve a more effective feature fusion. We will discuss the detail calculation formulas of the weights W_1 and W_2 in the experiment and discussion section.

Loss function design

To ensure the restoration ability of the network, it is necessary to select an appropriate loss function to optimize the network. In this paper, the reconstruction loss is used to calculate the difference between the pixels of the real text (the non-damaged text) image and the predicted text (the restored text) image. The reconstruction loss is defined as in Eq. (5):

$$L_{rec} = \frac{1}{N}\left|I_{gt} - I_{pred}\right|_1 \tag{5}$$

In Eq. (5), I_{gt} is the real text image and I_{pred} is the predicted text image. N is the pixel size of the image. We use the L1 norm to calculate the sum of the absolute difference values between them. Only using reconstruction loss is not enough to measure the structural similarity of the two images. Therefore, the perceptual loss is also used in this paper. It can be used to compare the differences of the deep features between the predicted text image and the real text image in the network. Traditional perceptual loss uses pre-trained VGG networks to extract features, which adds extra computation and space. In this paper, the features are extracted by discriminator to calculate the perceptual loss, which avoids introducing additional networks. The perceptual loss of the entire network can be calculated as follows:

$$L_{perceptual} = \sum_{i=1}^{N} \frac{1}{H_i W_i C_i}\left|\phi_i^{gt} - \phi_i^{pred}\right|_1 \tag{6}$$

In Eq. (6), ϕ_i^{gt} and ϕ_i^{pred} respectively represent the feature maps of the real text image and the predicted text image after the ith convolutional layer of the discriminator. H_i, W_i, C_i are the height, width, and channel of the feature map obtained from ith layer.

Finally, the network is a GAN structure, so it is necessary to use adversarial loss to calculate the probability of true or false of the predicted image. The adversarial loss of the network is defined as follows:

$$L_{adv} = E\left(\log(I_{gt})\right) + E(\log(1 - D(G(I_{mask}, Mask)))) \tag{7}$$

In Eq. (7), G is generator and D is discriminator. $E(*)$ is the expected value of the distribution function. I_{mask} is the input damaged image. *Mask* is the damaged area matrix (0 for valid pixels, 1 for invalid pixels).

In summary, the loss function of the entire network is:

$$L_{total} = \lambda_1 L_{perceptual} + \lambda_2 L_{rec} + \lambda_3 L_{adv} \tag{8}$$

where $\lambda_1, \lambda_2, \lambda_3$ are hyperparameters.

Damaged text image simulation

The training of the model requires abundant paired data of complete text and damaged text, which is difficult to obtain in reality. The traditional method of generating damaged text images is to randomly generate 0–1 matrix mask (0 is the region of effective pixels, 1 is the missing region) to simulate the damaged situation. To improve the restoration effectiveness of the model, we use a more real mask to generate simulated damaged text images.

Specifically, we invited relevant experts of ancient books to list some typical damage scenes. Then we extracted the damage regions and digitized these damage regions into mask matrixes composed of 0 and 1. Finally we summarized 12 common types of damage mask. The images of 12 type real masks are shown in Fig. 3. By translating and scaling the matrixes to increase the diversity of damage, the damaged text image can be simulated more realistically.

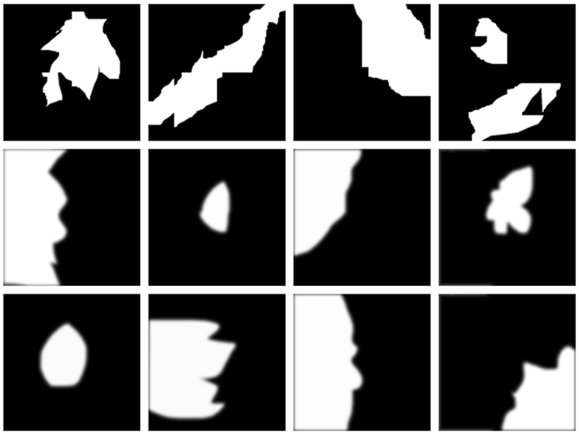


Fig. 3 The images of 12 type real masks

Experiment and discussion

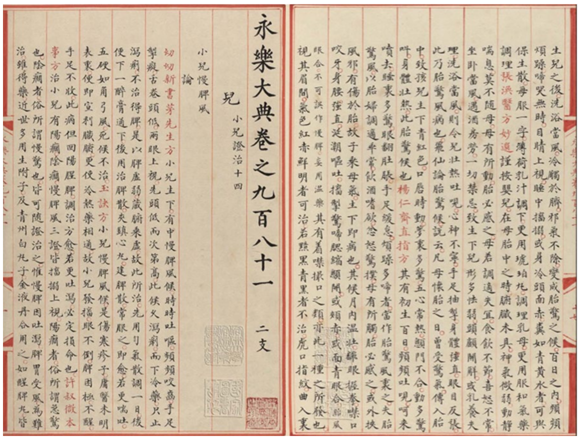
Dataset and experimental environment

In the existing public dataset, there is no suitable ancient Chinese text dataset for this experiment. Therefore, the dataset used in the experiment is the self-built dataset MSACCSD (Multi-Style Ancient Chinese Character Simulation Dataset). We obtained high quality digitized ancient books from the National Digital Library of China (http://www.nlc.cn/). By detecting and cutting the characters in ancient books *Yongle Dadian* (Fig. 4), 100 traditional characters with high frequency were selected. We did binary processing on these single-word pictures to remove the background, which facilitating the network training. To expand the type and style of the dataset, we collected the character TTF files in Regular script, Official script, Song style, Running script and other writing styles from Chinese Font Design website (https://chinesefontdesign.com). The final dataset contains 1600 commonly used simplified and traditional Chinese characters. We did data augmentation and damaged text image simulation. Then a dataset of 1600 types of 48,000 images including multiple font formats is generated. The size of the original character image in the dataset is 28 × 28. In the experiment, we scaled it to 128 × 128.

The network in this paper was trained using the Adam optimizer with a batch size of 8. The two parameters of Adam optimizer are 0.1 and 0.9. The learning rate during training is 0.0001. The size of the input image is 128 × 128. The hyperparameters λ_1, λ_2, λ_3 in the loss function are 6, 6, 1. Pytorch framework version 1.2.0 is used in all experiments.

Measure indicators

The quality of the restoration results is mainly measured by the structural differences between the restored text image and the original complete image. The human brain



生兒之後洗浴當風冷觸於臍邪氣不除變成胎驚之候百日之內頻頻
煩躁啼哭無時目睛上視睡中搐搦或身冷頭面赤囊如青黃水者可與
保生散母服一字薄荷乳汁調下更用琥珀丸調理乳母更用服和氣藥
調理張渙醫方妙選謹按嬰兒在母胎中之時臟腑未具神氣微弱動靜
喘息莫不隨母母有所動胎必感之母若調適失宜食飲不節喜怒不常
坐卧當風過酒房勞一切禁忌致生下兒形多怯弱頭顱開解或乳養失
理洗浴當風則令兒壯熱吐𠯗心神不寧手足抽掣身體強直眼目反張
此乃胎驚風病也衆仙論胎驚候說云凡母懷胎之日曾受驚氣傳入胎
中致孩兒生下青紅色口唇時動拳裹多驚五心常熱顖門不合動多驚
啼身體壯熱此胎驚候也楊仁齋直指方其有初生百日頻頻吐哯呵来
噴去睡裹多驚眼翻肚脹手足緩急煩躁多啼者當作胎驚風裹之夫胎
驚風以胎婦調適乖常飲酒嗜欲忿怒驚撲母有所觸胎必感之或外挾
風邪有傷於胎故子乘母氣生下即病也其候月內溫壯翻眼握拳噤口
咬牙身腰強直涎潮嘔吐搐掣驚啼腮縮顖開或頰赤或面青眼合胎風
眼合不可誤作慢脾妄用溫藥其有著噤撮口之類亦此一種之所發也
視其眉間氣色紅赤鮮明者可治若黯黑青黑者不治虎口指紋曲入裏

永樂大典卷之九百八十一　二支

兒　小兒證治十四

小兒慢脾風

論

幼幼新書茅先生方小兒生下有中慢脾風候時時吐嗅頻頻咬齒手足
掣瘲舌卷頭低兩眼上視先頭低而次第高此候久瀉痢而下冷藥只止
瀉痢不治得脾是以脾虛弱藏腑乘虛故此所治先用引氣散調一日後
便下一醉膏通下後用治脾散夾鎮心丸建脾散常服之即愈若更喘吐
五硬如角弓風死候不治玉訣方小兒慢脾風候是傷寒疹子膚瞖未明
表裏便即宜利臟腑更使冷熱藥相通故小兒發搐眼不倒脾困極不醒
手足不收此病但回陽醒脾調治方愈若更吐瀉必定損命也許叔微本
事方治小兒有陽癇陰癇慢脾風三證皆搐搦上視陽癇者俗所謂急驚
也陰癇者俗所謂慢驚也皆可隨證治之惟慢脾困吐瀉脾胃受風為難
治雖得藥近世多用生附子及青州白丸子金液丹合用之如醒脾丸皆

Fig. 4 The ancient book *Yongle Dadian*

can easily tell the difference between two images, but this is not easy for a computer. The measurement of image similarity is not simply obtained by distance calculation. At present, SSIM and PSNR are commonly used measure indicators. In recent years, it has been found that measuring the image similarity using deep features in neural networks is more effective and has higher accuracy than previous indicators. Then the learned perceptual image patch similarity LPIP [42] begun to apply to measure the similarity between two images. LPIPS is more consistent with human perception than traditional methods. Therefore, in addition to the basic SSIM and PSNR, LPIPS is used as the measure indicator. We use VGG [43] and AlexNet [44] respectively to extract features and calculate LPIPS value.

Comparative experiments

Qualitative and quantitative experiments

We compare the EA-GAN model with the current mainstream inpainting methods, including the globally and locally [15], PConv [10], GatedConv [16], EdgeConnect [30] and RFR-net [17]. Through qualitative analysis and quantitative analysis, it is proved that the method proposed in this paper has better restoration effects.

Qualitative analysis: Fig. 5 shows the comparison of experimental results between the proposed method and other mainstream methods in dataset MSACCSD. The columns in Fig. 5 from left to right show the damaged text (Input), the real complete text (GT), and then the restoration results of the globally and locally (GLCIC), Gated Conv (GC), PConv (PC), EdgeConnect (EC) and RFR-net (RFR) models in order. The last column is the restoration result of this paper (Ours).

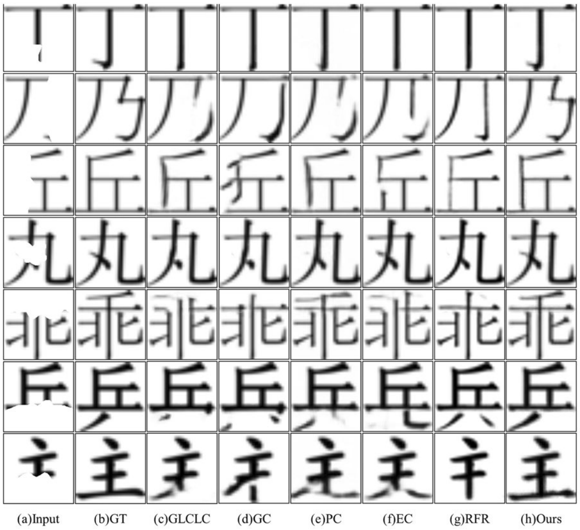


Fig. 5 Qualitative comparison results of different methods

The results show that the globally and locally, PConv, Gated Conv and EdgeConnect ensure visual consistency. However, they only infer a possible structure from the remaining information in the image, and the result is most likely wrong. Globally and locally uses both global and local context discriminators, compared with the other three methods, the text restoration effect will be improved. With the help of RFR module, RFR-net is able to improve the correctness of structure restoration to some extent by performing circular reasoning on large missing regions. But in most cases, RFR-net only relies on the information of the damaged text itself and cannot reason out the correct text structure. In contrast, the EA-GAN model proposed in this paper uses the Example Attention block to provide additional example guidance when restoring, so that the network can infer the correct text structure. Figure 5 also shows that the restoration results of the proposed method is basically consistent with the real text.

Quantitative analysis: This paper also compares the restoration effects of the proposed method and the current mainstream restoration methods from the quantitative analysis. The experimental results are measured by PSNR, SSIM, LPIPS (AlexNet) and LPIPS (VGG). LPIPS (AlexNet) and LPIPS (VGG) are LPIPS value calculated by extracting features using AlexNet [44] and VGG [43] networks respectively. The experimental results are shown in Table 1. The "↑" indicates that the larger of the number, the better of the effect. "↓" indicates that the lower of the number, the better of the effect. The best results are marked in bold. Compared with other methods, PSNR and SSIM increase by 9.82% and 1.82%, respectively, and LPIPS calculated by VGG and AlexNet decrease by 35.04% and 16.36%, respectively. In general, the proposed method is superior to other methods in every indicator.

Table 1 Quantitative comparison results of different methods on dataset MSACCSD ("↑" indicates that higher is better. "↓" indicates that lower is better. The best results are marked in bold black)

Method	PSNR↑	SSIM↑	LPIPS(AlexNet)↓	LPIPS(VGG)↓
GLCIC	23.15	0.905	0.0355	0.0589
PConv	24.19	0.919	0.0316	0.0567
GatedConv	22.98	0.907	0.0418	0.0617
EdgeConnect	25.45	0.924	0.0313	0.0617
RFR-net	25.24	0.933	0.0254	0.0428
Ours	**27.95**	**0.950**	**0.0165**	**0.0358**

Effect of damaged area on restoration effect

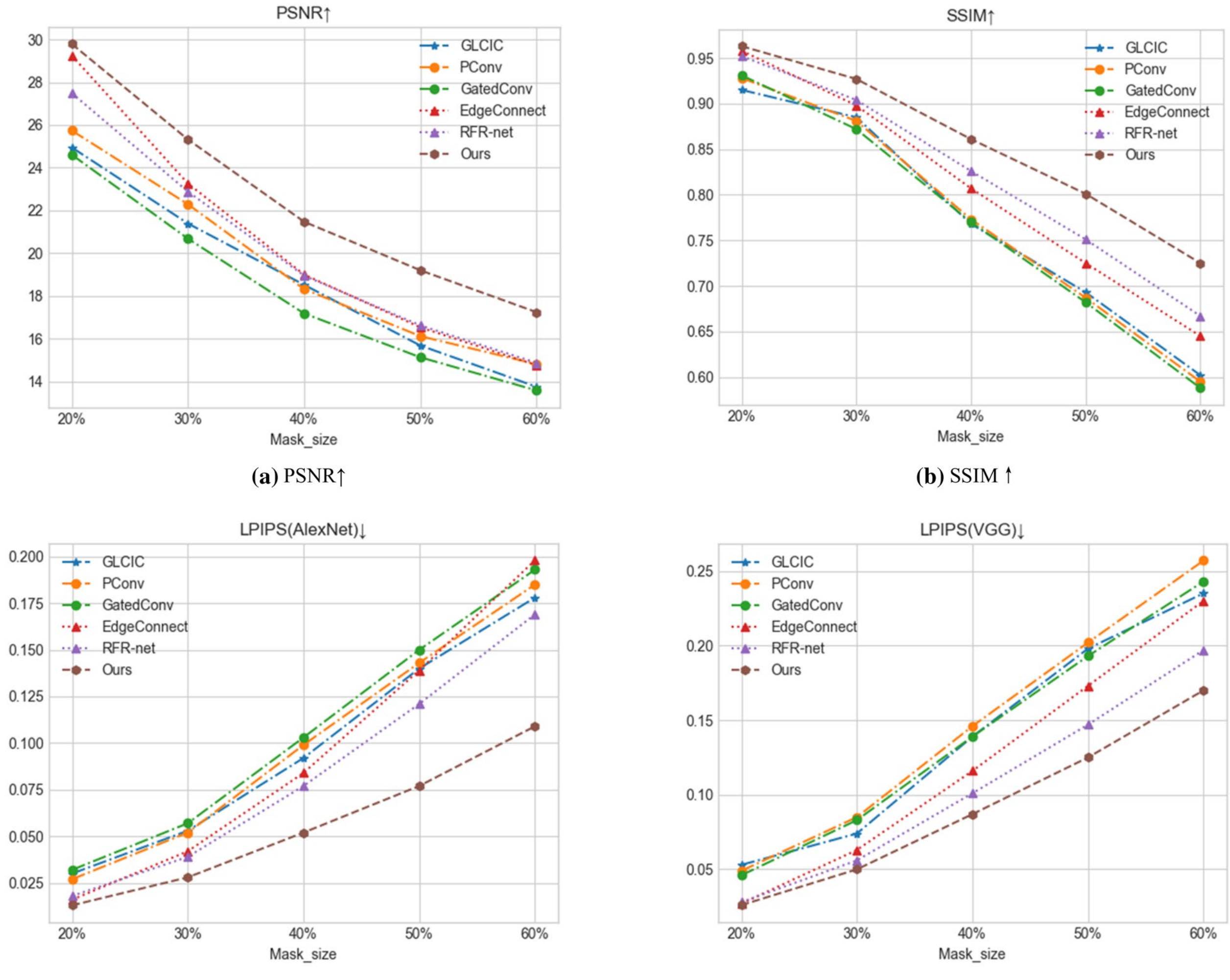
In this paper, the relationship between different damage area and restoration effect is studied. We use mask_size as a measure indicator of damage area. The experimental results are shown in Fig. 6 ("↑" means the larger of the number, the better of the effect; "↓" means the smaller of the number, the better of the effect). It can be observed that when the mask_size is small, the experimental results of our method is not much different from other methods. However, when the mask_size increases to 50%–60%, our method is significantly ahead of other methods in every indicator. It also fully proves that the proposed method is more effective in restoring large area damaged text.

Model robustness analysis

Restore the inexistent text in the training dataset

We assume that EA-GAN model can learn imitation writing rules after training, while other restoration methods mainly restore through inference rules learned from training. When faced with the training texts, they can restore the correct structure, but when faced with the inexistent text in the training dataset, they cannot reason out the right result. Based on this assumption, 250 inexistent text images are selected for the experiment. The experimental results are shown in Fig. 7 and Table 2.

Figure 7 shows the qualitative comparison results of different methods for restoring the inexistent texts in the training dataset. Each column in Fig. 7 from left to right is the damaged text (Input), the real complete text (GT), the globally and locally (GLCIC), Gated Conv (GC),



(a) PSNR↑

(b) SSIM ↑

(c) LPIPS(AlexNet) ↓

(d) LPIPS(VGG) ↓

Fig. 6 Restoration effect comparison of each method under different damage area ("↑" indicates that higher is better. "↓" indicates that lower is better.)

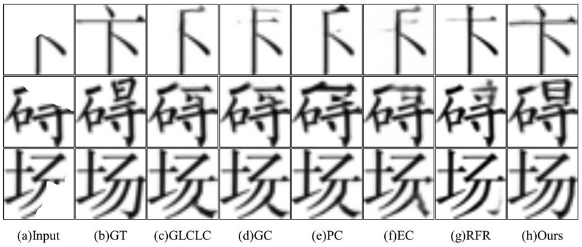


Fig. 7 **Q**ualitative **c**omparison results of different methods for restoring the inexistent texts in the training dataset

Table 2 Quantitative comparison results of different methods for restoring inexistent texts in the training dataset ("↑" indicates that higher is better. "↓" indicates that lower is better. The best results are marked in bold black)

Method	PSNR↑	SSIM↑	LPIPS(AlexNet)↓	LPIPS(VGG)↓
GLCIC	21.93	0.905	0.0458	0.0736
PConv	22.22	0.900	0.0502	0.0785
GatedConv	22.19	0.901	0.0482	0.0742
EdgeConnect	23.34	0.917	0.0441	0.0674
RFR-net	22.99	0.911	0.0467	0.0659
Ours	**24.66**	**0.925**	**0.0267**	**0.0559**

PConv (PC), EdgeConnect (EC), RFR-net (RFR) and the proposed method (Ours). The results show that the proposed method can restore the correct text structure of inexistent text in the training dataset, which cannot be done by other methods. And this also cites the correctness of the assumption from the side.

Table 2 shows the quantitative comparison results of different restoration methods for restoring inexistent texts in the training dataset, and the best results are marked in bold black. "↑" means the larger of the number, the better of the effect; "↓" means the smaller of the number, the better of the effect. The best results are marked in bold black. The results show that the proposed method is ahead of other methods in each indicator. From both quantitative analysis and qualitative analysis, the proposed method shows better restoration ability than other mainstream methods.

Effect of different examples on restoration results

To verify the effect of different examples on the restoration results, two Chinese characters are selected in this experiment. The experimental results are shown in Fig. 8. The simulated damaged images Input A/B are input into the E1 branch of the EA-GAN model (Fig. 1), and the corresponding high-resolution handwritten text Example A1/B1, the scaled handwritten text Example A2/B2 and the printed text Example A3/B3 are respectively input into the E2 branch of the EA-GAN model (Fig. 1) as examples. The corresponding output results Output A1/B1, Output A2/B2 and Output A3/B3 are obtained. Compared with the real image (GT A/B), it can be observed that the difference in pixel values of the input examples (Example A1/B1 and Example A2/B2) hardly affects the restoration results. The difference in the structural style of input examples (Example A2/B2 and Example A3/B3) greatly affects the restoration results.

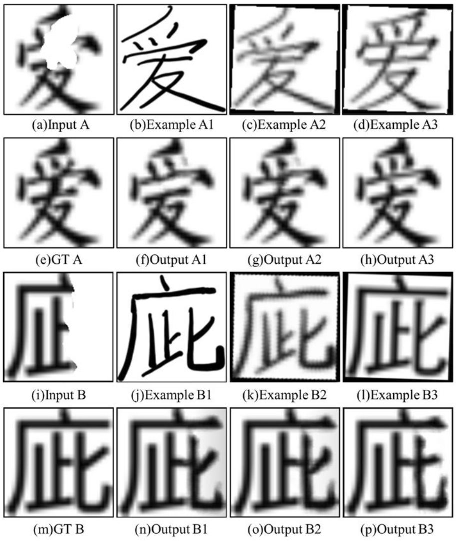


Fig. 8 Text restoration results of EA-GAN with different examples

Selection of weight calculation methods for Example Attention block

This experiment explores the effect of different weight calculation methods of feature fusion on the results. The overall feature fusion method in Example Attention block is shown in Eq. (3), and the final weight calculation method is shown in Eq. (4). To explore whether the spatial attention method will achieve better fusion effect, this experiment firstly compares it with the common hyperparameter method (M1 method). In addition, for the weight calculation method based on spatial attention mechanism, the original method extracts the attention from one input feature map, while this experiment generalizes it to attention weighting of two input feature map. Therefore, it is also necessary to discuss the feature input method of the spatial attention calculation method.

Finally, four different weight calculation methods are explored, which are denoted as M1, M2, M3 and M4. The M1 method is to set the weights as two updatable hyperparameters. The M2 method is to first calculate the global max pooling and average pooling on the feature

Table 3 Effect of different weight calculation method selection on experimental results ("↑" indicates that higher is better. "↓" indicates that lower is better. The best results are marked in bold black)

Method	PSNR↑	SSIM↑	LPIPS(AlexNet)↓	LPIPS(VGG)↓
M1	27.79	0.944	0.0174	0.0390
M2	27.81	0.948	0.0168	0.0365
M3	27.75	0.948	0.0169	0.0363
M4(Ours)	**27.95**	**0.950**	**0.0165**	**0.0358**

maps x and z respectively, and then concatenate the four results and use convolutions to get weight map. The M3 method is to separately calculate the global max pooling and average pooling of the two feature maps, and then concatenate the respective results and use convolution to obtain the respective weight maps. The M4 method is to concatenates x and z, then calculate the global max pooling and average pooling, and finally use convolutions to get weight map. The specific calculation methods of M1, M2, M3 and M4 are shown in Eqs. (9)–(12):

$$M1 : W_1 = \alpha, // W_2 = \beta \tag{9}$$

$$\begin{aligned} M2 : W_i =& \sigma(Fi^{7\times7}([AvgPool(x); MaxPool(x);\\ & AvgPool(z); MaxPool(z)])) \end{aligned} \tag{10}$$

$$\begin{aligned} M3 : W_1 &= \sigma(F_1^{7\times7}([AvgPool(x); MaxPool(x)]))\\ W_2 &= \sigma(F_2^{7\times7}([AvgPool(z); MaxPool(z)])) \end{aligned} \tag{11}$$

$$M4 : W_i = \sigma(F_i^{7\times7}([AvgPool([x; z]); MaxPool([x; z])])) \tag{12}$$

Table 3 shows the quantitative restoration results of EA-GAN model under four weight calculation methods, and the best results are marked in bold black. The results show that compared with the traditional hyperparameter fusion method, the use of spatial attention method for feature fusion can improve the restoration effect. And the restoration results are optimal when using the combination of the two feature maps as the input of the spatial attention calculation (M4). The combination of feature map x and feature map z enables the spatial attention to pay attention to the most valuable part of the damaged text and the most informative part of the fusion example features at the same time, and ignore the less important parts of the two, which can better guide the restoration process.

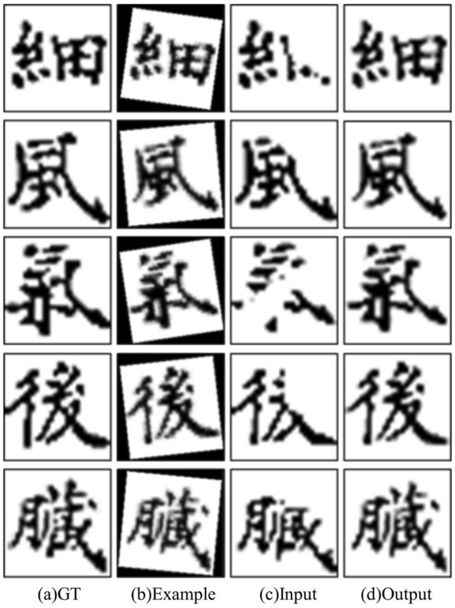


Fig. 9 The traditional Chinese character restoration results of *Yongle Dadian*

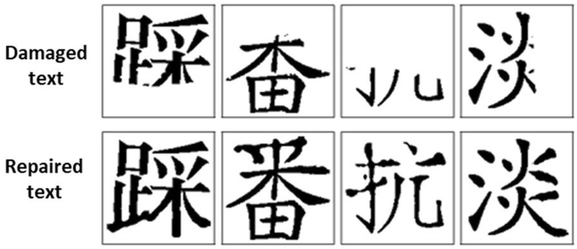


Fig. 10 Restoration results of damaged texts in torn papers

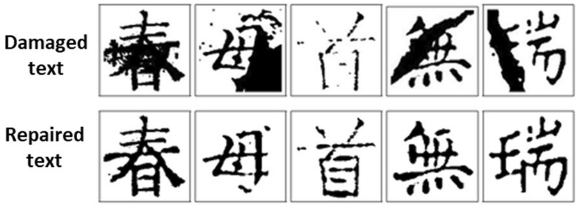


Fig. 11 Restoration results of damaged texts in ancient books and stone tablets

Restoration results for real scenes

We list the results of the traditional character restoration of the ancient book *Yongle Dadian*. The example character is generated by the scaling and rotating of the original character. The restoration results are shown in Fig. 9. The results show that the network can basically realize the restoration of the traditional Chinese character. Some strokes may have minor errors.

To further demonstrate the effectiveness of the proposed method, restoration experiments are conducted on high-resolution damaged text images in real scenes. Figures 10, 11 show the restoration results. The first row of the two figures is the damaged text image in the real scene, and the second row is the restored text image. The damaged texts in Fig. 10 come from torn papers in a real scene. The damaged texts in Fig. 11 come from ancient books and stone tablets, and the background and texture of the text have been removed. It can be observed that the restored texts basically guarantee the correctness of the structure. This also proves the effectiveness of the proposed method.

Conclusion

To improve the accuracy of text structure restoration, this paper proposes an example-based ancient Chinese book text restoration model EA-GAN. Based on the traditional generative adversarial network, the model introduces additional examples in the way of a two-branch generator to guide the restoration process and improve the accuracy of restoration. To solve the problems that the features may not be aligned and the convolution receptive field is limited when the damaged character features are fused with the example features, this paper proposes the Example Attention block. It can learn the global features from the reference example, and fully fuse the example text features and the damaged text features by the spatial Attention mechanism. Qualitative and quantitative comparison experiments are carried out on the dataset MSACCSD. The experimental results show that the proposed method has better restoration capability than the current mainstream inpainting methods, and still has better restoration effect on large-area damaged texts. It also has better robustness and generalization ability when faced with the untrained texts and real scene texts. However, we only focus on the structure during restoration, but do not consider the rich texture of ancient books. The restoration of texture is also the work we need to complete in the future.

Abbreviations

GAN	Generative Adversarial Network
VGG	Visual geometry group network
ReLU	Rectified Linear Unit
PSNR	Peak signal-to-noise ratio
EA	Example Attention
RFR-net	Recurrent Feature Reasoning Network
SENet	Squeeze and Excitation Networks
CBAM	Convolutional Block Attention Module
CNN	Convolutional Neural Network
SSIM	Structural similarity
LPIPS	Learned perceptual image patch similarity

Acknowledgements

This work was supported by Chongqing Language Research Project (yyk21223), research program of Chongqing Graduate education and teaching reform (yjg223027) and Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD-K202200203).

Author contributions

ZWJ designed the study, conducted the experiments and discussions, and mainly wrote the article; SBP participated in the construction of the model and part of the experiment; FRQ assisted in the query and sorting of the literature; PXH helped in the proofreading of the article; CSX provided overall guidance and supervision of the study and proposed an optimized protocol. All authors read and approved the final manuscript.

Funding

None.

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 4 October 2022 Accepted: 15 February 2023
Published online: 01 March 2023

References

1. Jian Z. Font processing standard for digitization of ancient books from the perspective of font database. China Publishing. 2021;22:55–9.
2. Jiajia Q. Research on the restoration and protection of painting and calligraphy cultural relics based on modern digital technology. Cult Relics Ident Apprec. 2019;01:106–7.
3. Wei Z, Xuben W, Ping J. Application of canny edge operator insimplified text repair. Microcomputer Inform. 2008;24(9):241–242250.
4. Na Z, Lujun C, Xuben W. Archaeological text restoration recognition method based on horizontal and vertical projection. Sci Technol Bull. 2014;30(06):185–7.
5. Song G, Li J, Wang Z. Occluded offline handwritten Chinese character inpainting via generative adversarial network andself-attention mechanism. Neurocomputing. 2020;415:146–56.
6. Ying D, Hua L, Yuquan Q, Qingzhi D. Research on irregular interference restoration algorithm for text image based on partial convolution. Computer Eng Sci. 2014;43(09):1634–44.
7. Shanxiong C, Shiyu Z, Hailing X, Fujia Z, Dingwang W, Yun L. A double discriminator gan restoration method for ancient yi characters. Acta Automatica Sinica. 2014;48(03):853–64.
8. Benpeng S, Xuxing L, Weize G, Ye Y, Shanxiong C. Restoration of ancient chinese characters using dual generative adversarial networks. Visual Informatics. 2022;6(1):26–34.
9. Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA. Context encoders: Feature learning by inpainting. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.2536–2544. 2016.
10. Liu G, Reda FA, Shih KJ, Wang T-C, Tao A, Catanzaro B. Image inpainting for irregular holes using partial convolutions. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 85–100. 2018.

11. Du W, Chen H, Yang H. Learning invariant representation for unsupervised image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp.14483–14492. 2020
12. Ning X, Li W, Liu W. A fast single image haze removal method based on human retina property. IEICE Trans Inf Syst. 2017;100(1):211–4.
13. Wan Z, Zhang B, Chen D, Zhang P, Chen D, Liao J, Wen F. Bringing old photos back to life. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2747–2757. 2020
14. Yi Z, Tang Q, Azizi S, Jang D, Xu Z. Contextual residual aggregation for ultra high-resolution image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7508–7517. 2020.
15. Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. ACM Trans Graphics (ToG). 2017;36(4):1–14.
16. Yu J, Lin Z, Yang J, Shen X, Lu X, Huang TS. Free-form image inpainting with gated convolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, p. 4471–4480; 2019.
17. Li J, Wang N, Zhang L, Du B, Tao D. Recurrent feature reasoning for image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p. 7760–7768. 2020.
18. Bertalmio M, Sapiro G, Caselles V, Ballester C. Image inpainting. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, p. 417–424. 2000.
19. Levin A, Zomet A, Weiss Y. Learning how to inpaint from global image statistics. In: ICCV, vol. 1, p. 305–312. 2003.
20. Kwatra V, Essa I, Bobick A, Kwatra N. Texture optimization forexample-based synthesis. In: ACM SIGGRAPH 2005 Papers, pp.795–802. 2005.
21. Barnes C, Shechtman E, Finkelstein A, Goldman DB. Patchmatch: A randomized correspondence algorithm for structural image editing. ACM Trans Graph. 2009;28(3):24.
22. Zhao H, Guo H, Jin X, Shen J, Mao X, Liu J. Parallel and efficient approximate nearest patch matching for image editing applications. Neurocomputing. 2018;305:39–50.
23. Qin Z, Zeng Q, Zong Y, Xu F. Image inpainting based on deeplearning: A review. Displays. 2021;69: 102028.
24. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. Adv Inneural Inform Process Syst. 2017;30:78.
25. Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X. Residual attention network for image classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3156–3164. 2017.
26. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. 2018.
27. Woo S, Park J, Lee J-Y, Kweon IS. Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 3–19. 2018.
28. Wang X, Girshick R, Gupta A, He K. Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision andPattern Recognition, pp. 7794–7803. 2018.
29. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghan, M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929. 2020.
30. Nazeri K, Ng E, Joseph T, Qureshi FZ, Ebrahimi M. Edgeconnect: Generative image inpainting with adversarial edge learning. arXiv preprint arXiv:1901.00212. 2019.
31. Li J, He F, Zhang L, Du B, Tao D. Progressive reconstruction ofvisual structure for image inpainting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019. p. 5962–5971.
32. Xiong W, Yu J, Lin Z, Yang J, Lu X, Barnes C, Luo J. Foreground-aware image inpainting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019; p.5840–5848.
33. Liao L, Xiao J, Wang Z, Lin C-W, Satoh S. Guidance and evaluation: Semantic-aware image inpainting for mixed scenes. In: European Conference on Computer Vision, Springer. p. 683–700; 2020.
34. Liao L, Xiao J, Wang Z, Lin C-W, Satoh S. Image inpainting guided by coherence priors of semantics and textures. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6539–6548; 2021.
35. Ho MM, Zhou J, He G. Rr-dncnn v2. 0: enhanced restoration-reconstruction deep neural network for down-sampling-based video coding. IEEE Transactions on Image Processing 30, 1702–1715; 2021.
36. Dogan B, Gu S, Timofte R. Exemplar guided face image super-resolution without facial landmarks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, p. 1814–1823. 2019.
37. Zhang Z, Wang Z, Lin Z, Qi H. Image super-resolution by neural texture transfer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p. 7982–7991; 2019.
38. Lu L, Li W, Tao X, Lu J, Jia J. Masa-sr: Matching acceleration and spatial adaptation for reference-based image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6368–6377; 2021.
39. Li X, Liu M, Ye Y, Zuo W, Lin L, Yang R. Learning warped guidance for blind face restoration. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 272–289; 2018.
40. Wanglong L, Hanli Z, Xianta J, Xiaogang J, Yongliang Y, Min W, Jiankai L, Kaijie S. Do inpainting yourself: Generative facial inpainting guided. arXiv preprint arXiv:2202.06358. 2022.
41. Liu T, Liao L, Wang Z, Satoh S. Reference-guided texture and structure inference for image inpainting. arXiv preprintarXiv:2207.14498. 2022.
42. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. Theunreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595. 2018.
43. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556; 2014.
44. Krizhevsky A, Sutskever I, Hinton GE. Image-net classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.
45. Yoo I, Kim H. Created era estimation of old Korean documents via deep neural network. Herit Sci. 2022;10:144. https://doi.org/10.1186/s40494-022-00772-9.
46. Lee H, Kwon H. Going deeper with contextual CNN for hyperspectral image classification. IEEE Trans Image Process. 2017;26(10):4843–55.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.