

# Research on Denoising Method of Chinese Ancient Character Image Based on Chinese Character Writing Standard Model

MIAO Yalin

Xi'an University of Technology

LI LIANG (✉ [1423627963@qq.com](mailto:1423627963@qq.com))

Xi'an University of Technology

JI Yichun

Xi'an University of Technology

LI Guodong

Xi'an University of Technology

---

## Research Article

**Keywords:** Ancient character, Chinese character writing standard model, Generative Adversarial, Ancient document simulation noise

**Posted Date:** February 4th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1272799/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Research on Denoising Method of Chinese Ancient Character Image Based on Chinese Character Writing Standard Model

MIAO Yalin<sup>①</sup> LIANG Li<sup>①\*</sup> JI Yichun<sup>①</sup> LI Guodong<sup>①</sup>

<sup>①</sup>(*School of Printing, Packaging and Digital Media, Xi'an University of Technology, Xi'an 710048, China*)

<sup>\*</sup>(*Correspondence author: LIANG Li (1423627963@qq.com)*)

**Abstract** Ancient documents are historical evidence of cultural inheritance, and the damage brought by natural and human factors to ancient documents is inevitable, resulting in the collected images of ancient Chinese characters containing a large amount of noise, which seriously affects the accuracy of subsequent image recognition and thus creates a great obstacle to the digitization of ancient documents. To address the complexity of ancient text structure, this paper proposes a Chinese ancient text image denoising method based on the Chinese character writing standard model. The method firstly adds four additional local branches based on the global branching, and uses the supplementary character detail information to weaken the phenomenon of strokes adhering to noise due to the lack of local details; secondly, it introduces the simulation noise of ancient documents to simulate the real ancient character image morphology, which can be used for the adversarial training of this method. In the training process, the minimum absolute value deviation, smoothing loss, structural consistency loss and the refined loss function formed by the adversarial loss are used to iteratively optimize the parameters. Finally, experiments prove that the model in this paper can increase the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of the image by at least 23.8% and 11.4%, and the user evaluation index (UV) has also reached more than 80%.

**Key words** :Ancient character; Chinese character writing standard model; Generative Adversarial; Ancient document simulation noise

## 0 Introduction

This subsection focuses on the necessity of targeting ancient texts for denoising. The ancient documents are the historical evidence of cultural heritage. Due to natural and human factors, different carriers of ancient documents have been damaged to different degrees, and the collected images contain a lot of noise. The progress and development of science and technology have brought convenience to society. With the power of technology, the noise in ancient Chinese character images can be effectively removed to help restore the original appearance of ancient Chinese texts, and at the same time, the integrity of the ancient character structure is preserved and sufficient preparation is made for the subsequent image processing work. In this paper, we will mainly focus on the problem of noise removal from ancient Chinese character images.

In the traditional sense, researchers have proposed a filter for image denoising. The premise is to know the specific types of noise in the image. The common types of noise are Gaussian noise, Poisson noise, speckle noise and salt and pepper noise. In 2016, amitha et al<sup>[1]</sup> proposed that adaptive Gaussian filter and median filter should be combined to apply to the inscription images containing various noise types. Although important phase information is retained, the effect of denoising was seriously affected by the large damage of some inscription images, which resulted in missing text strokes in the final denoising results. In the same year, Shi et al<sup>[2]</sup> analyzed the structure and characteristics of different types of inscription images and selected suitable filters to process them in order to highlight the structural part of Chinese characters, but the method is not universally applicable.

Deep learning has been widely used in image denoising due to its powerful feature learning capability<sup>[3]</sup>. In 2017,

Jiao et al<sup>[4]</sup> proposed to apply Deep Convolutional Neural Network (DCNN) to image denoising for feature extraction and learning. Experimental results showed that the network can effectively denoise images in high noise environments and has a greater advantage over traditional filter-based methods. In 2014, Dr. Goodfellow et al<sup>[5]</sup> introduced Generative Adversarial Network (GAN) into the field of deep learning, inspired by the idea of two-player zero-sum game in which the sum of the benefits of both players is constant and the loss of one player is equal to the gain of the other, so that it GAN<sup>[6]</sup> uses adversarial learning, where the discriminator learns the difference between the generative distribution and the true distribution, and then drives the generator to reduce the difference. In contrast to other generative networks that do not impose explicit restrictions on the distribution of the data, thus avoiding the need to manually design the network distribution. In 2020, Zhang et al<sup>[7]</sup> improved the generative adversarial network by combining convolutional neural networks to perform page-level denoising of calligraphy works, minimizing the artifacts introduced during denoising while preserving good Chinese character structure.

Since the beginning of our exposure to Chinese characters, we have also been exposed to the "Tianzig" model shown in Figure 1, which enables us to get a better sense of the radicals and overall structure of each character, and to standardize the steps of writing Chinese characters. In this paper, we propose a method for denoising images of ancient Chinese characters based on the model of Chinese character writing standard, drawing on "Tianzig". The main contributions are: (1) the noise in Chinese ancient character images is naturally formed, and no satisfactory original images have been preserved historically. To address the problem of the lack of datasets of real images of Chinese ancient character, this paper introduces an ancient document simulation noise to simulate the real ancient character image morphology and form a new paired dataset with the original target images for adversarial training of the network model; (2) For the complexity of ancient character structure, combined with the Chinese character writing specification model (Tianzig), this paper proposes a Chinese ancient character denoising model based on the Chinese character writing specification model. The model firstly introduces a channel attention mechanism to focus on the key regions of the image during feature extraction, secondly generates global and local Chinese character structure images after denoising, and finally fuses the two by a refinement processor to complete the complement of Chinese character detail information and weaken the phenomenon of strokes sticking to the noise due to the lack of local details.

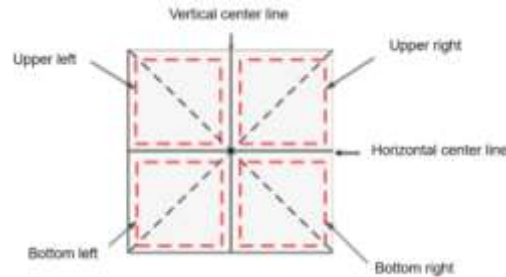


Fig.1 Tianzig

## 1 Related Work

This subsection introduces the related gan network to lead to the basic network cgan of this paper. The main advantage of GAN is that it goes beyond the traditional neural network classification and feature extraction to generate new data according to the characteristics of real data. As the two networks progress in the adversarial process and continue the adversarial process afterwards, the data obtained by the generative network becomes more and more perfect and approximates the real data, so that the desired data (pictures, sequences, videos, etc.) can be generated. The generator generates synthetic data from a given noise (generally a uniform or normal distribution). It tries to produce data closer to the real one. The most serious problem of neural networks is that they do not have the ability to explain their reasoning process and the basis of reasoning; they cannot ask the necessary queries to the user, and they cannot work when the data is not sufficient; they turn all the characteristics of the problem into numbers and all the reasoning into numerical calculations, which inevitably results in the loss of information; the theory and learning algorithms

need to be further improved and enhanced. So here we have chosen to generate an adversarial network and thus used the convolutional layer of the generator.

### 1.1 Generating adversarial networks

The network structure of the generative adversarial network is shown in Figure 2. Two game players are generator (G) and discriminative (D). Both of them make the data distribution of generator as close to the real data distribution as possible through continuous "game", thus "cheating" discriminator. Generator G is mainly used to capture the distribution of real data samples, and to obtain the generated image by receiving a random noise sampled from prior distribution. Discriminator D is used to distinguish the input data and output the probability that the input image is a real image. If the result is 1, it means that it is a real image, otherwise it means a faked image.

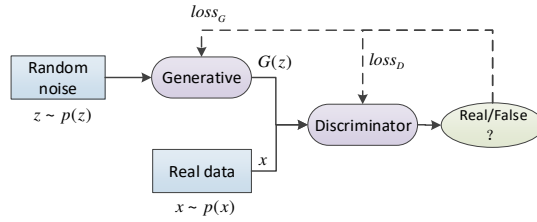


Fig.2 Generative Adversarial Network

In the training process, the purpose of judging network  $D$  is to judge true and false as much as possible. When a true image  $x$  is given, network  $D$  is expected to give a high score, while for the image  $G(z)$  generated by generator  $G$ , discriminator  $D$  is expected to give a low score as far as possible, that is, the calculated  $D(G(z))$  score is as small as possible. Therefore, the generation of the objective function of the adversarial network is the maximization and minimization process of the objective function, as shown in Equation (1) :

$$\begin{aligned} \min_{G} \max_{D} V(D, G) = & E_{x \sim P_{data}(x)} [\log D(x)] \\ & + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \end{aligned} \quad (1)$$

Where:  $P_{data}$  represents the distribution of real data;  $P_z$  is noise distribution;  $G(z)$  represents the false data generated by the generator model based on random noise  $z$ , which is close to the real data;  $D(x)$  represents the probability that the real data judged by the discriminant model is true;  $D(G(z))$  represents the probability that the model judges the fake data to be true.

### 1.2 Conditionally Generated Adversarial Network

GAN becomes uncontrollable in the face of larger images and complex data because the intermediate simulation is too free, so Mirza et al<sup>[8]</sup> proposed Conditional Generative Adversarial Nets (CGAN) in 2014, which introduces in the generator network and discriminator network Conditional Variable (CV) was introduced into the generator network and discriminator network, so that the GAN is trained using images with corresponding labels to make the generated data and the real data match better, and the network model is shown in Figure 3.

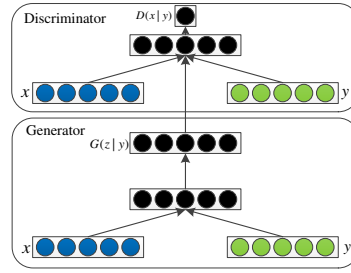


Fig.3 Conditional Generative Adversarial Nets

As can be seen from Figure 2-4, the network of CGAN does not change compared with the original GAN network, only the input data of generator  $G$  and discriminator  $D$  are changed. Therefore, CGAN can be embedded into other GAN networks as a general strategy. The loss function obtained is shown in Equation (2) :

$$\begin{aligned} \min_G \max_D V(D, G) = & E_{x \sim P_{\text{data}}(x)} [\log D(x | y)] \\ & + E_{z \sim P_z(z)} [\log(1 - D(G(z | y)))] \end{aligned} \quad (2)$$

In general, GAN can be applied in many fields because of its flexible learning ability. However, for GAN, it is often difficult to control the training and cannot generate the desired results. Conditional generative adversarial network CGAN in order to change this situation, so it is necessary to restrict the output and input of generative adversarial network, input the conditional label of control content into the generator and discriminator, or add other conditional information to restrict the input to the generator. Therefore, the generator network of the proposed denoising model in this paper will borrow the idea of CGAN and use a symmetric convolutional neural network structure to apply to the denoising of ancient Chinese character images.

## 2 Chinese Ancient Character image denoising method based on Chinese character writing standard model

This subsection focuses on the construction of the correlation model, the fusion of local and global branches, and the loss function.

### 2.1 Model Framework

In order to obtain high quality image denoising effect and reduce the noise residue near the stroke structure of ancient characters. In this paper, from the perspective of preserving the integrity of ancient script structure, the Chinese ancient character images are denoised by the Chinese character writing standard model. and uses the obtained local information to supplement the global information in detail. The overall framework is shown in figure 4, and the specific implementation process is as follows:

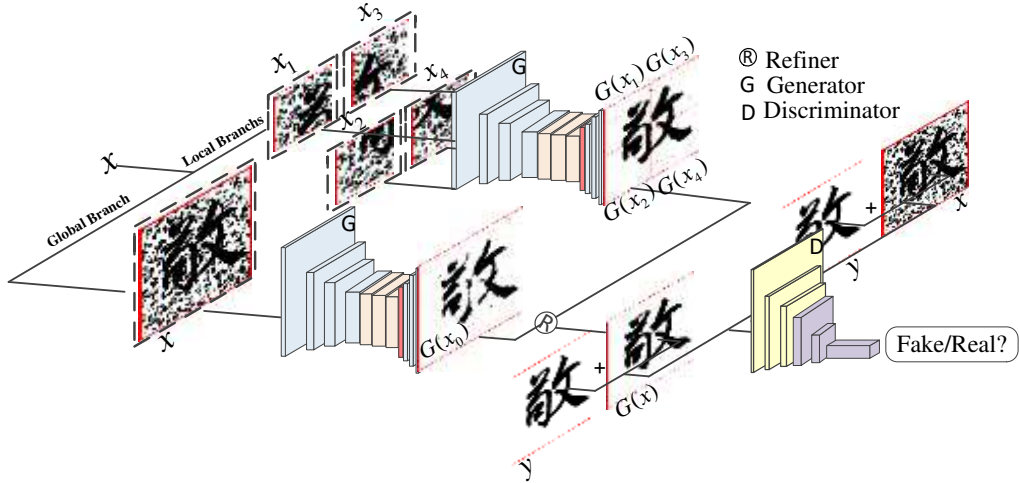


Fig.4 The overall structure of the Chinese character writing standard model

(1) The input of the model adds four additional local branches to the global branches, and the overall image of the Chinese ancient character is input to generator G to capture the most significant global noise and remove it, while the Chinese ancient character image is cropped in equal parts according to the midpoint of the width and height, and then input to generator G in a progressive manner to remove the noise near the local Chinese character stroke structure, where each generator G shares the same architecture, focusing on the focal region by adding a channel attention mechanism.

(2) The generated image  $(G(x_1), G(x_2), G(x_3), G(x_4))$  of the four local branches is stitched into an image according to their respective positions, and then the stitched image is merged with the generated image  $(G(x_0))$  of the global branch to generate the final denoising image  $G(x)$ . Finally,  $\{G(x), y\}, \{x, y\}$  are input to the discriminator D to determine the authenticity of the data composed of the source image and the target image respectively.

## 2.2 Generator Network

The generator network and discriminant network of CGAN can learn the representation hierarchy of objects from part to scene, add constraint conditions to the model with the added label information, guide the data generation process, and apply the learned features to new tasks to prove their applicability as general image representations. The generator network in this chapter draws on the idea of CGAN and adopts the symmetric convolutional neural network as shown in Figure 5 to learn end-to-end mapping directly from the input image and its corresponding real image, so as to improve the visual effect of the denoising image.

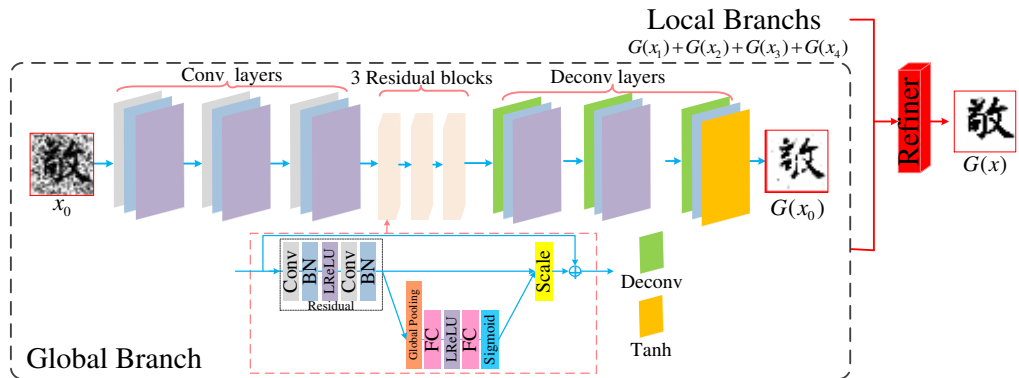


Fig.5 Generator network for Chinese character writing standard model

The overall structure of the generator network is divided into three main parts: the downsampling layer, the attention residual module, and the upsampling layer. In the downsampling part of the network, no pooling layer is used, and three convolutional layers are used to extract shallow features of the image. Each convolutional layer is followed by a batch normalization layer BN and an activation function layer LReLU to increase the nonlinearity of the network and thus accelerate the convergence. The core migration layer is constructed using residual blocks that introduce a channel focus mechanism to help the network focus on key regions and enhance the information transfer between layers so as to extract deeper hidden information. The final upsampling layer uses deconvolution, the first two groups both use the LReLU activation function, and the final output layer uses the Tanh activation function. Among them, the local branching can better preserve the local Chinese character structure details, but ignore the complete structure of the whole Chinese ancient character. In this paper, we introduce the refinement processor (Refiner) proposed in the literature [9], which is applied to face facial expression conversion, and firstly, the outputs of the four local branches are pixel-filled and stitched into one image according to their respective positions in the source Chinese ancient character image. Then, the stitched image and the output of the global branch are provided to the convolutional layer, which refines and fuses the two so that the local information it collects complements the global information in detail, and the final generated image is obtained after processing.

In convolutional neural network, the convolutional kernel is essentially the aggregation of spatial information and characteristic dimension information, which can help the network capture the global and local features of images. At present, most convolutional neural networks improve their performance through spatial dimension. This article uses the re-calibration approach, on the basis of the residual block introduced features channel attention module, make its focus on characteristics of the channel, the interdependent relationship between explicit modeling, and realign the leading feature of the adaptive response, learning the weight of each channel, according to the weight to the judgment of the next, as shown in figure 6.

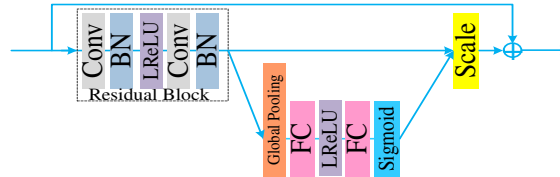


Fig.6 Residual Block of Fusion Channel Attention

### 2.3 Discriminator Network

The purpose of image denoising is to make the result of denoising comparable to real images in terms of visual perception and image quality. Therefore, this paper introduces a self-learning discriminator network, and gradually strengthens the ability to distinguish the authenticity of images through the iterative optimization of parameters in the network. The network structure of the discriminator is shown in Figure 7. In the whole discriminator network, the pooling layer is replaced by a convolutional layer with step size to avoid the loss of data information caused by the pooling layer and to enhance the description of the detailed features of Chinese ancient characters. Four convolutional layers with Switchable Normalization (SN) and Leaky ReLU activation functions are used to compute and learn the features, and the last layer is compressed using the Sigmoid function to obtain a regularized probability score between [0,1], which alleviates the gradient disappearance phenomenon and increases the model's stability.

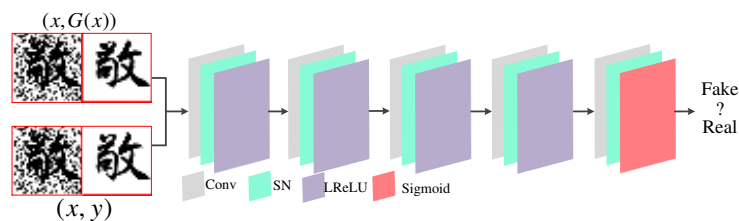


Fig.7 Discriminator network for Chinese character writing standard model

## 2.4 Refinement Processor

Local branches can better preserve local character structure details, but ignore the complete structure of the whole character. In this paper, we introduce a refinement processor (Refiner) proposed in the literature<sup>[17]</sup> for application to face facial expression transformation, where the outputs of four local branches are first pixel-filled and stitched into one image according to their respective positions in the source character image. Then, the stitched image and the output of the global branch are provided to the convolutional layer to refine and fuse the two, so that the local information it collects details and complements the global information, and the final generated image is obtained after processing, and the overall structure of Refiner is shown in Figure 8.

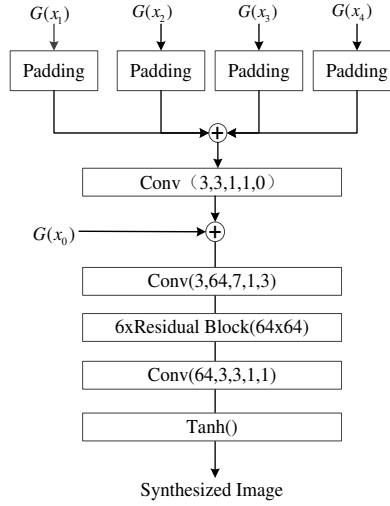


Fig.8 Refiner

## 2.5 Loss function

The loss function is an important indicator for the smooth training of the discriminator supervising generator. Traditional loss functions often use cross entropy as a pixel-level loss function, and update the generator network by obtaining the minimum value of the Mean Square Error(MSE). Although the image loss generated by MSE optimization has a high peak signal-to-noise ratio, the texture is often too smooth due to the lack of high-frequency information. In order to enable the network to generate clearer images of ancient chinese characters, this paper uses the hierarchical structure adversarial loss as the discriminative loss function. The minimum absolute error, structural consistency loss and smoothing loss are combined with appropriate weights to form a new refined loss function as the generator loss function. The definition of the refined loss function formed is shown in formula (3):

$$\begin{aligned} L &= L(D) + L(G) \\ &= L_{adv} + \lambda_1 L_1 + \lambda_S L_S + \lambda_{SSIM} L_{SSIM} \end{aligned} \quad (3)$$

Where:  $L_{adv}$ ,  $L_1$ ,  $L_S$ ,  $L_{SSIM}$  respectively represent the adversarial loss, minimum absolute error, smoothing loss, and structural consistency loss;  $\lambda_1$ ,  $\lambda_S$ ,  $\lambda_{SSIM}$  respectively represent the minimum absolute error weight, smoothing loss weight, and structural consistency loss weight.

The discriminator is used to discriminate whether the generated image is a real image. The adversarial loss function in this paper adopts a hierarchical discriminator, draws on the improved WGAN-GP loss function of the WGAN network<sup>[10]</sup>, and proposes to use a gradient penalty mechanism instead of weight clipping. The constraint function is gradient penalized by the differentiable function 1-Lipschitz, and the limitation of Lipschitz can be reflected by setting additional loss items. The input of the discriminator in this paper includes the source image  $x$  and target image  $y$  of ancient Chinese characters. The target image is used as label information  $y$  to guide the



sample generation of the generator. Its adversarial loss function is shown in equation (4):

$$L(D) = L_{\text{adv}} = E_{x \sim P_g} [D(x, G(x))] - E_{x \sim P_r} [D(x, y)] + \lambda E_{x \sim P_x} [(\|\nabla_{\tilde{x}} D(\tilde{x}, y)\|_2 - 1)^2] \quad (4)$$

Where,  $\lambda$  is set to 10,  $\tilde{x}$  and  $x_g$  the random difference value sampling on the and connection line,

$x_r \sim P_r, x_g \sim P_g, \varepsilon \sim \text{Uniform}[0,1]$ , is shown in formula (5):

$$x = \varepsilon x_r + (1 - \varepsilon) x_g \quad (5)$$

WGAN-GP GAN further stabilize the training process to ensure the quality of the resulting image. The loss function can be used to visually show the quality and convergence of the network training. The smaller the value of the discriminator's objective function, the more stable the network training and the better the generator optimization.

### 3 Data Set

This subsection focuses on creating the dataset for training and testing the model. In this paper, the dataset is obtained by self-constructing high-quality ancient character images in existing image database. Then image processing is performed and ancient document simulation noise is added according to the noise model, and the generated images containing noise and the target images form a paired dataset for the training of the network model.

#### 3.1 Ancient Documents Simulation Noise

In the process of model training, it needs to include the data pair of {source image, target image}, and the noise in the image of Chinese ancient characters in ancient documents is naturally formed, and no satisfactory original images has been preserved in history. This paper introduces ancient documents simulation noise<sup>[11]</sup> to simulate the noise naturally formed in Chinese ancient characters in ancient documents as shown in Figure 9. As can be seen from the figure, the shape of the real characters of the image noise block most of them are different, to closely mimic thereof herein by dots. For the original image, noise is randomly generated in a uniformly distributed manner, and then the final output noisy image and the original image are formed into a paired data set of {source image, target image} for network training.

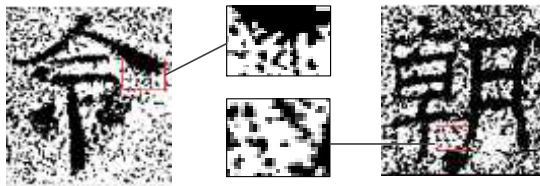


Fig.9 Real Chinese Ancient Characters Image

The specific method of adding noise is shown in formula (6):

$$\begin{cases} R_0 \sim U[1, 3] \\ (x_0, y_0) \in \{(x, y) \mid x \sim U[0, c], y \sim U[0, r]\}^n \\ n = 150, 300, 450, 600, 750, 900 \end{cases} \quad (6)$$

Where:  $(x, y)$  represents the coordinates of any pixel in the image,  $r, c$  represents the number of rows and columns in the image. In the image area, Draw a circle with a radius of  $R_0$  with  $(x_0, y_0)$  as the center. Each pixel in the circle has a probability of  $P$  may be a noise point. Here, set  $P$  to 0.8.  $R_0, x, y$  obeys a uniform random

distribution. In addition, in order to increase the diversity of noise, the noise in the rectangular area is added on the basis of the noise in the circular area. The specific expression is shown in (7):

$$\begin{cases} R_1 \sim U[1, 6] \\ (x_1, y_1) \in \{(x, y) | x \sim U[0, c], y \sim U[0, r]\}^n \\ n = 150, 300, 450, 600, 750, 900 \end{cases} \quad (7)$$

Where:  $R_1$  represents the boundary length of the rectangle as the center  $(x_1, y_1)$ ;  $P$  represents the probability that each pixel in the rectangle may be a noise point, set to 0.8;  $R_1, x, y$  obeys a uniform random distribution. As shown in Figure 10, the result of adding different intensities of noise to the original noise-free image.

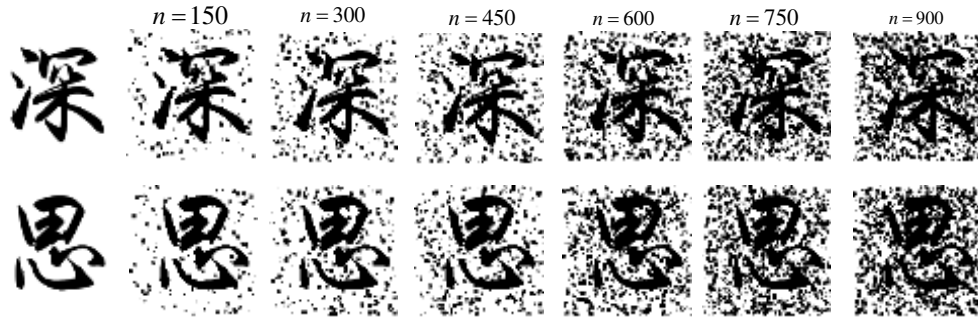


Fig.10 Simulation noise of different intensity of ancient document simulation

### 3.2 Experimental Data Set

This paper uses data generation technology to build a data set for the training of the image denoising model of Chinese ancient characters. The overall data set is shown in Figure 11.

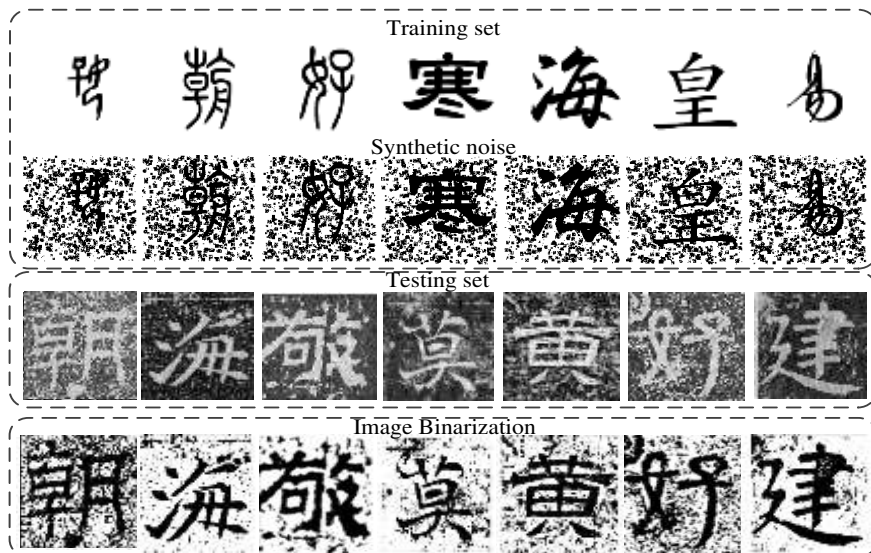


Fig.11 Data Set

The data set covers seven fonts (oracle bone inscriptions, bronze inscriptions, seal script, official script, running script, regular script, and cursive script) of the evolution of Chinese characters<sup>[12]</sup>. Each font contains 100 Chinese characters. The first line is the original data set. Introduce ancient documents simulation noise for each Chinese character image in the data set (the noise intensity is 600 in the figure) to simulate the real Chinese ancient character image of ancient documents, as shown in the second row, form a total of 700 pairs (1400) of training data sets with the original data set. The real ancient character images are shown in the third row, which are binarized and combined

with some of the datasets with added ancient documents simulation noise in a 1:1 ratio to form the final test set with a total number of 100.

### 3.3 Model Training

The training is to predict the denoised image from the noisy image, and the discriminator is to distinguish the match between the denoised image and the real image by outputting a probability value. When the generator generated image de-noising enough to deceive the discriminator, this indicates the end of training. Finally, the trained network is formally applied to the denoising processing of real Chinese ancient character images, as shown in Figure 12 for the training process of the model in this paper.

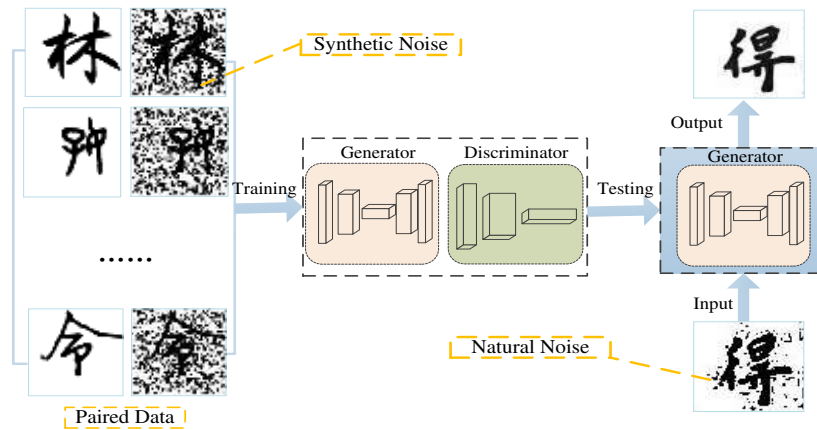


Fig.12 Training of Network Mode

As can be seen from the figure, first initialize the parameters of the two networks of generator G and discriminator D. Secondly,  $n$  samples are generated for the samples in the training set using the defined ancient literature simulation noise to form a paired data set; finally, the weights of the loss function are set to  $\lambda_1 = 10$ ,  $\lambda_S = 1$ ,  $\lambda_{SSIM} = 1$ , the parameters are updated using the Adam optimization algorithm, and the first-order exponential decay rate  $\beta_1$  and the second-order exponential decay rate  $\beta_2$  of the optimizer are set to 0.5 and 0.9, respectively. The learning rate is set to 0.0001, the number of batch samples is set to  $m=1$ , the number of iterations is set to  $T=200$ , and the entire training sample set is traversed according to the number of iterations.

## 4 Experimental Results and Discussion

This subsection focuses on the comparative experimental validation of different methods, different noise types, and different loss functions.

### 4.1 Experimental Environment

The model in this article is based on the open source deep learning library TensorFlow. Due to the large number of matrix operations in convolution operations, this article uses Graphics Processing Unit (GPU) instead of Central Processing Unit (CPU) to improve computing efficiency. The specific experimental environment parameters are shown in Table 1:

Tab.1 Experimental Environment Parameters

Name	Parameter	Name	Parameter
Cpu	Intel Core i7-7700	Running memory	32GB
GPU Graphics Card	Nvidia GeForce GTX 1080	disk space	4TB
Python version	3.7.7	CUDA version	10.0
Tensorflow version	1.14.0	Threads	Eight threads
Operating System	Ubuntu 20.04 LTS	CuDNN version	7.4

### 4.2 Evaluation Indicators

The evaluation indicator in this paper uses a combination of subjective and objective evaluation indicators to measure the effectiveness of the denoising model, where the subjective visual evaluation indicator uses user research evaluation (UV), and the objective evaluation indicator uses peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

#### (1) User survey and evaluation (UV)

User survey evaluation (User Version, UV) is a quantitative representation of visual evaluation, which intuitively reflects the user's subjective evaluation of the quality of the generated denoising image. This indicator evaluates and counts the denoising results generated by a certain number of users and different denoising models, and selects Chinese ancient character images with good denoising effects and visual aesthetics from the set number of generated results. Its quantified expression is shown in formula (8):

$$UV = \frac{n}{N \times M} \quad (8)$$

In the formula:  $n$  indicates the number of images rated as the best effect;  $N$  indicates that  $N$  user is selected for evaluation ( $N = 10$ );  $M$  indicates that the test results of each model are randomly

selected  $M = 20$ .

### (2) Peak Signal-to-noise Ration

Peak Signal to Noise Ratio (PSNR) is used to measure the difference between the corresponding pixels of the reconstructed image and the real image, and is one of the commonly used objective evaluation indicators for images. In ancient character image denoising, the larger the PSNR value, the more consistent the denoised image is with the target image, indicating that the denoising effect is better and the distortion effect is smaller. The measurement unit of PSNR is dB, and its calculation process is shown in formula (9):

$$PSNR = 10 \log \left( \frac{L^2}{MSE} \right) \quad (9)$$

In formula (9), MSE is the mean square error (Mean-Square Error), which is used to calculate the mean value of the square of the difference between the pixels at the same position in the real image and the denoised image. The specific calculation formula (10) is as follows:

$$MSE = \frac{\sum_{x=1}^M \sum_{y=1}^N [X(x, y) - \hat{X}(x, y)]^2}{MN} \quad (10)$$

Where:  $X$  represents the reference image;  $\hat{X}$  represents the image to be evaluated.

### (3) Structural similarity

Structural Similarity Index (SSIM) is one of the objective evaluation indicators used in this paper, which is mainly used to measure the degree of similarity between two images. It usually ranges from 0 to 1, and the result is 1 when the two images are the same, and the mathematical expression is shown in formula (11).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (11)$$

## 4.3 Experimental Results and Discussion

### (1) The influence of different intensities of noise on the denoising effect

Different simulation noise intensity will affect the denoising results of real Chinese ancient characters images. The network model is trained by adding noises of different intensities, and the test is performed on real images of Chinese ancient characters in ancient documents. The result is shown in Figure 13. The symbol  $n$  represents the noise intensity, and the red frame represents the details of the result image after denoising is enlarged. With the increase of noise intensity ( $n=150\sim 900$ , spacing 150), the effect of denoising also changes. In the initial stage, with the increase of noise intensity ( $n=150\sim 600$ ), the denoising effect is continuously enhanced, which can effectively reduce the stroke holes produced in the denoising process of Chinese ancient character images. After  $n > 600$ , the denoising effect tended to be stable or even decreased with the increase of noise intensity. It can be seen from Table 2 that when  $n=600$  is comparable to the SSIM value of  $n=750$ , but the PSNR value and the user evaluation index UV value are slightly higher., so this paper determines to add noise with an intensity of 600 to construct the paired data set.

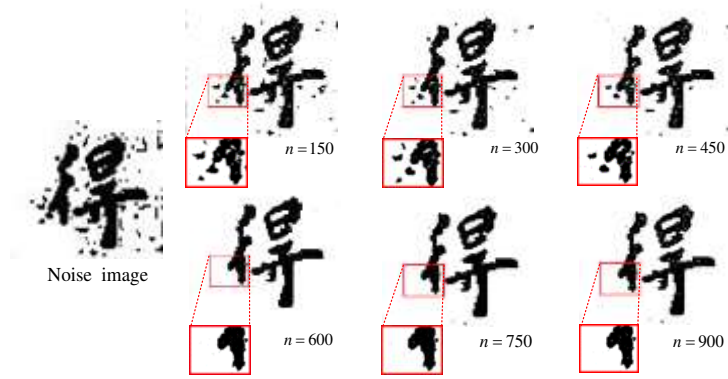


Fig.13 Denoising Results of Artificial Noise With Different Intensities After Model Training

Tab. 2 Index Evaluation Results of Different Intensities of Noise

Evaluation index	$n = 150$	$n = 300$	$n = 450$	$n = 600$	$n = 750$	$n = 900$
PSNR(dB)	16.0127	16.8549	17.8094	19.4697	19.3374	19.2137
SSIM	0.7053	0.6609	0.8283	0.8567	0.8635	0.8498
UV	46%	51%	70%	82%	78%	73%

## (2)The influence of different types of noise on the denoising effect

In order to verify the usability of the simulated noise of ancient documents introduced in this paper in the study of ancient character image denoising. On the basis of the previous experiment, Gaussian Noise with a variance of 0.001 and Salt and Pepper Noise with a noise value of  $prob = 0.12$  were added to the original data set for noise processing. A paired data set is formed for the training and testing of the model in this paper, and the test results are shown in Figure 14.

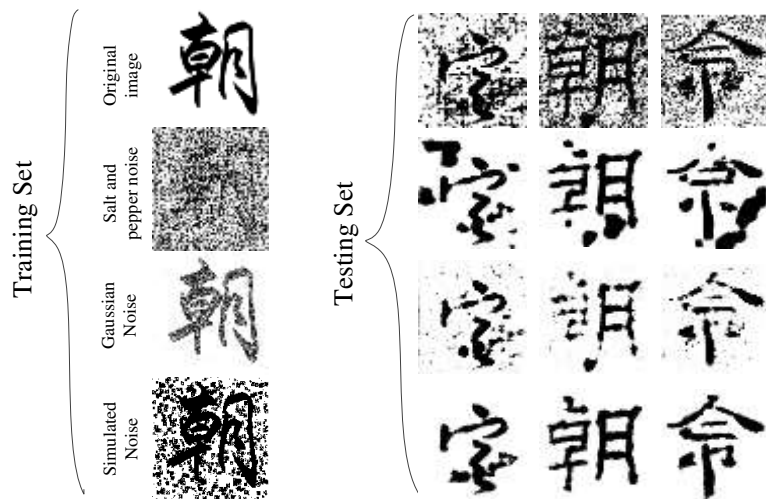


Fig.14 Denoising Effect of Different Noises after Model Training

As shown in the figure, the first column represents the processing results of adding different noises to the training set, and the second column and later represents the denoising results after adding different types of noises to the training network. It can be seen from the figure that the denoising effect of adding Gaussian noise makes the whole Chinese characters still have a lot of noise residues. At the same time, due to excessive denoising, there is a lack of strokes in the whole Chinese ancient character strokes, which runs counter to the idea of ensuring the integrity of the Chinese character structure in

ancient books; adding salt and pepper noise for network training, although the Chinese character structure is better preserved, but it makes the result figure A blocking noise that didn't exist before was generated. It can be seen from Table 3 that adding Simulation noise for network training denoising can increase PSNR and SSIM by at least 1.8529dB and 0.0730. The overall denoising effect is better, and the structure of Chinese characters is clearer by visual observation.

Tab. 3 Index Evaluation Results of Different Types of Noise

Noise type	PSNR(dB)	SSIM	UV
Salt and Pepper Noise	16.7842	0.7830	59%
Gaussian Noise	17.6168	0.6906	66%
Synthetic Noise	19.4697	0.8567	78%

### (3)Comparative experimental analysis of different models

The model in this paper is compared with other four image denoising models (BM3D<sup>[13]</sup>, DNCNN<sup>[14]</sup>, CycleGAN<sup>[15]</sup>, ID-CGAN<sup>[16]</sup>) through experiments. The experimental results are shown in Figure 15.



Fig.15 Denoising Effect of Different Methods

It can be seen from Figure 14 that the overall image after denoising by the BM3D model looks blurry. The analysis may be due to the smooth details generated in the process of joint filtering and inverse transformation resulting in the loss of information;DNCNN model can retain the outline structure of Chinese characters and noise well, but it will cause the overall loss of the internal structure of Chinese character strokes.The overall denoising effect of the CycleGAN model is significantly improved compared to the above two methods, but it removes the noise near the Chinese character structure and causes the loss of the local stroke structure; the ID-CGAN model is used for denoising, and the noise is still not completely eliminated but retained Some details of the structure of Chinese characters. The existence of image noise greatly affects visual perception and evaluation indicators. The model in this paper can effectively remove the noise that adheres to the structure of Chinese characters while retaining the complete outline and detailed structure of Chinese characters. It can be seen from Table 4 that compared with other models, the two evaluation indexes of PSNR and SSIM have increased by at least 4.2208dB and 0.0945 for the model in this paper, and the user evaluation

indexes have reached more than 80%.

Tab. 4 Evaluation results of metrics for different models of denoising

Model name	PSNR	SSIM	UV
BM3D	11.2175	0.5487	52%
DNCNN	9.4680	0.3605	36%
CycleGAN	16.8095	0.8282	69%
ID-CGAN	17.7276	0.8163	72%
Ours	21.9484	0.9227	86%

#### (4) Experimental analysis of loss function ablation

In order to ensure that the generated Chinese character denoised image can have a good visual effect and quality score, the model in this paper adds the minimum absolute error, smooth loss function  $L_S$  and structural consistency loss function  $L_{SSIM}$  based on the discriminator's anti-loss function to form a refined loss function. In this paper, the denoising effect of Chinese character image is investigated when any one of thinning loss functions is removed. The denoising results and index evaluation results are shown in Figure 16 and Table 5.

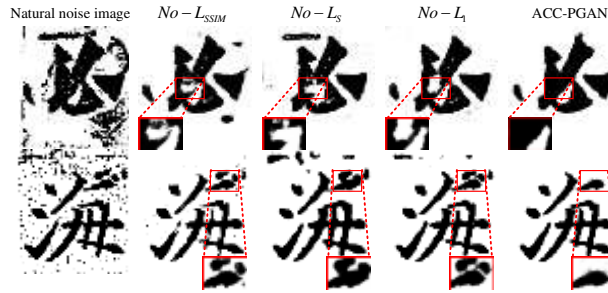


Fig.16 Ablation Experiment of Generator Loss Function

Tab. 5 Generation of Network Loss Function Ablation Experiment Evaluation Index

Loss function	PSNR	SSIM	UV
$No-L_{SSIM}$	16.8095	0.8282	57%
$No-L_S$	14.9921	0.7786	49%
$No-L_1$	17.0036	0.8362	62%
Ours	21.9484	0.9227	86%

The first column in Figure 16 is the real image of Chinese ancient characters in ancient books, After that, each column represents the denoising results when using different loss functions, and the red frame represents the enlarged details of the Chinese character image after denoising. It can be seen from the above that the lack of structural consistency loss function will lead to the fracture of the stroke structure of Chinese characters; the lack of smooth loss function will cause the residual noise of the overall Chinese character structure to stick to the strokes. Compared with the denoising results of the model in this paper, It is proved that the addition of smoothing loss function can ensure good image quality and structure characteristics.; the lack of  $L_1$  loss function can completely retain the overall



outline of the Chinese character, but it is easy to cause the loss of high-frequency information; the experimental results show that the composite thinning loss function used in this article is in three indicators. The evaluation results all reached the best, which can effectively eliminate noise while retaining the integrity of the structure of Chinese characters.

## 5. Conclusion

In this paper, we propose a method to denoise ancient Chinese characters based on the Chinese writing specification model, taking into account the complexity of the structure of ancient Chinese characters, the influence of the structural integrity of Chinese characters on the recognition of Chinese characters, and the "Tianzig" and generative adversarial networks. Firstly, on top of the global branches, four additional local branches are added according to the Chinese character writing standard model, a symmetric convolutional neural network is used as the generator network, and a feature channel attention mechanism is introduced in the feature extraction layer to focus on the key regions. Secondly, adaptive normalization is used in the discriminator network to prevent the gradient from disappearing due to the sudden change of the discriminator network during the training process, thereby improving the stability of the entire model; Use the simulated ancient literature to simulate noise again, and add noise to the data set to form a new data set, which is used for the model's adversarial training. Through subjective and objective evaluation indicators, comparative experiments are carried out on different noise intensities, different noise types, different methods, and different loss functions. The noise removal effect of this model is examined from different perspectives, and it is demonstrated that this model can remove the noise adhering to the stroke structure while weakening the holes existing in the strokes after the noise removal to ensure the integrity of the Chinese character structure.

## Funds

This work was supported by the Research on Calligraphy Culture Inheritance Technology of Ancient Inscription Based on Artificial Intelligence [grant number 62076200], National Natural Science Foundation of China.

This work was supported by the Application Research of Font Generation Technology Based on Artificial Intelligence [grant number 2020JM-468], Shaanxi Natural Science Foundation.

## Data Availability Statement

**- The datasets generated and/or analysed during the current study are not publicly available due [REASON WHY DATA ARE NOT PUBLIC] but are available from the corresponding author on reasonable request.**

Due to the raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study.

## References

[1] Amitha D, Raghu D. Robust Phase-Based Features Extracted From Image By A Binarization

- Technique[J]. *Journal of Computer Engineering*, 2016,18(4):10-14.
- [2] Shi Z, Xu B, Zheng X, et al. An integrated method for ancient Chinese tablet images denoising based on assemble of multiple image smoothing filters[J]. *Multimedia Tools & Applications*, 2016,75(19):12245-12261.
- [3] Miao Y, Ji Y, Peng E. Application of CNN-BiGRU Model in Chinese Short Text Sentiment Analysis[C], *Proceedings of the International Conference on Algorithms, Computing and Artificial Intelligence*. Sanya,China,2019: 510-514.
- [4] Jiao J, Tu W C, He S, et al. Formresnet: Formatted residual learning for image restoration[C], *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. Hawaii, USA,2017: 38-46.
- [5] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C], *Advances in neural information processing systems*. Montreal, Canada, 2014: 2672-2680.
- [6] Chen F, Zhu F, Wu Q. Xiao, et al. A review of research on generative adversarial networks and their applications in image generation[J]. *Journal of Computer Science*,2021,44(02):347-369.
- [7] Zhang J, Guo M, Fan J. A novel generative adversarial net for calligraphic tablet images denoising[J]. *Multimedia Tools and Applications*, 2020, 79(1/2):119-140.
- [8] Mirza M, Osindero S. Conditional Generative Adversarial Nets[J]. *Computer Science*, 2014:2672-2680.
- [9] Wu R, Zhang G, Lu S, et al. Cascade ef-gan: Progressive facial expression editing with local focuses[C], *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA,2020: 5021-5030.
- [10] Hu M, He M, Su W, et al. A TextCNN and WGAN-gp based deep learning frame for unpaired text style transfer in multimedia services[J]. *Multimedia Systems*, 2020,21:1-10.
- [11] Zhang J, Guo M, Fan J. A novel generative adversarial net for calligraphic tablet images denoising[J]. *Multimedia Tools and Applications*, 2020, 79(1/2):119-140.
- [12] Luo XF, Wang YF. A cultural perspective on the evolution of Chinese character forms[J]. *Journal of Shenyang Engineering College (Social Science Edition)*,2017,13(01):91-95.
- [13] Dabov K, Foi A, Katkovnik V, et al. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering[J]. *IEEE Transactions on Image Processing*, 2007, 16(08):2080-2095.
- [14] Zhang K, Zuo W, CHEN Y J, et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising[J]. *IEEE transactions on Image Processing*, 2017,26(7): 3142-3155.
- [15] Almahairi A, Rajeshwar S, Sordoni A, et al. Augmented cyclegan: Learning many-to-many mappings from unpaired data[C], *International Conference on Machine Learning*. Stockholm, Sweden,2018: 195-204.
- [16] Zhang H, Sindagi V, Patel V M. Image De-raining Using a Conditional Generative Adversarial Network[J]. *IEEE transactions on circuits and systems for video technology*, 2019, 30(11): 3943-3956.
- [17] Wu R, Zhang G, Lu S, et al. Cascade ef-gan: Progressive facial expression editing with local focuses[C], *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA,2020: 5021-5030.