



# Image Classification and Restoration of Ancient Textiles Based on Convolutional Neural Network

Sha Sha<sup>2,4</sup> · Yi Li<sup>1</sup> · Wantong Wei<sup>1</sup> · Yating Liu<sup>3</sup> · Cheng Chi<sup>1</sup> · Xuewei Jiang<sup>1</sup> · Zhongmin Deng<sup>2</sup> · Lei Luo<sup>2</sup> 

Received: 20 September 2023 / Accepted: 29 November 2023  
© The Author(s) 2024

## Abstract

Ancient textile images have a variety of styles and themes, and the classification of different types of textiles provides a reliable reference for the protection and restoration of cultural relics. Due to the low efficiency of traditional classification methods and the low accuracy of classification, the image restoration of textiles takes longer and the repair effect is poor. Therefore, this paper takes ancient textile images as the research object and selects YOLOv4–ViT collaborative identification network (YOLOv4–ViT network) and generative adversarial networks (GAN) restoration model from a variety of network models to classify and restore ancient textile images. In this work, YOLOv4–ViT network is used to recognize and classify pattern elements in ancient textile images. Then, according to the classification results, restoration training of ancient textiles was carried out using an improved GAN restoration model, for which the final classification accuracy reached 92.78% and the repair result even took only 1.5 s. On this basis, a reliable retrieval and restoration system is designed to realize the repair of damaged textile images, reduce the difficulty of repair, and help users retrieve and browse different categories of ancient textile images, thus solve the problems of slow retrieval speed in traditional retrieval methods and poor restoration effect of ancient textile images.

**Keywords** Convolutional neural network · Deep learning · Ancient textile image classification · YOLOv4–ViT network · GAN

## 1 Introduction

Judging from the textile cultural relics preserved today, each dynasty had different textile characteristics and patterns. Due to the long burial time and the influence of burial environments such as water erosion, soil, insects and others, the

unearthed textile relics often show problems such as distortion, folding, adhesion, damage and fading. The existence of the above problems increases the difficulty of finishing and repairing ancient textiles. Meanwhile, the textile restoration work is complicated and long-lasting, and the long-term restoration work will lead to a decrease in the fiber strength of textile cultural relics, which is prone to embrittlement and fracture. Therefore, improving the classification efficiency of textiles and quickly realizing the image restoration of ancient textile cultural relics are urgent problems to be solved to avoid secondary damage to textiles, and the restored image pattern elements also provide a lot of inspiration for art creators.

The convolutional neural network (CNN) and GAN are two pivotal models in the field of deep learning [1–6], which can not only effectively extract features from images, but also reduce the amount of computation, thus greatly improving the training efficiency. In terms of the model framework, CNN is a specialized neural network structure designed for processing image data, whereas GAN is an adversarial framework that utilizes adversarial training to generate

✉ Yating Liu  
sendohlyt@163.com

✉ Lei Luo  
leiluo@wtu.edu.cn

<sup>1</sup> School of Fashion, Wuhan Textile and Apparel Digital Engineering Technology Research Center, Wuhan Textile University, Wuhan 430073, Hubei, China

<sup>2</sup> State Key Laboratory of New Textile Materials and Advanced Processing Technologies, Wuhan Textile University, Wuhan 430200, Hubei, China

<sup>3</sup> School of Foreign Languages, Wuhan Textile University, Wuhan 430073, Hubei, China

<sup>4</sup> Institute of Design Innovation and Fiber Science, Wuhan Textile University, Wuhan 430073, Hubei, China

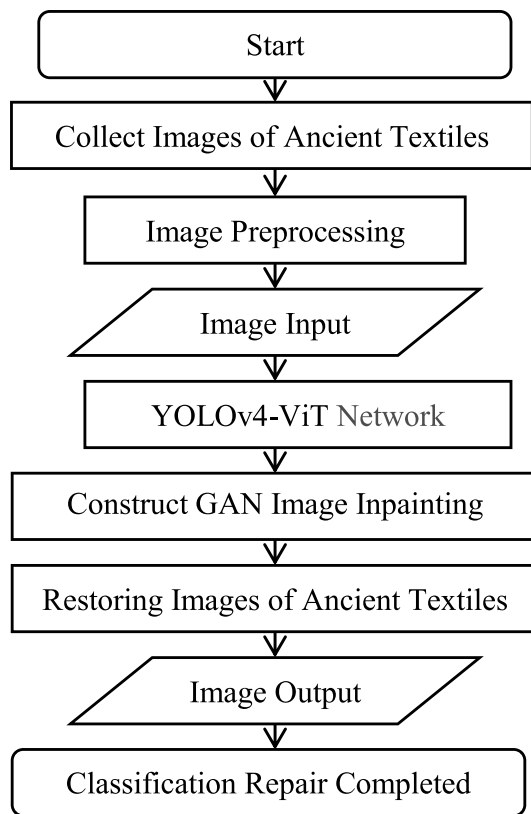
realistic samples. It has the capability to produce high-quality, diverse and authentic images or other types of data from random noise. Furthermore, during the training process, GAN employs a competitive learning strategy to update the discriminator iteratively through feedback signals. This iterative approach promotes continuous improvement of the model and facilitates the generation of more realistic output results. In terms of application areas, the use of CNN for detailed classification of ancient textile images lays a foundation for the research and dissemination of ancient textiles and also fills the gap of deep learning in the research field of ancient textile image retrieval [7–9]. CNN is widely used in field such as object detection, facial recognition and image classification, showing high efficiency and evaluation [10–15]. It helps users shorten the time of retrieving ancient textile images and thus improve the retrieval efficiency. Also, the automatic classification of ancient textile images can promote the construction of the digital platform of ancient textile art, which is of great significance to the protection and inheritance of national intangible cultural heritage [16–20]. GAN can be used for tasks such as image synthesis, super-resolution reconstruction, and image restoration. In the field of image synthesis, GAN is capable of generating images with specific attributes based on given conditions. When it comes to super-resolution reconstruction, GAN can transform low-resolution images into their high-resolution counterparts. Additionally, in the area of image restoration, GAN has the ability to automatically fill in missing or damaged parts.

Computer image classification technology has been progressively advancing and evolving, leading to its growing utilization in the classification of textile and clothing images [21–24]. Zhou et al. [25] proposed a clothing image classification method based on parallel convolutional neural network (PCNN) combined with optimized random vector functional link, which improved the stability and accuracy of classification. Dorozynski et al. [26] studied the problem of image retrieval in silk fabric database through CNN learning, exploiting available annotations to automatically generate training data and combining descriptor learning with auxiliary classification loss. Qian et al. [27] proposed that a model can be constructed to generate features according to the texture of embroidery stitch and simulate its embroidery style, so as to highlight the stereoscopic effect of embroidery works. Zhang et al. [28] extracted the color of Chinese traditional Yue embroidery by K-means clustering method and presented a method to judge the significance of color characteristics in cultural phenomena. Taking the costume patterns of the Tang tomb murals, Plain Unlined Silk and Spring Outing Painting of Madam Guo as the research objects, Liu et al.

[29, 30] realized the development and reverse reduction of clothing elements in the paintings through 3D interactive pattern-making technology and virtual simulation technology. The effectiveness of virtual clothing simulation is evaluated using the analytic hierarchy process and fuzzy comprehensive evaluation method. These research results provide new ideas for the rapid simulation and restoration of ancient Chinese clothing and provide certain reference for the redesign of ancient clothing. The above researches transformed the color elements of Yue embroidery in Chinese traditional cultural phenomena into storable and analyzable data resources, which is necessary for its sustainable development.

However, few researches have been done on the image restoration of ancient textiles currently. Common image restoration methods of ancient textile are all Criminisi algorithms based on texture restoration, which first calculate the priority to determine the order of damaged areas to be repaired, then lock the intact parts of the textile as sample blocks, which are then used to fill damaged areas of ancient textiles. While some images of ancient textiles can be restored in this way, the information of the restored textile images of the algorithm is shallow (less semantic, more noisy), and its deep features (less semantic, information abstraction) are not extracted. Moreover, large damage of textile relics can easily cause problems such as chaotic texture and incoherent edges of the restored images, resulting in poor effect of restoration. The task of ancient textile image classification is complicated, and since there are often many combinations of pattern elements in a single ancient textile image, manual recognition and classification are usually required, consuming a lot of time and labor costs. In the domain of computer vision, the YOLOv4 algorithm has exhibited remarkable efficacy in facilitating precise image recognition and classification outcomes [31, 32].

In addition, due to the huge amount of ancient textile image data, users need to spend a lot of time and energy to find the required pattern elements, which greatly reduces work efficiency. Therefore, we propose to classify and repair pattern elements in ancient textile images based on CNN. The collaborative identification network of YOLOv4–ViT is utilized to accurately identify and classify pattern elements in ancient textile images, followed by the application of an enhanced GAN restoration model to effectively restore damaged areas within these images. The experimental results show that the restored ancient textile images have complete structure and coherent patterns, the accuracy rate reached 92.78% in the classification results of some pattern element categories, and the model performance was better, which could quickly repair the damaged textile image in 1.5 s and



**Fig. 1** Ancient textile image classification and restoration flowchart

accurately find the ancient textile image that needed to be repaired in the retrieval system.

Figure 1 shows the process of classifying and repairing ancient textile images based on the YOLOv4–ViT network and GAN image restoration model. After collecting the ancient textile images, the preprocessing of the collected images is conducted for training the classification network model. Subsequently, the YOLOv4–ViT network is utilized to detect and recognize the processed dataset. Then, a GAN image restoration model is constructed for training image restoration. Finally, the fine classification of ancient textile

images and the restoration of damaged images are completed [33–35].

## 2 Image Characteristics Analysis of Ancient Textiles

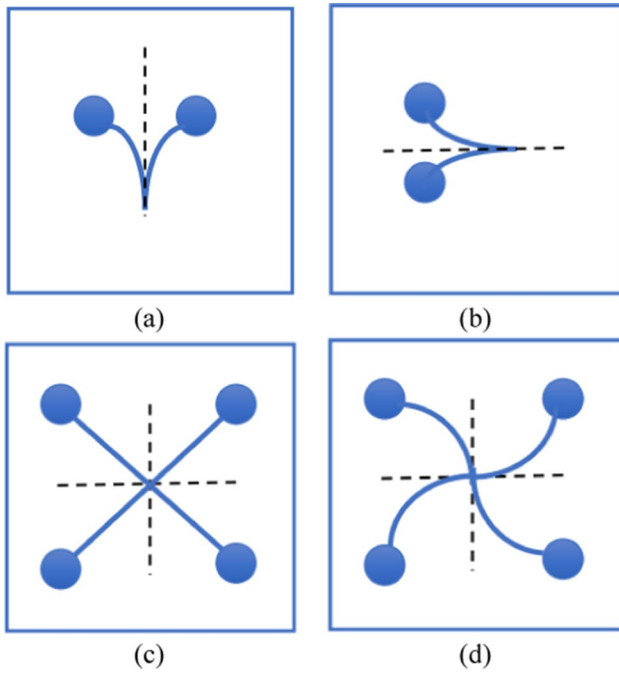
### 2.1 Pattern Acquisition and Analysis

To facilitate restoration, images of ancient textiles should first be collected, sorted and classified. Due to the lack of datasets for ancient textile images, we used two methods to collect samples of image data based on the Internet in this work: automatic crawling with web crawlers and manual search. To better classify the patterns of ancient textile images, relevant literature and books were consulted. By sorting out textile images in Jing chu area and summarizing archeological literature, combined with reference to the category concept of image dataset in the field of computer vision [36–39], the following three-layer concept system is obtained (Table 1).

The materials of ancient textiles mainly include cotton, linen, kudzu, silk, etc., and the patterns are mainly geometric patterns. Patterns on gross and silk are mainly in the form of embroidery, with patterns drawn on the ground silk fabric first, and then the pattern is painted on the silk fabric and woven embroidery. Different processes and patterns can have a variety of modeling effects. From the perspective of process, patterns of ancient textile image can be divided into weaving patterns and embroidery patterns, which are therefore put in the first layer. The second layer is based on the theme, shape and other characteristics of patterns, which can be summarized as geometry, animal, plant and landscape and implement. The third layer further classifies the pattern categories of the second layer, with the image types in ancient textile patterns basically covered. With the expansion of the dataset, the categories of the third layer can be added or deleted accordingly without affecting other layers.

**Table 1** Ancient textile image pattern type differentiation table

First layer	Second layer	Third layer
Weaving patterns (1) Embroidery pattern (2)	Geometry (1)	Unknown (00), strip (01), rhombus (02), cup shape (03), tower shape (04), rectangle (05), hexagon (06), cross (07), dot (08), triangle (09), S shape (10), ring (11), arc (12), z shape (13), shape (14), bow shape (15), fret (16), cloud shape (17), thunder shape (18)
	Animal (2)	Unknown (00), phoenix (01), dragon (02), tiger (03), deer (04), snake (05), kylin (06), horse (07), fish (08), dancer (09), warrior (10)
	Plant (3)	Unknown (00), flower (01), leaf (02), spray (03), corolla (04), tree (05)
	Landscape and implement (4)	Unknown (00), chariot (01), boat (02), mountain (03), water (04), sun (05), bow (06), arrow (07), sword (08), shield (09), cloud (10)



**Fig. 2** Symmetrical structure diagram: **a** left-and-right symmetry; **b** up-and-down symmetry; **c** multi-directional symmetry; **d** rotational symmetry

## 2.2 Structural Characteristics of Ancient Textile Image Patterns

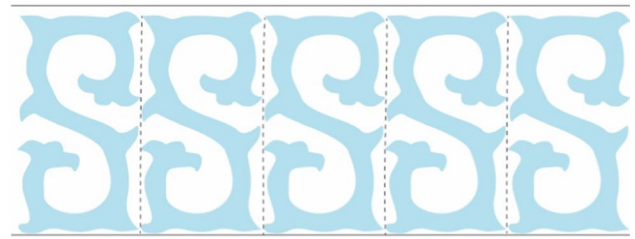
The image structure of ancient textiles is mainly reflected in the structure of the pattern, which can be divided into two categories: the single pattern structure and the consecutive pattern structure, the latter being the main pattern structure in ancient textiles.

### 2.2.1 Individual Pattern Structure

The single pattern structure, in which a unit pattern is arranged independently on ancient textiles, can generally be symmetrical or balanced. The symmetrical pattern structure can be seen in various forms, including left-and-right symmetry, up-and-down symmetry, multi-directional symmetry, rotational symmetry and so on, as shown in Fig. 2. The balanced pattern structure is not affected by symmetry axis or symmetry point, thus enjoying a high degree of freedom in composition.

### 2.2.2 Continuous Pattern Structure

In the consecutive pattern structure, patterns of multiple units are arranged regularly and repeatedly on ancient textiles, forming an infinite cycle of patterns. The consecutive pattern structure is generally divided into double-square



**Fig. 3** Double-square pattern

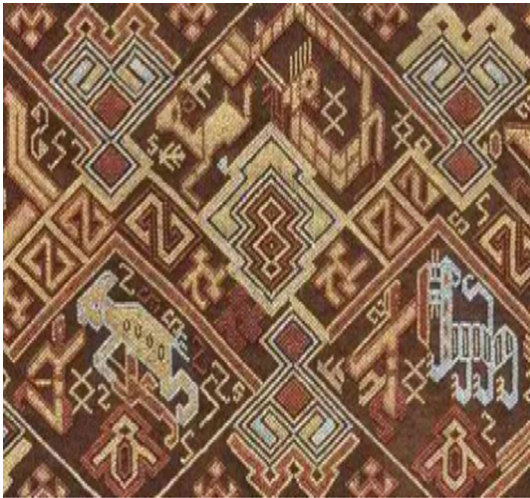


**Fig. 4** Four-square pattern diagram



**Fig. 5** Dragon coiling around phoenix embroidery patterns

structure and four-square structure. Among them, the double-square pattern refers to the pattern with a ribbon-shaped structure, in which a unit pattern is continuously arranged left and right or continuously up and down between two parallel lines, and it is generally used in long and narrow positions such as collar edge, sleeve edge, bottom edge and belt, as shown in Fig. 3. The four-square pattern is a pattern structure formed by the repeated arrangement of a unit pattern in the four directions: up, down, left and right, as shown



**Fig. 6** Rhombus-shaped geometric animal patterns

in Fig. 4. It can be extended to all sides and is typically used where there is a large area, such as a body part of a garment. As shown in Fig. 5, the embroidery of the dragon and phoenix phase is divided and scattered and then combined to form two dragons opposite each other on the back of the phoenix bird, so that the dragon body becomes part of the bird body, showing a kind of ethereal and magical beauty, which is a typical representative of the romantic Chu style. As shown in Fig. 6, a combined pattern is created by adding animal, plant and other patterns to the basic geometric pattern. It is used as a unit, which is arranged around in a shape of rhombus, forming a four-square structure.

To summarize, the ancient textile images predominantly exhibit cyclic and consecutive patterns in weaving, embroidery and textile processes. It is evident that most ancient textile images possess repetitive and regular characteristics. Building upon these features, this work employs the GAN models of deep learning to conduct subsequent research on restoring textile images.

### 3 Classification of Ancient Textile Images Based on YOLOv4–ViT Network

#### 3.1 Preliminary Preparation

##### 3.1.1 Dataset Preparation

In this experiment, 2057 images are divided into training set, validation set, and test set with a ratio of 8.1:0.9:1. The training set is used for the computer to learn and extract the features from ancient textile images. The validation set is used to properly tune the hyperparameters. The testing set

is used to verify the accuracy of the model. The dataset is placed in the JPEG Images folder in the VOC2007 folder contained in the VOC devkit folder before training.

##### 3.1.2 Preparation of XML Documents

Due to the supervised learning approach adopted in this work, it is necessary to label the images before image training. Label Img is the chosen image annotation tool selected for this experiment. The generated XML files follow the PASCAL VOC format, which is also convenient for post-training. This XML file is placed in the annotations folder within the VOC2007 folder, which is included in the VOC devkit folder before training.

#### 3.2 YOLOv4–ViT Model Framework

Aiming at the types of image patterns and structure types of ancient textiles, based on the existing object detection and recognition, the image combination enhancement and pre-training vision of transformer (ViT) classification network are introduced to realize the hierarchical labeling and detection of textiles. The structure of the model is shown in Figs. 7 and 8, and the combined enhanced YOLOv4–ViT collaborative recognition network (hereinafter referred to as YOLOv4–ViT) consists of two parts, including a pre-trained slice classification network and a backbone feature recognition network, which jointly complete the identification and detection target of textiles.

##### 3.2.1 Pre-trained Slice Classification Network

The pre-trained slice classification network is mainly divided into two modules: the slice module and the classification module, as shown in Fig. 7, where the slice module divides the input textile graph into different parts according to pixel size. The classification module adopts ViT network, combines parts as input into sequences, and uses multi-head self-attention (MSA) to extract features from the sequence, and finally classifies through CLS token position information.

The feature extraction part consists of a PE layer (position embedding) and a TE layer (transformer encoder). The PE layer divides the  $224 \times 224$  size input images into  $16 \times 16$  pixel blocks, each image generates 196 pixel blocks, and so the input sequence length is 196. Each pixel block consists of a tricolor matrix of red, green, and blue, with a size of  $16 \times 16 \times 3$ . The linear projection layer has a dimension size of  $768 \times N$ , so after passing the linear projection layer, there are 196 block features and the input dimension size is  $196 \times 768$ . Based on the location of each block feature plus the position coding CIs, the dimension of the input total feature is  $197 \times 768$ . The TE layer is a transformer encoder

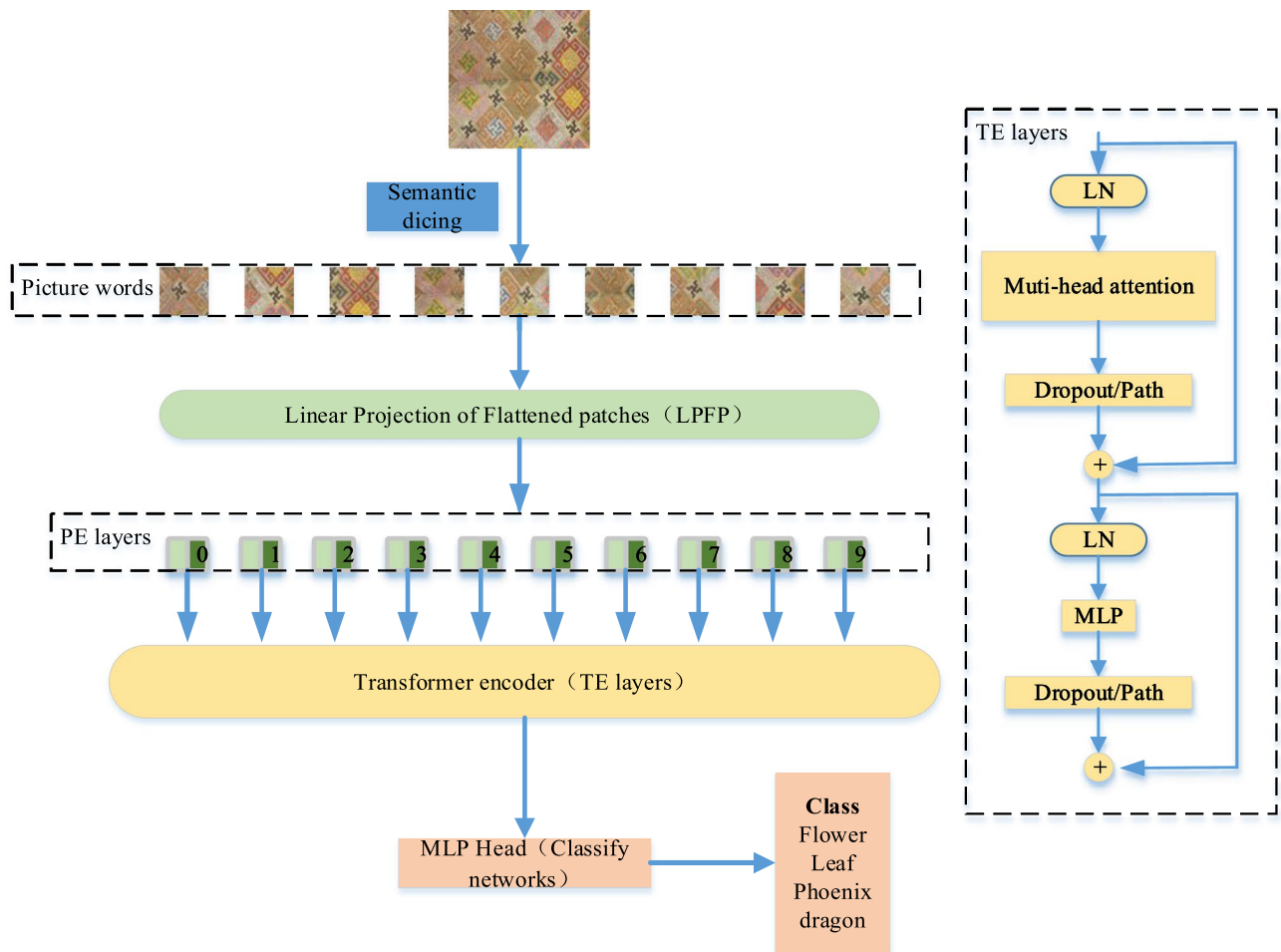


Fig. 7 ViT model framework

for feature extraction of individual feature blocks. This layer uses positional coding to interact feature blocks with other features, fusing features from different graph sequences. The self-attention feature extracted by the multi-head attention mechanism is used as the input features for the subsequent classification networks.

The classification network part is MLP Heat, which consists of four layers of TE-block stacking, and the input undergoes multiple layers of linear transformation to amplify the dimension to  $197 \times 768 \times 4$ . Then, the feature vector is reduced to  $197 \times 768$  by fully connected layer FC, and the image spatial convolution vector is calculated. Finally, the self-attention feature is classified by the regularized connection layer normalization (LN), and the specific calculation formula is as shown in Eqs. (1–4):

$$z_0 = [x_{\text{class}}; x_p^1 \dots x_p^n] + E_{\text{pos}} \quad E \in \mathbb{R}^{(P^2 \cdot C) \cdot D}, E_{\text{pos}} \in \mathbb{R}^{H \cdot W \cdot C}, \quad (1)$$

$$z'_\ell = \text{MSA}(\text{LN}(z_{\ell-1}) + z_{\ell-1}), \quad \ell = 1 \dots L, \quad (2)$$

$$z_\ell = \text{MLP}(\text{LN}(z'_\ell)) + z'_\ell, \quad \ell = 1 \dots L, \quad (3)$$

$$y = \text{LN}(Z_L^0), \quad (4)$$

where  $E_{\text{pos}}$  is the input feature sequence of the PE layer,  $Z_L^0$ ,  $z_0$ ,  $z_\ell$  is the convolution vector of image space in different states,  $x_{\text{class}}$  is a certain type of label object,  $x_p^n \in \mathbb{R}^{(P^2 \cdot C) \cdot D}$  is a two-dimensional pixel block, the upper left corner 1 ~ n represents the pixel block number, C is the number of channels, P is the pixel block size, there are a total of N pixel blocks,  $N = HW/P^2$ , and  $x \in \mathbb{R}^{H \cdot W \cdot C}$  is the input picture; MSA is a function of the long-headed attention mechanism; MLP is a multilayer perceptron function that is a shallow fully connected neural network; LN is the regularization calculation function, and y is the output classification result.

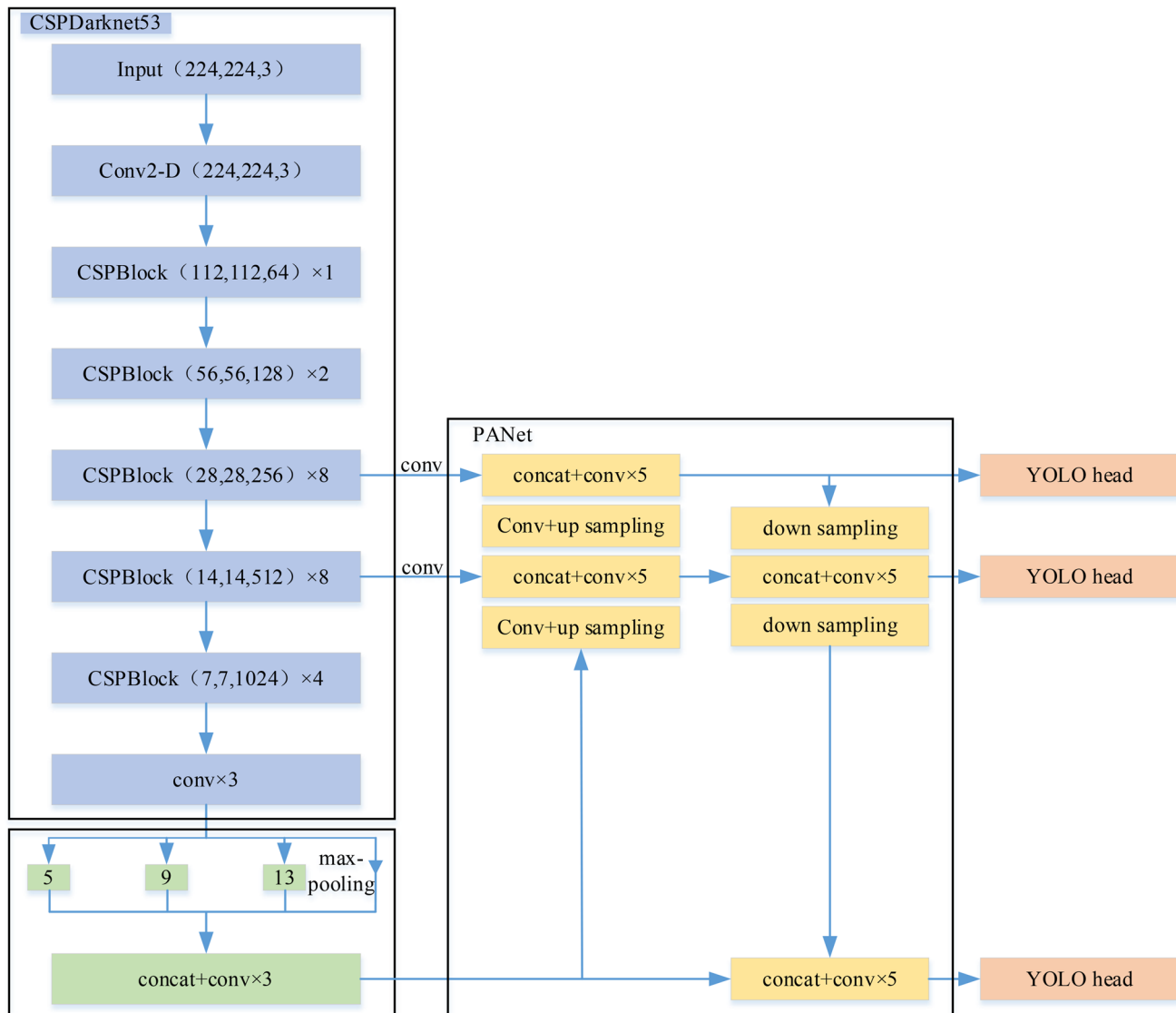


Fig. 8 YOLOv4 model framework

### 3.2.2 Backbone Feature Recognition Network

The backbone feature recognition network in this work is YOLOv4, which refers to the idea of cross-stage partial network (CSP Net), and uses CSPDarknet5, a cross-stage partial darknet (CSP Darknet) with stronger feature learning effect as the core network on the basis of YOLOv3. The CSPDarknet5 core network outputs three feature maps to ensure that the extracted image information is complete. Spatial pyramid pooling (SPP) stacks and convolves feature maps to expand the visual receptive field and facilitate global detection. The path aggregation network (PANet) samples the feature map up and down to fuse the extracted image information. Finally, YOLOv4 generates three detection heads to detect the target. The structure of YOLOv4 is

shown in Fig. 8, where conv stands for the convolutional layer, concat for concatenation, and conv 2-D for two-dimensional convolutional layer.

### 3.3 Classification of Ancient Textile Images

For object detection and classification, the YOLOv4-ViT network performs feature extraction from input images. The model's frozen training, which involves only fine-tuning the network to avoid corrupting the weights, occupies a minimal amount of video memory. However, upon thawing the backbone, the network's memory consumption increases significantly. Consequently, adjustments to the network's

parameters are required to facilitate the final implementation of network training and prediction of output results.

Step 1: Create new\_classes. The files, which specify the categories to be distinguished in this experiment, are shown in Table 1.

Step 2: Set hyperparameters, including batch size, learning rate, etc. The hyperparameters are set differently during freeze-out and after thawing. In object detection and recognition task, the features extracted from the backbone are generic, so freezing the training of the model can speed up the training and prevent the loss of weights. When the backbone is frozen, the occupied video memory is small and only fine-tuning of the network is required. After the backbone is thawed, the occupied video memory is large and the network parameters need to be changed. The parameter settings during freezing and after thawing are shown in Tables 2 and 3.

Step 3: Run the program and start training the network.

Step 4: Predict the training results and input the graph path for prediction.

### 3.4 Recognition Results

The evaluation indicators for evaluating object detection and recognition tasks are mean average precision (mAP), with larger mAP value indicating better effect of the object detection task, average precision (AP), F1 value, precision, and recall.

AP takes comprehensive consideration of recall and precision, indicating the quality of the detection of a certain category, and its expression is shown in Eq. (5).

$$AP = \int_0^1 P(R)dR, \quad (5)$$

where mAP is the average AP of each category, which is used to measure how well the model detects all categories, and its expression is shown in Eq. (6).

$$mAP = \frac{\sum AP}{N}. \quad (6)$$

After model training, the mAP obtained in this experiment is 92.78% as shown in Fig. 9, indicating that the YOLOv4–ViT network selected in this experiment performs

**Table 2** Hyperparameter setting during freezing

Parameter	Value setting
Learning rate	0.001
Batch_size	8
Init_epoch	0
Freeze_epoch	50

**Table 3** Hyperparameter settings after thawing

Parameter	Value setting
Learning rate	0.0001
Batch_size	4
Freeze_epoch	50
Epoch	100

well in the image recognition and classification of four pattern elements.

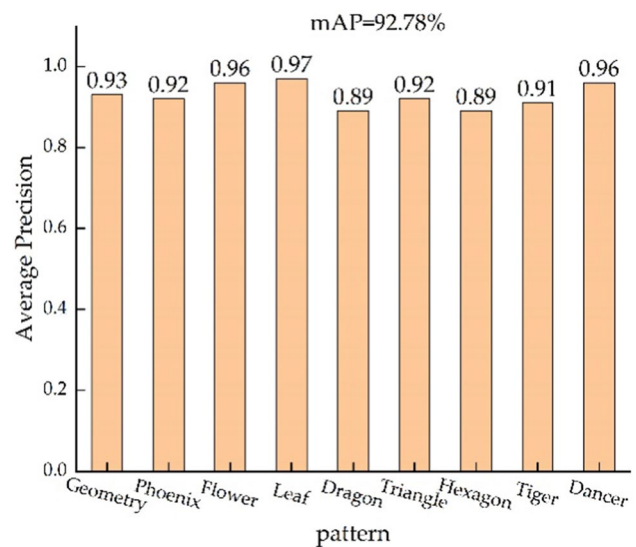
To validate the mAP obtained in this experiment, some images, both in and out of the dataset, are selected for testing, which can better test the generalization ability of the model. Some of the test results are shown in Fig. 10.

From the test results, YOLOv4–ViT network can identify and classify ancient textile images with multiple-label pattern elements, such as “flower2301” showing that the pattern belongs to the plant flowers in embroidery, and the recognition results are good. It not only identifies specific categories and corresponding codes of the intention, but also reflects the craftsmanship, subject, and shape information of the patterns, which is conducive to further recovering and training the ancient textile images of the same category.

## 4 Construction of Ancient Textile Image Restoration and Retrieval Restoration System

### 4.1 Constructing GAN Image Restoration Model

The task of image restoration of ancient textiles is to complete the incomplete parts in the image and repair the



**Fig. 9** Mean average precision (mAP)



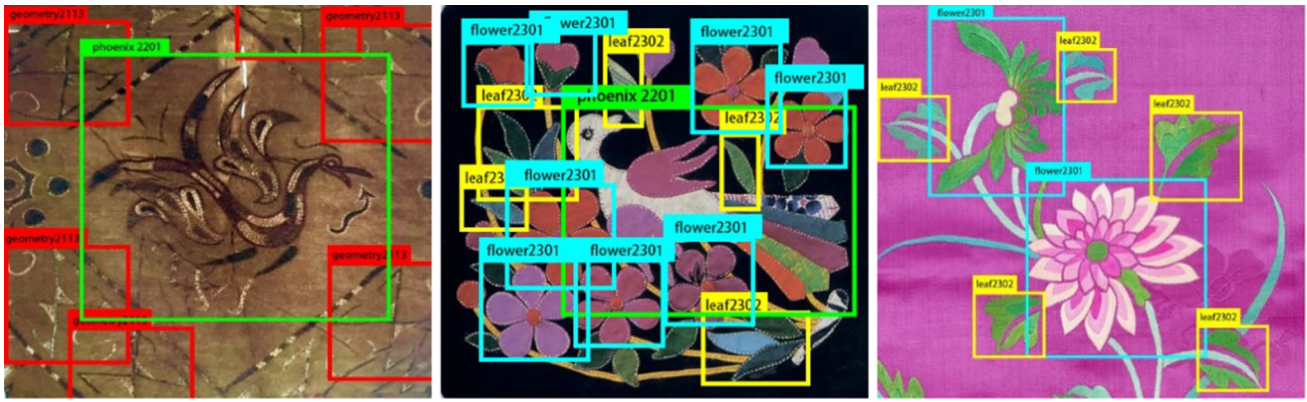


Fig. 10 Hybrid tag recognition results

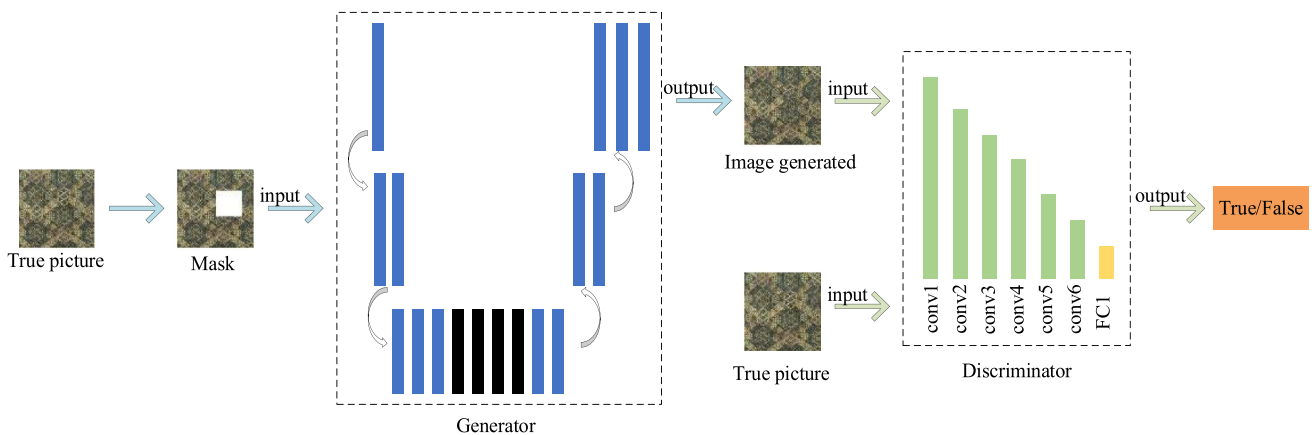


Fig. 11 Ancient textile image restoration network model architecture

broken locations in the pattern. In this work, a generative adversarial network (GAN) model based on U-Net network is constructed to carry out the restoration experiment of ancient textile images. The fundamental concept of the GAN model is to facilitate a game between the generator and the discriminator to achieve a Nash equilibrium, ultimately generating images that are highly reminiscent of the authentic ones. In its application to textile image restoration, due to the repetitiveness and regularity of ancient textile images, the data information close to the unknown areas of the original textile images can be generated by learning the input data information such as structural pattern features of the known areas in the textile image dataset, so that the incomplete textile images can be completed in structure and coherent in pattern. The specific network architecture is shown in Fig. 11.

#### 4.1.1 Generator Design

The generator of the ancient textile image restoration model is mainly based on the U-Net network, a typical encoder–decoder structure, and a well-known image segmentation network for processing medical images. Similar to the difficulty in obtaining data from medical images, it is also difficult to obtain data from ancient textile images, and the U-Net structure is thus used as a generator.

Step 1: U-Net backbone architecture design. The U-Net architecture extracts the features of textile images by four-fold upsampling via convolution and pooling, and then recovers the features by fourfold upsampling via transposed convolution, with sampling and downsampling exhibiting a U-shaped symmetric structure. In the middle, the extracted textile image features and the restored textile image features are spliced, which is also called skip connection, and the depth of the features is increased to fuse the feature maps.

Step 2: U-Net structural improvement. Due to the complexity of some textile patterns, this paper improves the

U-Net structure by reducing the times of sampling, and sampling and canceling the pooling layer so as to reduce the loss of textile image data information. Instead, two subsampling operations and two upsampling operations are carried out, respectively, to reduce the amount of calculation of the textile image restoration model.

Step 3: Selection of U-Net structure steps. The step size is selected according to the complexity of the textile image information and the size of the area to be repaired. The step size needs to be properly reduced when there are comparatively heavy details, complicated patterns, and large area to be repaired in the textile image. On the contrary, it can be appropriately increased to improve the computational efficiency when the information in the textile image is comparatively simple and the area to be repaired is small. In the sampling operation, convolution with adaptive step size is used to compress the data features of textile images. The feature restoration of textile images is conducted in the upsampling operation by transposed convolution with adaptive step size and a  $3 \times 3$  convolution kernel. The skip connection is canceled, and the atrous convolution is introduced, because part of the edge data may be lost in the process of image data subsampling, and reducing sampling times can effectively prevent data loss, also, introducing various convolution can widen the receptive field and reduce the loss of textile image information. ReLU is selected as the activation function of the generator whose network structure is shown in Fig. 12. The damaged and missing parts of the ancient textile images in the tombs of the State of Chu are repaired, thus preserving their original meaning.

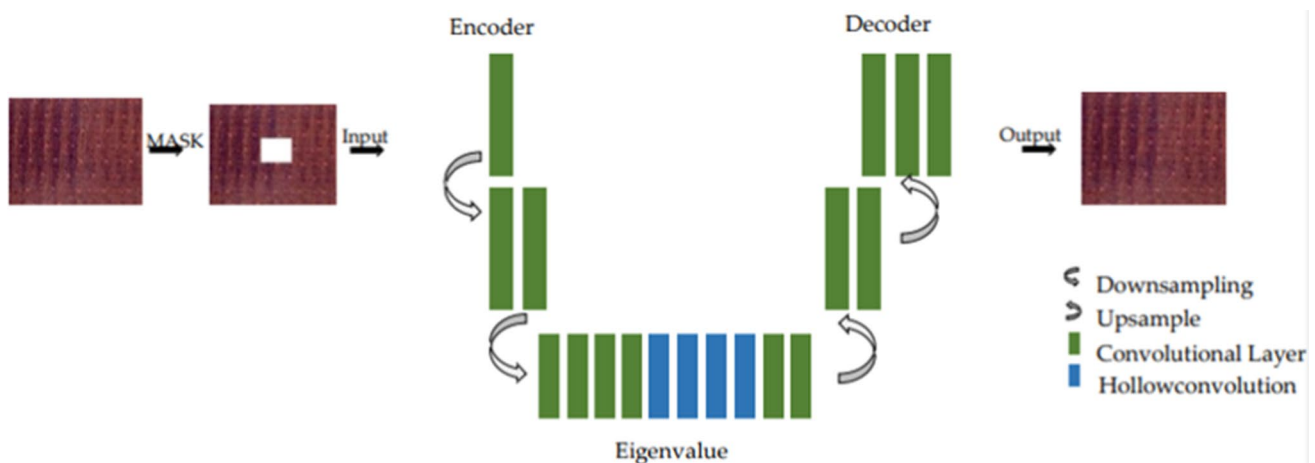
#### 4.1.2 Discriminator Design

The discriminator of the ancient textile image restoration model plays a very important role in image restoration. It

verifies the authenticity of the textile image after re-pairing, and is used to determine whether the result generated by the generator can be output as the final result. It can also point out the gap between the original ancient textile image and the generated one, thus training the generator to generate near-realistic images of ancient textiles. In this work, a seven-layer neural network structure is adopted for the discriminator, with six convolution layers and a totally connected layer. The convolution layer adopts  $5 \times 5$  convolution kernel with a step size of 1 and uses ReLU as the activation function. The fully connected layer uses the sigmoid as the activation function. The fact that every neuron in the fully connected layer is connected with the neurons in the previous layer enables it to integrate the information of all feature data used for classification. Classification can thereby be conducted to judge whether the input ancient textile image data is true or false. Sigmoid function is also often used in binary classification problems. The input data can be mapped between 0 and 1, the larger the value of the input positive data, the closer the function value is to 1, and the larger the value of the input negative data, the closer the function value is to 0. With the output result being 0 or 1, if the output is 0, the input ancient textile image is judged to

**Table 4** Discriminator structure table

Type of layer	Activation function
Layer of convolution 1	ReLU
Layer of convolution 2	ReLU
Layer of convolution 3	ReLU
Layer of convolution 4	ReLU
Layer of convolution 5	ReLU
Layer of convolution 6	ReLU
Fully connected layer 1	Sigmoid



**Fig. 12** Structure design of generator

be false, and if the output result is 1, the input ancient textile image is judged to be true. However, the convergence rate of the sigmoid function is relatively slow, so the RELU function with rapid convergence rate is selected as the activation function in the convolution layers. Experiments show that the discriminator can effectively discriminate the generated textile image data. After repeated games, the quality of the textile image generated by the generator is improved, which is close to the original ancient textile image. The discriminator structure is shown in Table 4.

### 4.1.3 Loss Function

The loss function is an important indicator to measure the quality of the GAN model, which is mainly used to show the difference between the generated textile image data and the original image data. In this work, the MSE loss function and the adversarial loss function are used as the loss functions of the generator and discriminator. The mean squared error (MSE) loss function, also called L2 loss function, is the average of the squared distance between the generated textile image and the original textile image, which is defined as Eq. (7).

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_1 - y'_i)^2, \quad (7)$$

where  $N$  represents the number of data samples of the input ancient textile image;  $y_i$  represents the image data of the original ancient textile;  $y'_i$  represents the generated predicted ancient textile image data. The network model guided by the MSE loss function can generate general information of ancient textile images, while obtaining detailed information requires the introduction of adversarial loss. The principle of the GAN network model is that the discriminator should try to identify the image generated by the generator, so that the output of the predicted image data is close to 0, and the output of the original image data is close to 1. Hence, the adversarial loss is defined as in Eq. (8).

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}} [\log D(x)] + E_{z \sim p_z} [\log(1 - D(G(z)))] \quad (8)$$

In the above,  $x$  is the raw ancient textile image data.  $P_{\text{data}}$  is data distribution of the original ancient textile images.  $P_z$  is the data distribution of the generated ancient textile images.  $D(x)$  denotes the discriminant result of the raw image data as closer to 1 as possible.  $D(G(z))$  denotes the discriminative result of the generated image data as close to 0 as possible. The adversarial loss function can optimize the generator and discriminator respectively, enabling the trained ancient textile image restoration model to achieve

better image restoration effect, making the restored ancient textile image close to the original one.

## 4.2 Experiment

This experiment is programmed in Python3.6 with the deep learning framework Pytorch and is performed in a Windows 10 environment with NVIDIA RTX 3060.

In this work, Anaconda is used to create a virtual environment with more than 180 packages and their dependencies such as Conda, Python, etc. This paper compiles the language in PyCharm, a kind of Python Integrated Development Environment (IDE), with tools such as debugging, code jump, unit testing, etc., which can assist Python in compiling and improve its efficiency.

### 4.2.1 Experimental Dataset

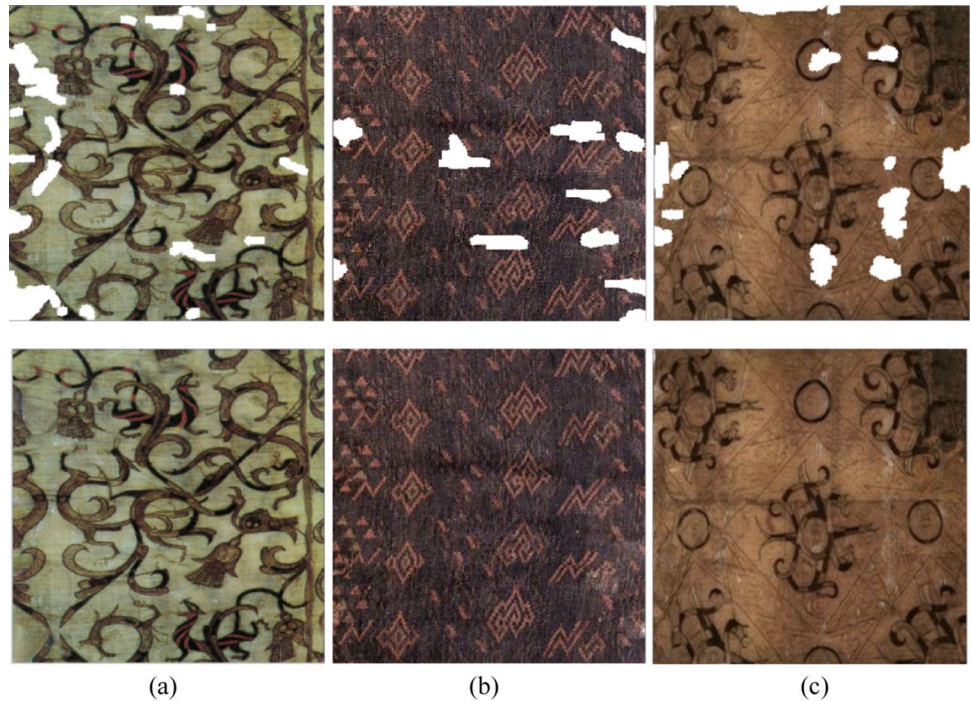
The main object of this work is the image of ancient textile relics. Due to the lack of sufficient image data on ancient textiles, there is no dataset on ancient textile images. Therefore, in the process of image collection, the main data source is the ancient textile images of Mashan Tomb No. 1, known as the "treasure house of silk", supplemented by the textile relics image of the State of Chu in other places. The textile images are scaled and cropped, one image is cropped into multiple images, and the image data volume is augmented by rotation and inversion. The textile image dataset is built by collecting, cropping, sorting, and classifying images.

### 4.2.2 Network Training

The collected image data of ancient textiles are adjusted into the size of  $256 \times 256$  by cropping. Some of them are used as test sets and the rest as training sets.

The process of textile image restoration based on GAN network model is as follows: (1) masks with a size of  $64 \times 64$  are added to random positions of the training set image data to realize partial occlusion; (2) the training data together with the masks are input into the generator, which extracts the eigenvalue of the original ancient textile images by subsampling with two convolutional operations; (3) the eigenvalue of the extracted and compressed ancient textile images is expanded by atrous convolution to increase the receptive field, and also reduce the loss of edge information of the original ancient textile images; (4) upsampling with two transposed convolutional operations are performed to generate the repaired ancient textile image data; (5) the included intact original ancient textile images are input into the discriminator together with the repaired ancient textile images generated by the generator; (6) loss function is used to calculate the gap between the generated image data and the original image data, and the generated images and the

**Fig. 13** Comparison before and after the restoration of ancient textile: **a** phoenix flower embroidery; **b** warring states brown rectangular brocade; **c** flat pattern embroidery



original images are discriminated by the discriminator; (7) the discriminant results are fed back to the generator to optimize its parameters, so that it can generate ancient textile images that are closer to the original ones; (8) the process of generating and discriminating is repeated for many times until the generated images are too close to the original ones to be distinguished by the discriminator. At this point, the GAN model is trained and the trained image restoration model can be used to restore ancient textile images.

#### 4.2.3 Experimental Results and Analysis

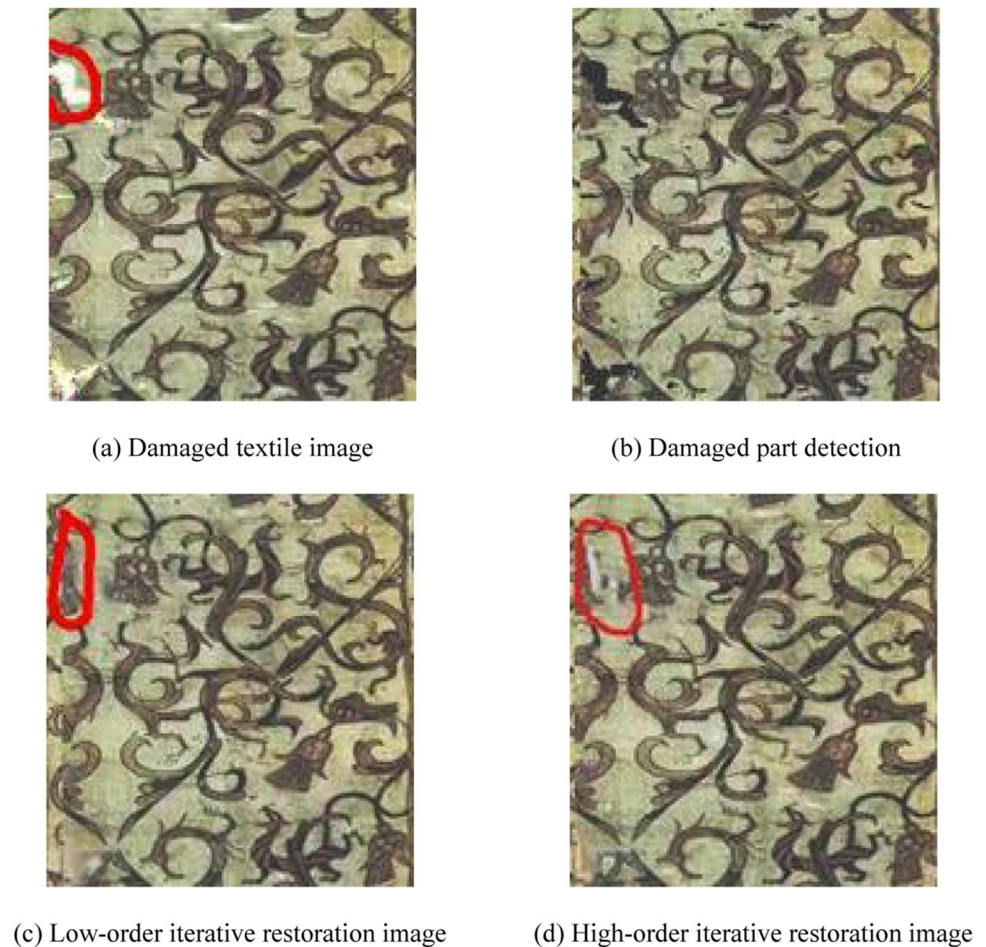
The set images are attached with free masks and fixed by the trained GAN network model. The input image data size is  $256 \times 256$ , and the repaired images are shown in Fig. 13.

To further test the effectiveness of the restoration model proposed in this work, the ancient textile images with complex patterns are selected, and DeepFill v2 algorithm, a typical image restoration method in recent years [33], is used for comparative experiment. The restoration results are shown in Fig. 14. There is an obvious contrast between the restoration effects of ancient textile images restored by the DeepFill v2 algorithm and the images restored by the proposed method. Compared to the methodology employed in this work, DeepFillv2 is a framework grounded on generative adversarial networks. The adaptive context encoder (ACE) possesses the capability to automatically identify missing regions in an image based on its contextual information. However, as observed at the marked red area, the restored image after multiple iterations exhibits particularly unsatisfactory

outcomes with noticeable white spots. Although there are no overt blemishes apparent in the image restored after a few iterations, it fails to effectively restore pattern information and appears visually indistinct and disordered. Additionally, there is color blending between distinct regions. It can be seen that the images recovered using the DeepFill v2 algorithm show overall blurring, artifacts, noise, and other problems, and the recovered pattern information is incoherent. In contrast, the image repaired using the method described in this work (Fig. 13) has a clearer restored region, restoring the structure of the textile image and a portion of the pattern. The restored content is well connected to the original parts of the fabric, and the pattern structure is integrated with the remaining part of the image without obvious color boundaries. The results were satisfactory, and the restored image had a certain amount of ornamental value.

The proposed GAN network model is quantitatively compared with the DeepFill v2 algorithm repair results. The quantitative comparison results are shown in Table 5, which shows that GAN network has the best evaluation indicators in time, peak-signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). As a widely used index in image reconstruction quality evaluation, PSNR measures the peak-signal-to-noise ratio of the generated image and the original image. SSIM means the structural similarity between the generated and real images based on their brightness, contrast and structure. The higher the PSNR value, the better is the image quality; the larger the SSIM value, the stronger is the comparability of the structural characteristics. The method in this work achieves the repair effect of textiles

**Fig. 14** Image restored by DeepFill v2 method



**Table 5** Comparison of the DeepFill v2 method with the method in this work

	Time(s)↓	PSNR↑	SSIM↑
DeepFill v2 method	2.35	27.85	0.86
This work	1.50	31.37	0.94

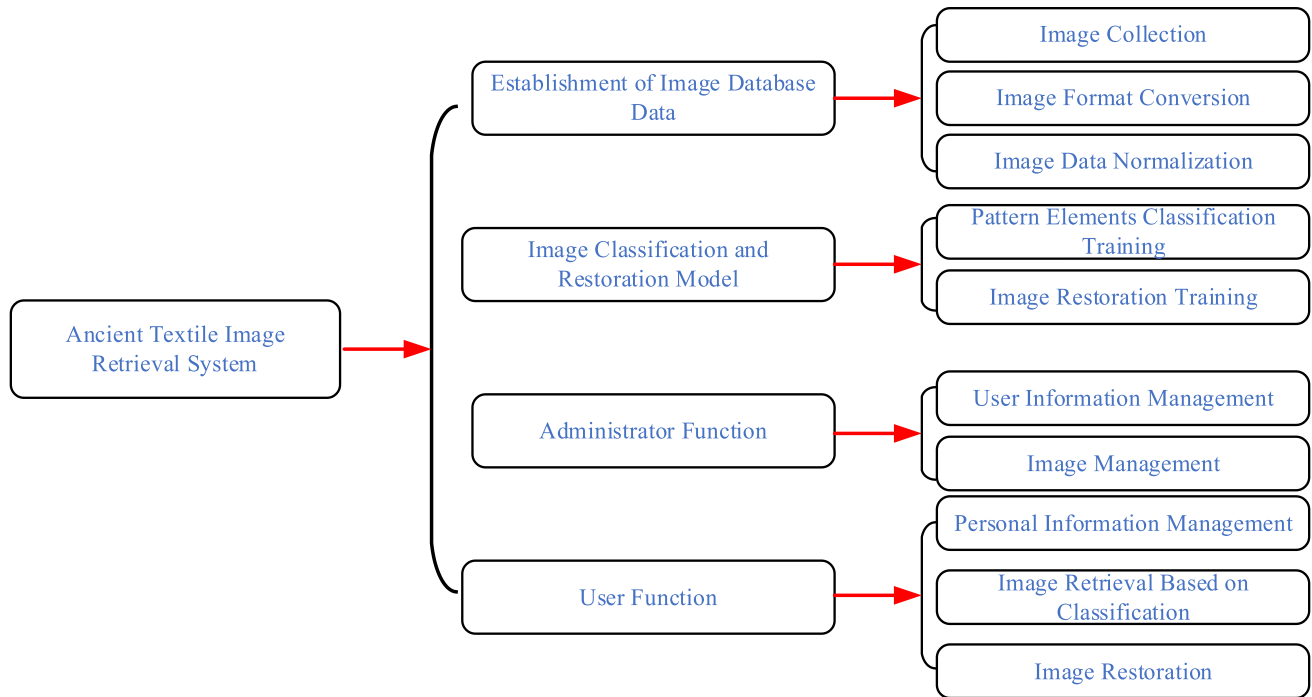
in only 1.5 s, which reduces the time by 35.71% compared with the DeepFill v2 algorithm. There is also some improvement in the PSNR and SSIM indicators, indicating that the optimized GAN model can repair and generate more reasonable textile images.

By analyzing the above experimental results, it can be seen that the existing image repair algorithms are trained based on face, architecture, landscape and other image datasets, while the image restoration model proposed in this work is constructed and trained based on the characteristics of repetitiveness and regularity of Chu textile images, learning the features of Chu textile images to further restore them. The method proposed in this work is necessary and very effective for the restoration of textile images from the Chu tombs.

### 4.3 Construction of Image Retrieval and Restoration System

Based on the above experimental results, to facilitate the classification-based retrieval and restoration of ancient textile image content, this paper designs an ancient textile image retrieval and restoration system, and the specific functional module design is shown in Fig. 15.

The whole retrieval system is divided into four modules. The first module is to establish an ancient textile image database to collect, convert and normalize the ancient textile image data so that the data can be trained. The second module is the training of classification and restoration model. First, the YOLOv4–ViT network is used to classify the pattern elements, then the improved GAN restoration model is used for image restoration training. After that, the classification and restoration model can be applied to image retrieval and restoration of ancient textiles. The third is the administrator module. The administrator manages users' personal information, as well as the ancient textile image database, importing the images uploaded by users into the database, extracting the features of image data, and establishing the image feature database to facilitate future retrieval and



**Fig. 15** Ancient textile image retrieval and restoration system

restoration. The fourth module is the user function module. Users can use the ancient textile image retrieval system only after registering their personal information, and then they can upload the ancient textile images to be retrieved or restored, perform feature extraction before similarity matching with the data in the feature library, and finally retrieving or restoring the desired ancient textile images.

## 5 Conclusion

This work mainly presents an image classification and restoration method based on deep learning, which combines the YOLOv4–ViT collaborative identification network and an improved GAN image restoration model to solve the problem of incomplete structure and broken patterns of ancient textile images. The YOLOv4–ViT network was able to recognize and classify each pattern element, and the accuracy of the ancient textile image classification model reached 92.78%. Classified ancient textile images can be trained for image restoration based on species, and the modified adaptive one-step GAN model is used for training, which improves the training efficiency of the model by achieving restoration of even damaged textiles within 1.5 s. With its diverse of pattern elements, ancient textile imagery is a rich source of inspiration for art creators. To facilitate art creators to quickly retrieve the pattern elements in ancient textile images, this paper also designs a simple ancient textile

image retrieval and restoration system that can retrieve and restore the images in both in terms of patterns and content and greatly improve the retrieval and restoration efficiency. At the same time, a database of ancient textile images has been established in the design of the retrieval and restoration system, which also provides some technical support for the digital protection and inheritance of ancient textile images.

The calculation of the method proposed in this work poses a significant challenge due to its heavy reliance on an extensive amount of datasets and computing resources. In future studies, it is necessary to explore more lightweight model structures and algorithms to alleviate the computational difficulty. Additionally, the effectiveness of the proposed methods might be limited under extremely low light conditions or complex backgrounds. Therefore, further research needs to improve lighting techniques and background modeling methods to enhance the system's performance and robustness across various application scenarios.

**Author Contributions** SS: conceptualization, editing. YL: writing, editing. WW: validation. YL: writing, editing. CC: validation. XJ: supervision. ZD: supervision. LL: supervision, editing.

**Funding** This work is supported by the National Natural Science Foundation of China (No. 61802285), the Scientific Research Project of Hubei Provincial Department of Education (No. D20201704), Hubei Province Technical Innovation Special Project (No. 2019AAA005), Philosophy and Social Science Research Project of Hubei Province (No. 22ZD083), Municipal Science and Technology Bureau Support

Special Project (No. 2022013988065214), School-Land cooperation Project (No. 20220602), and Open Subject of Wuhan Textile and Apparel Digital Engineering Technology Research Center (No. 0100000).

**Data Availability** Data will be made available on request.

## Declarations

**Conflict of Interest** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alzubaidi, L., et al.: Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **8**(1), 53 (2021)
- Wang, W., Deng, N., Xin, B.: Sequential detection of image defects for patterned fabrics. *IEEE Access* **8**, 174751–174762 (2020)
- Chan, T.F., Shen, J.: Nontexture inpainting by curvature-driven diffusions. *J. Vis. Commun. Image Represent.* **12**(4), 436–449 (2001)
- Chan, T.F., Kang, S.H., Shen, J.: Euler's elastica and curvature-based inpainting. *SIAM J. Appl. Math.* **63**(2), 564–592 (2002)
- Fang, Y., Li, Y., Tu, X., et al.: Face completion with Hybrid Dilated Convolution. *Signal Process.: Image Commun.* **80**, 115664 (2020)
- Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *J. Sci.* **313**(5786), 504–507 (2006)
- Deldjoo, Y., et al.: Content-based video recommendation system based on stylistic visual features. *J. Data Semant.* **5**(2), 99–113 (2016)
- Liu, W., et al.: A survey of deep neural network architectures and their applications. *Neurocomputing* **234**, 11–26 (2017)
- Alom, M.Z., et al.: A state-of-the-art survey on deep learning theory and architectures. *Electronics* **8**(3), 292 (2019)
- Wu, X., et al.: Recent advances in deep learning for object detection. *Neurocomputing* **396**, 39–64 (2020)
- Sarraf, A., et al.: A comprehensive review of deep learning architectures for computer vision applications. *Am. Sci. Res. J. Eng. Technol. Sci.* **77**(1), 1–29 (2021)
- Xiao, Y., et al.: A review of object detection based on deep learning. *Multimed Tools Appl.* **79**(33), 23729–23791 (2020)
- Shrestha, A., Mahmood, A.: Review of deep learning algorithms and architectures. *IEEE Access.* **7**, 53040–53065 (2019)
- Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. *Commun. ACM* **60**(6), 84–90 (2017)
- Yang, W., et al.: Deep learning for single image super-resolution: a brief review. *IEEE Trans. Multimed.* **21**(12), 3106–3121 (2019)
- Zhang, K., et al.: Deep unfolding network for image super-resolution. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 3217–3226 (2020)
- Rouse, D. M., Hemami, S.S.: Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM. *Human Vision and Electronic Imaging XIII.* 680615 (2008)
- Criminisi, A., et al.: Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.* **13**(9), 1200–1212 (2004)
- Goodfellow, I.J., et al.: *Generative Adversarial Nets.* MIT Press (2014)
- Lahiri, A. et al.: Prior guided gan based semantic inpainting. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 13696–13705 (2020)
- Ronneberger, O., et al.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds.) *Medical image computing and computer-assisted intervention—MICCAI 2015.* 9351. Springer, Cham (2015)
- Zhao, L., et al.: Uctgan: Diverse image inpainting based on unsupervised cross-space translation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 5741–5750 (2020)
- Pathak, D., et al.: Context encoders: feature learning by inpainting. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2536–2544 (2016)
- Leene, J.E.: Restoration and preservation of ancient textiles, and natural science. *Stud. Conserv.* **6**(sup1), 190–191 (2014)
- Zhou, Z., et al.: Classification of clothing images based on a parallel convolutional neural network and random vector functional link optimized by the grasshopper optimization algorithm. *Textile Res. J.* **92**(9–10), 1415–1428 (2021)
- Dorozynski, M., Rottensteiner, F.: Deep descriptor learning with auxiliary classification loss for retrieving images of silk fabrics in the context of preserving European Silk Heritage. *ISPRS Int. J. Geo Inf.* **11**(2), 82 (2022)
- Qian, W., et al.: Aesthetic art simulation for embroidery style. *Multimed. Tools Appl.* **78**(1), 995–1016 (2019)
- Zhang, Z., et al.: Research on big data analysis technology of Chinese traditional culture Yue embroidery color network. *J. Phys. Conf. Ser.* **1345**(2), 022021 (2019)
- Liu, K., Wu, H., et al.: Archaeology and restoration of costumes in tang tomb murals based on reverse engineering and human-computer interaction technology. *Sustainability* **14**(10), 6232 (2022)
- Liu, K., Zhao, J., et al.: Research on digital restoration of plain unlined silk gauze gown of Mawangdui Han Dynasty Tomb based on AHP and human-computer interaction technology. *Sustainability* **14**(14), 8713 (2022)
- Dhillon, A., Verma, G.K.: Convolutional neural network: a review of models, methodologies and applications to object detection. *Prog Artif Intell.* **9**(2), 85–112 (2020)
- Marti, A., WordNet.: An electronic lexical database. *Automated Discovery of Wordnet Relations,* 5 (1998)
- Liu, G., et al.: Image inpainting for irregular holes using partial convolutions. In: *Proceedings of the European conference on computer vision (ECCV).* 85–100 (2018)
- Cybulska, M.: To see the unseen. *Computer graphics in visualisation and reconstruction of archaeological and historical textiles.* 17 (2012)
- Yang, F, et al.: Learning texture transformer network for image super-resolution. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* 5791–5800 (2020)
- Nair, V., Hinton, G. E.: Rectified linear units improve restricted Boltzmann machines Vinod Nair. In: *International Conference on Machine Learning.* 807–814 (2010)
- Yu, J., et al.: Free-form image inpainting with gated convolution. *IEEE/CVF International Conference on Computer Vision (ICCV),* Seoul, Korea (South), 4470–4479 (2019)

38. Kheira, L., Nagham, S.: A new vision of a simple 1D Convolutional Neural Networks (1D-CNN) with Leaky-ReLU function for ECG abnormalities classification. *Intell.-Based Med.* **6**, 100080 (2022)
39. Zeiler, M.D., et al.: Adaptive deconvolutional networks for mid and high level feature learning. In: *International Conference on Computer Vision*. Barcelona, Spain, 2018–2025 (2011)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.