Heritage Science

RESEARCH

Open Access

ConvSRGAN: super-resolution inpainting of traditional Chinese paintings

Qiyao Hu[1,3†], Xianlin Peng[2,4†], Tengfei Li[1], Xiang Zhang[1*], Jiangpeng Wang[2] and Jinye Peng[1,5]

Abstract

Existing image super-resolution methods have made remarkable advancements in enhancing the visual quality of real-world images. However, when it comes to restoring Chinese paintings, these methods encounter unique challenges. This is primarily due to the difficulty in preserving intricate non-realistic details and capturing comple semantic information with high dimensionality. Moreover, the preservation of the original artwork's distinct style and subtle artistic nuances further amplifies this complexity. To address these challenges and effectively restore traditional Chinese paintings, we propose a Convolutional Super-Resolution Generative Adversarial Network for Chinese landscape painting super-resolution, termed ConvSRGAN. We employ Enhanced Adaptive Residual Module to delve deeply into multi-scale feature extraction in images, incorporating an Enhanced High-Frequency Retention Module that leverages an Adaptive Deep Convolution Block to capture fine-grained high-frequency details across multiple levels. By combining the Multi-Scale Structural Similarity loss with conventional losses, our ConvSRGAN ensures that the model produces outputs with improved fidelity to the original image's texture and structure. Experimental validation demonstrates significant qualitative and quantitative results when processing traditional paintings and murals datasets, particularly excelling in high-definition reconstruction tasks for landscape paintings. The reconstruction effect showcases enhanced visual fidelity and liveliness, thus affirming the effectiveness and applicability of our approach in cultural heritage preservation and restoration.

Keywords Cultural relic image restoration, Traditional Chinese paintings, Image super-resolution, Deep learning, MS-SSIM loss

Introduction

The cultural significance of traditional Chinese paintings is undeniable. They serve as a testament to the changing times and are imbued with rich cultural meanings. However, the passage of time and natural factors have often led to damage and deterioration of these valuable works of art. As a result, it is crucial to take measures to protect and restore them. Currently, some restoration experts attempt to restore these paintings through manual drawing, but the varying personal drawing styles of the restorers make it challenging to reproduce the authentic brushstrokes and style of the original artwork.

With the continuous improvement of deep learning in image processing, image super-resolution techniques have been used to reconstruct high-resolution details from blurry natural images. Especially in cases where

†Qiyao Hu and Xianlin Peng have contribute equally to this work.

*Correspondence:
Xiang Zhang
xiangz@nwu.edu.cn
[1] School of Information Science and Technology, Northwest University, Xi'an 710127, Shannxi, China
[2] School of Art, Northwest University, Xi'an 710127, China
[3] State-Province Joint Engineering and Research Center of Advanced Networking and Intelligent Information Services, Xi'an 710127, China
[4] Shaanxi Key Laboratory of Higher Education Institution of Generative Artificial Intelligence and Mixed Reality, Xi'an 710127, China
[5] Shaanxi Silk Road Cultural Heritage Digital Protection and Inheritance Collaborative Innovation Center, Xi'an 710127, China

direct physical restoration of the original artwork is not feasible, utilizing non-contact restoration methods like image super-resolution techniques becomes particularly important. This technology enables a more precise restoration of the details in ancient traditional paintings, contributing to the protection and inheritance of China's precious cultural heritage. For example, Xiao et al. [1] construct a global correlation graph based on pre-extracted features from all samples and utilize a graph convolutional neural network with maximum mean discrepancy loss to approximate the feature distributions in two domains. This model effectively preserves the structural information between samples. A large number of single-image super-resolution (SISR) approaches heavily rely on supervised learning. However, to reduce dependency on supervised learning, Prajapati et al. [2] achieve certain results by employing unsupervised learning in a GAN framework. They also introduce a new loss function based on Mean Opinion Score (MOS) to evaluate the quality of the generated images. EaSRGAN [3] improves upon SRGAN by incorporating multi-stage training for the generator and discriminator, with a focus on edge and flat region enhancement. This approach pays attention to the perceptual edge information, resulting in fewer artifacts and higher image quality. Zhao et al. [4] propose a multi-level semantic progressive restoration approach for painting images. This method gradually shifts attention from high-level and large-scale information to increasingly fine scales, yielding better results compared to other one-step restoration methods. Although existing methods have achieved excellent performance in super-resolution of real-world images.

However, traditional Chinese paintings typically exhibit complex layout structures and abstract representations of objects and scenes. The challenge lies in the faithful reconstruction of the details of original artwork while maintaining its unique artistic style. In summary, the key difficulties encountered in restoring traditional Chinese paintings include:

1. Traditional Chinese paintings, including ink paintings and meticulous paintings, emphasize the variations in brushstrokes and lines, which deviate from the objective and realistic representation found in real-world images. The texture information embedded in these artworks encapsulates the distinctiveness of brushstrokes and the stylistic characteristics that define the artwork. Preserving their original forms during model inference is of paramount importance.
2. Chinese traditional paintings encompass a multitude of abstract elements and symbolism. Modeling such irregular and highly abstract content presents a significant challenge. The super-resolution process may introduce distortions or deviations from the original artistic style, further complicating the reconstruction process.
3. The development of traditional Chinese painting reconstruction techniques is hindered by the limited availability of datasets that align with high-resolution and low-resolution traditional paintings. This scarcity of data inhibits progress in this field.

To overcome the above limitations, we introduce a novel method to facilitate the super-resolution of traditional Chinese painting images, termed ConvSRGAN. Our contribution is threefold:

- We propose a novel dataset specifically for Chinese landscape painting super-resolution, termed SRCLP. It facilitate further research and exploration in this field by providing an extensive and high-quality dataset. The dataset is available at https://github.com/LPDLG/SRCLP-Dataset
- We propose an EARM to extract high-level abstract features. Additionally, to address the issue of lost contour information in painting images, we design an EHRM within the EARM to enhance the edges and textures of different-level feature maps. Furthermore, we introduce an ADCB in the EHRM to model large-scale spatial dependencies, allowing the model to better understand and reproduce the global layout and brushstroke trends in traditional Chinese paintings.
- We introduce a combination of MS-SSIM loss ($\mathcal{L}_{M-S}$) and traditional loss weighting, which pay more attention to contours and pixel differences at various scales and suppressing color and brightness distortions.

Related work

Traditional Chinese painting restoration

Recently, significant advancements have been made in Chinese painting style transfer [5, 6], poetry-to-image [7], image-to-image translation [8], and traditional Chinese painting image generation techniques [9]. These developments have provided valuable support for the preservation and continuation of traditional art, while also paving the way for new opportunities in digital art evolution. For example, SAPGAN (Sketch-And-Paint GAN) [10] first employs SketchGAN to generate sketches of landscape paintings and then uses PaintGAN to transform the sketches into Chinese landscape paintings. Zhang et al. [11] propose a generative adversarial network-based model for automatically generating Chinese landscape paintings with styles closely resembling traditional

Chinese paintings, which plays a significant role in the preservation of digital cultural heritage.

At the same time, there are also some research methods focusing on the restoration and super-resolution of traditional Chinese landscape paintings. Shi et al. [12] build the Ref-ZSSR network based on generative adversarial networks (GAN) to extract and apply the global information of images from the painting itself, successfully achieving restoration of damaged ancient paintings. Nagar et al. [13] apply diffusion models to the artistic restoration of mural images, effectively addressing various degradation issues such as noise, blur, and fading. It is worth mentioning that Lyu et al. [14] apply the diffusion model to Chinese landscape paintings, propose CLDiff, and introduce an attention mechanism. This approach achieves good performance in super-resolution tasks of traditional Chinese landscape paintings, providing high-resolution results with clear ink texture.

Although there has been some headway in utilizing deep learning for traditional Chinese landscape painting, research that specifically concentrates on super-resolution techniques for traditional Chinese paintings remains limited. Traditional Chinese painting is renowned for its unique artistic language, techniques, and aesthetic characteristics, including ink charm, vividness of artistic conception, and non-realistic composition. These attributes present new challenges for existing general image super-resolution algorithms. The process of super-resolution on landscape paintings can accurately restore the artistic effects of traditional Chinese painting, which is of great significance for the preservation of digital cultural heritage. Therefore, we will conduct comprehensive research on the super-resolution of Chinese landscape paintings.

Single image super-resolution

Since the introduction of the CNN-based super-resolution algorithm by Dong [15], deep learning methods have gained significant popularity in addressing image super-resolution tasks, leading to groundbreaking advancements in this domain. Ledig et al. [16] use a generative adversarial approach to train their network and define a content loss function, achieving superior results compared to traditional methods. After that, ESRGAN [17] builds upon this work by stacking multiple dense blocks to restore HR images. It also introduces the concept of perceptual loss to reconstruct images that closely resemble human perception. While dense connections help with feature reuse, the challenge lies in the complexity of training the model and the requirement for a large amount of high-quality HR-LR data for supervised training. Real-ESRGAN [18] utilizes a high-order degradation model to simulate complex degradation distributions in images, allowing for the generation of paired HR and low-resolution LR images. However, when dealing with more complex textures and details, there may be instances of distortion and blurring.

Although CNN-based methods have achieved success in super-resolution tasks, they still face inherent limitations, and learning long-range dependencies in images has always been a key issue in the field of computer vision. Since the introduction of Vision Transformers [19], many visual tasks [20–22] have demonstrated the excellent performance of Transformers in addressing this issue. The SwinIR [23] model leverages the power of the Swin Transformer [24] to enhance image super-resolution. By utilizing local attention and window-based shifts, the model gains a better understanding of the overall image structure, enabling it to effectively capture long-range dependencies and improve its performance in super-resolution tasks. ESRT [25] combines the strengths of the CNN and Transformer architectures, using a CNN network to learn deep image features and introducing an efficient multi-head attention mechanism called EMSH in the Transformer to capture dependencies between similar tokens. This approach reduces network parameters while improving feature representation. DAT [26] achieves feature aggregation and captures global contextual information by alternating between spatial and channel self-attention mechanisms within consecutive Transformer blocks. This approach aims to enhance the quality of super-resolution images. Despite the enhanced capability of Transformer-based super-resolution models in capturing global dependency information in images, the significant computational complexity arising from the Hadamard product of self-attention matrices presents considerable challenges, particularly when dealing with large-scale high-resolution artistic images. To circumvent the complex operations in Transformers, we propose a novel module that achieves similar effects to attention mechanisms. This module enables learning of both image semantics and local textures in artistic images while avoiding complex calculations.

These advancements have improved the super-resolution quality of real-world images to varying extents. However, traditional Chinese paintings, which are not real but rather contain complex layout structures and element arrangements, involve intricate texture details within mountains, rocks, and vegetation that require progressive learning. To this end, we concatenate EARM to progressively extract high-level information. Additionally, to address the differences in structure and texture information caused by scale variations, we introduce the $\mathcal{L}_{M-S}$ in the combining function to improve the quality of super-resolution reconstruction.

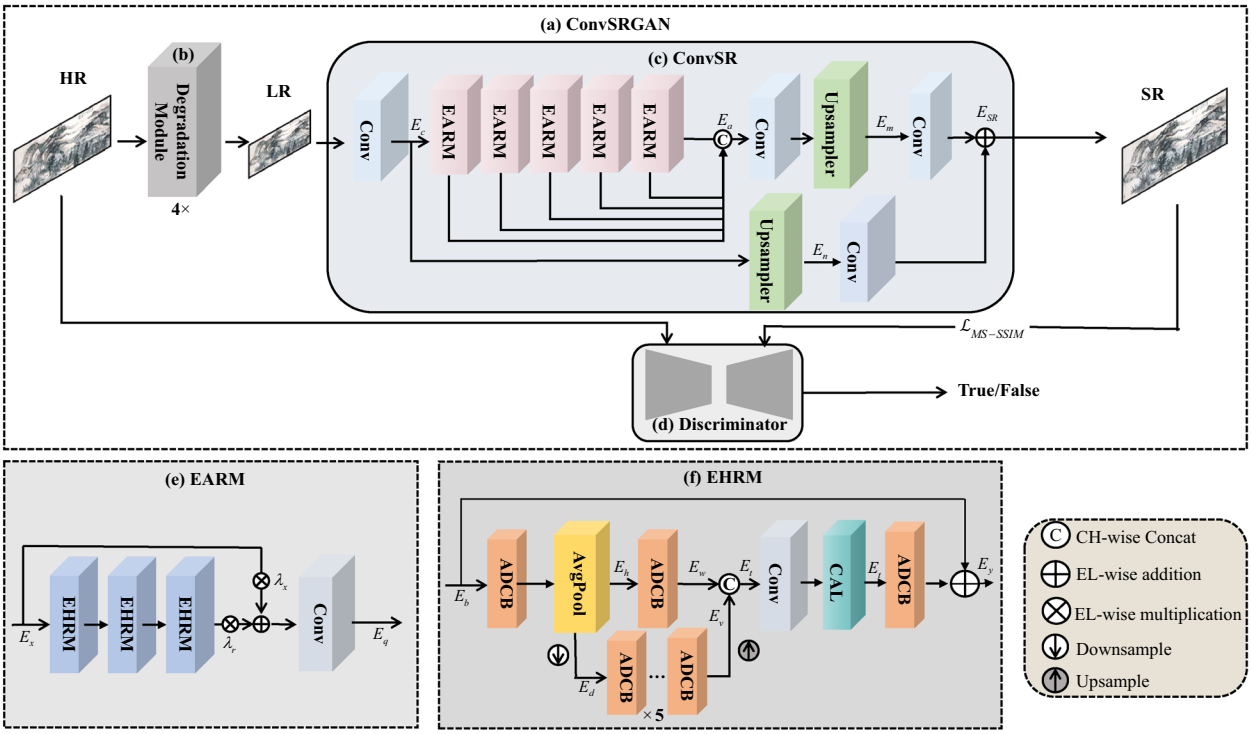

Fig. 1 Overview of our framework. **a** ConvSRGAN network. **b** Image degradation Module. **c** ConvSR network. **d** Discriminator. **e** Enhanced Adaptive Residual Module. **f** Enhanced High-frequency Retention Module

Vision expend

While Transformers have shown remarkable ability in capturing long-distance dependencies in visual tasks, the high computational requirements and intricate inference procedures have incentivized researchers to incorporate larger convolutional kernels in convolutional operations. It aims to expand the model's field of view, promoting the acquisition of improved contextual information. ConvNeXt [27] introduces a pure ConvNet model that achieves comparable performance to the Swin Transformer by optimizing training strategies and utilizing large-sized convolutional kernels. This demonstrates the effectiveness of large-kernel convolutions. RepLKNet [28] utilizes reparameterized depthwise convolutions to design high-performance large-kernel CNNs. By employing a 31 × 31 oversized convolutional kernel, it achieves improved results in various typical downstream tasks while maintaining lower latency. VapSR [29] utilizes depth-wise separable convolutions instead of dense connection layers and implements pixel-level attention allocation with large convolutional kernels in the attention branch. This approach aims to enhance the resolution of generated images. LKDN [30] builds upon the BSRN [31] baseline structure and introduces a more efficient large-kernel attention (LKN) module to learn global image features and improve image clarity. It also reduces computational costs by distilling networks through an analysis of the computational efficiency of both BSRN and VapSR.

Inspired by these works, we consider replacing the Transformer architecture with large-kernel convolutions to avoid complex attention operations while capturing more feature information. To this end, we combine the characteristics of large-kernel convolutions and depth-wise separable convolutions to design the ADCB. This block provides the network with more contextual information, allowing for the preservation of the coherence of the painting structure while also finely retaining the artistic style and unique techniques of the artwork.

Methodology

Overall structure

As shown in Fig. 1, the overview of our model. (a) ConvSRGAN network: It comprises shallow feature extraction, deep feature extraction, and feature fusion. (b) Image degradation module: Our model incorporates a module specifically designed to emulate the degradation effects typically encountered in real-world landscape paintings. (c) ConvSR network: The core architecture of our model revolves around a streamlined process that begins with receiving input data, progresses through a sophisticated feature extraction phase, and ultimately leads to a

refined feature super-resolution. (d) Discriminator: It is trained to distinguish between the synthesized images and genuine samples. By iteratively refining this process, the generator learns to produce images that increasingly resemble authentic instances, thus enhancing the super-resolution output. (e) Enhanced Adaptive Residual Module (EARM): It adjust the residual connections and main path weights to automatically select and retain the relevant feature information. (f) Enhanced High-frequency Retention Module (EHRM): It preserves high-frequency information at the current resolution by adding average pooling layers.

The ConvSRGAN begins by taking high-resolution painted images (HR) as input. It simulates the degradation distribution of the real world through a second-order degradation module, resulting in low-resolution images (LR) in $4\times$ resolution.

Then, low-resolution images are further restored in ConvSR, where they are mapping to the feature space through a convolutional layer. The resulting features are separated into deep and shallow features, with the shallow features preserving the overall structure and arrangement of the painted images, while the deep features restore the high-level texture details.

Moreover, the preliminary features are dynamically extracted using the concatenated EARM to capture different depths of image texture features. The initial features are also processed through convolutional layers to obtain shallow features, which are then combined with the deep features through long skip-connection to generate a high-resolution painted image.

In addition, after generating the super-resolution image, it is compared to the original high-resolution image to calculate loss fuction using the discriminator network. We use a U-Net discriminator with spectral normalization to provide more accurate gradient feedback for local textures.

And finally, $\mathcal{L}_{M-S}$ is introduced to account for the characteristics of landscape painting images. We use four loss functions, which is Adversarial loss, Perceptual loss, MS-SSIM Loss and L1 loss to minimizes the feature vector distances between the super-resolution image and the original high-resolution image. The network parameters are updated through gradient back-propagation. Detailed explanations of each component of the network will be provided in the following sections.

Degradation module

Traditional paintings undergo various forms of deterioration over time, including climate erosion, pigment fading, and other factors, which ultimately lead to the degradation of the image quality, resulting in blurry and distorted representations. Meanwhile, when digitally preserving these images, the use of different storage methods and sharpening techniques often introduces undesirable artifacts. Basic techniques such as Blur, Resize, Noise, JPEG compress prove inadequate in accurately simulating the intricate degradation patterns observed in traditional painted images. Consequently, a substantial disparity exists between artificially synthesized low-resolution images (LR) and genuine degraded images.

Inspired by Real-ESRGAN [18], we have constructed a model that simulates the actual degradation process. As shown in Fig. 1b, the model can reflect various complex degradation phenomena that may occur in traditional paintings after long-term preservation, including but not limited to color loss, texture blurring, and structural distortion. The first-order degradation is contained four operations of degradation process, which are Blur, Resize, Noise, JPEG. The second-order degradation means secondary process of the first-order degradation. The formulation of this model is represented as Eq. (1):

$$\begin{aligned} &I_L = M^2(I_H) \\ &M \in \{Blur, Resize, Noise, JPEG\} \end{aligned} \tag{1}$$

where M^2 represents second-order degradation. The first-order degradation is contained four operations of degradation process. I_H and I_L respectively represent the images before and after degradation process.

By using degradation module, the generated degraded images are closer to real-world traditional paintings that have been damaged over time. This helps to bridge the significant gap between simulated images generated solely based on basic degradation techniques and actual degraded images.

ConvSR network

The ConvSR network comprises shallow feature extraction, deep feature extraction, and feature fusion. Given the degraded low-resolution image (LR) as input, a convolution layer with the kernel size of 3×3 is used to extract the structure features E_c. Simultaneously, the low-resolution image is mapping to the feature space. This process can be formulated as:

$$E_c = \Phi_c(I_L) \tag{2}$$

where $\Phi_c(\cdot)$ represents the process of the first convolutional layer.

To capture the distinctive traits of Chinese traditional painting, including brushstroke techniques, composition structure, and element arrangement, ConvSR performs two separate branchs on the input shallow features.

As shown in Fig. 1c, one branch focuses on texture feature extraction. The other branch involves convolutional operations to preserve the structural feature. Then, the

two separate branches perform the corresponding element-wise addition during the fusion process. Specifically, we extract features at different depths by adjusting the number of proposed EARM.

By combining the output of different EARM at different levels of image features, we obtain the deep features E_d.

$$E_i = \xi_i(E_{i-1}), \quad i = 1, 2, 3, 4, 5 \tag{3}$$

$$E_a = concat(E_1, E_2, E_3, E_4, E_5) \tag{4}$$

where i denotes the number of EARM. $\xi_i(\cdot)$ denotes the operation of EARM. E_a denotes the output features in terms of the fusion of Modules of EARM with different depths.

As for the deep features E_a, they are first passed through a single convolutional layer to reduce the number of channels and then upsampled. The upsampled deep features are integrated using another convolutional layer to capture information from different depths. This process is defined as follows:

$$E_m = Upsample(\Phi(E_a)) \tag{5}$$

where $\Phi(\cdot)$ denotes the convolution layer and $Upsample(\cdot)$ denotes the Upsample layer.

To preserve the structure and layout features of the painting, the other branch directly upsamples the shallow feature E_s to the original size.

$$E_n = Upsample(E_c) \tag{6}$$

Finally, the deep features and shallow features are merged to obtain the super-resolution image. The feature fusion is through element-wise addition process.

$$E_{SR} = \Phi(E_m) + \Phi(E_n) \tag{7}$$

where E_{SR} denotes the feature matrix of super-resolution image.

It can be concluded that during the learning of image features in the ConvSR, the shallow structural features obtained through a single convolutional layer preserve the overall layout and shape of the landscape painting images. On the other hand, processing the deep features can enhance the details and texture of the image, resulting in a super-resolution image that is clearer and more realistic.

EARM

As shown in Fig. 1e, EARM has three EHRM in the main path and utilizes skip connections to learn residual information from the input. In this process, we dynamically adapt and adjust the residual connections and main path weights to automatically select and retain the relevant feature information. This allows the model to flexibly adjust and adapt to different painting styles, improving its ability to handle various styles of traditional painted images and enhancing its generalization capability.

By incorporating the EARM into the model, we ensure that it not only enhances the quality of the artwork at the pixel level but also captures and conveys the artistic spirit and aesthetic mood of the original piece. Our objective is to preserve the artistic essence of Chinese traditional painting throughout the digital processing, allowing the model to reflect the true artistic essence inherent in traditional artwork. This process can be represented as follows:

$$E_q = \Phi\left(\lambda_r \cdot \delta^3(E_x) + \lambda_x \cdot E_x\right) \tag{8}$$

where $\delta^3(\cdot)$ denotes the operation through three EHRM. λ_r and λ_x are the adaptive weights of the two paths respectively. Finally, we use a convolution layer to adjust the output dimension.

EHRM

High-frequency information refers to the details and textures in an image that has high variation frequencies, such as leaves, petals, and mountain folds. Depicting these details and textures is essential for representing the imagery in a painting. To extract subtle features in painting images, such as textures, lines, and shadows, we propose the Enhanced High-frequency Retention Module, termed EHRM.

It preserves high-frequency information at the current resolution by adding average pooling layers. Specifically, we introduce the Adaptive Deep Convolutional Block, termed ADCB. Increasing the size of the convolution kernel improves the limitations of traditional CNNs in terms of their field of view and enables the extraction of more contextual information.

As shown in Fig. 1f, it is assumed that the feature of the input EHRM is E_b, the first ADCB extracts features that serve as input to the high-pass filter. The high-pass filter calculates the high-frequency information of these features, denotes E_h.

$$E_h = A_v(\kappa_a(E_b)) \tag{9}$$

where κ_a is the operation of the ADCB. A_v denotes the Average Pooling layer.

After obtaining E_h, we decrease the size of the feature map to reduce computational cost and feature redundancy. The downsampled feature map is represented as E_d.

$$E_d = Downsample(E_h) \tag{10}$$

We utilize five ADCB to explore its potential information, with weight sharing to reduce parameters. Simultaneously, an ADCB is used in the feature space to align E_h with E_d, yielding E_w.

$$E_w = \kappa_a(E_h) \tag{11}$$

$$E_u = \kappa_a^5(E_d) \tag{12}$$

After feature extraction, E_u is upsampled to the original size using bilinear interpolation.

$$E_v = Upsample(E_u) \tag{13}$$

Then, we concat E_v and E_w to retain the original details and obtain the feature E_t. This operation can be represented as follows:

$$E_t = concat(E_w, E_v) \tag{14}$$

To strike a satisfactory balance between model complexity and performance, we use five ADCBs in our experiments, which are denoted as κ_a^5.

To extract more image features, E_t is input into the Channel Attention Layer (CAL) Module after convolution operation.

$$E_r = \omega(\Phi(E_t)) \tag{15}$$

where $\omega(\cdot)$ denotes the operation of the CAL Module.

Finally, in order to maintain the shallow features of the image and further extract the deep features, E_r is added with E_b after ADCB module to obtain the output result of EHRM.

$$E_y = E_b + \kappa_a(E_r) \tag{16}$$

ADCB

As shown in Fig. 2, Adaptive Deep Convolutional Block (ADCB) is composed of Deep Residual Block (DRB) and Channel Attention Layer (CAL) Module, which can better capture the complex features of traditional painted images.

Deep Residual Block. It is composed of a DepthWise Convolutional (DW Conv) layer and two Pointwise convolutional (PW Conv) layers, and will perform element-wise addition operations on the features before and after these three convolutions.

Inspired by ConvNeXt [27], we use a 7×7 kernel size in the DRB instead of the more common 3×3 size to provide a larger receptive field and achieve effects similar to non-local attention mechanisms. This design helps the model learn the spatial depth created by distance and shading in traditional paintings, as well as understand the layout and spatial structure of the artwork (Fig. 4).

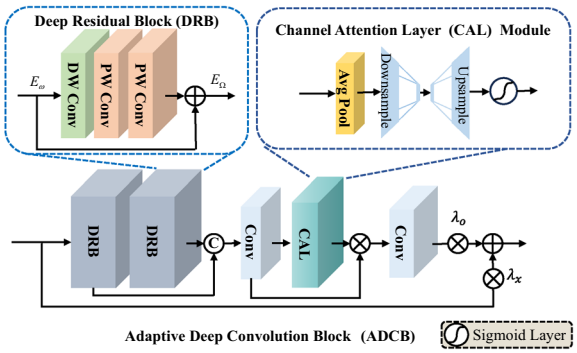

Fig. 2 Overview of Adaptive Deep Convolution Block. We employ a 1×1convolutional layer to reduce the number of its channels. The Channel Attention Layer Module is used to highlight channels with high activation values

To balance the increase in model parameters caused by large convolution kernel, we use Depth Separable convolution. Depth Separable convolution consists of a DW Conv layer and a PW Conv layer, which are used in DRB to extract input features, thus reducing the complexity of the model.

E_ω represents the output feature of the first DRB. The process of the second DRB can be represented as:

$$E_k = \Phi_d(E_\omega) \tag{17}$$

$$E_z = \Phi_p(E_k) \tag{18}$$

$$E_\Omega = \Phi_p(E_z) + E_\omega \tag{19}$$

where E_Ω represents the output feature of the second DRB. Φ_p represents pointwise convolution. Φ_d represents depthwise convolution.

Channel Attention Layer Module. It is composed of average pooling layer, downscaling layer, upscaling layer, and activation via sigmoid fuction. In addition, Element-wise multiplacation is performed on the features before and after the CAL module. CAL Module is used to learn crucial channel information, focusing on important feature information in the input image, and enhancing the expression of feature information to improve the accuracy of image super-resolution results.

Then, the output of the two DRB are connected through a convolutional layer to reduce the number of channels. Finally, the weights of different paths are adjusted in an adaptive manner to better utilize hierarchical features.

Discriminator details

Considering the high-order degradation models used in our study, especially Real-ESRGAN [18], the degradation space becomes extensive and intricate. As shown in Fig. 1d, we use a U-Net discriminator with spectral normalization, which has stronger discriminative power and can provide more accurate gradient feedback for local textures. At the same time, the use of spectral normalization not only helps to reduce artifacts and oversharpening issues in GAN training but also makes the training process more stable.

Furthermore, the $\mathcal{L}_a$ is iteratively refined through a comparative computation of the super-resolution (SR) images produced by the ConvSR network with their corresponding high-resolution (HR) original images. This iterative optimization process serves to elevate the discriminative capacity of the model, thereby enhancing the authenticity of the images generated by the generator, ensuring a heightened level of fidelity in the super-resolution process.

Loss function

Traditional paintings often have rich colors and gradients, and traditional loss functions cannot capture the pre-processing steps of low-pass filtering and color space conversion that simulate the human visual system. They also fail to capture the visual perception and artistic style of restoring the original image. When dealing with artistic works such as landscape paintings that are rich in details, layers, and expressive concepts, using a combination of losses helps improve the performance of the network. Specifically, we consider that traditional $\mathcal{L}_1$ distance provides pixel-level differences, the perceptual loss based on the VGG network improves the visual effect of the super-resolution image. Additionally, we introduce the $\mathcal{L}_{M-S}$ to capture structural information at different resolution levels, thus better preserving the overall structure and layout of the original image.

During the training of the ConvSR network, we use the $\mathcal{L}_1$ and $\mathcal{L}_{M-S}$ loss functions for training. During the training of the ConvSRGAN, in addition to the previous two loss functions, we also use the $\mathcal{L}_a$ and $\mathcal{L}_p$ loss functions for training in the discriminator.

MS-SSIM Loss. We introduce $\mathcal{L}_{M-S}$, which can be represented by the specific calculation formula as:

$$\mathcal{L}_{M-S} = 1 - \prod_{m=1}^{M} \left(\frac{2\mu_S\mu_H + c_1}{\mu_S^2 + \mu_H^2 + c_1} \right)^{\beta_m} \left(\frac{2Cov(S,H) + c_2}{\sigma_S^2 + \sigma_H^2 + c_2} \right)^{\gamma_m} \tag{20}$$

where M represents different scales. μ, σ represent mean, standard deviation, respectively. $Cov(\cdot)$ represent covariance operation. β_m and γ_m denote the relative importance of two items.

Adversarial loss We use adversarial loss to perform adversarial training between the super-resolution results generated by the generator and the original HR images, in order to optimize the generation performance of the generator. The adversarial loss can be represented as follows:

$$\mathcal{L}_a = \sum_{n=1}^{N} -\log D_{\tau_d}(G(I_L)) \tag{21}$$

where $D(\cdot)$ and $G(\cdot)$ respectively represent discriminator and generator. $D_{\tau_d}(G(\cdot))$ represents the probability that the super-resolution image matches the ground truth. We achieve better super-resolution performance by minimizing $\mathcal{L}_a$.

Perceptual loss Inspired by SRGAN [16], we use perceptual loss for training. Perceptual loss is defined as a weighted combination of content loss and adversarial loss:

$$\mathcal{L}_p = \mathcal{L}_{\text{vgg}} + 10^{-3}\mathcal{L}_a \tag{22}$$

where $\mathcal{L}_{vgg}$ is calculated based on a pre-trained VGG19 network [32]. $\mathcal{L}_a$ represents the adversarial loss.

Joint loss During the training of the ConvSRGAN network, using a combination of losses helps improve the performance of the network. To achieve better super-resolution results, we optimize the weight hyper parameters of each loss function during training.

$$\mathcal{L}_J = \mathcal{L}_p + \alpha\mathcal{L}_a + \beta\mathcal{L}_1 + \sigma\mathcal{L}_{M-S} \tag{23}$$

where α, β, σ are hyperparameters that balance the different loss terms, we set $\alpha = 0.1$, $\beta = 1.0$ and $\sigma = 1.0$.

Experiments

Datasets

In the experiments, we mainly train and test on the SRCLP dataset. We validate the generalization ability of our method on different datasets. Additionally, we also conduct testing on Mural, Painter By Numbers [33] and Flickr2K [34] datasets. For each dataset, it should be pointed out that we select only a subset of data for testing without participating in the training process.

SRCLP. In this paper, we construct a high-quality dataset called Super-Resolution Chinese Landscape Painting, termed **SRCLP**. All images in this dataset are sourced from collaborating institutions and digital art databases. We enlisted professional artists to meticulously classify the collected ancient paintings, considering different styles, dynasties, and color features to obtain diverse style information. The selected paintings have undergone

Table 1 Comparison results on ConvSR and CovnSRGAN

Method	GAN-Based			w/o GAN		
	PSNR ↑	SSIM ↑	LPIPS ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Bicubic	24.801	0.710	0.3622	–	–	–
RRDBNet [17]	–	–	–	28.722	0.819	0.3008
SRResNet [16]	27.318	0.772	0.2874	28.235	0.799	0.3189
EDSR [37]	27.349	0.768	0.2660	28.515	0.799	0.3106
ESRT [25]	27.258	0.780	0.2429	**28.977**	0.811	0.2919
Real-ESRGAN [18]	27.545	0.795	0.2549	–	–	–
BSRN [31]	26.301	0.769	0.3158	28.701	0.809	0.2947
LKDN [30]	25.106	0.750	0.3445	28.523	0.804	0.2928
VapSR [29]	26.529	0.771	0.2840	28.374	0.808	0.2962
ConvSR	–	–	–	28.916	**0.820**	**0.2914**
ConvSRGAN	**28.281**	**0.803**	**0.2334**	–	–	–

For these comparison models, we using our dataset to training separately. In addition, in order to make a fair comparison with ConvSR, we only train the generators of these comparison models. The output images of the generators are used for metrics computation. ↑ Higher values are better, ↓ Lower values are better

*Optimal results are displayed in bold, while suboptimal results are underlined

careful screening and categorization to ensure the diversity and representativeness of datasets.

The SRCLP dataset consists of 900 high-resolution Chinese traditional painting images. During the training phase, we applied techniques such as random rotation and cropping for data augmentation. This increased the number of images from the original 900 traditional landscape paintings to 2175 images. Figure 3 illustrates the composition of our original dataset, which encompasses a broad range of traditional paintings hailing from diverse historical dynasties and showcasing the unique styles of several esteemed artists. Among them, 2079 images were used for training, and a subset of 96 images was selected for testing purposes. Building upon

(a) **(b)** **(c)** **(d)**

Fig. 3 Examples of Traditional Chinese Paintings. **a** is a painting of *'Autumn Colors on the Wutong Tree'* in *Ming* dynasty, **b** is a painting of *'Joy in the Heart of Heaven'* by artist *Emperor Shunzhi* in *Qing* dynasty, **c** is a piece imitated by artist *Li Wu* in the *Qing* dynasty, **d** is a painting of *'Secluded Dwelling on Mount Shu'* by the modern and contemporary artist *Daqian Zhang*. These works of art range from expressive to realistic, each displaying unique artistic characteristics and technical expressions

Fig. 4 Examples of Mural datasets. From left to right, *Cave*, *Temple*, *Tomb*, and *Thangka*

the existing foundation, we have meticulously curated a collection of supplementary paintings originating from disparate dynasties and the portfolios of diverse artists. These thoughtfully chosen pieces have been incorporated into our test dataset, thereby augmenting its scope and diversity to encompass a broader chronological and stylistic spectrum.

Mural We proposed Mural dataset includes four different types of mural images: *Cave*, *Temple*, *Tomb*, and *Thangka*. We have collected this data from relevant websites and museums. These images originate from various locations and consist of over 1200 high-resolution images. For the experiment, we selected 99 of the most representative images from the dataset for testing purposes.

Painter By Numbers [33]. Painter By Numbers is a dataset sourced from the Google Arts & Culture project, consisting of over 100,000 images of artworks. These images encompass various art styles and painting techniques. Therefore, we selected a subset of images from different artistic styles for qualitative comparisons, as they were not specifically used for training or quantitative evaluation.

Flicker2K [34]. Flickr2K is a commonly used image super-resolution reconstruction dataset that is especially suitable for studying high quality (2K resolution) image restoration tasks. It contains 2650 high quality images in 2K resolution. The images cover different subjects, including people, animals, landscapes and more. Therefore, we carefully selected and used 2000 of these images to train and test on our ConvSRGAN.

Implementation details

In our experiments, the model implemented by PyTorch is trained on a NVIDIA GeForce RTX 4090 GPU. Before training, we applied a series of data augmentation techniques to improve the model's performance and robustness. This included resizing, cropping, rotating, mirroring, and adding noise to the images. The original painting images were resized to a unified resolution of 512×512, and the batch size was set to 2. To stabilize the model training and achieve better performance, we employed a weighted combination of multiple losses. The weights for $\mathcal{L}_1$, $\mathcal{L}_p$, $\mathcal{L}_{M-S}$, and $\mathcal{L}_a$ were set to 1, 1, 1, and 0.1, respectively.

In the experiment, the High-Resolution painting (HR) is first passed through a simulated degradation network to obtain the Low-Resolution painting (LR), which is then input into the network for training.

Training process We employ a two-stage training strategy. The training time is about 25 h.

In the first stage, the ConvSR network is trained to quickly converge and generate high-resolution painting (SR). The training is carried out for 200,000 iterations,

Table 2 Comparison results with kernel size, loss function and loss weight

Method	Model	PSNR ↑	SSIM ↑	LPIPS ↓
Kernel Size	ConvSRGAN w/ 3 × 3	28.020	0.800	0.2443
	ConvSRGAN w/ 5 × 5	27.949	0.799	0.2411
	*ConvSRGAN w/ 7 × 7	**28.281**	**0.803**	**0.2334**
Loss Function	ConvSRGAN w/ $\mathcal{L}_{GV}$	26.733	0.789	0.2464
	ConvSRGAN w/ $\mathcal{L}_{LDL}$	27.149	0.788	0.2475
	ConvSRGAN w/o $\mathcal{L}_{M-S}$	27.783	0.787	0.2412
	*ConvSRGAN w $\mathcal{L}_{M-S}$	**28.281**	**0.803**	**0.2334**
$\mathcal{L}_{M-S}$ Weight	ConvSRGAN w/ δ=2.0	28.002	0.799	0.2458
	ConvSRGAN w/ δ=0.5	28.133	0.803	0.2349
	*ConvSRGAN w/ δ=1.0	**28.281**	**0.803**	**0.2334**

* Represents our model, optimal results are displayed in bold

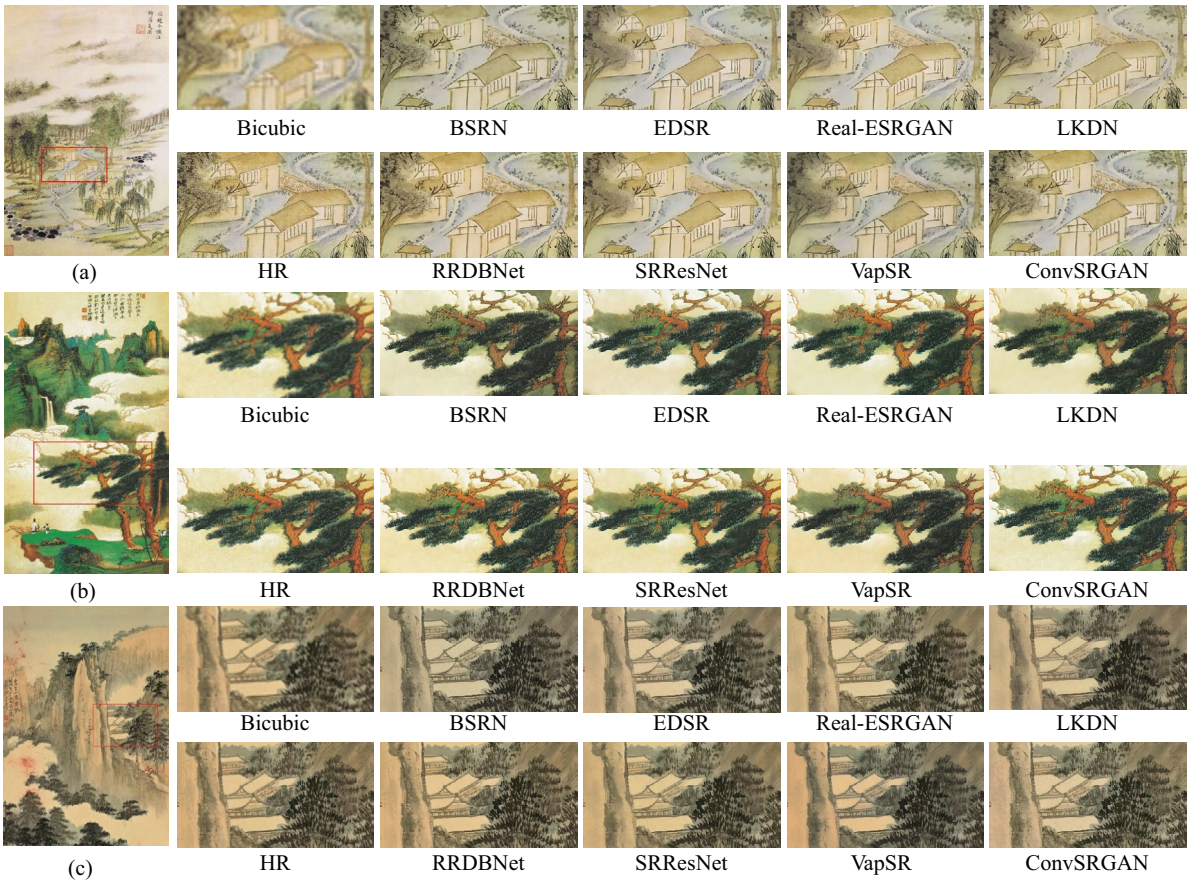

Fig. 5 Visual comparisons of ConvSRGAN. Both (**a**, **c**) are painted by *Shimin Wang*, a *Qing* Dynasty artist known for his delicate brushwork and the harmonious blend of form and spirit in his paintings. **b** is a painting of *'Leisurely Strolling with a Walking Stick'*. These landscape images are from our SRCLP dataset and are used for testing. Compared with the SOTA methods, the method we proposed preserves the form of brush strokes consistent with the original landscape painting and exhibits a more refined form of expression. Zoom in for best view

with an initial learning rate of 2×10^{-4}. The learning rate is decayed in multiple steps, with the learning rate halving at each step. The decay stages are set at [100,000, 150,000, 175,000].

In the second stage, the model trained in the first stage is used as a pre-trained model to train the ConvSRGAN network. We introduce a discriminator with the aim of calculating $\mathcal{L}_a$ and $\mathcal{L}_p$ between the high-resolution images SR generated in the first stage and the HR images. This helps to restore the original painting style features and provide more details, ultimately optimizing the generation performance of the ConvSR generator network. The initial learning rate is set to 1×10^{-4} during the second stage of training. Similarly, the learning rate is adjusted using multiple step decay, with the learning rate halving at each step. The decay stages are set at [100,000, 150,000].

Inference process Given a traditional painting LR image that is blurry due to weather erosion, the ConvSR network is used to output a restored and complete super-resolution image. In the inference process, we used 100 images to calculate the inference time, it took 110.18 s.

Evaluation metrics

To evaluate the quality of the SR images obtained by our method, we adopted three metrics for assessment, including: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [35], and Learned Perceptual Image Patch Similarity (LPIPS) [36]. Among them, the PSNR and SSIM metrics are used for evaluating the texture and structural integrity of the SR images, while the LPIPS metric is used for assessing the visual effects of the images.

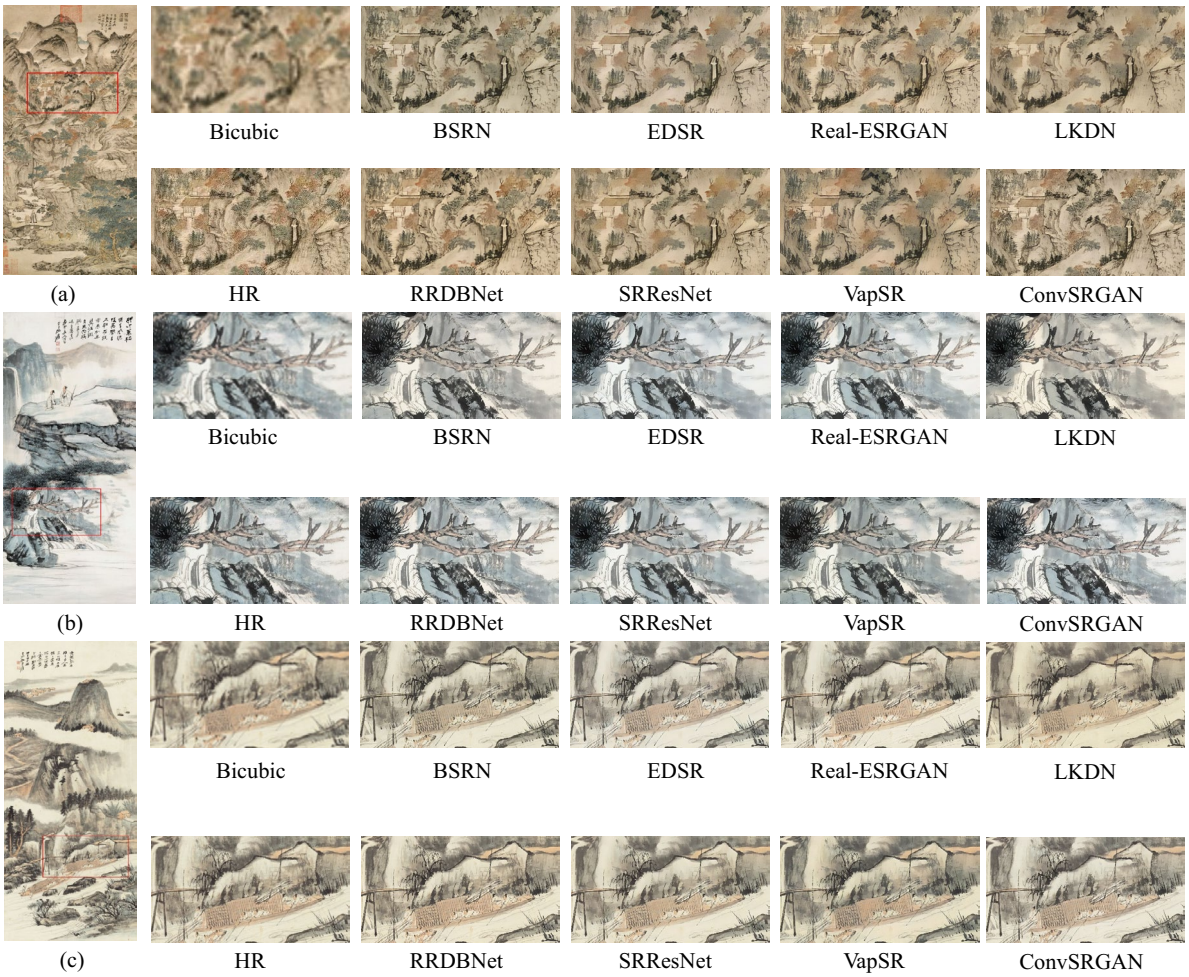

Fig. 6 Visual comparisons of ConvSRGAN. **a** is a painting of *The Relocation of Ge Zhichuan* from the Yuan Dynasty. **b** is a painting of *Strolling with a Walking Stick*, illustrating a waterfall cascading down into a lake at the foot of the mountain gorge. **c** is a painting of *Boating in a Rapid Stream* by *Daqian Zhang*. These landscape images are from our SRCLP dataset and are used for testing. Compared with other models, our method not only deals with the texture in landscape painting, but also retains the original color and painting style of the landscape. Zoom in for best view

Comparison with state-of-the-art method

To demonstrate the effectiveness of our proposed ConvSRGAN framework, we conducted quantitative and qualitative comparisons with several state-of-the-art methods on the SRCLP dataset we constructed.

Baselines

We conduct qualitative and quantitative comparisons with several baseline methods to demonstrate the effectiveness of the method.

1. **RRDBNet** [17]: It introduces the Residual-in-Residual Dense Block without batch normalization as the basic network building unit.
2. **SRResNet** [16]: It is the first framework capable of inferring photo-realistic natural images for 4× upscaling factors.
3. **EDSR** [37]: The significant performance improvement of this model is due to optimization by removing unnecessary modules in conventional residual networks.
4. **ESRT** [25]: It is a hybrid model, which consists of a Lightweight CNN Backbone and a Lightweight Transformer Backbone.
5. **Real-ESRGAN** [18]: A high-order degradation modeling process is introduced to better simulate complex real-world degradations.
6. **BSRN** [31]: It contains two efficient designs. One is the usage of blueprint separable convolution, which takes place of the redundant convolution operation.

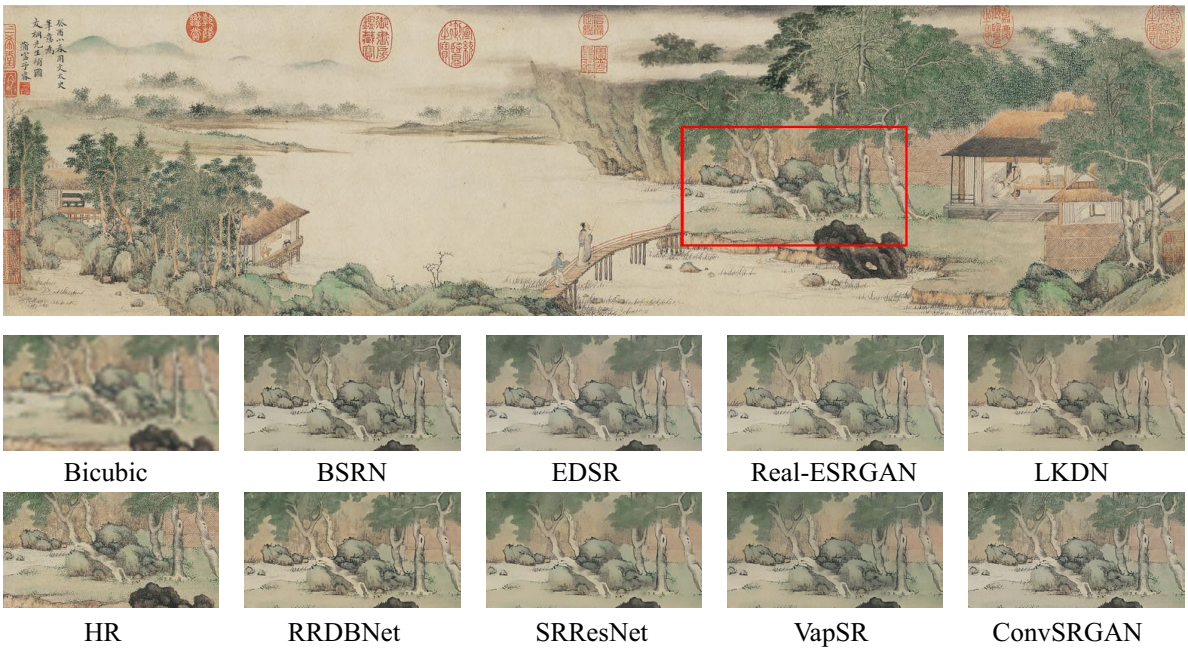

Fig. 7 Visual comparisons on an ancient painting. It is a painting of *Visiting a Friend with a Qin* by the *Qing* Dynasty painter *Rui Shang*. The painting showcases a stream flowing horizontally, with a secluded pavilion and elegant buildings on the opposite bank. The composition includes a bridge with rich brushstroke details. In this painting, our method has a superior performance in the processing of leaf texture compared to other methods. Zoom in for best view

The other is to enhance the model ability by introducing more effective attention modules.

7. **LKDN** [30]: It simplifies the model structure and introduces more efficient attention modules to reduce computational costs while also improving performance.
8. **VapSR** [29]: The large receptive field pixel attention mechanism is used with parameter reduction, pixel normalization, and intermediate attention conversion steps to enhance super-resolution performance in a lightweight manner.

Comparison analysis

We conducted quantitative comparisons ConvSR and ConvSRGAN with other SOTA methods. The quantitative results are shown in Table 1. Two versions of our model, ConvSR and ConvSRGAN, all achieved satisfactory results.

In addition, in order to make a fair comparison with ConvSR, we only train the generators of these comparison models. The output images of the generators are used for metrics computation and comparison.

In the comparative assessment against GAN-based models, specifically when pitted against Real-ESRGAN, our ConvSRGAN outperforms with a notable 0.736 dB increase in PSNR, achieving an impressive score of 28.281 dB. Furthermore, it surpasses in SSIM by 0.008, reaching 0.803, indicating a closer alignment with the structural information of the original images. Additionally, in the evaluation of LPIPS, our model demonstrates a superior performance, edging past ESRT with an increase of 0.0095 to attain a score of 0.2334. These metrics collectively affirm the enhanced fidelity, structural retention, and perceptual quality of our model's outputs in the restoration and enhancement of images.

When contrasting ConvSR with w/o GAN architectures, particularly in comparison to ESRT, although there is a marginal difference of 0.061 dB, our model attains a commendable PSNR of 28.916 dB, indicative of high reconstruction fidelity. It is also noteworthy that our model edges ahead on the LPIPS metric by a slim margin of 0.0005, achieving 0.2914, suggesting a closer alignment with human perception of image quality. Moreover, in terms of SSIM, which evaluates structural similarity, our model excels with a score of 0.820, surpassing RRDBNet by 0.001. This underscores the effectiveness of our model in preserving structural integrity while maintaining a competitive edge in overall image quality assessments.

It is worth noting that considering the artistic characteristics of the Chinese landscape painting images, we attempted to train the ConvSR component with a GAN

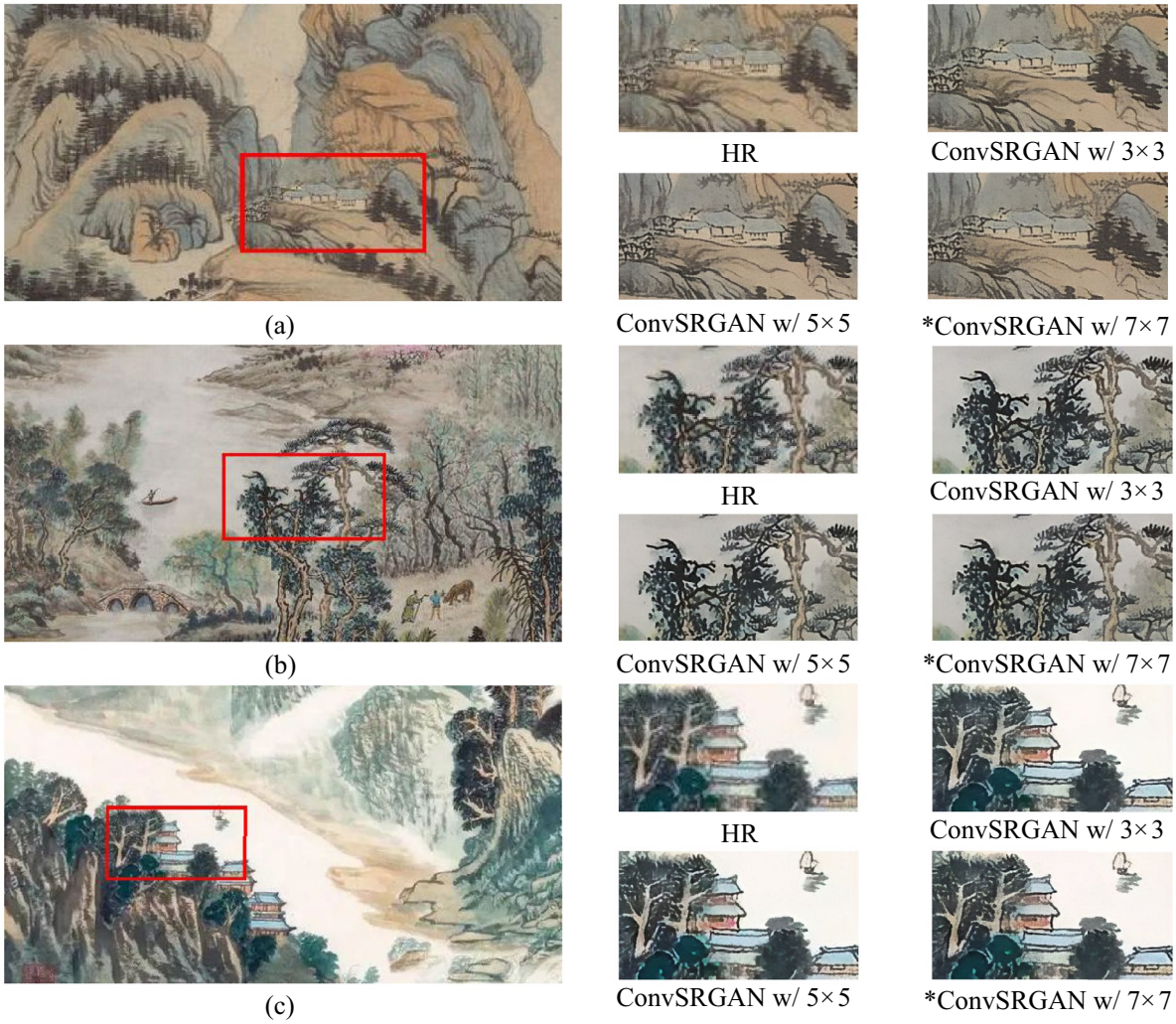

Fig. 8 Visual comparison of the model trained with different kernel size. All of (**a**, **b**, **c**) are from our SRCLP dataset. These three landscape pictures have different and prominent color styles. *ConvSRGAN w/ 7 × 7 is our model. The reliability of the kernel size selection on the training of the ConvSRGAN method was verified and evaluated by ablation experiments. Zoom in for best view

to achieve a more visually pleasing perceptual effect. This was aimed at providing more visual information during the super-resolution process and thereby obtaining more accurate evaluation results.

LPIPS evaluates image quality based on learned human perception, focusing more on the subjective perception of images by the human eye. It can better reflect the human eye's perception of artistic works compared to the PSNR and SSIM metrics. Therefore, our goal is to optimize the performance of our model on the LPIPS metric.

Based on the experimental findings, it is evident that the complete ConvSRGAN network demonstrated a 0.058 improvement in the LPIPS evaluation metric. Moreover, it attained a prominent LPIPS value in comparison to other state-of-the-art techniques, although this trade-off may have led to more modest enhancements in other performance metrics. Nevertheless, we achieved super-resolution outcomes that align better with human perceptual preferences.

In addition, compared with Real-ESRGAN, our method showed a 0.736 dB improvement in the PSNR metric. Additionally, compared with ESRT, our method showed a 0.0095 improvement the LPIPS metric. From the values of various metrics, it can be observed that our model achieved better super-resolution performance (Fig. 4).

Visualization

To perform a visualization comparison of the performance between ConvSRGAN and other SOTA methods, we selected a set of representative painting images from the SRCLP dataset for testing and inference.

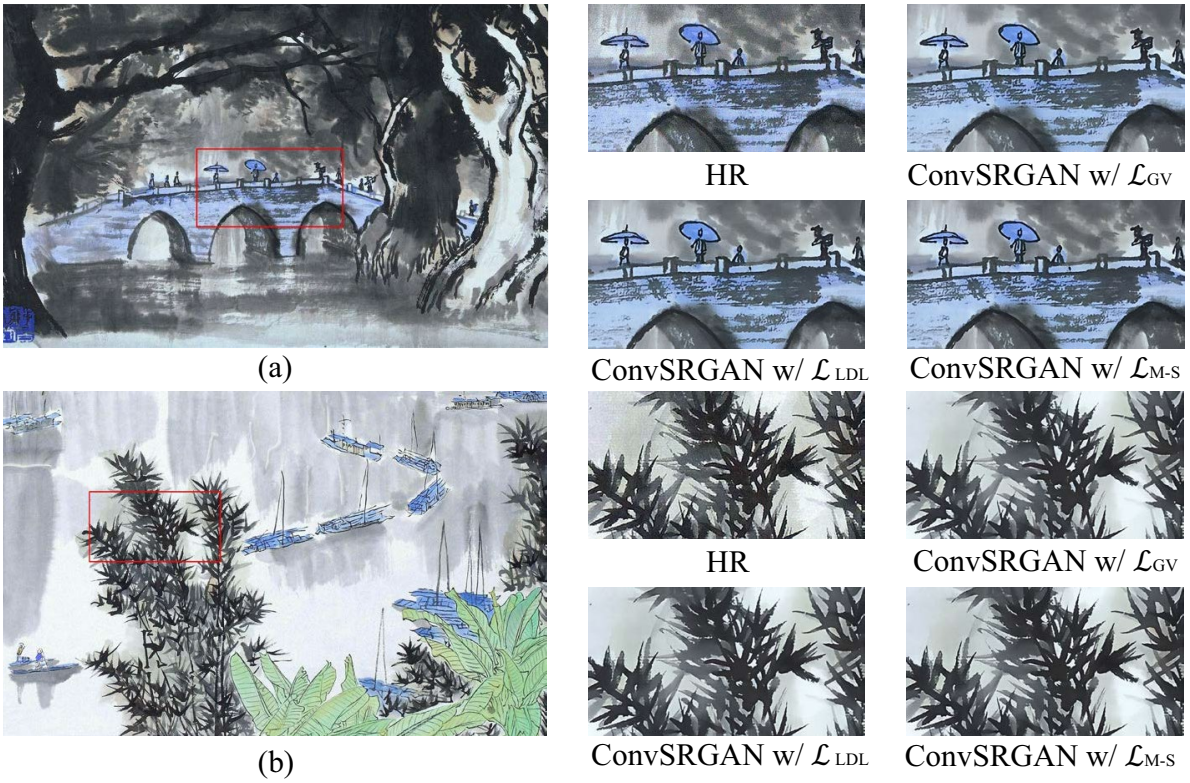

Fig. 9 Visual comparison of the model trained with different loss functions. Both (**a**, **b**) are from our SRCLP dataset. We chose two landscape pictures with different texture characteristics for comparison. *ConvSRGAN w/ $\mathcal{L}_{M-S}$ is our model. The effectiveness of the joint loss on the training of the ConvSRGAN method was verified and evaluated by ablation experiments on the loss function. Zoom in for best view

Specifically, from the pavement texture in Figs. 5a, 6f, the tree trunk in Fig. 5b, and the mountain in Fig. 6d, it can be observed that methods such as EDSR [37], SRResNet [16], and LKDN [30] have lost some of the fine texture details in their results. In contrast, ConvSRGAN preserves the brushstroke forms consistent with the original artwork and exhibits a more delicate representation.

Furthermore, color is an essential means of expressing emotions, atmosphere, and artistic conception in artwork. In the case of the mountain in Figs. 5c, 6d, the result from BSRN [31] appears to have darker colors. It can be seen that ConvSRGAN effectively restores the color, saturation, and brightness information of the artwork, enhancing the artistic expression of the image and conveying richer emotions and artistic conception.

ConvSRGAN bring the image closer to the artistic characteristics and rhythmic style of the original artwork. Chinese traditional painting emphasizes the use of brushstrokes to depict the form and structure of the painting through contours and edges, thus giving the artwork a sense of space and layers.

Moreover, as can be seen in Fig. 5b, Real-ESRGAN [18] and LKDN fail to model the finer textures of the branches and leaves, resulting in blurred results. In contrast, our proposed EHRM block enhances the high-frequency details in the image while preserving the contour of painting through long skip connections. It also indicated that ConvSRGAN not only maintains the different forms of painting elements but also enhances the clarity of the edge lines, making the image more vivid and three-dimensional.

As can be seen in Fig. 7, the super-resolution results from methods like LKDN lose the texture and quality of the rocks, while there are partial artifacts in the result from RRDBNet [18]. The brushstrokes and techniques are crucial for representing the natural form of objects and creating a sense of depth in traditional paintings. It can be seen that our model can learn the interweaving brushwork in the artwork, enhancing the subtlety and artistic effects of the image.

Ablation study

Comparison analysis

By comparing the visual results with various advanced methods, it can be observed that our proposed approach not only improves the loss of high-frequency information in the reconstruction process of painting images but also achieves more accurate super-resolution results. Additionally, our method captures a broader range of spatial

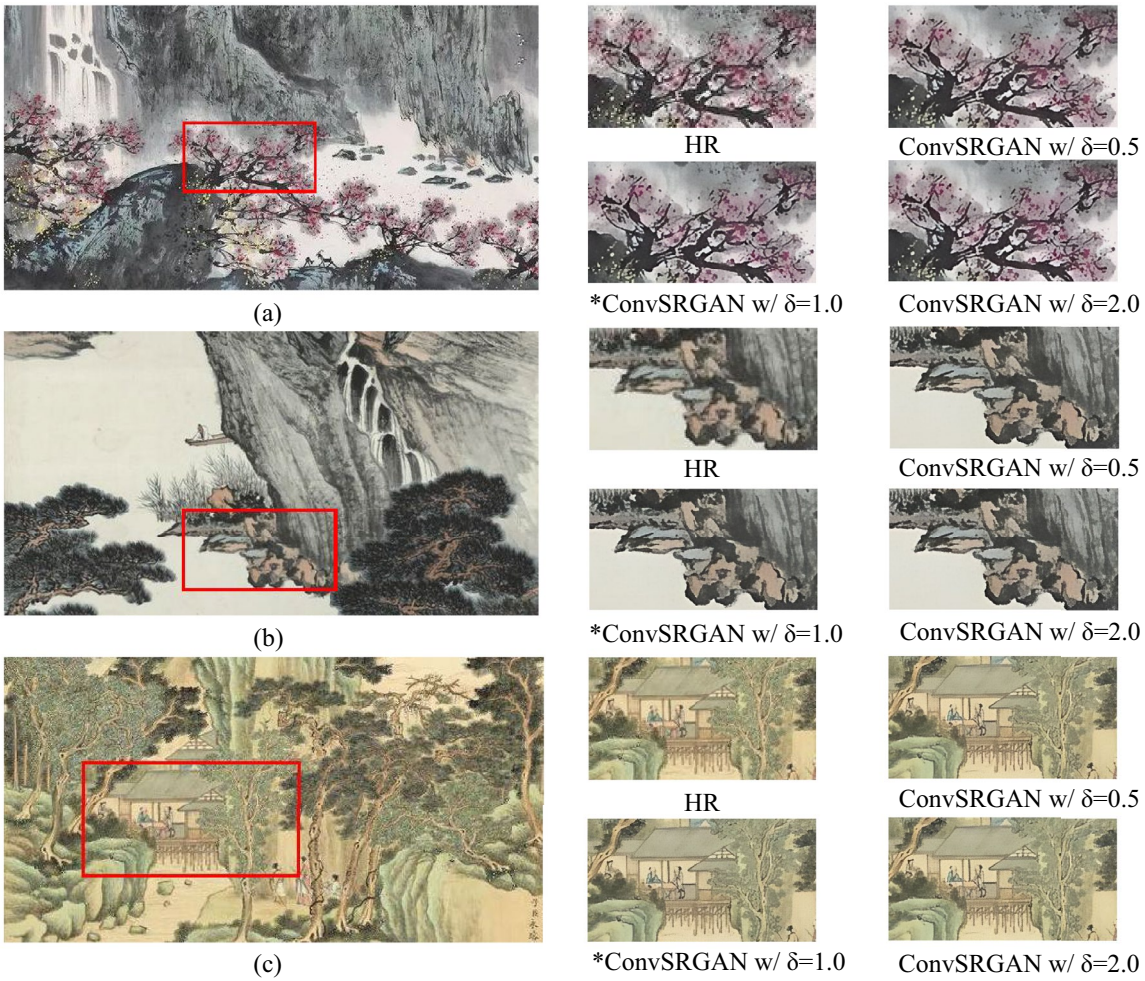

Fig. 10 Visual comparison of the model trained with different weight parameters. All of (**a**, **b**, **c**) are from our SRCLP dataset. These three landscape pictures have different and prominent color styles. *ConvSRGAN w/ δ=1.0 is our model. The reliability of the δ selection on the training of the ConvSRGAN method was verified and evaluated by ablation experiments. Zoom in for best view

Table 3 Comparison of ConvSRGAN on different datasets

Dataset	PSNR ↑	SSIM ↑	LPIPS ↓
Mural	24.122	0.724	0.2726
Painter By Numbers [33]	25.633	0.681	0.3079
Flickr2K [34]	23.733	0.710	0.3079
SRCLP(Ours)	24.675	0.770	0.2409
Song Dynasty	27.529	0.633	0.3695
Yuan Dynasty	25.800	0.618	0.3649
Ming Dynasty	25.797	0.648	0.3582
Qing Dynasty	24.734	0.657	0.3539
Painter *Daqian Zhang*	25.226	0.741	0.2854

In the SRCLP Dataset, we also testing on the paintings of different dynasties and the paintings of the painter *Daqian Zhang*

dependencies, which is crucial for understanding and reproducing visual characteristics such as the overall layout, brushstroke trends, and color rhythms present in Chinese traditional painting.

Table 4 Comparison results of ConvSRGAN on Mural

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Bicubic	22.267	0.570	0.4493
SRResNet [16]	24.122	0.724	0.2726
ESRT [25]	23.980	0.729	0.2652
VapSR [29]	22.489	0.707	0.3033
LKDN [30]	22.937	0.682	0.3363
Real-ESRGAN [18]	<u>24.229</u>	**0.771**	<u>0.2490</u>
ConvSRGAN	**24.675**	<u>0.770</u>	**0.2409**

* Optimal results are displayed in **bold**, while suboptimal results are <u>underlined</u>

To enhance the credibility of our model, we have devised three ablation experiments: the kernel size in the ADCB, the $\mathcal{L}_{M-S}$, and the weight of the $\mathcal{L}_{M-S}$. As illustrated in Table 2.

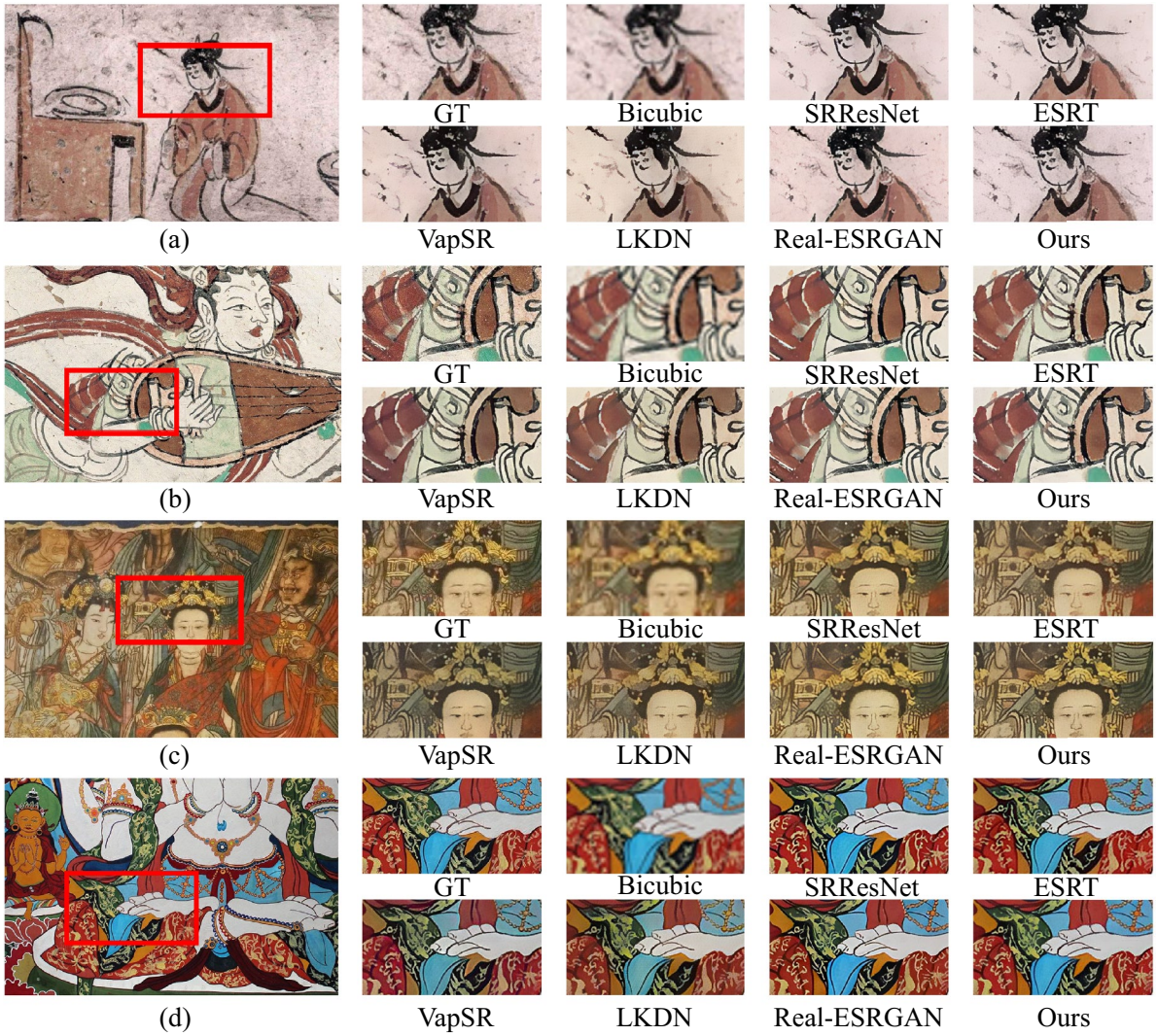

Fig. 11 Visual comparison on Mural. **a** *Cave*, **b** *Temple*, **c** *Tomb*, **d** *Thangka*. Our method preserves the overall color while processing the local lines more finely on Mural. Zoom in for best view

Comparison analysis on kernel size

For comparison analysis, we set three models about kernel sizes, which is ConvSRGAN w/ 7×7, ConvSRGAN w/ 5×5, and ConvSRGAN w/ 3×3.

In terms of evaluating metrics, ConvSRGAN w/ 7×7 kernel size achieved the best results across all three metrics. Compared with ConvSRGAN w/ 3×3 convolution kernel size, our model improves PSNR by 0.261 dB, reaching 28.281dB, and SSIM by 0.003. Additionally, when compared to the ConvSRGAN w/ 5×5 kernel size, our model achieves a 0.0077 improvement in LPIPS.

It should be noted that increasing the size of the convolutional kernel results in a larger sensory field. It can cover a larger region of the image, allowing our model to better understand the global structure and contextual information of the image. In addition, a larger convolutional kernel is able to extract richer features, including image texture, edge information, and shape. ConvSR network can more accurately deal with the details of the image, which is crucial for super-resolution processing of landscape paintings.

Comparison analysis on MS-SSIM loss

Initially, we performed ablation experiments by comparing ConvSRGAN with ConvSRGAN w/o MS-SSIM loss ($\mathcal{L}_{M-S}$). As can be seen in Table 2, the PSNR value increases by 0.498dB and other metrics are also improved when using $\mathcal{L}_{M-S}$. Meanwhile, we compare the results with three kinds of loss function: $\mathcal{L}_{M-S}$, Gradient

variance loss ($\mathcal{L}_{GV}$) [38] and Local Discriminative Learning loss ($\mathcal{L}_{LDL}$) [39].

$\mathcal{L}_{GV}$ aims to minimize the distance between the variance maps, resulting in clearer images. $\mathcal{L}_{LDL}$ stabilizes the model training process by computing artifact maps in the reconstructed images. In the comprehensive evaluation against four baseline models, ConvSRGAN w/ $\mathcal{L}_{M-S}$ exhibits significant advancements. Specifically, unlike ConvSRGAN w / o $\mathcal{L}_{M-S}$, our model achieves a substantial improvement of 0.498 dB in PSNR, resulting in an impressive score of 28.281 dB. Compared to ConvSRGAN w / $\mathcal{L}_{GV}$, our model surpasses its SSIM by 0.014, indicative of superior preservation of structural similarity. Collectively, these metrics highlight the exceptional restoration and enhancement capabilities of our model across various dimensions of image quality assessment. Our findings demonstrate that $\mathcal{L}_{M-S}$ preserves the structure and layout of painting images in the super-resolution task by considering the structural similarity of images at different scales.

Comparison analysis on loss weight

In our study, a comparative experiment was conducted among three models, with their loss function weights systematically varied to δ=0.5, δ=1.0, and δ=2.0, respectively. Subsequently, a rigorous index testing protocol was employed for each model configuration to quantitatively evaluate this impact. The detailed findings from these assessments have been compiled and are presented in Table 2, offering insights into how the strategic tuning of the loss function weight δ can be utilized to optimize model performance and mitigate overfitting issues.

Quantitative evaluation has revealed that ConvSRGAN w/ δ=1.0 outperforms ConvSRGAN w/ δ=0.5 with a notable increase of 0.147 dB in PSNR. Furthermore, ConvSRGAN w/ δ=1.0 excels with the highest SSIM of 0.803. Collectively, these metrics affirm our model's superiority across multiple dimensions of image assessment.

Visualization

As depicted in Figs. 8, 9, 10, we visualized analysis the performance on kernel size, MS-SSIM loss and loss weight.

Visualization on kernel size

Indeed, incorporating a larger convolution kernel, expands the receptive field of the model, allowing it to capture more extensive contextual information within the input data. This enhancement in the model's horizon can lead to improved feature learning and a heightened ability to model complex patterns, thereby augmenting its fitting capacity.

As depicted in Fig. 8a, the effect of ConvSRGAN w/ 7×7 kernel size shows the reconstructed building textures in scenes were more natural and realistic. However, ConvSRGAN w/ 3×3 kernel size produced excessively smooth outcomes. Meanwhile, in the mountain texture depicted in Fig. 8b, ConvSRGAN w/ 5×5 kernel size led to color distortion.

Based on the results of our comparative tests and the theoretical understanding of larger convolution kernels, we can confidently conclude that the adoption of larger kernel sizes, such as our 7×7 convolution kernel, indeed bolsters the reconstruction capabilities of our model. The expanded receptive field enables the model to better grasp the broader context within images, thereby improving its ability to restore fine details and intricate structures present in landscape paintings.

Visualization on MS-SSIM loss

The loss function can quantify the distance or structural differences between images. Meanwhile, loss function plays a crucial role in guiding model fitting and ultimately affects the effectiveness and performance of the model during training. To evaluate the efficacy of our proposed method, we perform ablation experiments on the ConvSRGAN w/ $\mathcal{L}_{M-S}$ to validate the improvement in model performance.

To further test the effectiveness of the ConvSRGAN w/ $\mathcal{L}_{M-S}$, we set up some ablation models and analyzed the resulting changes. As shown in the Fig. 9b, the use of other loss functions failed to recover the shadowed parts of the bamboo leaves and resulted in brightness distortion. However, the ConvSRGAN w/ $\mathcal{L}_{M-S}$ effectively resolved this issue.

In the process of image reconstruction, it is imperative to account for two pivotal loss functions: $\mathcal{L}_1$ and $\mathcal{L}_{M-S}$. The former, $\mathcal{L}_1$, centers on quantifying pixel-wise discrepancies, ensuring a precise replication of individual elements. Conversely, $\mathcal{L}_{M-S}$ prioritizes the maintenance of structural similarity across varied resolutions, thereby safeguarding the coherence and integrity of image structures. Achieving an optimal equilibrium between these two measures is fundamental, as it enables the preservation of both intricate details and the overarching compositional structure, which is vital for the faithful and visually coherent restoration of images.

Visualization on loss weight

To examine the impact of the weighting coefficient of the $\mathcal{L}_{M-S}$ on the reconstruction process, we conducted experiments using three distinct weighting parameters with corresponding δ values. This systematic approach aimed to meticulously gauge the repercussions of

different loss function weightings on the extent of overfitting. As depicted in the Fig. 10, ConvSRGAN w/ δ =1.0 exhibits superior performance in retaining intricate details and texture fidelity during the processing phase. This visual evidence reinforces our earlier quantitative findings, demonstrating a heightened capability to preserve the subtle nuances and fine elements of the original content, thereby enhancing the overall quality and authenticity of the processed images.

It can be identified that if the σ value is too large, the model may sacrifice some finer details. However, if the value of σ is too small, an excessive focus on details may lead to a loss of overall coherence. Our findings show that a σ value of 1.0 yields the most favorable results.

Comparison with different dataset

Comparison analysis

We conducted some comparative experiments on the ConvSRGAN model with three datasets: Mural, Painter by Numbers [33] and Flickr2K [34], respectively. Moreover, we also compare with other models on the Mural dataset.

Comparison analysis on different dataset

We conducted some comparative experiments on the ConvSRGAN model with different types of dataset to verify its performance. We carefully selected 2000 images on Painter By Numbers [33] and Flickr2K [34], respectively, as a training set. In addition, we selected 100 images as the testing set. As illustrated in Table 3.

Moreover, in the SRCLP Dataset, we selected 100 images of different dynasties: the *Song* Dynasty, *Yuan* Dynasty, *Ming* Dynasty, *Qing* Dynasty, for testing and analysis. In addition, we selected 68 images of the paintings of famous painter *Daqian Zhang*. As the one of the greatest painters in Chinese modern art history, *Daqian Zhang* have a high reputation at home and abroad. All results show that our network performs well in different styles of datasets.

Comparison analysis on mural

As illustrated in Table 4, we testing and compare with six state-of-the-art methods. Table 4 shows that in direct comparison with Real-ESRGAN, our model showcases a significant enhancement of 0.446 dB in PSNR, elevating the score to 24.675 dB. Although the difference in SSIM is marginal at 0.001, our model maintains a strong score of 0.770, indicative of comparable or slightly improved structural preservation. Moreover, our model distinguishes itself by achieving the optimal outcome in LPIPS, surpassing the baseline performance. Collectively, these metrics validate our model's advancements in image restoration and enhancement. Meanwhile, the performance of our method is comparable to the state-of-the-art Real-ESRGAN, demonstrating its applicability on Mural super-restoration inpainting.

Visualization

We visualize and analyze on these three datasets: Mural, Painter by Numbers [33] and Flickr2K [34], separately. Moreover, we also visualized some landscape paintings on the paintings of different dynasties and the paintings of the painter *Daqian Zhang*.

Visualization on mural

To further validate the effectiveness of our model, we have conducted testing experiments on Mural dataset. Figure 11 visually illustrates the results, effectively demonstrating how our method delivers more aesthetically pleasing outcomes compared to Real-ESRGAN. The comparison between the ESRT and Real-ESRGAN results shows that using inappropriate loss functions, high-pass filtering, and pooling operations can result in excessive smoothing of image details, leading to less realistic super-resolution results. In contrast, our method handles the image more naturally, recovering more accurate image texture details, as shown in the facial details in Fig. 11c. Overall, our approach proves to be highly effective for restoring mural images.

Visualization on painter by numbers

Furthermore, we have conducted a qualitative comparison of the visual effects between ConvSRGAN and other advanced methods on Painter By Numbers dataset. As shown in Fig. 12, it can be observed that other methods still exhibit shortcomings when restoring non-real-world images. In Fig. 12a, it is apparent that BSRN, VapSR, and LKDN lose color information in their results, while the super-resolution result from RRDBNet is overly smooth, exhibiting low fidelity in the lines of buildings and the texture of flowers, thus compromising the original brushstrokes and artistic style.

However, from the experimental results, our method did not achieve ideal results in Fig. 12a with blurred edges of the petals. We consider that this may be attributed to improper parameter adjustment in the EHRM, resulting in the loss of high-frequency information. Nonetheless, our method has achieved natural visual effects on other images, closely aligning with the artistic style and technical characteristics of the original paintings. It has demonstrated good visual performance, proving the applicability of ConvSRGAN to other artistic images.

Fig. 12 Visual comparisons on Painter By Numbers. All of (**a**, **b**, **c**, **d**) are from dataset Painter By Numbers. The visual effects of our method handle the color of the painting better than the SOTA model, and the lines are cleaner and clearer. Zoom in for best view

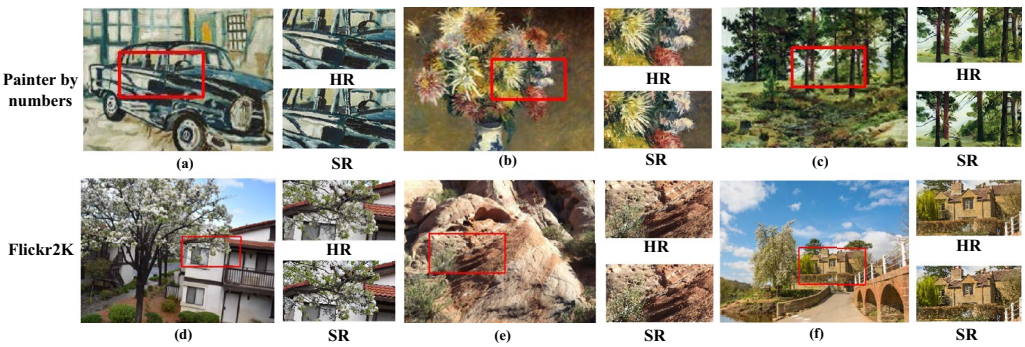

Fig. 13 Visualization on Flickr2K. The visual effects of our method handle the color of the painting better than the SOTA model, and the lines are cleaner and clearer. Zoom in for best view

Visualization on Flickr2K

As illustrated in Fig. 13, our model performs well on a dataset of natural images, Flickr2K. SR represents the super-resolution images produced by the ConvSRGAN. HR represents the high-resolution original images, as the ground truth.

Upon examining the visual evidence presented in Fig. 13, it becomes evident that our model excels in preserving details across categories (a), (b), (e), and (f), demonstrating a commendable capability in handling intricate texture components. For example, it is relatively good for the contours of distant houses in Fig. 13f. Nevertheless, it

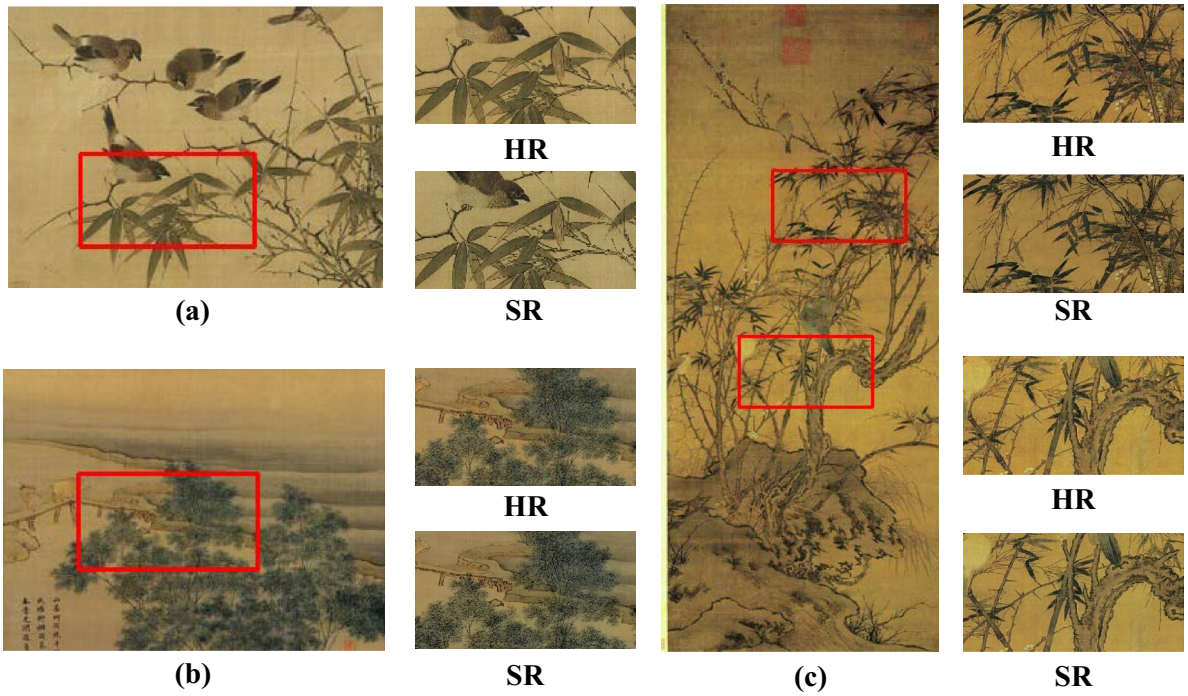

Fig. 14 Visual comparisons on the paintings of *Song* dynasty. They are paintings of *'Frost Shinohan'*, *'Twilight Return'*, and *'plum and bamboo gathering birds'*. They all show the relatively delicate brushwork and expressive expression of the *Song* Dynasty painters

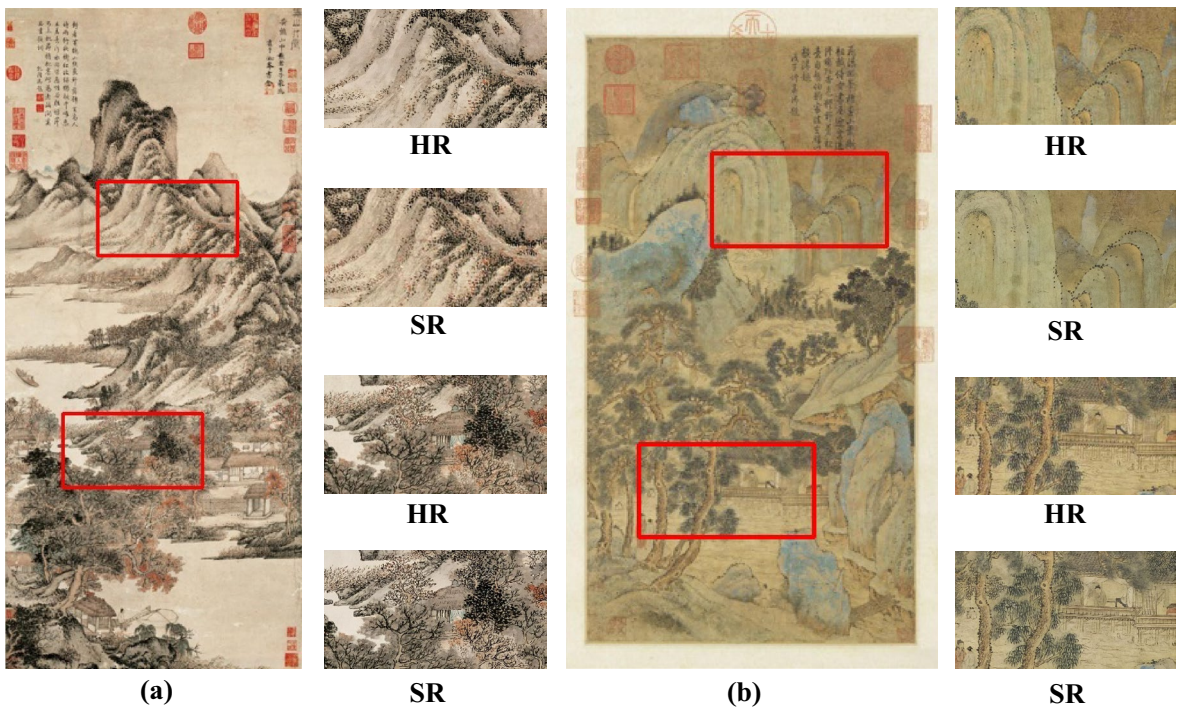

Fig. 15 Visual comparisons on the paintings of *Yuan* Dynasty. They are paintings of *'The Qiushan Grass Hall'* by Yuan Dynasty artist *Meng Wang* and *'The mountains'* by artist *Mengxu Zhao*. They all have the overall composition of landscape scenes and the depiction of local details

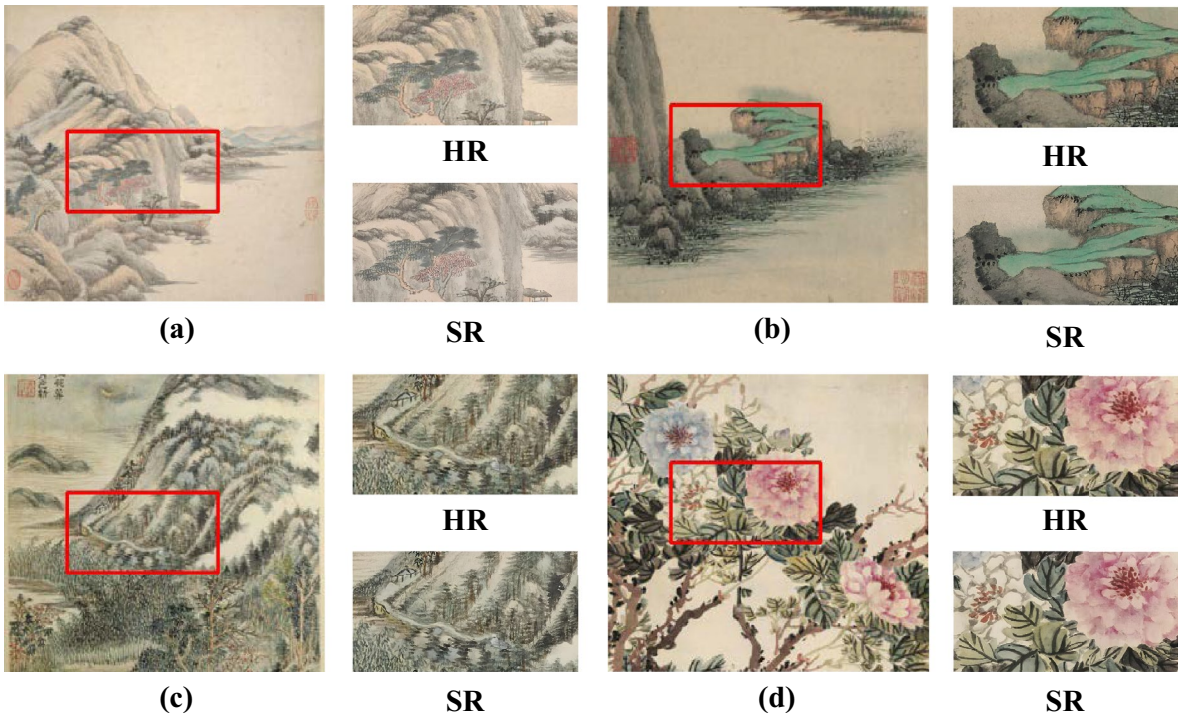

Fig. 16 Visual comparisons on the paintings of *Ming* and *Qing* Dynasty. Both (**a**, **b**) are the paintings of *Ming* dynasty artist *Jian Wang*, illustrating the profound capacity of *Ming* artists to comprehend and depict the aesthetic essence of landscape painting's artistic conception. Meanwhile, (**c**, **d**) represent the *Qing* Dynasty through the painting of *'Du Fu's Poetry in Paint'* of artist *Shimin Wang* and the painting of *'Peony Blossoms'* of artist *Zhiqian Zhao*, respectively. These latter two pieces exemplify meticulous detail and the rich artistic conception that is paramount to the expressive lexicon of Chinese painting tradition

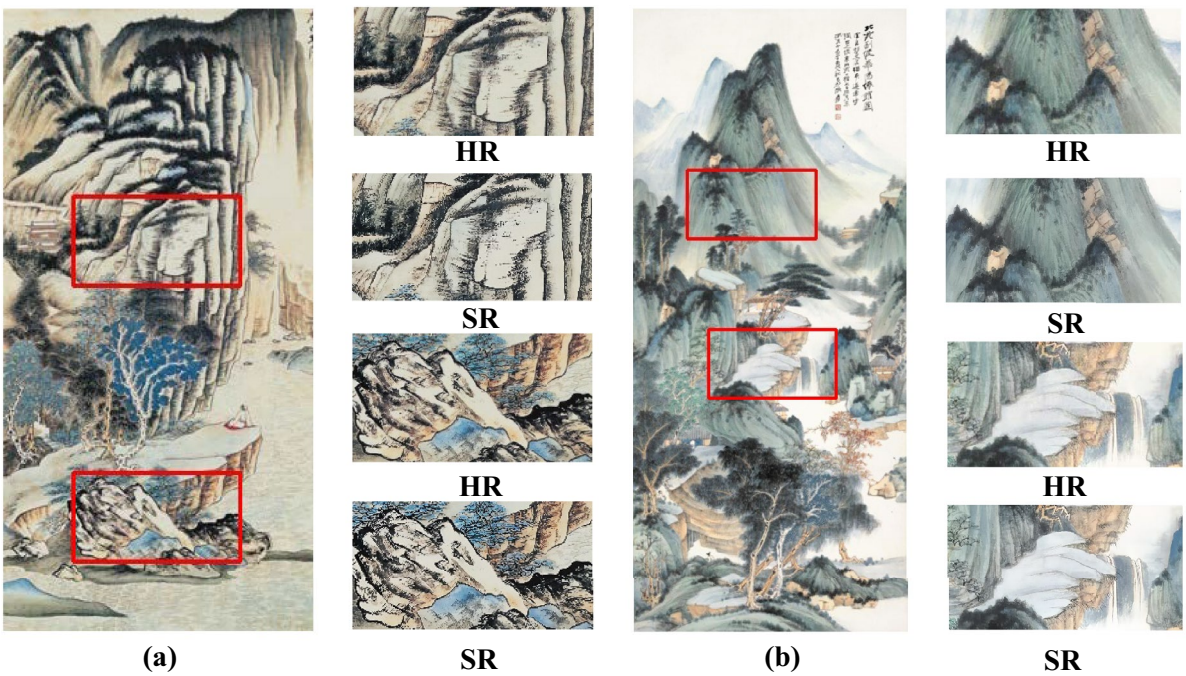

Fig. 17 Visual comparisons on the paintings of painter *Daqian Zhang*. **a** depicts the painting of *'Prime Minister amidst Mountains with Immortal Essence'*. **b** showcases *Daqian Zhang's* interpretation of *Yuanhua Dong's* artistic styles, jointly illustrate the profound depth of understanding of *Daqian Zhang* in the realm of landscape painting. His proficiency goes beyond meticulous attention to textural details. It extends to encompassing a grand and awe-inspiring aesthetic of artistic conception, highlighting his comprehensive mastery over the art form

is imperative to acknowledge that there exists room for enhancement, particularly in the processing of categories (c) and (d). A case in point is the suboptimal restoration of the floral feature depicted in Fig. 13d, which, upon thorough analysis, we attribute to the similarity in hue between the floral subject and its surrounding background. This low contrast scenario poses a challenge for the network in effectively isolating and extracting distinct textural features, thereby necessitating further optimization strategies to mitigate such instances.

By comparing art images in Painter By Numbers with natural images in Flickr2K, exemplified in Fig. 13(b) depicting a floral arrangement alongside and in Fig. 13d presenting a real-world floral scene, our model demonstrates proficiency in handling relatively intricate textural details. However, it encounters challenges in identifying and reconstituting image characteristics that exhibit low contrast, highlighting a limitation in feature extraction for such subtler visual elements.

Visualization on different dynasties

As illustrated in Figs. 14, 15, 16, these diverse selections of Chinese paintings, spanning various dynasties, encompass not only breathtaking landscapes but also intimate portrayals of flora in detailed close-ups. A comparative analysis reveals that our model excels in achieving a nuanced classification of ancient paintings across epochs, all while meticulously preserving the original works' details and inherent artistic essence intact. This demonstrates a high level of fidelity in maintaining the unique aesthetic sensibilities and spiritual depth embedded within each piece, thereby testifying to the efficacy and sensitivity of our approach in handling such culturally and aesthetically rich content.

Visualization on the painter Daqian Zhang

As illustrated in Fig. 17, the artwork of *Daqian Zhang* is renowned for its distinctive bright, elegant, and graceful aesthetic. Our model, when put to the test through comparative analyses, has proven capable of adeptly managing the intricate details characteristic of his pieces, ensuring the preservation of *Zhang*'s signature style. This highlights the efficacy of our model in accurately capturing and reproducing the refined nuances and aesthetic hallmarks integral to *Zhang*'s artistic legacy.

Conclusion

In this paper, we propose an innovative framework for super-resolution inpainting of Traditional Chinese Paintings, termed ConvSRGAN. We utilize a series of Enhanced Adaptive Residual Module to progressively learn the depth information of the images. In particular, within the EARM, we introduce an Enhanced High-frequency Retention Module to preserve high-frequency details through a specially designed Adaptive Depthwise Convolution Block and pooling operations that broaden the model's receptive field.

To ensure that the model achieves more realistic and nuanced texture restoration, we incorporate the $\mathcal{L}_{M-S}$ in the training with a combined loss function for supervised learning. Overall, the ConvSRGAN framework presented in this paper aims to address the challenges specific to traditional Chinese painting images and provides a novel solution for enhancing image resolution while preserving the artistic style and details of the paintings.

The experimental results demonstrate that ConvSRGAN achieves significant performance in handling traditional painting and mural datasets, particularly in high-definition restoration tasks for landscape paintings, showing remarkable visual fidelity and vividness. This validates its effectiveness and universality in the field of cultural heritage preservation and restoration. Furthermore, the model achieves excellent visual results on other artistic datasets while preserving the unique artistic style of the paintings, further confirming its robustness and generalizability in artistic image super-resolution tasks.

Discussion

The future research plan will focus on deepening exploration in two key areas: cultural heritage conservation and utilization, and optimization of modeling technology.

Firstly, in terms of cultural heritage conservation, we will continue to study the application of image super-resolution models in the field of cultural heritage, including but not limited to improving the high-definition restoration ability of the models for traditional artworks. We will also develop more refined image restoration algorithms specifically tailored to the material and age characteristics of cultural artifacts.

Secondly, on the technical level, we will explore new network architectures or loss functions to achieve better inference results. Additionally, we will further investigate the performance of the models, including improving the quality of super-resolution images and the speed of model inference.

Acknowledgements

This research was supported by the National Key Research and Development Program of China (No. 2023YFF0715103), National Natural Science Foundation of China (Grant No. 62,306,237 and No. 62,173,270), Key Research and Development Program of Shaanxi (No. 2024GX-YBXM-149), Northwest University Graduate Innovation Project (No. CX2023194), Natural Science Foundation of Shaanxi (No. 2023-JC-QN-0750).

Author contributions

Qiyao Hu: Conceptualization, software, validation, resources, data curation, Formal analysis. Xianlin Peng: Software, Investigation, Validation. Tengfei Li: Preparation, methodology; Xiang Zhang: Writing - Review & Editing,

Supervision, Project administration. Jiangpeng Wang: Project administration. Jinye Peng: Project administration.

Availability of data and materials
The datasets used or analysed during the current study are available from the e-mail huqiyao@nwu.edu.cn on reasonable request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors have no competing interests to declare that are relevant to the content of this article.

Received: 8 February 2024 Accepted: 14 May 2024
Published online: 31 May 2024

References

1. Xiao J. Research on super-resolution relationship extraction and reconstruction methods for images based on multimodal graph convolutional networks. Math Probl Eng. 2022. https://doi.org/10.1155/2022/1016112.
2. Prajapati K, Chudasama V, Patel H, Upla K, Ramachandra R, Raja K, Busch C. Unsupervised single image super-resolution network (usisresnet) for real-world data using generative adversarial network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; 2020. p. 464–465. https://doi.org/10.1109/cvprw50498.2020.00240.
3. Das B, Roy SD. Edge-aware image super-resolution using a generative adversarial network. SN Comput Sci. 2023;4(2):146. https://doi.org/10.1007/s42979-022-01561-8.
4. Zhao L, Lin S, Lin Z, Ding J, Huang J, Xing W, Lin H. Progressive multilevel feature inpainting algorithm for chinese ancient paintings. J Comput-Aided Des Comput Graph. 2023. https://doi.org/10.3724/SP.J.1089.2023.19544.
5. Qiao T, Zhang W, Zhang M, Ma Z, Xu D. Ancient painting to natural image: A new solution for painting processing. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV); IEEE; 2019. p. 521–530. https://doi.org/10.1109/wacv.2019.00061/
6. Wang Z, Zhang J, Ji Z, Bai J, Shan S. Cclap: controllable Chinese landscape painting generation via latent diffusion model. arXiv preprint. 2023. https://doi.org/10.1109/icme55011.2023.00362.
7. Yuan S, Dai A, Yan Z, Liu R, Chen M, Chen B, Qiu Z, He X. Learning to generate poetic Chinese landscape painting with calligraphy. arXiv preprint. 2023. https://doi.org/10.2496/ijcai.2022/696.
8. Gui X, Zhang B, Li L, Yang Y. Dlp-gan: learning to draw modern Chinese landscape photos with generative adversarial network. Neural Comput Appl. 2023. https://doi.org/10.1007/s00521-023-09345-8.
9. Xu Z, Shang H, Yang S, Xu R, Yan Y, Li Y, Huang J, Yang HC, Zhou J. Hierarchical painter: Chinese landscape painting restoration with fine-grained styles. Vis Intell. 2023;1(1):19. https://doi.org/10.1007/s44267-023-00021-y.
10. Xue A. End-to-end chinese landscape painting creation using generative adversarial networks. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2021. p. 3863–3871. https://doi.org/10.1109/wacv48630.2021.00391.
11. Zhang G, Zhang J, Song J, Guo J, Zhou C. Automatic generation model of Chinese landscape painting based on confrontation generation network. Comput Telecommun. 2020. https://doi.org/10.1596/j.cnki.dnydx.2020.03.001.
12. Shi H, Xu D, Zhang H, Yue Y. A single historical painting super-resolution via a reference-based zero-shot network. Int J Comput Intell Syst. 2021;14(1):1577–88. https://doi.org/10.2991/ijcis.d.210503.002.
13. Nagar S, Bala A, Patnaik SA. Adaptation of the super resolution sota for art restoration in camera capture images. In: 2023 International Conference on Emerging Techniques in Computational Intelligence (ICETCI), IEEE; 2023. p. 158–163. https://doi.org/10.1109/icetci58599.2023.10331102.
14. Lyu Q, Zhao N, Yang Y, Gong Y, Gao J. A diffusion probabilistic model for traditional Chinese landscape painting super-resolution. Heritage Sci. 2024;12(1):4. https://doi.org/10.1186/s40494-023-01123-y.
15. Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image super-resolution. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13, Springer; 2014. p. 184–199. https://doi.org/10.1007/978-3-319-10593-2_13.
16. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017. p. 4681–4690. https://doi.org/10.1109/cvpr.2017.19.
17. Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao, Y, Change Loy C. Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops; 2018. https://doi.org/10.1007/978-3-030-11021-5_5.
18. Wang X, Xie L, Dong C, Shan Y. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 1905–1914. https://doi.org/10.1109/iccvw54120.2021.00217.
19. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv preprint. 2020. https://doi.org/10.4855/arXiv.2010.11929.
20. Chen X, Hsieh C-J, Gong B. When vision transformers outperform resnets without pre-training or strong data augmentations. arXiv preprint. 2021. https://doi.org/10.4855/arXiv.2106.01548.
21. Wang W, Xie E, Li X, Fan D-P, Song K, Liang D, Lu T, Luo P, Shao L. Pyramid vision transformer: A versatile backbone for dense prediction without convolutions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 568–578. https://doi.org/10.1109/iccv48922.2021.00061.
22. d'Ascoli S, Touvron H, Leavitt ML, Morcos AS, Biroli G, Sagun L. Convit: Improving vision transformers with soft convolutional inductive biases. In: International Conference on Machine Learning, PMLR; 2021. p. 2286–2296. https://doi.org/10.1088/1742-5468/ac9830.
23. Liang J, Cao J, Sun G, Zhang K, Van Gool L, Timofte R. Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 1833–1844. https://doi.org/10.1109/iccvw54120.2021.00210.
24. Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2021. p. 10012–10022. https://doi.org/10.1109/iccv48922.2021.00986.
25. Zhisheng L, Hong L, Juncheng L, Linlin Z. Efficient transformer for single image super-resolution. arXiv preprint. 2021. https://doi.org/10.4855/arXiv.2108.11084.
26. Chen Z, Zhang Y, Gu J, Kong L, Yang X, Yu F. Dual aggregation transformer for image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision; 2023. p. 12312–12321. https://doi.org/10.1109/iccv51070.2023.01131.
27. Liu Z, Mao H, Wu C-Y, Feichtenhofer C, Darrell T, Xie S. A convnet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 11976–11986. https://doi.org/10.1109/cvpr52688.2022.01167.
28. Ding X, Zhang X, Han J, Ding G. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 11963–11975. https://doi.org/10.1109/cvpr52688.2022.01166.
29. Zhou L, Cai H, Gu J, Li Z, Liu Y, Chen X, Qiao Y, Dong C. Efficient image super-resolution using vast-receptive-field attention. In: European Conference on Computer Vision, Springer; 2022. p. 256–272. https://doi.org/10.1007/978-3-031-25063-7_16.
30. Xie C, Zhang X, Li L, Meng H, Zhang T, Li T, Zhao X. Large kernel distillation network for efficient single image super-resolution. In: Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2023. p. 1283–1292. https://doi.org/10.1109/cvprw59228.2023.00135.
31. Li Z, Liu Y, Chen X, Cai H, Gu J, Qiao Y, Dong C. Blueprint separable residual network for efficient image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 833–843. https://doi.org/10.1109/cvprw56347.2022.00099.
32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint. 2014. https://doi.org/10.4855/arXiv.1409.1556.
33. Duck WK. Painter by Numbers. Kaggle; 2016. https://kaggle.com/competitions/painter-by-numbers.
34. Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: Dataset and study. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops; 2017.
35. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process. 2004;13(4):600–12. https://doi.org/10.1109/tip.2003.819861.
36. Zhang R, Isola P, Efros AA, Shechtman E, Wang O. The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p. 586–595. https://doi.org/10.1109/cvpr.2018.00068.
37. Lim B, Son S, Kim H, Nah S, Mu Lee K. Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2017. p. 136–144. https://doi.org/10.1109/cvprw.2017.151.
38. Abrahamyan L, Truong AM, Philips W, Deligiannis N. Gradient variance loss for structure-enhanced image super-resolution. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE; 2022. p. 3219–3223. https://doi.org/10.1109/icassp43922.2022.9747387.
39. Liang J, Zeng H, Zhang L. Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2022. p. 5657–5666. https://doi.org/10.1109/cvpr52688.2022.00557.

Publisher's Note