Heritage Science

RESEARCH Open Access

A diffusion probabilistic model for traditional Chinese landscape painting super-resolution

Qiongshuai Lyu[1*], Na Zhao[2,3], Yu Yang[1], Yuehong Gong[1] and Jingli Gao[1]

Abstract

Traditional Chinese landscape painting is prone to low-resolution image issues during the digital protection process. To reconstruct high-quality images from low-resolution landscape paintings, we propose a novel Chinese landscape painting generation diffusion probabilistic model (CLDiff), which is similar to the Langevin dynamic process, and realizes the transformation of the Gaussian distribution into the empirical data distribution through multiple iterative refinement steps. The proposed CLDiff can provide ink texture clear super-resolution predictions by gradually transforming the pure Gaussian noise into a super-resolution landscape painting condition on a low-resolution input through a parameterized Markov Chain. Moreover, by introducing an attention module with an energy function into the U-Net architecture, we turn the denoising diffusion probabilistic model into a powerful generator. Experimental results show that CLDiff achieves better visual results and highly competitive performance in traditional Chinese Landscape painting super-resolution tasks.

Keywords Chinese landscape painting, Denoising diffusion probabilistic model, Attention mechanism, U-Net, Super-resolution

Introduction

In the long history and cultural development of China, traditional landscape painting is a very important form of cultural and artistic expression. Traditional Chinese landscape painting not only shows the beauty of the Chinese land, but also integrates the painter's thinking and emotional sustentation of the universe, nature, society, and life, which perfectly embodies the aesthetic thoughts in traditional Chinese ancient philosophy. Figure 1 shows traditional Chinese landscape paintings with diverse styles and unique charm. However, due to unpredictable factors such as natural, human, and equipment, this kind of art treasures with Oriental characteristics in the process of digital protection can lead to problems such as low resolution and semantic loss. This seriously hinders the inheritance of Chinese excellent history and culture.

At present, research on the super-resolution task of traditional Chinese landscape painting is rare and mainly focuses on the field of image generation and image translation. To exploit the multi-scale image information, Lin et al. [1] proposed a multi-scale generative adversarial network (GAN) to transform the sketch into Chinese paintings. To evaluate the quality of Chinese landscape paintings generated by different strategies, Lv et al. [2] investigated the influence of the network model, loss function, and training objective on the quality of generated Chinese landscape paintings in conditional generative adversarial networks. Zhou et al. [3] proposed an interactive and generative framework based on cycle-GAN, which can generate Chinese landscape paintings from input sketches. A recent work is SAPGAN [4] (Sketch-And-Paint GAN), which first employs SketchGAN to generate the sketches of

*Correspondence:
Qiongshuai Lyu
4354@pdsu.edu.cn
[1] School of Software, Pingdingshan University, Pingdingshan 467000, China
[2] School of Journalism and Communication, Pingdingshan University, Pingdingshan 467000, China
[3] School of Communication, East China Normal University, Shanghai 200241, China

Fig. 1 Some traditional Chinese landscape paintings

landscape paintings, and then uses PaintGAN to realize the transformation from the sketches to Chinese landscape paintings.

In addition, for the image super-resolution task, some scholars have proposed a variety of different solutions. To model the local structure of complex images and reduce the time cost, an adaptive sparse domain selection and adaptive regularization method [5] is proposed. Considering the non-local self-similarity property of images, a simple and effective non-local centralized sparse representation method [6] is proposed to solve the problem of image super-resolution. These methods achieve appealing super-resolution performance but often require solving a complex iterative optimization problem, and the model lacks prior knowledge learned from large-scale datasets when solving. Recent works have shown that deep learning methods have achieved excellent performance in learning complicated empirical distributions of images. By combining the advantages of convolutions with Transformers, a strong baseline model [7] is proposed for image super-resolution. To utilize image perception information, a generative adversarial network is proposed for image super-resolution [8] (SRGAN). By improving SRGAN, the later proposed ESRGAN [9] and Real-ESRGAN [10] further improved the performance of image SR. However, GAN-driven methods are prone to mode collapse [11], resulting in no diversity in the generated images. Additionally, the training process of GAN-driven methods is unstable and prone to the vanishing gradient problem [12] or exploding gradient problem [13].

Very recently, the diffusion probabilistic model [14] has shown great potential in various low-level vision tasks [15–19]. The diffusion probabilistic model (DM) is a parameterized Markov chain with a variational inference process, which includes a diffusion process and a reverse process [20]. The diffusion process converts data samples x_0 into random noise $x_t, t \in [1, \ldots, T]$ by gradually adding noise σ, i.e., $x_0 \rightarrow x_1 \rightarrow \cdots \rightarrow x_t \rightarrow \cdots \rightarrow x_{T-1} \rightarrow x_T$. The reverse process is the opposite direction of the diffusion process, and the generation of data samples is achieved by repeatedly executing the inverse transformation of sampling, i.e., $x_{t-1} = f(x_t)$. The DM is trained by optimizing the variational lower bound on negative log-likelihood, it does not require regularization and optimization techniques to avoid optimization instability and mode collapse.

In this paper, we present a novel diffusion probabilistic model for traditional Chinese landscape painting super-resolution (CLDiff) to enhance the visual effect of the reconstructed image. Some methods [2, 4] only consider the advantage of the GAN while neglecting the potential difficulties in training, while other methods cannot balance the perceptual performance of the image. Unlike these methods, our proposed CLDiff is inspired by the denoising diffusion probabilistic model [20]. CLDifff is a condition image generation model that learns to convert a standard normal distribution to a Chinese landscape painting data distribution through an iterative refinement step. To sum up, the main contributions of this paper are as follows: (1) a novel denoising diffusion probabilistic model for the super-resolution task of traditional Chinese landscape painting (CLDiff) is proposed. CLDiff adopts a process similar to Langevin dynamics and utilizes parameterized Markov chain trained using variational inference to generate traditional Chinese landscape paintings gradually. (2) To further enhance the visual effect of traditional Chinese landscape painting super-resolution, an attention mechanism with an energy function is proposed based on insights from visual neuroscience. By introducing the proposed attention mechanism into the U-Net framework, CLDiff becomes a very effective super-resolution model for traditional Chinese landscape painting. (3) Different from the existing methods based on GAN, the proposed CLDiff avoids the problems of mode collapse and training instability.

Methodology

Related study

Denoising diffusion probabilistic model

Denoising diffusion probabilistic models are used to achieve high-quality image processing tasks. Moreover, there have been works indicating that the quality of the generated images has exceeded GAN. Recently, denoising diffusion probabilistic models have been widely used in image super-resolution and image inpainting. Saharia et al. [16] proposed a repeated refinement image super-resolution diffusion model, which achieves high-quality image super-resolution effects through an iterative refinement process. Li et al. [15] proposed a super-resolution diffusion probabilistic model for the face, which transforms a pure noise image into a face super-resolution result through a Markov chain. Saharia et al. [21] implemented four different image translation tasks using diffusion models and investigated the impact of loss functions and attention mechanisms on model performance. Whang et al. [17] proposed a conditional diffusion model, which uses the predict-and-refine strategy to make the sampling more effective and improve the quality of image deblurring. To address the image inpainting problem, Lugmayr et al. [18] achieved free-form inpainting only by improving reverse diffusion iterations. Additionally, diffusion models have also been successfully applied to medical image generation and object detection. Inspired by the above works, we extend the denoising diffusion probabilistic model to the super-resolution task of traditional Chinese landscape painting for the first time.

Image Super-resolution

Image super-resolution is a low-level vision task that aims to recover a high-resolution image from a low-resolution version. As a classical ill-posed inverse problem in the field of image processing, various solutions [22] have emerged in recent years. Dong et al. [23] first proposed a deep convolutional neural network for end-to-end low-resolution to high-resolution mapping. Ma et al. [24] proposed to use GAN for super-resolution tasks. This method utilizes the structural information of images to generate visually pleasing detail information. To advantage of neural architecture search (NAS), Pan et al. [25] proposed a Gaussian process based on NAS that won first place in the image super-resolution task. Considering the computational cost, Zhou et al. [26] proposed an SRFormer for image super-resolution, which can enjoy performance while also reducing resource consumption. These methods have excellent performance in super-resolution scenarios of natural images, which has great inspiration for the design of our model.

Attention mechanism

The proposal of attention mechanism reflects the application of biological mechanisms in artificial intelligence. Moreover, some studies [27–29] have achieved great success in applying attention mechanisms to low-level visual tasks. By adaptively adjusting the interdependencies between channels, the channel attention mechanism is introduced into the residual block to form a deep residual channel attention network [30], which realizes image super-resolution tasks. To refine the quality of image generation, the self-attention mechanism is integrated into the generative adversarial network [31] to improve the resolution of the generated image. Considering the neurons should be adjusted dynamically based on the context information, a context reasoning attention network [32] was put forward and realized the appealing image super-resolution effect. The latent diffusion model [19] introduces a cross-attention layer into the model architecture to improve the quality of generated images and the flexibility of the model. Different from these works, we design a novel attention mechanism to improve the quality of reconstructed images.

CLDiff

Inspired by the denoising diffusion probabilistic model [20], the proposed CLDiff is a conditional image generative diffusion model, which guides the reconstruction of high-quality traditional Chinese landscape paintings by conditional input low-resolution images. Given an input and output dataset $\{x, y\}$ of traditional Chinese landscape paintings, where x is the Chinese landscape painting, y is its corresponding low-resolution painting. We aim to train the model to learn an approximate conditional probability distribution $p(x|y)$. When the model training is completed, the pure noisy image is transformed into a Chinese landscape painting image through the iterative refinement process under the guidance of the conditional input low-resolution painting. Specifically, the proposed

CLDiff contains two processes: a forward Gaussian diffusion process and a reverse generation process, see Fig. 2.

The forward Gaussian diffusion process starts with a high-quality Chinese landscape painting image x_0, and gradually adds noise to x_0 through a T-step iterative process. This process is a forward Markov chain that transforms the data distribution $q(x_0)$ into a latent variable distribution $q(x_T)$. It can be defined as:

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}), \tag{1}$$

where a single-step diffusion model is defined as a Gaussian distribution:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t|\sqrt{\alpha_t}x_{t-1}, (1-\alpha_t)I), \tag{2}$$

where $\alpha_t \in (0,1), t \in [1,\ldots,T]$ is hyper-parameters, it controls the strength of the noise added each time. However, the efficiency of the single-step diffusion process is relatively low, and to improve the diffusion efficiency, it can be directly estimated x_t through a series of equation transformations:

$$q(x_t|x_0) = \mathcal{N}\left(x_t|\sqrt{\rho_t}x_0, (1-\rho_t)I\right), \tag{3}$$

where $\rho_t = \prod_{i=1}^{t}\alpha_i$. There are no unknown variables to learn in Eq. (3). This allows us to obtain the intermediate hidden variable x_t at any timestep,

$$x_t = \sqrt{\rho_t}x_0 + \sqrt{1-\rho_t}z, z \sim \mathcal{N}(0,I), \tag{4}$$

Therefore, in the forward Gaussian diffusion process, we use Eq. (3) to obtain x_T. When the timestep T is large enough, x_T can be seen as indistinguishable from pure Gaussian noise.

The reverse generation process is a stochastic denoising process that starts from the pure noise image $x_T \sim \mathcal{N}(0,I)$ and iteratively refines the image through a T-step reverse Markov chain. This process transforms the data distribution $p(x_T)$ of the latent variable into the data distribution $p(x_0)$ of the Chinese landscape painting. In the equation transformation of the forward Gaussian diffusion process, if x_0 and x_t are given, the posterior probability distribution of x_{t-1} can be obtained

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}|\mu_t(x_t, x_0), \beta_t I), \tag{5}$$

where

$$\mu_t(x_t, x_0) = \frac{\sqrt{\rho_{t-1}}(1-\alpha_t)}{1-\rho_t}x_0 + \frac{\sqrt{\alpha_t}(1-\rho_{t-1})}{1-\rho_t}x_t, \tag{6}$$

$$\beta_t = \frac{(1-\rho_{t-1})(1-\alpha_t)}{1-\rho_t}, \tag{7}$$

Equation (6) indicates that $\mu_t(x_t, x_0)$ depends on x_t and x_0. There are no variables in Eq. (7), so β_t is a deterministic value. Combining Eqs. (6) and (7), a one-step reverse Markov chain be obtained by sampling a slightly less noisy image x_{t-1} from x_t. According to Eq. (5), it seems that the high-quality Chinese landscape painting x_0 can be obtained through the reverse generation step T times. However, this is impractical because x_0 is unknown in Eq. (5), x_0 is exactly what we need to estimate. As shown by the red fork in Fig. 2. To solve this problem, the reverse generation process was successfully carried out to estimate x_0. Referring to [17, 19], we designed a denoising network f_θ to estimate the high-quality Chinese landscape painting $\bar{x}_0 = f_\theta(x_t, \rho_t)$ from the latent variable noisy image x_t. Therefore, we can utilize the estimate $f_\theta(x_t, \rho_t)$ to replace x_0 in Eq. (5), and the reverse generation process can be expressed as:

$$p(x_{t-1}|x_t) = q\left(x_{t-1}|x_t, f_\theta(x_t, \rho_t)\right), \tag{8}$$

To guide this reverse image super-resolution process, we take the conditional input low-resolution image y and the hidden variable $x_t = \sqrt{\rho_t}x_0 + \sqrt{1-\rho_t}\epsilon, \epsilon \sim \mathcal{N}(0,I)$ as the input of the denoising network. Equation (8) can be rewritten as follows:

$$p\left(x_{t-1}|x_t, y\right) = q\left(x_{t-1}|x_t, f_\theta\left(x_t, \rho_t, y\right)\right), \tag{9}$$

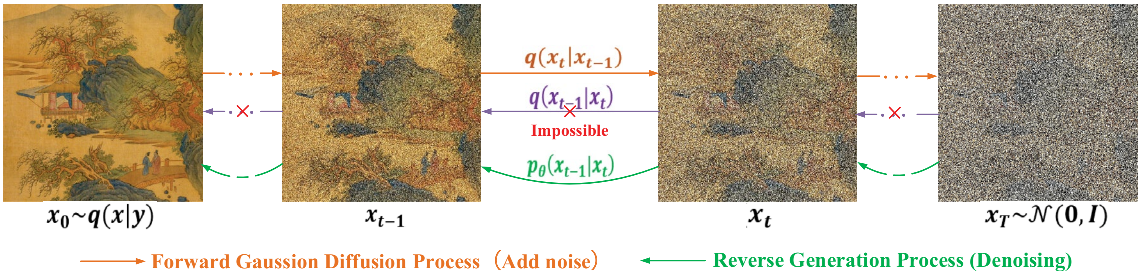

$q(x_t|x_{t-1})$
$q(x_{t-1}|x_t)$
Impossible
$p_\theta(x_{t-1}|x_t)$
$x_0 \sim q(x|y)$
x_{t-1}
x_t
$x_T \sim \mathcal{N}(0,I)$
Forward Gaussion Diffusion Process（Add noise）
Reverse Generation Process (Denoising)

Fig. 2 The forward Gaussian diffusion process and the reverse generation process in CLDiff

Figure 3 shows the single-step training process and the single-step inference process of our proposed model. Therefore, the reverse image super-resolution process procedure of our proposed model depends on the input condition y. Finally, the training objective is:

$$L(\theta) = E||f_\theta(\sqrt{\rho_t}x_0 + \sqrt{1-\rho_t}, \rho_t, y) - \epsilon||_1, \quad (10)$$

Based on the above analysis, the denoising network f_θ is an important part of the proposed model. Inspired by [19, 33], we adopt the attention mechanism to improve the U-Net architecture. The denoising network architecture in CLDiff is shown in Fig. 4.

CLDiff transforms the diffusion probabilistic model into a conditional traditional Chinese landscape painting super-resolution model by enhancing the U-Net backbone with the proposed attention mechanism. Existing attention mainly learns a weighted feature combination along the channel or spatial dimension to refine features. Channel attention generates 1-D weights and spatial attention generates 2-D weights. However, these two attention mechanisms do not fully conform to the principles of human visual neuroscience. In fact, human visual neurons are very sensitive to important features, and stimulated neurons can suppress surrounding neurons to highlight their importance [34]. Therefore, a novel attention mechanism is

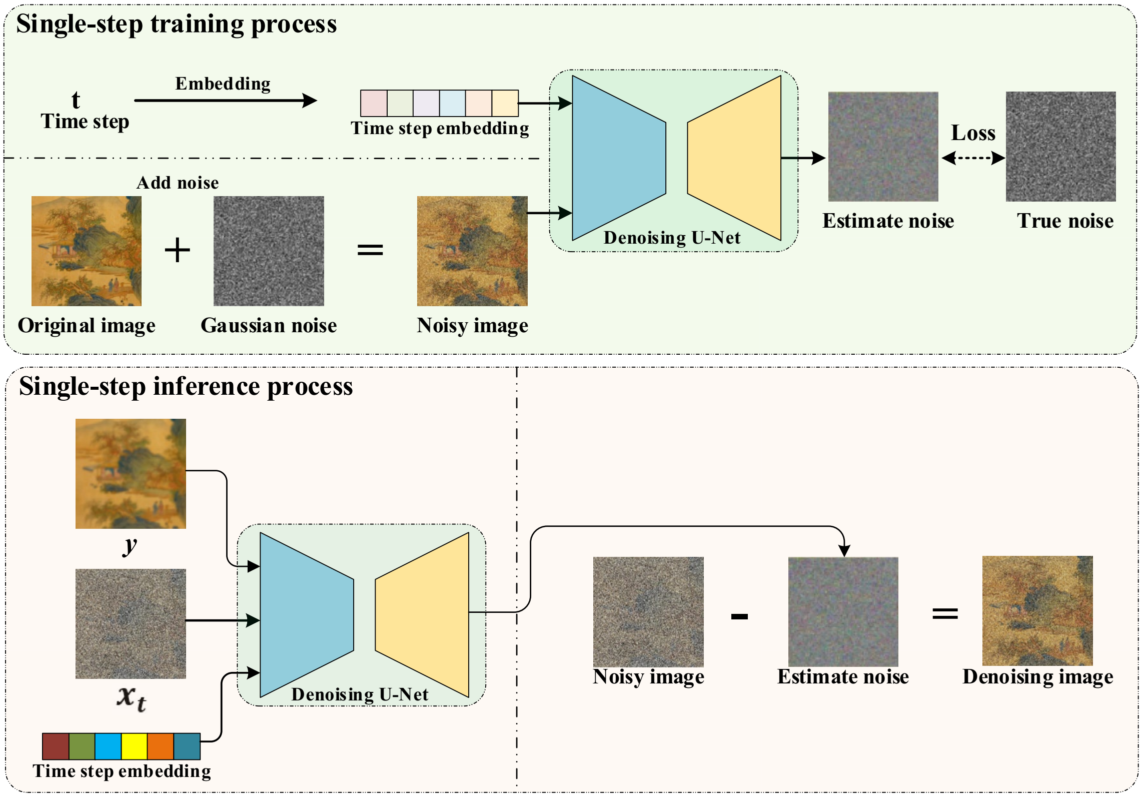


Fig. 3 Single-step training process and single-step reverse inference process

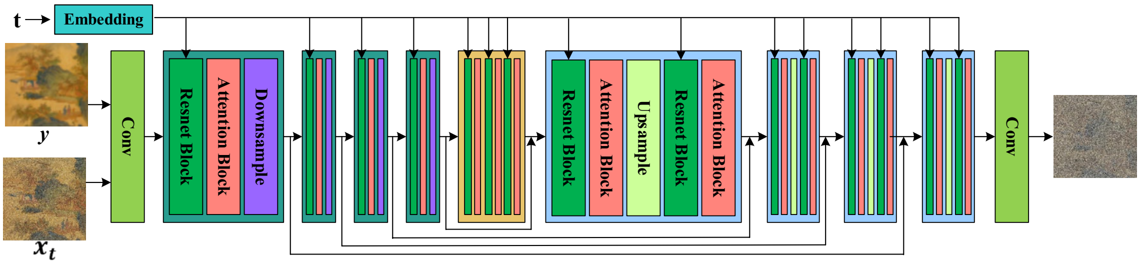


Fig. 4 The framework of the proposed CLDiff

proposed to enhance the ability of the U-Net to capture key features.

The proposed attention mechanism also includes the channel attention branch and spatial attention branch. The difference is that inspired by human visual neurons [35–37], we have introduced an energy function into the channel attention mechanism and the spatial attention mechanism, respectively. The energy function can further help the attention mechanism to enhance the key features while weakening the secondary features. We adopted the same strategy as in Refs. [38, 39] to parallel the two kinds of attention, and then the element-wise addition operation merges the two attention mechanisms. Figure 5 shows the structure of the attention mechanism with an energy function. The proposed attention mechanism can be expressed as:

$$f_{EA}(X) = f_{EC}(X) + f_{ES}(X), \tag{11}$$

where $X \in \mathbb{R}^{c \times h \times w}$ is the input tensor, f_{EC} represents the channel attention mechanism operation with energy function, $f_{EC}(X) = f_E(f_{CA}(X))$, and f_{ES} represents the spatial attention mechanism operation with energy function, $f_{ES}(X) = f_E(f_{SA}(X))$. f_{CA} is channel attention mechanism operation, it can be expressed as:

$$f_{CA}(X) = F_{SG}\left(W_k\left(W_v X \times F_{SM}\left(W_q X\right)\right)\right) \odot X, \tag{12}$$

and f_{SA} is spatial attention mechanism operation, it can be expressed as:

$$f_{SA}(X) = F_{SG}(F_{SM}(F_{GP}(W_q X)) \times W_v X) \odot X, \tag{13}$$

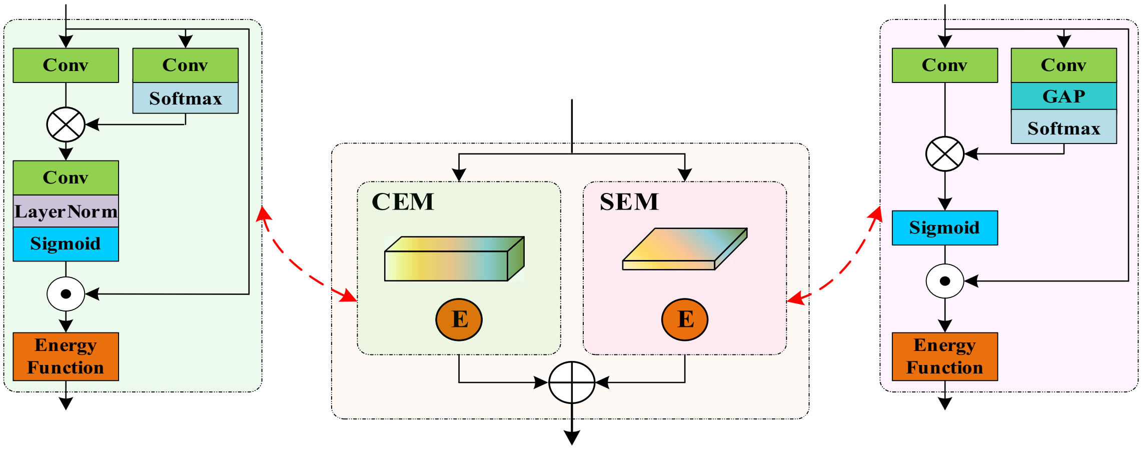
where, W_k, W_v, W_q are 1×1 convolution operations.F_{SG} is sigmoid operation, F_{SM} is softmax operation, F_{GP} is global average pool operation. F_E is energy function [34, 37]. To simplify and prevent overfitting, it can be expressed as a binary classification function with a regularization term:

$$F_E(.) = (1 - \hat{o}_t)^2 + \frac{1}{N-1}\sum_{i=1}^{N-1}(-1 - \hat{o}_i)^2 + \lambda J(w), \tag{14}$$

where $\hat{o}_t$ and $\hat{o}_i$ represents the output of the target neuron and surrounding neurons in a single channel of the input tensor X, respectively. $\hat{o}_t = w_t o_t + b_t$, $\hat{o}_i = w_t o_t + b_t$, w_t and b_t denote weight and bias. $N = h \times w$ is the number of neurons on the current channel. λ is the regularization parameter. $J(w)$ is the regularization term, it is the l_2 -norm of the parameter vector, i.e., $||w||_F^2$. Equation (14) represents the linear separability between the target neuron and the surrounding neurons. The stochastic gradient descent algorithm can reduce the computational burden of the energy function in each channel. This allows linearly separable operations to be implemented in deep learning frameworks. With the pixels in each channel following the same distribution, the minimum energy can be obtained by algebraic transformation:

$$e_t = \frac{4(\sigma^2 + \lambda)}{(o_t - \mu)^2 + 2\sigma^2 + 2\lambda}, \tag{15}$$

where $\mu = 1/N\sum_{i=1}^{N} o_i$, $\sigma^2 = 1/N\sum_{i=1}^{N}(o_i - \mu)^2$. The larger energy of neuron o_t can be obtained by $1/e_t$. The larger energy $1/e_t$, the neuron o_t is more important for capturing key features. To simulate the regulatory effect of mammalian attention mechanisms, the sigmoid function f_s was used to scale extreme data:



(a) Left **(b) Middle** **(c) Right**

Fig. 5 Attention mechanism with energy function

$$X = f_s\left(\frac{1}{E}\right) \odot X, \quad (16)$$

where E groups all e_t across the channel and spatial dimensions.

Experiments

Dataset and setting

The proposed model is implemented with the Pytorch framework and runs on a platform with two Nvidia RTX2080ti GPUs. The training dataset of the model is the traditional Chinese landscape painting dataset. Part of this dataset [4] is based on four open-access museum galleries: the Smithsonian Freer Gallery, Harvard University Art Museum, Princeton University Art Museum, and Metropolitan Museum of Art. The other part is collected from the Baidu image search engine. We used the crawler technology to obtain 1000 images from the Baidu search engine and selected 300 images with high quality by manual means. The data augmentation technique is applied to the Chinese landscape painting images to better train the proposed model. Figure 6 lists some examples of data augmentation effects.

Our model was trained for 1e6 epochs with a mini-batch size of 1. We set the timestep T = 2000. We set the forward Gaussian diffusion process to constants increasing linearly from 1e-6 to 1e-2. The U-Net adopts Adam optimizer with a learning rate of 3e-6. The trained U-Net was used to represent the reverse generation process.

Performance comparison and results

In terms of qualitative comparison, the corresponding SR results are shown in Figs. 7 and 8. Figure 7 shows the super-resolution (× 2) result at 256 × 256 → 512 × 512. One can see that the overall visual effect of all methods is good. Due to the lack of prior knowledge learned on large-scale datasets, ASDS [5], and NCSR [6] blur the details of the image and destroy the local semantic information of the image, while SwinIR [7] and CLDiff make the SR image texture clearer and the visual effect

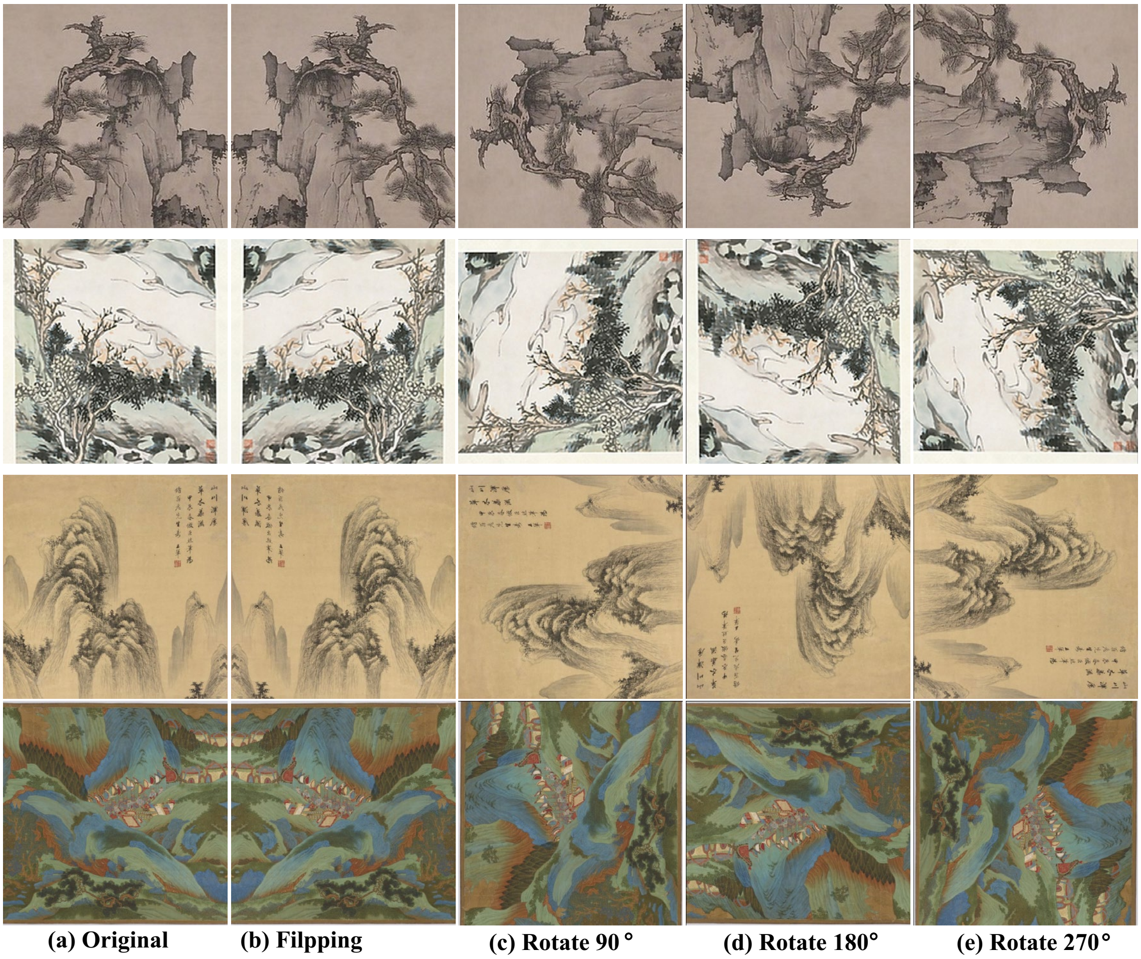


Fig. 6 Examples of the data augmentation effects

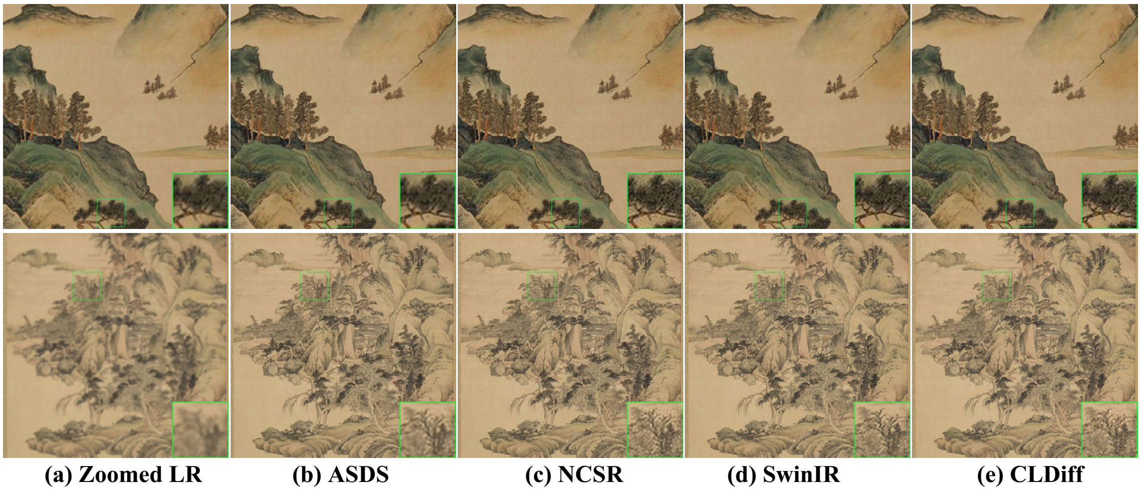


Fig. 7 Qualitative comparisons with different methods on the × 2 super-resolution task

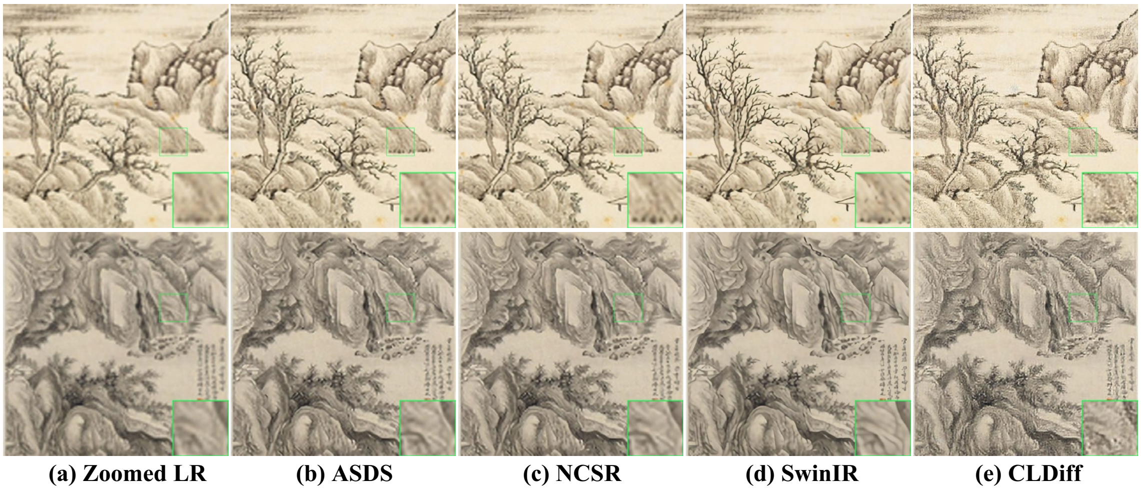


Fig. 8 Qualitative comparisons with different methods on the × 4 super-resolution task

better. Figure 8 shows the super-resolution (× 4) result at 128 × 128 → 512 × 512. From Fig. 8, it can be seen that for large-scale image super-resolution tasks, ASDS, and NCSR over-smooth the image content and lose the detail information of the image, while SwinIR and CLDiff have relatively good semantic information. Compared with the comparison methods, whether the scale is × 2 or × 4 SR tasks, the image lines and dot ink textures of the proposed CLDiff are more harmonious and clearer, with appropriate ink color and soft lines.

In terms of quantitative comparison, PSNR (peak signal-to-noise ratio), and SSIM (structural similarity) are used as quantitative metrics. PSNR evaluates the mean square error between the reconstructed image and the original image, and a larger value indicates a better quality of the reconstructed image. SSIM evaluates the similarity between the reconstructed image and the original image in terms of brightness, contrast, and structure. SSIM values range from 0 to 1. Fig shows the average PSNR and SSIM results. From Fig. 9, we can see that the proposed CLDiff has higher PSNR, the main reason is that reverse diffusion inference involves the iterative denoising process. Although the value of SSIM did not completely surpass the comparison method, this did not affect the quality of the reconstructed image. A similar conclusion has also been found in Ref. [40, 41]. Moreover, Fig. 10 shows that the image super-resolution quality and visual effect of the proposed CLDiff are better than or close to the original image.

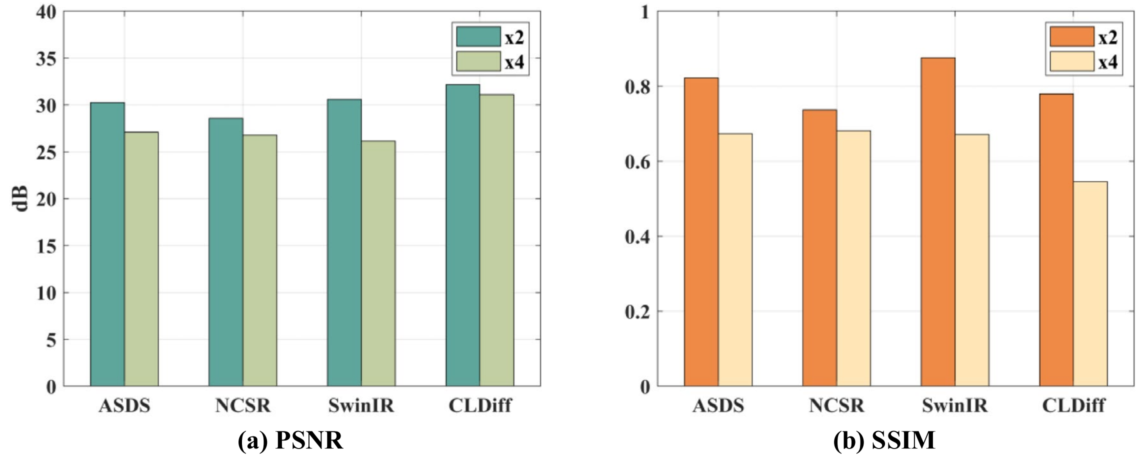
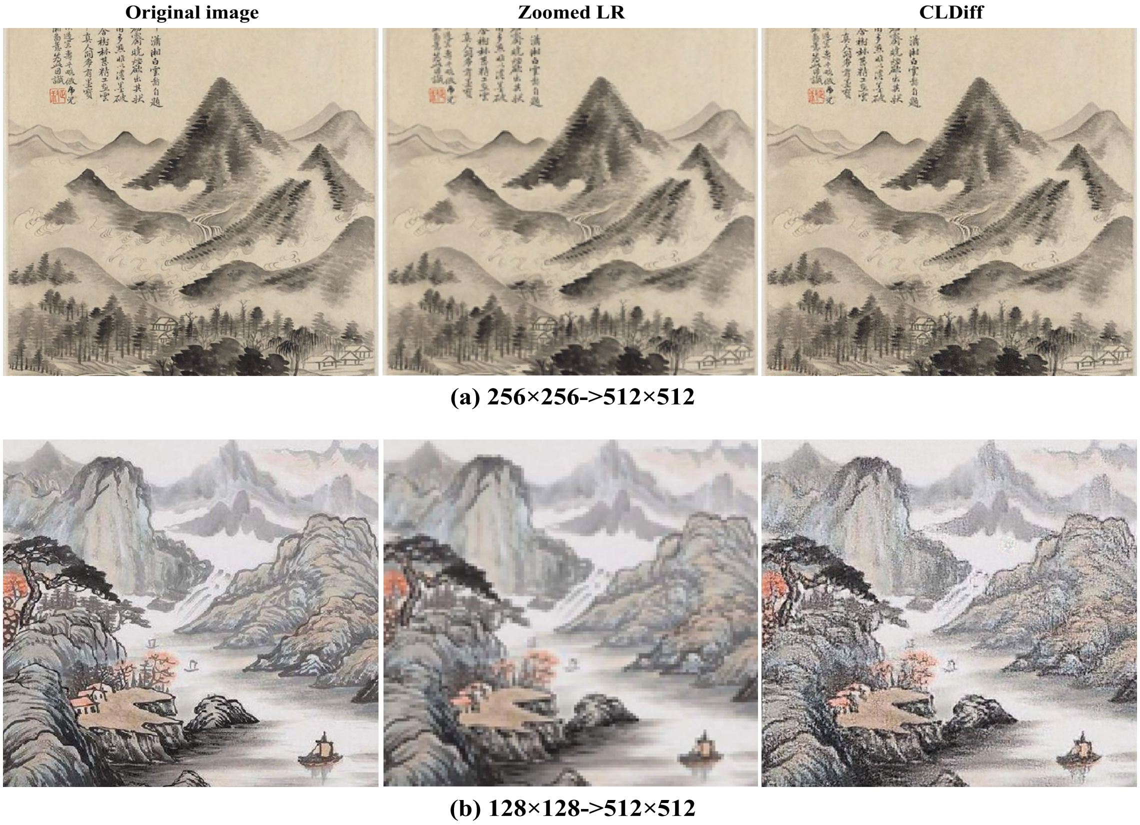


Fig. 9 Quantitative comparisons with different methods



(a) 256×256->512×512

(b) 128×128->512×512

Fig. 10 The results of CLDiff

Ablation studies

To evaluate the proposed attention mechanism, we conduct ablation experiments. We remove the proposed attention mechanism while ensuring that other parameters of the model remain unchanged during training, named CLDiff*. As can be seen from Fig. 11, the attention mechanism has an impact on the visual effect of image reconstruction. Removing the attention mechanism reduces the image reconstruction performance of the model. CLDiff* cannot reconstruct the image well, and there are obvious color spots on the reconstructed image. CLDiff* results in a significant

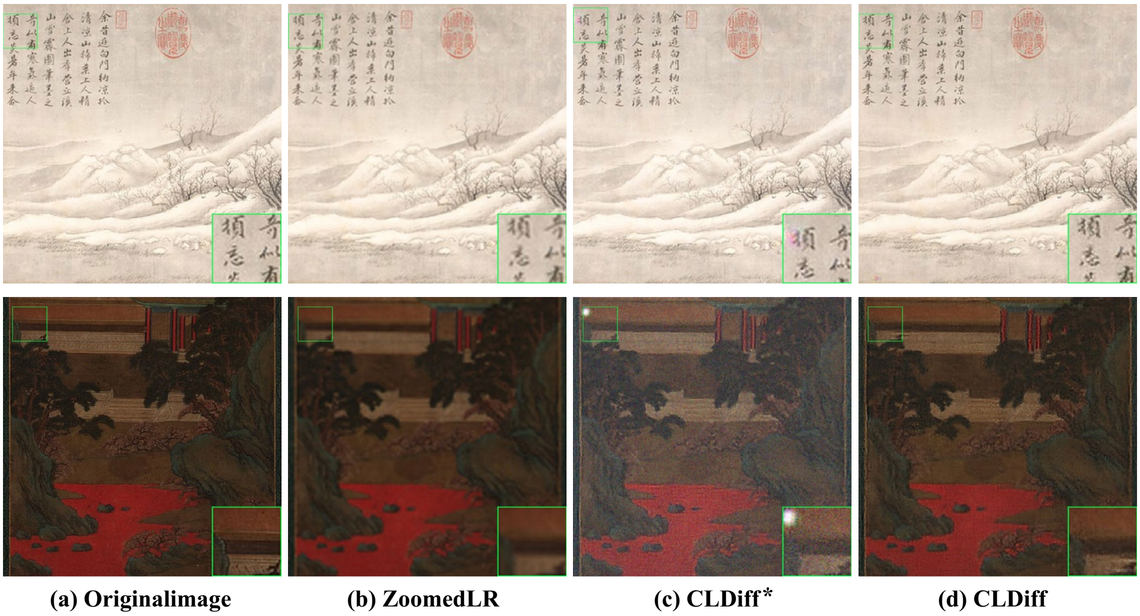


Fig. 11 The results of CLDiff* and CLDiff

color difference between the super-resolution image and the original image. CLDiff not only restores the color and texture of the original image but also has almost no color difference from the original image. Moreover, from Fig. 12, it can be seen that removing the attention mechanism does indeed affect the quantitative metrics of the model. After removing the attention mechanism, CLDiff* reduces the performance of the model compared with CLDiff.

Conclusion

To protect excellent traditional Chinese landscape paintings and alleviate the problem of low resolution in the digitization process of landscape paintings, we propose a diffusion model-based super-resolution method for traditional Chinese landscape paintings. The proposed CLDiff is similar to Langevin dynamics, which exploits a parameterized Markov chain to transform Chinese landscape paintings to latent variable distribution and then reconstruct super-resolution paintings in the reverse generation process which iteratively denoises the latent using an improved U-Net conditioned on low-resolution

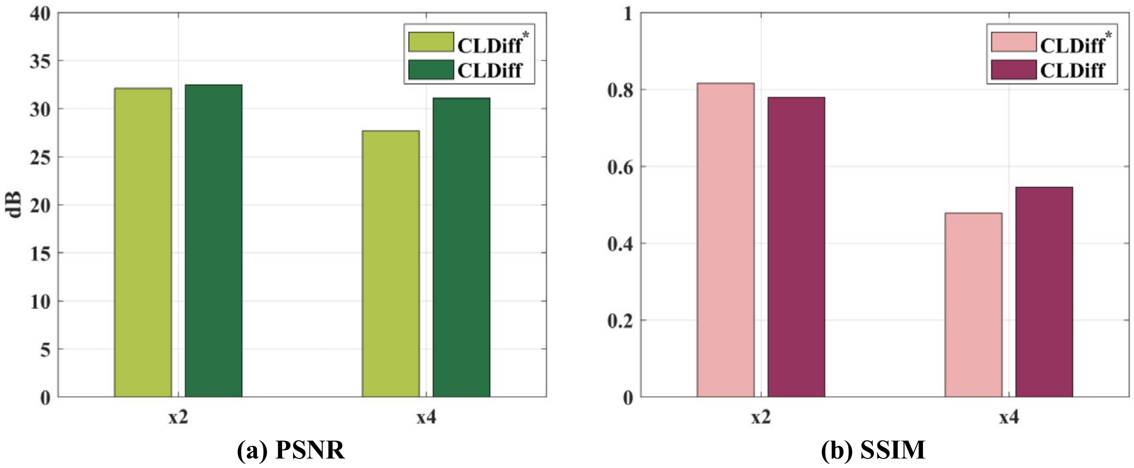


Fig. 12 The PSNR and SSIM results of CLDiff*and CLDiff

image. The proposed attention mechanism is helpful to enhance the visual quality of super-resolution images. Extensive experiments demonstrate that our proposed method has a good image super-resolution visual effect and avoids the problems of image over-smoothing and model training instability.

In the future, we will conduct further research from two aspects: (1) improving the performance of diffusion models and accelerating the inference speed of models. (2) We further explore the research of the diffusion model in the restoration and editing of traditional Chinese landscape paintings.

Abbreviations

LR	Low-resolution
HR	High-resolution
SR	Super-resolution
GAN	Generative adversarial network
DDPM	Denoising diffusion probabilistic model

Acknowledgements
None.

Author contributions
QL performed program design, and experiments and wrote the manuscript, NZ contributed to the review and revision of the manuscript, YY contributed to data management; YG polished the manuscript, JG performed the analysis with constructive discussions.

Funding
This work was supported by the Henan Province Key R&D Promotion Special Project (No. 232102210011, No. 232102521002), the Doctoral Research Start-up Fund of Pingdingshan University (No. PXY-BSQD-2023019), the key scientific research projects of universities in Henan Province (No. 24A520034, No. 24A520033), Henan Province colleges and universities young backbone teacher training program, Project of International Cultivation of Henan Advanced Talents.

Availability of data and materials
The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request. The code used for data analysis in this study can be obtained from the corresponding author upon reasonable request.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 7 November 2023 Accepted: 16 December 2023
Published online: 02 January 2024

References

1. Lin D, Wang Y, Xu G, Li J, Fu K. Transform a simple sketch to a chinese painting by a multiscale deep neural network. Algorithms. 2018;11(1):4. https://doi.org/10.3390/a11010004.
2. Lv X, and Zhang X. Generating Chinese classical landscape paintings based on cycle-consistent adversarial networks. 2019 6th International Conference on Systems and Informatics (ICSAI), Shanghai, China, 2019; 1265–1269. https://doi.org/10.1109/ICSAI48974.2019.9010358.
3. Zhou L, Wang QF, Huang K, Lo CH. An interactive and generative approach for Chinese Shanshui Painting Document, 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, NSW, Australia, 2019; 819–824. https://doi.org/10.1109/ICDAR.2019.00136.
4. Xue A. End-to-End Chinese landscape painting creation using generative adversarial networks, 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), 2021; 3862-3870. https://doi.org/10.1109/WACV48630.2021.00391
5. Dong W, Zhang L, Shi G, Wu X. Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Trans Image Process. 2011;20(7):1838–57. https://doi.org/10.1109/TIP.2011.2108306.
6. Dong W, Zhang L, Shi G, Li X. Nonlocally centralized sparse representation for image restoration. IEEE Trans Image Process. 2013;22(4):1620–30. https://doi.org/10.1109/TIP.2012.2235847.
7. Liang J, Cao J, Sun G, Zhang K, Van Gool L. and Timofte R. SwinIR: image restoration using swin transformer. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021;1833–1844. https://doi.org/10.1109/ICCVW54120.2021.00210.
8. 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. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017; 105–114, https://doi.org/10.1109/CVPR.2017.19.
9. Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y, Loy CC. ESRGAN: enhanced super-resolution generative adversarial networks. European Conference on Computer Vision (ECCV). 2018;11133:63-79. https://doi.org/10.1007/978-3-030-11021-5_5
10. Wang X, Xie L, Dong C. and Shan Y. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021;1905–1914. https://doi.org/10.1109/ICCVW54120.2021.00217.
11. Ravuri S, Vinyals O. Classification accuracy score for conditional generative models. 2019. https://doi.org/10.48550/arXiv.1905.10887.
12. Arjovsky M , Chintala S , Bottou L. Wasserstein GAN. 2017. https://doi.org/10.48550/arXiv.1701.07875.
13. Wu YL, Shuai HH, Tam ZR and Chiu HY. Gradient normalization for generative adversarial networks. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021; 6353–6362. https://doi.org/10.1109/ICCV48922.2021.00631.
14. Sohl-Dickstein J, Weiss E, Maheswaranathan N, Ganguli S. Deep unsupervised learning using nonequilibrium thermodynamics. Proceedings of the 32nd International Conference on Machine Learning (ICML), 2015; 37:2256–2265. https://proceedings.mlr.press/v37/sohl-dickstein15.html.
15. Li H, Yang Y, Chang M, Chen S, Feng H, Xu Z, Li Q, Chen Y. SRDiff: Single image super-resolution with diffusion probabilistic models. Neurocomputing. 2022;479:47–59. https://doi.org/10.1016/j.neucom.2022.01.029.
16. Saharia C, Ho J, Chan W, Salimans T, Fleet DJ, Norouzi M. Image super-resolution via iterative refinement. IEEE Trans Pattern Anal Mach Intell. 2023;45(4):4713–26. https://doi.org/10.1109/TPAMI.2022.3204461.
17. Whang J, Delbracio M, Talebi H, Saharia C, Dimakis AG and Milanfar P. Deblurring via stochastic refinement. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022;16272–16282, https://doi.org/10.1109/CVPR52688.2022.01581.
18. Lugmayr A, Danelljan M, Romero A, Yu F, Timofte R and Van Gool L. RePaint: inpainting using denoising diffusion probabilistic models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022;11451–11461. https://doi.org/10.1109/CVPR52688.2022.01117.
19. Rombach R, Blattmann A, Lorenz D, Esser P and Ommer B. High-resolution image synthesis with latent diffusion models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022;10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042.
20. Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. Adv Neural Inf Process Syst. 2020;33:6840–51.
21. Saharia C, Chan W, Chang H, Lee CA, Ho J, Salimans T, Fleet DJ and Norouzi M. Palette: image-to-image diffusion models. arXiv e-prints, 2021. https://doi.org/10.48550/arXiv.2111.05826.

22. Al-Mekhlafi H, Liu S. Single image super-resolution: a comprehensive review and recent insight. Front Comput Sci. 2024;18: 181702. https://doi.org/10.1007/s11704-023-2588-9.
23. Dong C, Loy CC, He K, Tang X. Learning a deep convolutional network for image super-resolution. European conference on computer vision (ECCV). 2014; 184–199.
24. Ma C, Rao Y, Lu J, Zhou J. Structure-preserving image super-resolution. IEEE Trans Pattern Anal Mach Intell. 2022;44(11):7898–911. https://doi.org/10.1109/TPAMI.2021.3114428.
25. Pan Z, Li B, Xi T, Fan Y, Zhang G, Liu J, Han J, Ding E. Real image super resolution via heterogeneous model ensemble Using GP-NAS. European conference on computer vision (ECCV). 2020; 423–436. https://doi.org/10.1007/978-3-030-67070-2_25.
26. Zhou Y, Li Z, Guo C-L, Bai S, Cheng M-M, Hou Q. SRFormer: permuted self-attention for single image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).2023;12780–12791.
27. Shi B, Darrell T and Wang X. Top-Down visual attention from analysis by synthesis. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023;2102–2112, https://doi.org/10.1109/CVPR52729.2023.00209.
28. Ouyang D. He S, Zhang G, Luo M, Guo H, Zhan J, Huang Z. Efficient multi-scale attention module with cross-spatial learning. ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023;1–5, https://doi.org/10.1109/ICASSP49357.2023.10096516.
29. Xiang J, Chen J, Liu W, Hou X, Shen L. RamGAN: region attentive morphing GAN for region-level makeup transfer. European Conference on Computer Vision (ECCV). 2022;13682:719-735. https://doi.org/10.1007/978-3-031-20047-2_41.
30. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y. Image super-resolution using very deep residual channel attention networks. European Conference on Computer Vision (ECCV). 2018; 11211:294-310. https://doi.org/10.1007/978-3-030-01234-2_18
31. Zhang H, Goodfellow I, Metaxas D, Odena A. Self-attention generative adversarial networks. 2018. https://doi.org/10.48550/arXiv.1805.08318.
32. Zhang Y, Wei D, Qin C, Wang H, Pfister H and Fu Y. Context reasoning attention network for image super-resolution. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 4258–4267, https://doi.org/10.1109/ICCV48922.2021.00424.
33. Liu Y, Wang Y, Li N, Cheng X, Zhang Y, Huang Y, Lu G. An attention-based approach for single image super resolution. 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 2018; 2777–2784. https://doi.org/10.1109/ICPR.2018.8545760.
34. Webb BS, Dhruv NT, Solomon SG, Tailby C, Lennie P. Early and late mechanisms of surround suppression in striate cortex of macaque. J Neurosci. 2005;25(50):11666–75. https://doi.org/10.1523/JNEUROSCI.3414-05.2005.
35. Carrasco M. Visual attention: the past 25 years. Vision Res. 2011;51(13):1484–525. https://doi.org/10.1016/j.visres.2011.04.012.
36. Aubry M, Russell BC, Sivic J. Painting-to-3D model alignment via discriminative visual elements. ACM Trans Graphics. 2014;33(2):1–14. https://doi.org/10.1145/2591009.
37. Yang L, Zhang R, Li L and Xie X. SimAM: A Simple, Parameter-free attention module for convolutional neural networks. Proceedings of the 38th International Conference on Machine Learning (ICML). 2021; 139:11863–11874. https://proceedings.mlr.press/v139/yang21o.html.
38. Liu H, Liu F, Fan X, Huang D. Polarized self-attention: towards high-quality pixel-wise mapping. Neurocomputing. 2022;506:158–67. https://doi.org/10.1016/j.neucom.2022.07.054.
39. Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. European Conference on Computer Vision. 2018; 11211:3–19. https://doi.org/10.1007/978-3-030-01234-2_1.
40. Berthelot D, Milanfar P, Goodfellow I. Creating high resolution images with a latent adversarial generator. 2020. https://doi.org/10.48550/arXiv.2003.02365.
41. Menon S, Damian A, Hu S, Ravi N and Rudin C. PULSE: self-supervised photo upsampling via latent space exploration of generative models. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020; 2434–2442, https://doi.org/10.1109/CVPR42600.2020.00251.

Publisher's Note

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