Journal of Imaging

MDPI

Article

Progressive Upsampling Generative Adversarial Network with Collaborative Attention for Single-Image Super-Resolution

Haoxiang Lu [1,2,3,4], Jing Zhang [5,*], Mengyuan Jing [1,2,3], Ziming Wang [4] and Wenhao Wang [4]

1 Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Sciences, Guangzhou 510080, China; hxlu1005@163.com (H.L.); jingmengyuan@gdph.org.cn (M.J.)
2 Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Southern Medical University, Guangzhou 510080, China
3 Guangdong Provincial Key Laboratory of Artificial Intelligence in Medical Image Analysis and Application, Guangzhou 510080, China
4 School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541004, China; worthyman@guet.edu.cn (Z.W.); whwang2018@163.com (W.W.)
5 School of Business, Guilin Institute of Information Technology, Guilin 541100, China
* Correspondence: zhangjing@guit.edu.cn

Abstract

Single-image super-resolution (SISR) is an essential low-level visual task that aims to produce high-resolution images from low-resolution inputs. However, most existing SISR methods heavily rely on ideal degradation kernels and rarely consider the actual noise distribution. To tackle these issues, this paper presents a progressive upsampling generative adversarial network with collaborative attention mechanism called PUGAN. Specifically, the residual multiscale blocks (RMBs) based on stacked mixed-pooling multiscale structures (MPMSs) is designed to make full use of multiscale global–local hierarchical features, and the frequency collaborative attention mechanism (CAM) is used to fully dig up high- and low-frequency characteristics. Meanwhile, we design a progressive upsampling strategy to guide the model's learning better while reducing the model's complexity. Finally, the discriminator is also used to evaluate the reconstructed high-resolution images for balancing super-resolution reconstruction and detail enhancement. Our PUGAN can yield comparable PSNR/SSIM/LPIPS values for the NTIRE 2020, Urban 100, and B100 datasets, whose values are 33.987/0.9673/0.1210, 32.966/0.9483/0.1431, and 33.627/0.9546/0.1354 for the scale factor of $\times 2$ as well as 26.349/0.8721/0.1975, 26.110/0.8614/0.1983, and 26.306/0.8803/0.1978 for the scale factor of $\times 4$, respectively. Extensive experiments demonstrate that our PUGAN outperforms state-of-the-art SISR methods in qualitative and quantitative assessments for the SISR task. Additionally, our PUGAN shows the potential benefits to pathological image super-resolution.

Keywords: single-image super-resolution; generative adversarial network; progressive upsampling; attention mechanism

Check for updates

Academic Editor: Gabriel Cristobal

Received: 12 January 2026
Revised: 4 February 2026
Accepted: 10 February 2026
Published: 11 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Low-resolution (LR) images encounter uncomfortable visual quality (e.g., blurry details and outlines, low peak signal-to-noise ratio, etc.) and compromised unreliable delivery of information [1,2]. The former brings about an unsatisfactory visual experience, while the latter may be harmful to the advanced visual tasks, such as inaccurate image understanding and object detection. Therefore, promoting the resolution of the images has proven highly effective for image understanding. Single-image super-resolution (SISR) can reconstruct a

visually high-resolution (HR) images with clearer details and outlines from the corresponding LR versions, which has been widely applied in computer-aided diagnosis (CAD) [3], advanced driver assistance systems (ADAS) [4], remote sensing [5,6], and other real-world practical applications. Pathological images contain tumor microenvironment (TME), including tumor epithelial, tumor-infiltrating lymphocytes (TILs), tumor-associated stroma, etc., clinically related to the occurrence, development, and metastasis of tumors. In clinical diagnosis, pathologists conduct a step-by-step analysis of the pathological images from low magnification to high magnification. But the scanning of high-magnification pathological images is time-consuming and labor-intensive, and the storage is also challenging. Many researchers [7–9] have proven that the SISR technology can yield high-magnification pathological images from their low-magnification versions, whereas the inherently ill-posed nature of SISR makes it a challenging problem.

In earlier times, interpolation-based [10], reconstruction-based [11,12], example-based [11], dictionary learning-based [13,14], and other traditional SSIR approaches have been proposed for yielding HR images. Among them, the interpolation-based methods [10] (such as bilinear interpolation, bicubic interpolation, etc.) and the reconstruction-based methods (such as maximum a posteriori probability estimation, convex set projection, etc.) can estimate the pixels of the HR images based on the LR inputs, but they encounter poor robustness. The latter two methods struggle to construct the knowledge base regarding the mapping from the LR inputs to their HR versions. However, they typically contain resource-intensive operations and inevitably introduce observable handcraft-halos in the HR images. Benefiting from the rapid development of deep learning technology and computing resources, many researchers have proposed learning-based SISR approaches, including convolutional neural networks (CNNs) [15,16], generative adversarial networks (GANs) [17,18], diffusion model [19], etc., for exploring a reliable nonlinear mapping between the LR images and their corresponding HR ones. For example, Dong et al. [15] first illustrated that traditional sparse-coding-based SR methods can be regarded as a deep CNN, and further presented a CNN-based SISR method to directly learn an end-to-end mapping between the LR/HR paired images. Umer et al. [20] employed an adversarial strategy to train the model with pixel-wise supervision in the HR domain from its generated LR counterpart. Shang et al. [16] further proposed a diffusion probabilistic model (DPM) based on residual structure, which utilizes a CNN to restore primary low-frequency components and a DPM to predict the residual between the ground truth image and the CNN-predicted image. Although most of learning-based SISR approaches can generate visually pleasing HR images, they inevitably generate artifact-halos and blurry details due to the ideal degradation means (i.e., bicubic down-sampled, etc.) and rarely take the real noise (i.e., inherent sensor noise, stochastic noise, etc.) into account.

To solve the above-listed problems, we present a feasible and effective SISR method named PUGAN. This SISR method contains three main components: a noise collection and frequency decomposition, a frequency collaborative progressive generator (FCPG), and an image perceptual discriminator (IPD). In the first stage, a sliding window is employed to collect noise from the LR input to construct a noise pool, and the convolutional Gaussian filtering is used to extract the high- and low-frequency information of the input. In the FCGN stage, we introduce some residual multiscale blocks (RMBs) based on the mixed-pooling multiscale structure (MPMS) to fully explore multiscale features at global or local levels, and the collaborative attention mechanism (CAM) can make full use of the complementarity of high- and low-frequency information. Meanwhile, we introduce a progressive upsampling strategy to better guide the model's learning while reducing the model's complexity. In the IPG stage, the discriminator is hired to evaluate the reconstructed HR images for balancing super-resolution reconstruction and detail enhancement. The

validation experiments conducted on the public benchmarks have demonstrated that our proposed method works better than state-of-the-art SISR methods.

The main contributions of this work are summarized as follows:

(1) We propose a progressive upsampling generative adversarial network with collaborative attention (called PUGAN) for the SISR task. Extensive experiments demonstrate that our method achieves the state-of-the-art performance and enjoys high efficiency. Meanwhile, our PUGAN also generalizes well to pathological images.

(2) We design an FCPG containing RMBs and CAM to generate the HR images. The RMB built upon MPMS in a dual-patch manner to fully explore multiscale global and local features, further enhancing the model's multiscale representation capability. Meanwhile, the CAM is employed to explore the complementarity of high- and low-frequency features.

(3) We construct a noise pool using the sliding window to collect noise from the LR input and further select noise randomly to simulate real noise. Meanwhile, the progressive upsampling strategy is further introduced to our PUGAN to guide the model's learning better while reducing the model's complexity.

We demonstrate the organization of the remainder in this paper as follows. In Section 2, we review the previous work related to our method, including learning-based SISR methods and attention mechanisms. In Section 3, the motivation and architecture of our proposed method are presented in detail. In Section 4, we give the implementation details and experimental settings, as well as an ablation study and comparisons with state-of-the-art SISR methods. In the end, we present the conclusion of our work in Section 5.

2. Related Works

Over the past decade, many SISR approaches have been proposed. We outline some related learning-based SISR methods and attention mechanisms in this section.

2.1. Learning-Based SISR Methods

Ranging from the first SISR network SRCNN [15], learning-based SISR approaches relying on the powerful feature extraction and representation capability have spurred dramatic improvements from different perspectives. Lim et al. [21] injected the residual learning techniques into the CNN to develop an enhanced deep SR network (EDSR) for reconstructing HR images of different upscaling factors in a single model. Kong et al. [22] further proposed a ClassSR to improve image resolution using different methods based on the difficulty of sub-images. However, these CNN-based methods primarily focus on local structures and details, inevitably leading to unclear details in some enhanced images. Hence, Zhao et al. [23] employed a spatial shuffle multi-head self-attention for global pixel dependency modeling. Li et al. [24] employed the high-frequency enhancement residual block to extract high-frequency information, and further enhanced the global and local features by the shift rectangle window attention and the hybrid fusion blocks. These have proven that the vision transformer (ViT) can capture global information by effectively extracting long-range dependencies, while overlooking the importance of high-frequency features. So, Wu et al. [25] combined CNN with transformer to propose a hybrid SP-MISR network for sequential images with fixed sub-pixel shifts. Liu et al. [26] proposed a hybrid Mamba–Transformer model for SISR for effectively leveraging both Mamba and Transformer architectures. However, these methods confront a heavy computational burden and overly rely on paired images, limiting their real-world applications.

To reduce the reliance on synthetic datasets, Wang et al. [27] proposed an enhanced super-resolution GAN (ESRGAN) based on the residual-in-residual dense block (RRDB) without batch normalization, adversarial loss, and perceptual loss. Notably, the LR images obtained through known degradation (for instance, bicubic downsampling) from the

HR images cannot accurately describe the real image degradation. To tackle this issue, Prajapati et al. [28] designed a new direct unsupervised super-resolution method using GAN (called DUS-GAN) to accomplish the SR task without degradation estimation of real-world LR data. Ma et al. [18] presented a detail-enhanced generative adversarial network to better reconstruct the image details even with a limited number of training samples. Dong et al. [29] performed a cross dropout-based dynamic network (CDDNet), using the degradation of LR images using degradation weights as the global attention, for multi-degradation blind SR. Cho et al. [30] leveraged the connection between degradation kernel shapes and the frequency-domain characteristics of LR images to simplify the kernel estimation process.

2.2. Attention Mechanism

In human perception, attention is the process by which the visual system prioritizes a sequence of information, selectively focusing on salient stimuli [31,32]. Inspired by this mechanism, some researchers proposed attention mechanisms to further promote the feature extraction and representation ability of learning-based methods, which have been widely applied in natural language processing, medical image analysis, natural image processing, and other computer vision applications [33–35]. Park et al. [36] proposed the bottlenet attention module (BAM), which dynamically selects features in a self-inclusion and adaptive manner. Subsequently, the convolutional block attention module (CBAM) [37] was proposed by combining channel attention with spatial attention to explore the relationship between internal channels and enhance the implicit information correlation. Zhang et al. [38] integrated the attention mechanism with the residual idea to design the residual channel attention mechanism module to optimize the model's ability to extract details. Sun et al. [35] replaced the ordinary convolutions in the attention mechanism module with multiscale dilated convolution to make it extract feature images at different receptive field scales. Wang et al. [39] designed a multiscale large kernel attention (MLKA), including multiscale structure and gate schemes to obtain the abundant attention map at various granularity levels. Wu et al. [40] integrated the channel attention and self-attention mechanisms to design a hybrid attention calibration mechanism. Guo et al. [41] developed a hybrid attention-dense connected Transformer network containing an effective dense Transformer block (EDTB) and a hybrid attention block (HAB) for effective performance on SR tasks with magnification factors of 2, 3, and 4. Malkocoglu et al. [42] designed a deep channel attention SR model for promoting the performance of object detection. Su et al. [43] introduced a concise yet effective soft thresholding operation to obtain high-similarity-pass attention. Zhang et al. [44] proposed an efficient shuffle attention (SA) module to address this issue, which adopts Shuffle Units to combine two types of attention mechanisms effectively. However, most of the existing attention modules suffer from low computational efficiency and rarely consider valuable structural priors.

3. Methodology

In this section, we first briefly introduce the overall network architecture of our proposed PUGAN. Subsequently, the noise collection and frequency decomposition, FCPG, IPD, as well as loss functions are demonstrated in detail.

3.1. Network Architecture

This paper proposes a progressive upsampling generative adversarial network with collaborative attention named PUGAN, and the workflow of our PUGAN is demonstrated in Figure 1. Our PUGAN involves three main components: noise collection and frequency decomposition, FCPG, and IPD. Following reference [45], we introduce a discriminator

network D_{θ_D}, which is optimized alternately with the generator G_{θ_G} to solve the adversarial min–max problem:

$$\min_{\theta_G}\max_{\theta_D} \mathrm{E}_{\mathbf{I}^{\mathrm{HR}}\sim p_{\mathrm{train}}(\mathbf{I}^{\mathrm{HR}})}\left[\log D_{\theta_D}\left(\mathbf{I}^{\mathrm{HR}}\right)\right] + \mathrm{E}_{\mathbf{I}^{\mathrm{LR}}\sim p_G(\mathbf{I}^{\mathrm{LR}})}\left[\log\left(1 - D_{\theta_D}\left(G_{\theta_G}\left(\mathbf{I}^{\mathrm{LR}}\right)\right)\right)\right] \quad (1)$$

The core idea of this formulation is to train a generative model G to deceive a differentiable discriminator D, which is itself trained to distinguish super-resolved images from real ones. Therefore, G learns to produce outputs so realistic that they become difficult for D to classify.

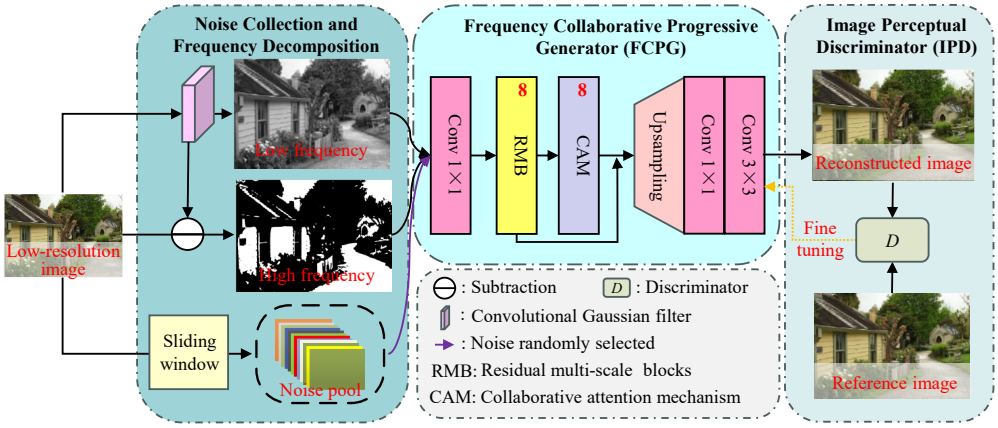


Figure 1. The workflow of our PUGAN. In first stage, the sliding window is used to construct noise pool and further extract high/low-frequency information from the LR input. In our FCPG, it first employs eight residual multiscale blocks (RMBs) built upon a mixed-pooling multiscale structure (MPMS) to fully explore multiscale local–global features. Then, eight collaborative attention mechanism (CAM) blocks are further used to utilize the complementarity of high/low-frequency information. In our IPG, we employ a discriminator to balance super-resolution reconstruction and detail enhancement.

3.2. Noise Collection and Frequency Decomposition

Currently, most existing SISR methods usually adopt the ideal bicubic sampling as the degradation kernels, which cannot accurately describe the image degradation, resulting in artifact-halos in some reconstructed HR images. Hence, inspired by [46,47], we perform a slide window to collect noise from the LR input, and we further randomly select noise to simulate the distribution of the image noise captured in the real world. Specifically, given a global patch P_i with the size of $d \times d$ and a local patch Q_j^i (notably, each global patch P_i is extracted by scanning the noisy LR input with a stride s_g, and each Q_j^i can be obtained by scanning inside P_i with a stride s_l). We determine whether a given P_i is a smooth patch by evaluating the differences in mean and variance between P_i and each of its local patches Q_j^i. More precisely, two constraints are first defined as follows:

$$\begin{cases} \left|Mean\left(Q_j^i\right) - Mean(P_i)\right| \le \alpha \cdot Mean(P_i) \\ \left|Var\left(Q_j^i\right) - Var(P_i)\right| \le \beta \cdot Var(P_i) \end{cases} \quad (2)$$

where $Mean(\cdot)$ and $Var(\cdot)$ calculate the mean and variance of a global patch P_i and a local patch Q_j^i, and $\alpha, \beta \in (0, 1)$. If both constraints are simultaneously satisfied for every local patch Q_j^i, the global patch P_i is classified as a smooth patch and added to the set $S = \{s_1, s_2, \cdots s_t\}$.

So, the noise pool Ω can be derived by

$$\Omega = \underset{i \in \{1,2,\cdots,t\}}{\text{collect}} (s_i - Mean(s_i)) \tag{3}$$

where collect$(\cdot)$ denotes the noise collection operation. Notably, we perform the anisotropic Gaussian kernel on the paired synthetic data to simulate the generation of real noise. For unpaired data, the KernelGAN is regarded as a degraded kernel to estimate the distribution of the image noise under the constraints of Equation (4).

$$G(\mathbf{I}_{LR}) = \arg\min_{G} \max_{D} \left\{ \underset{x \in patches(I_{LR})}{E} \left[(D(X) - 1)^2 + D(G(x))^2 \right] + R \right\} \tag{4}$$

where R is regular terms, X is image patch, and $\mathbf{I}_{LR}$ is the input LR images.

Generally, the characteristics of the image manifest differently at different frequency spaces [24,48]. The content of the image is mainly expressed in its low-frequency components, while details, contours, noise, etc., are mainly expressed in high-frequency components. Based on this concept, we perform the convolutional Gaussian filter on the LR input to extract high- and low-frequency features. This stage can be formulated as

$$\begin{cases} \mathbf{X}_{\mathrm{L}} = \mathbf{W}_{\mathrm{L,Gaussian}} * \mathbf{X}_{\mathrm{LR}} \\ \mathbf{X}_{\mathrm{H}} = \mathbf{X}_{\mathrm{LR}} - \mathbf{X}_{\mathrm{L}} \end{cases} \tag{5}$$

where $\mathbf{X}_{\mathrm{LR}}$ is the input LR image, and $\mathbf{X}_{\mathrm{L}}$ and $\mathbf{X}_{\mathrm{H}}$ are low- and high-frequency features of the input. $\mathbf{W}_{\mathrm{L,Gaussian}}$ represents the convolutional Gaussian filter. $*$ represents the convolution operation. After that, we utilize the FCPG to process these high- and low-frequency components to make details clearer and remove inherent noise simultaneously. Finally, the generated images are further fed into the IPD for evaluating image quality.

3.3. Frequency Collaborative Progressive Generator

In our carefully designed PUGAN, the FCPG is used to generate HR images with clearer details and visually comfortable experience, and the structure of our FCPG is shown in Figure 1. The FCPG employs the RMB to explore multiscale features and the CAM to explore the complementarity and correlation of high/low-frequency features.

RMB: The characteristics of the image in different scale spaces manifest differently. Hence, we design a high-efficiency RMB [see Figure 2] by stacking mixed-pooling multiscale-structure (MPMS) blocks in a dual-residual path manner (reported in [49]), which can make full use of the intermediate features.

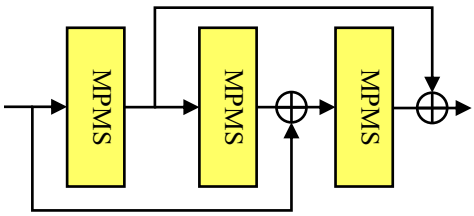


Figure 2. The structure of the RMB containing stacked three MPMSs using the dual-residual path.

Specifically, the features extracted by the j_{th} MPMS block in the i_{th} RMB can be expresed as

$$F_{i,j} = H_{i,j}^{\mathrm{MPMS}}(F_{i,j-1}) \tag{6}$$

where $H_{i,j}^{\mathrm{MPMS}}(\cdot)$ denotes the j_{th} MPMS blocks in the i_{th} RMB. $F_{i,j}$ and $F_{i,j-1}$ are the output of j_{th} and $j-1_{th}$ MPMS block in the i_{th} RMB. After that, we further perform the dual-residual path to fully explore the intermediate informative features, namely

$$F_i = F_{i,j-1} + F_{i,j+1}\left(F_{i-1} + F_{i,j}\right) \tag{7}$$

where F_i and F_{i-1} denote the output generated by the i_{th} and $i-1_{th}$ RMBs. This carefully designed dual-residual path connection can guarantee more information to be bypassed.

MPMS: The image features exhibit different representations at different scales and levels. Therefore, we employ the MPMS, a trip-branch structure (see Figure 3), to fully dig up such multiscale and hierarchical features. In the top branch, we first perform the parametric rectified linear unit (PReLU) function on the input, and further concatenate multiscale features extracted by a depthwise convolution with a dilation rate $r \in \{1,2,3\}$. This stage can be defined as

$$F'_{\mathrm{in}} = \mathrm{cat}\left(\underset{r\in\{1,2,3\}}{\mathrm{DWConv}_3^r}\left(H_{\mathrm{PReLU}}(F_{\mathrm{in}})\right)\right) \tag{8}$$

where $\underset{r\in\{1,2,3\}}{\mathrm{DWConv}_3^r}(\cdot)$ represents a depthwise convolution with a dilation rate $r \in \{1,2,3\}$, and $H_{\mathrm{PReLU}}(\cdot)$ denotes parametric rectified linear unit function. Subsequently, the output is processed by the 1×1 Conv, and further adds the feature extracted by the 27×27 Conv to it. Finally, the added result is processed by the squeeze-and-excitation networks (SE Net) block and PReLU to obtain the multiscale features F'_{multi}, i.e.,

$$F'_{\mathrm{multi}} = H_{\mathrm{PReLU}}\left(H_{\mathrm{SE}}\left(\mathrm{Conv}_1\left(F'_{\mathrm{in}}\right) \oplus \mathrm{Conv}_{27}(F_{\mathrm{in}})\right)\right) \tag{9}$$

where $\mathrm{Conv}_1(\cdot)$ and $\mathrm{Conv}_{27}(\cdot)$ denote a convolution withe the size of 1×1 and 27×27, respectively. $H_{\mathrm{SE}}(\cdot)$ denotes the SENet block.

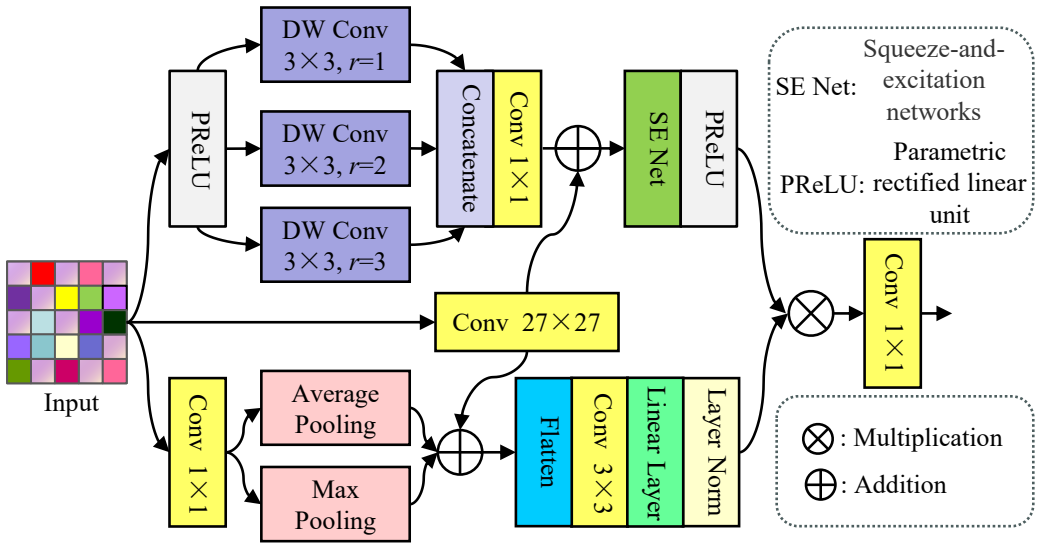


Figure 3. The structure of the MPMS, which is a trib-branch structure. The top branch mainly consists of depthwise convolution with different dilation rates, the middle branch only consists of a convolution layer with the size of 27×27, and the bottom branch mainly consists of average and max pooling. These carefully designed structures can fully explore multiscale features at local or global levels.

In the bottom branch, we first perform 1×1 Conv to extract low-level features, and they are fed into average and max pooling to dig up hierarchical features. After that, these features are added to the features yielded by the 27×27 Conv, i.e.,

$$F'_{\text{hie}} = H_{\text{AP}}(\text{Conv}_1(F_{\text{in}})) \oplus H_{\text{MP}}(\text{Conv}_1(F_{\text{in}})) \oplus \text{Conv}_{27}(F_{\text{in}}) \tag{10}$$

where $H_{\text{AP}}(\cdot)$ and $H_{\text{MP}}(\cdot)$ denote average and max pooling, respectively. Finally, we employ successive Flatten, 3×3 Conv, Linearization, and Normalization to generate hierarchical features. And we further integrate hierarchical and multiscale features using the pixel multiplication to generate the mixed features F_{mix}.

$$F_{\text{mix}} = F'_{\text{mul}} \otimes F'_{\text{hie}} \tag{11}$$

where $\otimes$ stands for the multiplication. Overall, the RMB employs the MPMS to multiscale hierarchical features, and the dual-patch skip connection is used to stack three MPMS to make full use of multilevel features.

CAM: In our designed FCPG, the CAM is used to fully explore the complementarity and correlation of high-frequency and low-frequency characteristics. The structure of the CAM is illustrated in Figure 4, which is a dual-branch structure containing high- and low-frequency processing branch. For the high-frequency information, we first employ the 1×1 Conv and detail-preserving pooling to extract features with abundant details, and further fed them into the parallel channel attention (CA) and spatial attention (SA) to explore the correlation of features at the channel and spatial level. And we concatenate them for aggregating these features and further processed by the 3×3 Conv and normalization layer, i.e.,

$$\begin{cases} \hat{F}_h = \text{cat}(H_{\text{SA}}(\text{Conv}_1(H_{\text{DP}}(F_h))), H_{\text{CA}}(\text{Conv}_1(H_{\text{DP}}(F_h)))) \\ \hat{F}'_h = \text{Conv}_3(H_{\text{Norm}}(F_h)) \end{cases} \tag{12}$$

where $H_{\text{DP}}(\cdot)$ denotes the detail preserving pooling. $H_{\text{CA}}(\cdot)$ and $H_{\text{SA}}(\cdot)$ stand for the channel attention (CA) and spatial attention (SA), respectively. $H_{\text{Norm}}(\cdot)$ denotes the normlization layer. $\text{Conv}_3(\cdot)$ denotes a convolution with the size of 3×3. After that, the weight F^l_{Kspa} yielded by the K-spaCA from the low-frequency branch multiplied by it, and the high frequency information F_h processed by the 1×1 Conv is added. Finally, the output is processed by the successive MLP, 3×3 Conv, and linear layers to generate the final high-frequency features F'_h, and it can be formulated as

$$F'_h = H_{\text{Linear}}\left(\text{Conv}_3\left(H_{\text{MLP}}\left(\left(\hat{F}' \odot F^l_{\text{Kspa}}\right) \oplus \text{Conv}_1(F_h)\right)\right)\right) \tag{13}$$

where $\oplus$ and $\odot$ represent pixel addition and multiplication, $H_{\text{MLP}}(\cdot)$ denotes the multi-layer perceptron, and $H_{\text{Linear}}(\cdot)$ denotes linear layer.

For the low-frequency information, we employ the successive 1×1 Conv, 3×3 Conv, and SENet block to extract low-level features from the low-frequency information $\hat{F}_l$, and the output is further fed to the K-Sparse channel attention (K-spaCA). The K-spaCA employed two scoring strategies to select the most important K_1 and K_2 channels, and we further add them by weight factors α and β (empirically set them as 0.1 and 0.3, respectively.) and processed by the softmax. Subsequently, the output F^l_{Kspa}, the feature $\text{Conv}_1(F_l)$ extracted by the 1×1 Conv from the high frequency information F_h, and the features F^h_{MLP} extracted by the MLP are integrated by the pixel multiplication operation. This stage can be expressed as

$$\begin{cases} \hat{F}_l = H_{\mathrm{SE}}(\mathrm{Conv}_3(\mathrm{Conv}_1(F_l))) \\ F^l_{\mathrm{Kspa}} = H_{\mathrm{soft}}\left(0.1F_{K_1}\left(\hat{F}_l\right) \oplus 0.3F_{K_2}\left(\hat{F}_l\right)\right) \\ \hat{F}' = F^l_{\mathrm{Kspa}} \odot \mathrm{Conv}_1(F_l) \odot F^h_{\mathrm{MLP}} \end{cases} \tag{14}$$

where $H_{\mathrm{soft}}(\cdot)$ denotes the softmax function. Finally, the integrated features $\hat{F}'$ is processed by the 3×3 Conv, and further add it to the feature extracted by the 1×1 Conv from the high-frequency information F_h. The final low-frequency features can be obtained by the linear operation, i.e.,

$$F'_l = H_{\mathrm{Linear}}\left(\hat{F}' \oplus \mathrm{Conv}_1(F_l)\right) \tag{15}$$

where F'_l is the final low-frequency feature. In all, the CAM is a dual-branch structure, which can fully explore the complementary and correction of features at different frequency levels. The top branch mainly contains detail-preserving pooling (DP Pooling), self-attention (SA), and channel attention (CA) used to process the high-frequency information. Meanwhile, the bottom branch mainly contains SENet and K-spare channel attention to enhance the low-frequency information.

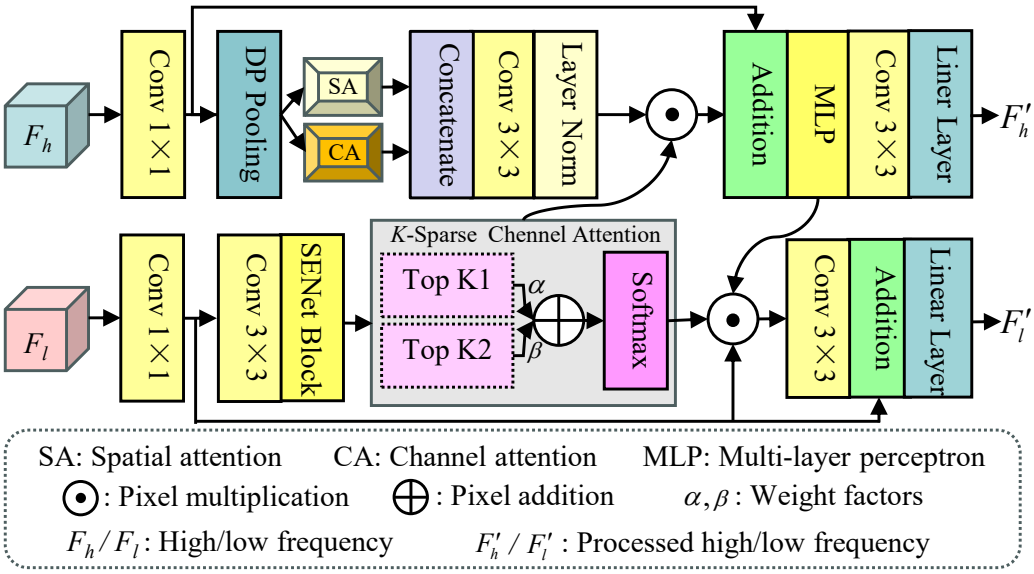


Figure 4. The structure of the CAM, and it consists of two branches for exploring the complementary nature of high/low-frequency information.

3.4. Image Perceptual Discriminator

In GAN-based methods, the discriminator plays a crucial role, which can assess the reconstructed image quality to drive the generated image quality toward reality. Hence, we design an image perceptual discriminator in accordance with [50] to solve the maximization problem in Equation (1). The structure of our discriminator is shown in Figure 5. It can easily be seen that our design is inspired by the architecture of the VGG network, and our discriminator mainly contains four convolutional layers with a increasing number of filter kernels. First, the convolutional layer is used to obtain shallow features. Then, we employ the strided convolutions to increase the image resolution each time. Finally, the feature maps of results are further processed by successive convolutional layer, batch normal (BN), and a sigmoid activation function to obtain a probability for sample classification.

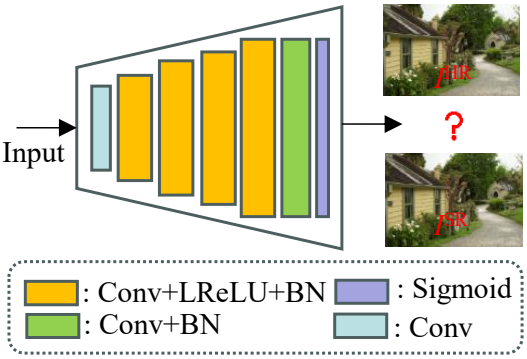


Figure 5. The archetecture of the discriminator used in our PUGAN.

3.5. Loss Function

For training our proposed PUGAN, we introduce a set of differentiable loss functions to guarantee the reconstructed HR images with clearer details and visually pleasing appearance.

Adversarial loss. The adversarial loss can encourage the network to produce solutions that reside on the natural image manifold by attempting to fool the discriminator. We formally define the adversarial loss across based on the probabilities of the FCPG for all training images,

$$L_{\text{adv}} = \sum_{n=1}^{N} -\log\left(D_{\theta_D}\left(G_{\theta_G}\left(I^{\text{HR}}\right)\right)\right) \tag{16}$$

where N is the total number of training images, and $D_{\theta_D}\left(G_{\theta_G}\left(I^{\text{LR}}\right)\right)$ denotes the probability that the reconstructed image is a visually pleasing HR image.

Spatial consistency loss. The pixel-wise MSE loss widly used in learning-based SISR methods, while the reconstructed HR images typically exihibit perceptually unsatisfying and overly smooth textures. Following Ledig et al. [50], we define the spatial consistency loss L_{spa} based on the ReLU activation layers of the pre-trained 19-layer VGG network $\phi_{\text{VGG}}(\cdot)$ to measure the difference in neighboring regions between the reconstructed HR image Y and its corresponding LR version I. The L_{spa} can be defined as

$$L_{\text{spa}} = \frac{1}{K}\sum_{i=1}^{K}\sum_{j\in\Omega(i)}\left(\left|\phi_{\text{VGG}}(Y_i) - \phi_{\text{VGG}}(Y_j)\right| - \left|\phi_{\text{VGG}}(I_i) - \phi_{\text{VGG}}(I_j)\right|\right)^2 \tag{17}$$

where K is the number of local areas. $\Omega(i)$ is the window size centered at the region i, and its size typically set as 4×4.

Color constancy loss. To make our reconstructed HR images exhibit better color fidelity, we introduce the color constancy loss function. This function relying on the Gray-World color constancy hypothesis can fully explore the relationships among R, G, and B channels of the reconstructed images. The L_{col} can be defined as

$$L_{\text{cor}} = \sum_{\forall(p,q)\in\varepsilon}\left(J^p - J^q\right)^2, \varepsilon \in \{(R,G),(R,B),(G,B)\} \tag{18}$$

where J^p and J^q denote the average pixels value of channel p and q in reconstructed images.

Perceptual loss. It can assess a solution with respect to perceptually relevant characteristics. Hence, the perceptual loss formulated as a weighted sum of a spatial consistency loss, a color constancy loss, and an adversarial loss component can assess the quality of HR images from the aspects of spatial structure and color. The L_{Per} can be expressed as

$$L_{\text{Per}} = L_{\text{spa}} + \lambda_1 L_{\text{col}} + \lambda_2 L_{\text{adv}} \tag{19}$$

where the weights λ_1 and λ_2 are applied to balance the details enhancement and color correction, empirically setting them as 0.1 and 10^{-3}, respectively.

4. Experimental Results and Analysis

This section first presents the implementation details and experimental setting. Subsequently, the validation experiments, ablation study, generalization, and computational complexity are demonstrated. Finally, we discuss the limitations and future works related to our PUGAN.

4.1. Implementation Details

Four common publicly used benchmarks, including three paired datasets—NTIRE 2020 [51], Urban 100 [52], and B 100 [53]—and a real-world dataset DPED [54], are used in our work. Among them, three paired public datasets contain synthesized RGB images, and we perform the anisotropic Gaussian kernel on them to simulate the generation of real noise. The DPED [54] dataset contains unpaired images captured by a smartphone, and we apply the KernelGAN on DPED [54] to estimate the distribution of the noise. During the training, we randomly select 80% images from NTIRE 2020 [51], and the remaining 20% of the data is the test setting. Additionally, we further perform our PUGAN and other state-of-the-art SISR methods on the Urban 100 [52], B 100 [53], and DPED [54] datasets to verify their effectiveness in real-world applications.

In our validation experiments, we employ the commonly used technologies, including rotating by 90°, horizontal flipping, and scaling, to augment the training datasets (randomly select 80% images from NTIRE 2020 [51]) containing RGB images. The input RGB images are randomly cropped into a specific size of 64×64. Our proposed PUGAN and other compared SISR methods are implemented in the Pytorch framework on an NVIDIA Tesla P100 GPU to test their performance. For better training of our SISR method, the batch size is set to 16, the number of iterations is set to 60,000, and the learning rate is set to 1×10^{-4}. We further use the ADAM optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$, and $\varepsilon = 10^{-8}$ to optimize the parameters of our carefully designed PUGAN.

4.2. Experimental Settings

Comparison Methods: We select some state-of-the-art SISR approaches as comparison methods to verify the performance of our proposed PUGAN for the SISR vision task. These compared methods contain ESRGAN [27], SDSR [48], TDSR [48], RealSR [55], DASR [56], DUSGAN [28], IDMBSR [57], FASR [58], DMGSR [59], and SRMamba-T [26]. To ensure the fairness of the comparison, all above-listed SISR comparison methods adopt publicly available source code with recommended parameters.

Evaluation Metrics: Three commonly used evaluation metrics including peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS) are applied to the quantitative assessment of image quality. And these metrics can be calculated by

$$\mathrm{PSNR} = 10 \times \log_{10} \frac{max^2(\hat{\mathrm{I}})}{\mathrm{MSE}(\mathrm{I}, \hat{\mathrm{I}})} \tag{20}$$

where I and $\hat{\mathrm{I}}$ are the ground truth and reconstructed HR images. $\mathrm{MSE}(\cdot)$ represents the solution of variance, and $max^2(\hat{\mathrm{I}})$ denotes the maximum pixel of reconstructed HR images $\hat{\mathrm{I}}$.

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{21}$$

where μ_x and μ_y denote the average pixels value of the ground truth I and reconstructed HR image $\hat{I}$. σ_x and σ_y are the pixel variances of these two images, and σ_{xy} is the pixel covariance. C_1 and C_2 are constant.

$$\text{LPIPS} = \sum_{l} \frac{\sum_{h,w} \left\| w_l \odot \left(y^l_{hw} - \hat{y}^l_{hw} \right) \right\|_2^2}{H^l W^l} \tag{22}$$

where y^l_{hw} and $\hat{y}^l_{hw}$ denote the input and reconstructed images' features of the l_{th} layer in the trained CNN-based model. $H \times W$ is the height and width of the image.

Among them, PSNR can measure the differences at the pixel-level, and a higher PSNR score suggests a better quality while exhibiting relatively weak correlation with human visual perception. SSIM can measure the similarity between reconstructed HR results and their corresponding ground truth in terms of brightness, contrast, and structure, and a higher SSIM score suggests a better structure preservation. LPIPS is a learning-based image quality assessment method, which can measure the differences in feature maps extracted by a pre-trained CNN model. A lower LPIPS score suggests a better visual perception.

4.3. Comparisons on the Synthetic Datasets

We first perform our PUGAN and compared SISR approaches on the Urban 100 [52], B 100 [53], and DPED [54] datasets to test their performance in the SISR task from the qualitative and quantitative evaluations.

Qualitative Evaluation. Figure 6 demonstrates the HR images randomly selected from the NTIRE 2020 [51]. We can easily observe that SDSR [48] and FASR [58] inevitably introduce observable noise in some reconstructed HR images (e.g., the first row of the Ex. 1 and Ex. 4 in Figure 6). In addition, DASR [56], FASR [58], and IDMBSR [57] can make the details clearer for some reconstructed HR images, but they fail in removing artifact-halos. FASR [58] also shows poor performance in noise suppression. DASR [56] yields unnatural visual experience (the first row of the Ex. 4 in Figure 6). ESRGAN [27] and TDSR [48] can generate noise-free HR images. However, these two SISR methods are often considered inferior to other compared SISR models in detail enhancement. DMGSR [59] shows better performance in removing significant noise, while it is unsatisfactory for local details and edges preservation (e.g., the first row of the Ex. 1 in Figure 6, and the second row of the Ex. 3 in Figure 6). In contrast, our PUGAN can effectively yield visually satisfactory and noise-free HR images with clearer details and edges.

Quantitative Evaluation. Table 1 presents the average PSNR, SSIM, and LPIPS scores of different SISR approaches for $\times 2$ and $\times 4$ on the Urban 100 [52], B 100 [53], and DPED [54] datasets. From the quantitative evaluation scores of three public benchmarks in Table 1, we can easily find that our PUGAN yields comparable and higher values of PSNR and SSIM than compared SISR methods. For the LPIPS, the proposed PUGAN outperforms other listed state-of-the-art SISR models, suggesting that our method can reconstruct HR images with a pleasant visual experience. Overall, the qualitative and quantitative comparisons demonstrate that our carefully designed PUGAN exhibits superior performance for generating noise-free and artifact-free images with clearer details.

4.4. Comparisons on the Real-World Datasets

To verify the application on real-world data, we further perform all the above-mentioned SISR approaches on the DPED [54] dataset and assess their performance from both quantitative and qualitative perspectives.

Input (a) (b) (c) (d) (e) (f) (g) (h) (i)

Figure 6. Vision comparison of different SISR methods on the NTIRE 2020 [51]. The HR image reconstructed by (**a**) ESRGAN [27], (**b**) SDSR [48], (**c**) TDSR [48], (**d**) RealSR [55], (**e**) DASR [56], (**f**) IDMBSR [57], (**g**) FASR [58], (**h**) DMGSR [59], and (**i**) Ours.

Table 1. Quantitative analysis of different comparison methods on paired benchmarks. Red/blue text stands for the best/second-best performance.

Methods	Scale	NTIRE 2020			Urban 100			B 100		
		PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
ESRGAN [27]	×2	30.989(±2.387)	0.9179(±0.1963)	0.4435(±0.1100)	29.507(±2.301)	0.8946(±0.2073)	0.5107(±0.1206)	29.987(±2.375)	0.9056(±0.1824)	0.4789(±0.1326)
SDSR [48]	×2	31.218(±2.112)	0.9011(±0.1788)	0.3574(±0.0999)	31.075(±2.300)	0.8956(±0.1998)	0.3689(±0.1203)	31.173(±2.212)	0.8997(±0.1314)	0.3613(±0.0954)
TDSR [48]	×2	31.369(±1.428)	0.9237(±0.1368)	0.2917(±0.1016)	31.101(±1.306)	0.9063(±0.1208)	0.3104(±0.1251)	31.238(±1.511)	0.9113(±0.1474)	0.2988(±0.1026)
RealSR [55]	×2	32.015(±1.289)	0.9318(±0.1710)	0.2540(±0.0310)	31.997(±1.355)	0.9219(±0.1604)	0.2678(±0.0294)	32.000(±1.747)	0.9267(±0.1567)	0.2588(±0.0234)
DASR [56]	×2	30.586(±1.575)	0.9112(±0.1603)	0.2309(±0.0177)	30.054(±1.462)	0.9084(±0.1412)	0.2113(±0.0237)	30.510(±1.428)	0.9109(±0.1379)	0.2029(±0.0199)
DUSGAN [28]	×2	30.111(±1.554)	0.9099(±0.1097)	0.2619(±0.0192)	30.093(±1.274)	0.9001(±0.1724)	0.2310(±0.0210)	30.108(±1.426)	0.9085(±0.1257)	0.2403(±0.0199)
IDMBSR [57]	×2	31.268(±1.274)	0.9326(±0.1667)	0.2213(±0.0174)	31.001(±1.363)	0.9254(±0.1248)	0.2271(±0.0169)	31.217(±1.578)	0.9300(±0.1780)	0.2100(±0.0210)
FASR [58]	×2	32.385(±1.302)	0.9384(±0.1107)	0.2190(±0.0134)	32.107(±1.425)	0.9289(±0.1007)	0.2037(±0.0348)	31.589(±1.478)	0.9412(±0.1694)	0.2095(±0.0484)
DMGSR [59]	×2	32.867(±1.177)	0.9334(±0.0989)	0.1901(±0.0235)	31.271(±1.546)	0.9129(±0.1001)	0.1945(±0.0197)	31.968(±1.701)	0.9299(±0.1092)	0.1934(±0.0218)
SRMamba-T [26]	×2	33.247(±1.439)	0.9591(±0.1027)	0.1393(±0.0472)	33.103(±1.538)	0.9483(±0.1230)	0.1536(±0.0589)	33.119(±1.672)	0.9522(±0.1196)	0.1496(±0.0503)
Ours	×2	33.987(±1.101)	0.9673(±0.0799)	0.1210(±0.0200)	32.966(±1.239)	0.9483(±0.0983)	0.1431(±0.0376)	33.627(±1.377)	0.9546(±0.0865)	0.1354(±0.0274)
ESRGAN [27]	×4	23.745(±2.179)	0.6852(±0.1207)	0.2071(±0.0989)	23.379(±2.237)	0.6671(±0.1331)	0.2173(±0.0899)	23.401(±2.101)	0.6798(±0.1123)	0.2239(±0.0783)
SDSR [48]	×4	22.909(±2.4810)	0.6854(±0.1179)	0.4384(±0.0899)	22.764(±2.324)	0.6672(±0.1228)	0.4697(±0.0675)	22.849(±2.3783)	0.6917(±0.1023)	0.4462(±0.0666)
TDSR [48]	×4	21.998(±2.079)	0.4358(±0.0998)	0.4609(±0.0418)	21.590(±2.111)	0.4083(±0.1001)	0.4731(±0.0098)	21.783(±1.985)	0.4576(±0.9597)	0.4547(±0.0107)
RealSR [55]	×4	24.989(±1.997)	0.6919(±0.0579)	0.2270(±0.0071)	24.869(±1.939)	0.6704(±0.0662)	0.2346(±0.0082)	24.953(±1.879)	0.6888(±0.0710)	0.2241(±0.0074)
DASR [56]	×4	25.461(±1.679)	0.7992(±0.1023)	0.2013(±0.0064)	25.307(±1.705)	0.7597(±0.1084)	0.2039(±0.0059)	25.436(±1.652)	0.7793(±0.1123)	0.2010(±0.0045)
DUSGAN [28]	×4	24.218(±1.811)	0.6172(±0.0992)	0.5625(±0.1998)	24.192(±1.707)	0.5423(±0.1010)	0.5512(±0.2000)	24.203(±1.910)	0.5593(±0.1869)	0.5104(±0.1118)
IDMBSR [57]	×4	25.429(±1.232)	0.8079(±0.1027)	0.2000(±0.0100)	25.379(±1.407)	0.7540(±0.1005)	0.2017(±0.0998)	25.400(±1.376)	0.7998(±0.0571)	0.2028(±0.0646)
FASR [58]	×4	25.971(±1.223)	0.8113(±0.1116)	0.1996(±0.0109)	25.648(±1.392)	0.8023(±0.1401)	0.2208(±0.0210)	25.946(±1.245)	0.8099(±0.1327)	0.2175(±0.0119)
DMGSR [59]	×4	26.078(±1.444)	0.7782(±0.1228)	0.1979(±0.0174)	25.976(±1.652)	0.8347(±0.1000)	0.2002(±0.0971)	26.013(±1.535)	0.8498(±0.1219)	0.2113(±0.0163)
SRMamba-T [26]	×4	26.201(±1.457)	0.8603(±0.1238)	0.1980(±0.0358)	26.119(±1.676)	0.8417(±0.2018)	0.1999(±0.0758)	26.198(±1.469)	0.8541(±0.1556)	0.2054(±0.0754)
Ours	×4	26.349(±1.139)	0.8721(±0.1015)	0.1975(±0.0236)	26.110(±1.458)	0.8614(±0.1298)	0.1983(±0.0307)	26.306(±1.367)	0.8803(±0.1201)	0.1978(±0.0286)

Qualitative Evaluation. We show the HR images reconstructed by our PUGAN and compared SISR methods in Figure 7. As easily observed from the reconstructed HR results, ESRGAN [27], SDSR [48], RealSR [55], and DASR [56] introduce observable noise into some HR images. In addition, RealSR [55] fails in making details and edges clearer. TDSR [48] is unsuccessful in removing hazy-like appearance and noise suppression (e.g., the first row of the Ex. 1 in Figure 7, and the second row of the Ex. 2 in Figure 7). IDMBSR [57] shows satisfactory performance in detail enhancements but yields undesired artifact-halos in some HR images (e.g., the second row of the Ex. 3 in Figure 7). DMGSR [59] cannot make the details and contours clearer, with some DMGSR-reconstructed HR images suffering

from blurry details (e.g., the first and second row of the Ex. 4 in Figure 7) and hazy-like appearance (e.g., the second row of the Ex. 2 in Figure 7). In comparison, our carefully designed PUGAN is superior to the compared SISR methods in detail enhancement and noise suppression, and it also works better in removing artifact-halos and hazy-like appearance.

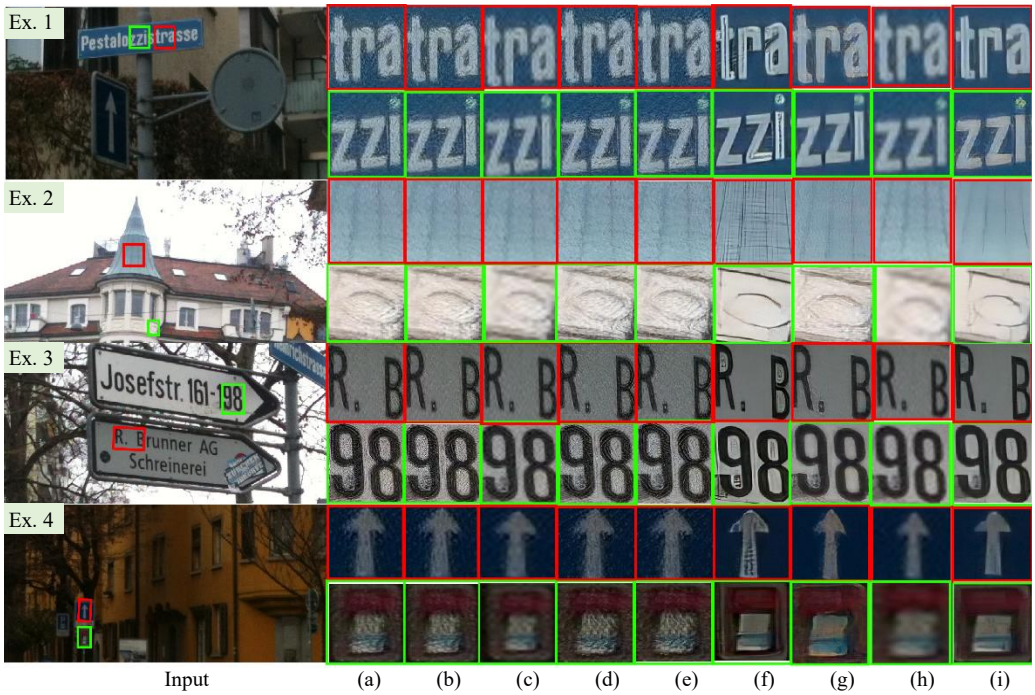


Figure 7. Vision comparison of different SISR methods on the DPED [54]. The HR image reconstructed by (**a**) ESRGAN [27], (**b**) SDSR [48], (**c**) TDSR [48], (**d**) RealSR [55], (**e**) DASR [56], (**f**) IDMBSR [57], (**g**) FASR [58], (**h**) DMGSR [59], and (**i**) Ours.

Quantitative Evaluation. We also perform the LPIPS to evaluate the HR images reconstructed by different SISR methods, and their average LPIPS score for the DPED [54] dataset are illustrated in Table 2. From Table 2, it can easily be observed that our PUGAN can yield more satisfactory and comparable LPIPS scores than compared SISR approaches. The quantitative and qualitative assessment results suggest that our proposed method achieves the state-of-the-art performance for the SISR vision task.

Table 2. The average LPIPS scores of different SISR methods on the DPED [54] for the scale of ×2 and ×4. **Bold** text denotes the best performance.

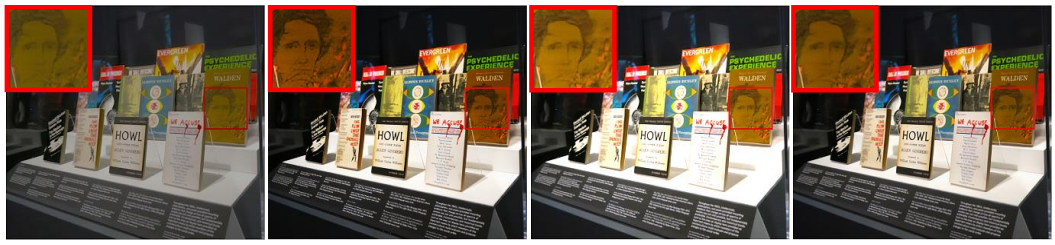
Scale	ESRGAN	SDSR	TDSR	RealSR	DUSGAN
×2	0.3579 (±0.0311)	0.2917 (±0.0296)	0.3001 (±0.0248)	0.2785 (±0.0415)	0.2711 (±0.0340)
×4	0.3784 (±0.0203)	0.3037 (±0.0279)	0.3209 (±0.0240)	0.2699 (±0.0246)	0.2868 (±0.0124)
Scale	**IDMSBSR**	**FASR**	**DMGSR**	**SRMamba-T**	**Ours**
×2	0.2432 (±0.0211)	0.2347 (±0.0209)	0.2137 (±0.0295)	0.1758 (±0.0204)	**0.1560** (±0.0109)
×4	0.2559 (±0.0191)	0.2466 (±0.0158)	0.2394 (±0.0100)	0.2003 (±0.0097)	**0.1884** (±0.0089)

4.5. Ablation Study

To fully test the effect of each component in our PUGAN, we perform ablation studies including frequency decomposition (FD), upsampling operations (UO), and CAM on the test datasets.

(1) Study of the FD: The FD operation can separate noise, details, and content from the LR images. For verifying its effectiveness, we first remove the FD operation from our PUGAN (-*w/o* FD), then the guided filter (GF) and bilateral filter (BF) are used to

extracted frequency features (Ours_{GF} and Ours_{BF}). Figure 8 presents HR images generated by our method with different FD operations. It can be observed that *-w/o* FD fails in details boosting. In contrast, our PUGAN can generate more visually pleasing images with clearer details than other FD operations. Furthermore, we present average PSNR and SSIM scores of our PUGAN with different FD operations in Table 3. It can easily be found that our method yields higher PSNR and SSIM scores than ablated models. Overall, the qualitative and quantitative analyses show that the FD operation plays an indispensable role in our designed PUGAN.



(*a*) -w/o FD (*b*) Ours_{GF} (*c*) Ours_{BF} (*d*) Ours

Figure 8. Vision comparison of our PUGAN with different FD operations. (**a**) *-w/o* FD, (**b**) Ours_{GF} , (**c**) Ours_{BF}, (**d**) Ours.

Table 3. The average PSNR and SSIM scores of our PUGAN with different operations on the test datasets. **Bold** text stands for the best performance.

Operations	Scale	PSNR	SSIM	LPIPS
-w/o FD	×4	$25.997_{(\pm 2.018)}$	$0.8104_{(\pm 0.1857)}$	$0.2317_{(\pm 0.0876)}$
Ours_{BF}	×4	$26.010_{(\pm 1.488)}$	$0.8494_{(\pm 0.1540)}$	$0.2011_{(\pm 0.0548)}$
Ours_{GF}	×4	$26.101_{(\pm 1.387)}$	$0.8500_{(\pm 0.1421)}$	$0.1998_{(\pm 0.0410)}$
Ours	×4	$\mathbf{26.349}_{(\pm 1.139)}$	$\mathbf{0.8721}_{(\pm 0.1015)}$	$\mathbf{0.1975}_{(\pm 0.0236)}$
Post	×4	$26.001_{(\pm 1.569)}$	$0.8499_{(\pm 0.1526)}$	$0.2010_{(\pm 0.0483)}$
Prog.	×4	$\mathbf{26.349}_{(\pm 1.139)}$	$\mathbf{0.8721}_{(\pm 0.1015)}$	$\mathbf{0.1975}_{(\pm 0.0236)}$

(2) Study of the UO: Figure 9 presents the flowcharts of post-upsampling and progressive upsampling, named post and prog., respectively. To analyze their effectiveness, we perform our PUGAN with different UO on test datasets, and the generated HR images are demonstrated in Figure 10. It can easily be seen that our PUGAN with prog. means can make local details clearer than the post means. The likely reason is that the prog. means with partial skip connections can alleviate information loss and ambiguity. We further present their PSNR and SSIM scores in Table 3. Intuitively, our method with prog. means outperforms post means in yielding higher PSNR and SSIM scores.

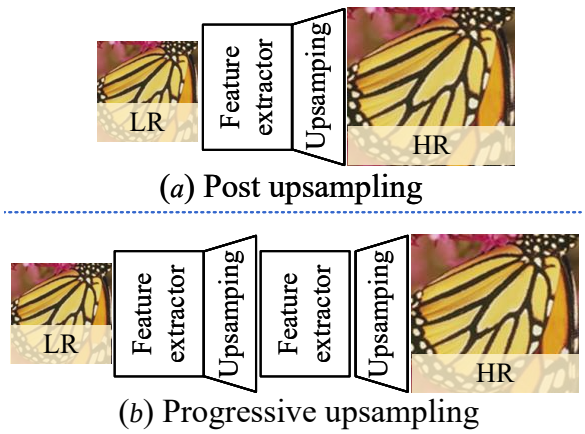
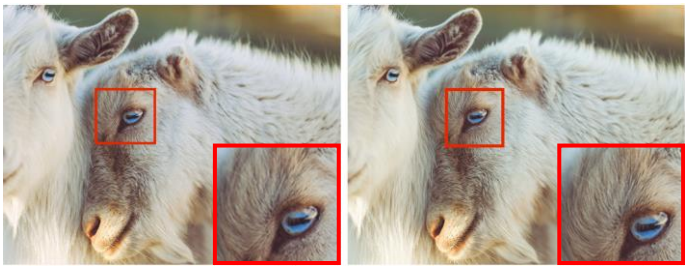


Figure 9. The flowchart of post upsampling (**a**) and progressive upsampling (**b**).



(*a*) Post (*b*) Prog.

Figure 10. Vision comparison of our PUGAN with different UO. Among them, the results yielded by (**a**) Post and (**b**) Prog. upsampling operations.

(3) Study of the CAM: Figure 11 presents the HR images yielded by our PUGAN with CAM and CBAM. It can be observed that our PUGAN with CAM can generate visually comfortable images with clearer details. Furthermore, their PSNR and SSIM scores are presented in the Table 4. We can also find that the CAM can bring absolute improvements in PSNR and SSIM scores for our SISR method. That is to say, our PUGAN with CAM outperforms the latter in terms of qualitative and quantitative assessments, benefiting from our designed CAM.

(a) Input (b) CAM (c) CBAM

Figure 11. (**a**) Input, Vision comparison of our proposed PUGAN with (**b**) CAM and (**c**) CBAM.

Table 4. The average PSNR and SSIM scores of our PUGAN with different operations on the test datasets. **Bold** text stands for the best performance.

Methods	Scale	PSNR	SSIM	LPIPS
CAM	×4	**26.349**$_{(\pm 1.139)}$	**0.8721**$_{(\pm 0.1015)}$	**0.1975**$_{(\pm 0.0236)}$
CBAM	×4	26.001$_{(\pm 1.786)}$	0.8341$_{(\pm 0.1128)}$	0.2103$_{(\pm 0.0672)}$
C → R	×4	26.130$_{(\pm 1.987)}$	0.8512$_{(\pm 0.1389)}$	0.1987$_{(\pm 0.0670)}$
R → C	×4	**26.349**$_{(\pm 1.139)}$	**0.8721**$_{(\pm 0.1015)}$	**0.1975**$_{(\pm 0.0236)}$

There are two main combinations of the sequence of CAM and RMB, including C → R (CAB-RMB) and R → C (RMB-CAB). For determining the optimal combination sequence, we evaluate them on the test dataset and their HR images are shown in Figure 12. From Figure 12, we can easily see that the latter combination, i.e., R → C (RMB-CAB), can make HR images exhibit clearer details and outlines. Table 4 further illustrates their corresponding PSNR and SSIM scores. We observe that R → C (RMB-CAB) outperforms C → R (CAB-RMB) in PSNR and SSIM evaluation metrics. In conclusion, the qualitative and quantitative analyses have demonstrated that the R → C (RMB-CAB) is the optimal combination sequence.

(a) Input (b) C→R (c) R→C

Figure 12. Vision comparison of our proposed PUGAN with defferent combinations. (**a**) Input, the reconstructed HR images by (**b**) C → R (CAB-RMB) and (**c**) R → C (RMB-CAB).

4.6. Comparison of Computational Complexity

The computational complexity of learning-based applications determines their real-world applications. Hence, we compare our carefully designed PUGAN with other state-of-the-art SISR methods in terms of Parameters and Flops to verify their work efficiency. Their parameters (Param) and Flops are demonstrated in Table 5. From Table 5, it can be easily observed that most compared SISR methods exhibit more Parameters and Flops, leading to a heavy computational burden and a narrow practical application. By contrast, our proposed PUGAN enjoys fewer parameters and higher inference speed than the above-listed SISR methods. This result mainly occurs because the progressive upsampling can decompose large-scale input into smaller-scale ones to reduce the model's learning complexity.

Table 5. Computational complexity comparison of existing SISR methods. Red/blue text stands for the best/second-best performance.

Method	Param (M)	Flops (G)
ESRGAN [27]	16.71	12.02
SDSR [48]	0.02	0.15
TDSR [48]	6.20	2.01
RealSR [55]	1.70	5.30
DASR [56]	1.48	5.00
DUSGAN [28]	1.97	3.03
IDMBSR [57]	1.71	3.33
FASR [58]	1.28	6.80
DMGSR [59]	1.51	3.09
SRMamba-T [26]	1.26	4.22
Ours	0.23	1.72

4.7. Application on Pathological Images

Digital pathological images are the gold standard for cancer diagnosis, which contain tumor microenvironment (TME), including tumor epithelial, tumor-infiltrating lymphocytes (TILs), tumor-associated stroma, etc., clinically related to the occurrence, development, and metastasis of tumors [8,60,61]. However, the scanning of high-magnification pathological image is time-consuming and difficult to store, limiting its application in clinical diagnosis. To address this issue, we train our carefully designed PUGAN on the HistoSR dataset [8] to examine its generalization. We divided the HistoSR dataset at a ratio of 7:3 into the training and test sets, and we set the initial parameters of our PUGAN to the previously determined values. Notably, all mentioned comparison methods are retrained in the same way as our proposed model.

Figure 13 presents the HR pathological image patches generated by our PUGAN and other state-of-the-art SISR methods. The following observation can be seen from Figure 13: Most SISR methods struggle to yield visually satisfactory HR pathological image patches, and some generated HR images confront artifact-halos and blurry details [e.g., the bottom

picture of Figure 13c]. Additionally, FASR [58] and DMGSR [59] are unsuitable for removing inherent noise. DASR [56] reconstructs HR pathological image patches, while introducing an uncomfortable visual experience. In contrast, the proposed PUGAN effectively removes artifact-halos and makes structural details clearer with the benefit of our carefully designed method. Additionally, we also present the average PSNR and SSIM scores of different SISR methods in Table 6. It can be seen that our PUGAN is superior to other compared SISR methods in generating more satisfactory PSNR and SSIM scores. Overall, the quantitative and qualitative evaluations suggest that our carefully designed SISR approach consistently produces noise-free and artifact-free HR pathological image patches with observable and clearer details. The experiments further demonstrate that our method exhibits robust generalization ability.

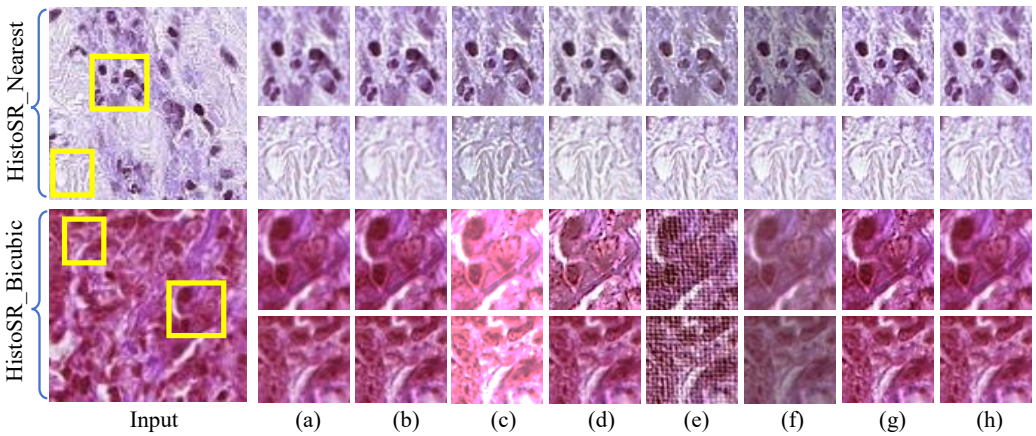


Figure 13. Vision comparison of our PUGAN on the HistoSR [8] with the nearest and bicubic degradations. The HR pathological images reconstructed by (**a**) SDSR [48], (**b**) RealSR [55], (**c**) DASR [56], (**d**) IDMBSR [57], (**e**) FASR [58], (**f**) DMGSR [59], (**g**) Ours, and (**h**) Ground Truth.

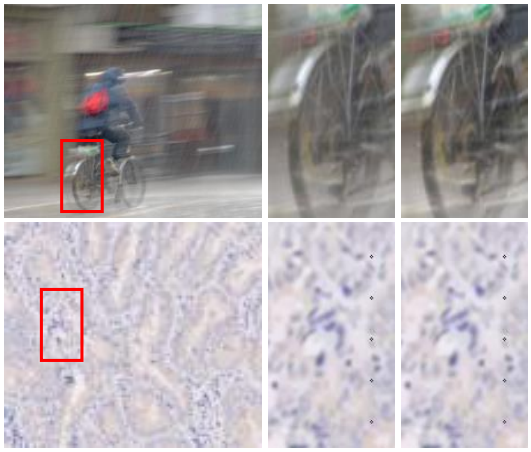
Table 6. Quantitive comparison of our proposed PUGAN on the HistoSR with the nearest and bicubic degeradations. **Bold** text stands for the best performance.

Methods	Histo_Nearest			Histo_Bicubic		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
ESRGAN [27]	31.984$_{(\pm 1.113)}$	0.8867$_{(\pm 0.0174)}$	0.2496$_{(\pm 0.0011)}$	31.998$_{(\pm 0.961)}$	0.8718$_{(\pm 0.1312)}$	0.2319$_{(\pm 0.0101)}$
SDSR [48]	32.004$_{(\pm 1.166)}$	0.9287$_{(\pm 0.0191)}$	0.2667$_{(\pm 0.0172)}$	32.901$_{(\pm 2.183)}$	0.9413$_{(\pm 0.0137)}$	0.2579$_{(\pm 0.0075)}$
TDSR [48]	32.500$_{(\pm 2.143)}$	0.9238$_{(\pm 0.1001)}$	0.2419$_{(\pm 0.0111)}$	33.010$_{(\pm 1.967)}$	0.9589$_{(\pm 0.1347)}$	0.2274$_{(\pm 0.0401)}$
RealSR [55]	32.499$_{(\pm 1.428)}$	0.9301$_{(\pm 0.1547)}$	0.2398$_{(\pm 0.0018)}$	33.0299$_{(\pm 1.557)}$	0.9574$_{(\pm 0.1168)}$	0.2299$_{(\pm 0.0121)}$
DASR [56]	32.489$_{(\pm 2.010)}$	0.9000$_{(\pm 0.0153)}$	0.2501$_{(\pm 0.0026)}$	31.887$_{(\pm 1.993)}$	0.9153$_{(\pm 0.0067)}$	0.2243$_{(\pm 0.0009)}$
DUSGAN [28]	32.413$_{(\pm 1.329)}$	0.9236$_{(\pm 0.0276)}$	0.2517$_{(\pm 0.0243)}$	33.002$_{(\pm 1.568)}$	0.9564$_{(\pm 0.0386)}$	0.2287$_{(\pm 0.0015)}$
IDMBSR [57]	32.350$_{(\pm 3.002)}$	0.9148$_{(\pm 0.1760)}$	0.2418$_{(\pm 0.0124)}$	33.098$_{(\pm 2.510)}$	0.9583$_{(\pm 0.1548)}$	0.2235$_{(\pm 0.0101)}$
FASR [58]	32.217$_{(\pm 1.111)}$	0.8997$_{(\pm 0.1007)}$	0.2467$_{(\pm 0.0201)}$	32.763$_{(\pm 0.989)}$	0.9401$_{(\pm 0.0984)}$	0.2268$_{(\pm 0.0115)}$
DMGSR [59]	32.203$_{(\pm 1.856)}$	0.9023$_{(\pm 0.1452)}$	0.2971$_{(\pm 0.0026)}$	33.111$_{(\pm 1.138)}$	0.8999$_{(\pm 0.1142)}$	0.3105$_{(\pm 0.0957)}$
SRMamba-T [26]	32.323$_{(\pm 1.859)}$	0.9317$_{(\pm 0.1046)}$	**0.2370**$_{(\pm 0.0115)}$	33.104$_{(\pm 1.642)}$	0.9599$_{(\pm 0.0984)}$	0.2310$_{(\pm 0.0710)}$
Ours	**32.601**$_{(\pm 1.010)}$	**0.9398**$_{(\pm 0.0842)}$	0.2376$_{(\pm 0.0257)}$	**33.138**$_{(\pm 0.979)}$	**0.9609**$_{(\pm 0.0271)}$	**0.2217**$_{(\pm 0.0004)}$

4.8. Limitation and Discussion

SISR plays a crucial role in low-level computer vision, which can yield a visually HR image from its corresponding LR version and further promote the performance of image content understanding, image classification, and other advanced computer vision tasks. Experiments have proven that our PUGAN works well on the SISR tasks for natural and pathological images in most situations. However, the proposed method fails in LR images with complex degradations, such as motion blur, severe compression artifacts, etc. For

example, we present some failure instances reconstructed by our proposed PUGAN in Figure 14. It can easily be seen that the reconstructed HR images exhibit unsatisfactory details and undesired artifact-halos. The reason may be that our PUGAN only collects inherent noise from the LR input and does not consider the LR images captured in the real world, which exhibit multiple and complex degradation. In the future, we will inject promote learning and specialized mixture-of-experts (MoE) for tackling these challenging issues.



(a) Input (b) GT (c) Ours

Figure 14. Some failure instances of our proposed PUGAN. From left to right, (**a**) Input, (**b**) Ground Truth (GT), and (**c**) Ours.

5. Conclusions

This paper proposes a feasible and effective SISR method based on the progressive upsampling generative adversarial network with collaborative attention named PUGAN. This method first employs a sliding window to collect noise from the inputs to construct a noise pool and further randomly selects noise to simulate real noise. Subsequently, the convolutional Gaussian filtering is used to extract the high- and low-frequency information of the LR input, and the FCPG further performs the residual multiscale blocks (RMBs) and collaborative attention mechanism (CAM) to fully explore the features at different scales and frequencies for yielding artifact-free and noise-free images with clearer details. Meanwhile, the progressive upsampling strategy is introduced to reduce the model's complexity. Finally, the discriminator is used to judge the differences between the reconstructed images and their corresponding images for generating visually pleasing and artifact-free images with clearer details. Although our method shows better performance in the SISR task, it still cannot make the details clearer in the partial darkness of the image. We leave this challenging case as our future work.

Author Contributions: Conceptualization, H.L., J.Z. and Z.W.; methodology, H.L.; software, H.L. and W.W.; validation, H.L.; investigation, M.J. and J.Z.; data collection, J.Z. and M.J.; writing—original draft preparation, H.L. and J.Z.; writing—review and editing, W.W. and Z.W.; visualization, M.J. and J.Z.; supervision, Z.W. and J.Z.; funding acquisition, Z.W. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under grants (82502451, 82272075, 62462017), the Guangxi Natural Science Foundation under grant (2025GXNSFBA069390), and the Guangxi Regional Innovation Capacity Improvement Program under Grant (XT2503960034).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Review Committee of Guangdong Provincial People's Hospital (protocol code KY2025 1117 0 and approval date 24 November 2025).

Informed Consent Statement: Patient consent was waived due to the data used were obtained from the public databases.

Data Availability Statement: The data presented in this study are available in NTIRE dataset at https://aistudio.baidu.com/datasetdetail/216023 (accessed on 15 November 2025), reference number [44], Urban 100 dataset at https://tianchi.aliyun.com/dataset/88706/ (accessed on 15 November 2025), reference number [45], DPED dataset at https://aiff22.github.io/#dataset (accessed on 15 November 2025), reference number [46], and HistoSR dataset at https://pan.baidu.com/s/1H4TFsKNKZTno8Sz6IOs38A?pwd=7rqv (accessed on 15 November 2025), reference number [52].

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Su, H.; Li, Y.; Xu, Y.; Fu, X.; Liu, S. A review of deep-learning-based super-resolution: From methods to applications. *Pattern Recognit.* **2025**, *157*, 110935. [CrossRef]
2. Yue, Z.; Liao, K.; Loy, C.C. Arbitrary-steps image super-resolution via diffusion inversion. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 23153–23163.
3. Jiang, Q.; Sun, H.; Chen, Q.; Huang, Y.; Li, Q.; Tian, J.; Zheng, C.; Mao, X.; Jiang, X.; Cheng, Y.; et al. High-resolution computed tomography with 1,024-matrix for artificial intelligence-based computer-aided diagnosis in the evaluation of pulmonary nodules. *J. Thorac. Dis.* **2025**, *17*, 289. [CrossRef] [PubMed]
4. Brandt, M.; Chave, J.; Li, S.; Fensholt, R.; Ciais, P.; Wigneron, J.P.; Gieseke, F.; Saatchi, S.; Tucker, C.; Igel, C. High-resolution sensors and deep learning models for tree resource monitoring. *Nat. Rev. Electr. Eng.* **2025**, *2*, 13–26. [CrossRef]
5. Jiang, K.; Yang, M.; Xiao, Y.; Wu, J.; Wang, G.; Feng, X.; Jiang, J. Rep-Mamba: Re-parameterization in vision Mamba for lightweight remote sensing image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5637012. [CrossRef]
6. Zhu, C.; Liu, Y.; Huang, S.; Wang, F. Taming a diffusion model to revitalize remote sensing image super-resolution. *Remote Sens.* **2025**, *17*, 1348. [CrossRef]
7. Zhang, J.; Cheng, S.; Liu, X.; Li, N.; Rao, G.; Zeng, S. Cytopathology image super-resolution of portable microscope based on convolutional window-integration transformer. *IEEE Trans. Comput. Imaging* **2025**, *11*, 77–88. [CrossRef]
8. Chen, Z.; Guo, X.; Yang, C.; Ibragimov, B.; Yuan, Y. Joint spatial-wavelet dual-stream network for super-resolution. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2020; pp. 184–193.
9. Xu, X.; Kapse, S.; Prasanna, P. SuperDiff: A diffusion super-resolution method for digital pathology with comprehensive quality assessment. *Med. Image Anal.* **2025**, *107*, 103808. [CrossRef]
10. Zhang, Y.; Yu, W. Comparison of DEM super-resolution methods based on interpolation and neural networks. *Sensors* **2022**, *22*, 745. [CrossRef] [PubMed]
11. Khan, R.; Sablatnig, R.; Bais, A.; Khawaja, Y.M. Comparison of reconstruction and example-based super-resolution. In *2011 7th International Conference on Emerging Technologies*; IEEE: Piscataway, NJ, USA, 2011; pp. 1–6.
12. Yu, M.; Shi, J.; Xue, C.; Hao, X.; Yan, G. A review of single image super-resolution reconstruction based on deep learning. *Multimed. Tools Appl.* **2024**, *83*, 55921–55962. [CrossRef]
13. Patel, R.; Thakar, V.; Joshi, R. Dictionary learning-based image super-resolution for multimedia devices. *Multimed. Tools Appl.* **2023**, *82*, 17243–17262. [CrossRef]
14. Meng, K.; Zhao, M.; Cattani, P.; Mei, S. Single image super-resolution based on Bendlets analysis and structural dictionary learning. *Results Phys.* **2024**, *57*, 107367. [CrossRef]
15. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 184–199.
16. Shang, S.; Shan, Z.; Liu, G.; Wang, L.; Wang, X.; Zhang, Z.; Zhang, J. Resdiff: Combining cnn and diffusion model for image super-resolution. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; Volume 38, pp. 8975–8983.
17. Xu, Y.; Zhou, Y.; Ma, H.; Yang, H.; Wang, H.; Zhang, S.; Li, X. Wavelet-based dual discriminator GAN for image super-resolution. *Knowl.-Based Syst.* **2025**, *317*, 113383. [CrossRef]
18. Ma, C.; Mi, J.; Gao, W.; Tao, S. DESRGAN: Detail-enhanced generative adversarial networks for small sample single image super-resolution. *Neurocomputing* **2025**, *617*, 129121. [CrossRef]
19. Chen, B.; Li, G.; Wu, R.; Zhang, X.; Chen, J.; Zhang, J.; Zhang, L. Adversarial diffusion compression for real-world image super-resolution. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 28208–28220.

20. Umer, R.M.; Foresti, G.L.; Micheloni, C. Deep generative adversarial residual convolutional networks for real-world super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*; IEEE: Piscataway, NJ, USA, 2020; pp. 438–439.
21. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*; IEEE: Piscataway, NJ, USA, 2017; pp. 136–144.
22. Kong, X.; Zhao, H.; Qiao, Y.; Dong, C. ClassSR: A general framework to accelerate super-resolution networks by data characteristic. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2021; pp. 12016–12025.
23. Zhao, L.; Gao, J.; Deng, D.; Li, X. SSIR: Spatial shuffle multi-head self-attention for single image super-resolution. *Pattern Recognit.* **2024**, *148*, 110195. [CrossRef]
24. Li, A.; Zhang, L.; Liu, Y.; Zhu, C. Exploring frequency-inspired optimization in Transformer for efficient single image super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2025**, *47*, 3141–3158. [CrossRef]
25. Wu, Q.; Zeng, H.; Zhang, J.; Li, W.; Xia, H. A hybrid network of CNN and transformer for subpixel shifting-based multi-image super-resolution. *Opt. Lasers Eng.* **2024**, *182*, 108458. [CrossRef]
26. Liu, C.; Zhang, D.; Lu, G.; Yin, W.; Wang, J.; Luo, G. SRMamba-T: Exploring the hybrid mamba-transformer network for single image super-resolution. *Neurocomputing* **2025**, *624*, 129488. [CrossRef]
27. 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 Workshops*; Springer: Cham, Switzerland, 2019; pp. 63–79.
28. Prajapati, K.; Chudasama, V.; Patel, H.; Upla, K.; Raja, K.; Ramachandra, R.; Busch, C. Direct unsupervised super-resolution using generative adversarial network (DUS-GAN) for real-world data. *IEEE Trans. Image Process.* **2021**, *30*, 8251–8264. [CrossRef] [PubMed]
29. Dong, Y.; Zhou, H.; Zheng, L.; Wang, X.; Ma, J. Cross dropout based dynamic learning for blind super resolution. *Neurocomputing* **2025**, *620*, 129234. [CrossRef]
30. Cho, S.; Cho, N.I. Blind image super-resolution with efficient network design using frequency domain information. *IEEE Signal Process. Lett.* **2025**, *32*, 2524–2528. [CrossRef]
31. Niu, Z.; Zhong, G.; Yu, H. A review on the attention mechanism of deep learning. *Neurocomputing* **2021**, *452*, 48–62. [CrossRef]
32. Soydaner, D. Attention mechanism in neural networks: Where it comes and where it goes. *Neural Comput. Appl.* **2022**, *34*, 13371–13385. [CrossRef]
33. Zhong, J.; Tian, W.; Xie, Y.; Liu, Z.; Ou, J.; Tian, T.; Zhang, L. PMFSNet: Polarized multi-scale feature self-attention network for lightweight medical image segmentation. *Comput. Methods Programs Biomed.* **2025**, *261*, 108611. [CrossRef]
34. Hu, L.; Wang, X.; Liu, Y.; Liu, N.; Huai, M.; Sun, L.; Wang, D. Towards stable and explainable attention mechanisms. *IEEE Trans. Knowl. Data Eng.* **2025**, *37*, 3047–3061. [CrossRef]
35. Sun, X.; Zhang, B.; Wang, Y.; Mai, J.; Wang, Y.; Tan, J.; Wang, W. A multiscale attention mechanism super-resolution confocal microscopy for wafer defect detection. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 1016–1027. [CrossRef]
36. Park, J.; Woo, S.; Lee, J.Y.; Kweon, I.S. BAM: Bottleneck attention module. *arXiv* **2018**, arXiv:1807.06514. [CrossRef]
37. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19.
38. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 286–301.
39. Wang, Y.; Li, Y.; Wang, G.; Liu, X. Multi-scale attention network for single image super-resolution. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2024; pp. 5950–5960.
40. Wu, Q.; Yang, Z.; Zeng, H.; Zhang, J.; Xia, H. Super-resolution reconstruction of sequential images based on an active shift via a hybrid attention calibration mechanism. *Eng. Appl. Artif. Intell.* **2025**, *144*, 110178. [CrossRef]
41. Guo, Y.; Tian, C.; Liu, J.; Di, C.; Ning, K. HADT: Image super-resolution restoration using Hybrid Attention-Dense Connected Transformer Networks. *Neurocomputing* **2025**, *614*, 128790. [CrossRef]
42. Malkocoglu, A.B.V.; Samli, R. A novel model for higher performance object detection with deep channel attention super resolution. *Eng. Sci. Technol. Int. J.* **2025**, *64*, 102003. [CrossRef]
43. Su, J.N.; Gan, M.; Chen, G.Y.; Guo, W.; Chen, C.L.P. High-similarity-pass attention for single image super-resolution. *IEEE Trans. Image Process.* **2024**, *33*, 610–624. [CrossRef]
44. Zhang, Q.L.; Yang, Y.B. SA-Net: Shuffle attention for deep convolutional neural networks. In *Proceedings of the ICASSP 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: Piscataway, NJ, USA, 2021; pp. 2235–2239.
45. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*.
46. Chen, J.; Chen, J.; Chao, H.; Yang, M. Image blind denoising with generative adversarial network based noise modeling. In *IEEE Conference on Computer Vision and Pattern Rrecognition*; IEEE: Piscataway, NJ, USA, 2018; pp. 3155–3164.

47. Zhou, R.; Susstrunk, S. Kernel modeling super-resolution on real low-resolution images. In *IEEE/CVF International Conference on Computer Vision*; IEEE: Piscataway, NJ, USA, 2019; pp. 2433–2443.
48. Fritsche, M.; Gu, S.; Timofte, R. Frequency separation for real-world super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop*; IEEE: Piscataway, NJ, USA, 2019; pp. 3599–3608.
49. Lan, R.; Sun, L.; Liu, Z.; Lu, H.; Pang, C.; Luo, X. MADNet: A fast and lightweight network for single-image super resolution. *IEEE Trans. Cybern.* **2020**, *51*, 1443–1453. [CrossRef]
50. 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 *IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2017; pp. 4681–4690.
51. Lugmayr, A.; Danelljan, M.; Timofte, R. Ntire 2020 challenge on real-world image super-resolution: Methods and results. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*; IEEE: Piscataway, NJ, USA, 2020; pp. 494–495.
52. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2015; pp. 5197–5206.
53. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 898–916. [CrossRef] [PubMed]
54. Ignatov, A.; Kobyshev, N.; Timofte, R.; Vanhoey, K.; Van Gool, L. Dslr-quality photos on mobile devices with deep convolutional networks. In *IEEE International Conference on Computer Vision*; IEEE: Piscataway, NJ, USA, 2017; pp. 3277–3285.
55. Ji, X.; Cao, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F. Real-world super-resolution via kernel estimation and noise injection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*; IEEE: Piscataway, NJ, USA, 2020; pp. 466–467.
56. Wei, Y.; Gu, S.; Li, Y.; Timofte, R.; Jin, L.; Song, H. Unsupervised real-world image super resolution via domain-distance aware training. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2021; pp. 13385–13394.
57. Zhang, Y.; Dong, L.; Yang, H.; Qing, L.; He, X.; Chen, H. Weakly-supervised contrastive learning-based implicit degradation modeling for blind image super-resolution. *Knowl.-Based Syst.* **2022**, *249*, 108984. [CrossRef]
58. Zhang, Y.; Liu, Z.; Liu, S.; Sun, Y. Frequency aggregation network for blind super-resolution based on degradation representation. *Digit. Signal Process.* **2023**, *133*, 103837. [CrossRef]
59. Liu, Z.; Huang, J.; Wang, W.; Lu, H.; Lan, R. Learning distinguishable degradation maps for unknown image super-resolution. *IEEE Trans. Multimed.* **2025**, *27*, 2530–2542. [CrossRef]
60. Aggarwal, A.; Bharadwaj, S.; Corredor, G.; Pathak, T.; Badve, S.; Madabhushi, A. Artificial intelligence in digital pathology—Time for a reality check. *Nat. Rev. Clin. Oncol.* **2025**, *22*, 283–291. [CrossRef]
61. Liu, Y.; Yang, P.; Wang, T.; Lei, B. Utilizing state space model for diffusion processes in breast tumor pathological image super-resolution. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–4.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.