

Article

Image Super-Resolution Reconstruction Based on Hierarchical Feature Aggregation and Laplacian High-Frequency Compensation

Kangliang Xiao , Shaozhang Xiao * , Bolun Chen , Yuanyuan Wang  and Raees ul Haq Muhammad 

Huai'an University, Huai'an 223001, China

* Correspondence: xiaoshaozhang@hau.edu.cn; Tel.: +86-139-5239-9010

Abstract

Existing image super-resolution methods still suffer from limitations in edge-structure restoration, high-frequency texture preservation, and artifact suppression, which may lead to blurred contours and unnatural textures. To address these issues, this paper proposes an image super-resolution method based on hierarchical feature aggregation and Laplacian high-frequency compensation. First, a Hierarchical Feature Aggregation Attention Block (HFAB) is designed in the generator to progressively extract image features at different levels through multiple convolutional layers. A High-Frequency Variance Adaptive Channel Attention Block (HFVB) is further introduced to adaptively enhance key texture and edge information. Second, a Laplacian Adaptive Upsampling (LAU) module is developed to combine low-frequency content reconstruction with high-frequency detail compensation, thereby strengthening edge contours, preserving fine textures, and reducing artifacts. Finally, a Dissimilarity Structural Similarity Index Measure (DSSIM) loss is incorporated into the loss function to constrain local structural consistency and further improve the structural preservation and perceptual quality of reconstructed images. Experimental results on Set5, Set14, BSD100, and Urban100 show that, compared with SRGAN, the proposed method improves PSNR by 0.31 dB, 0.18 dB, 0.14 dB, and 0.14 dB, respectively, while reducing LPIPS by 0.0234, 0.0183, 0.0222, and 0.0222. These results indicate that the proposed method provides consistent improvements over SRGAN and achieves modest, metric-dependent gains over ESRGAN, suggesting an incremental enhancement in reconstruction accuracy and perceptual quality on both natural image benchmarks and complex urban scene datasets.

Keywords: image super-resolution reconstruction; hierarchical feature aggregation; high-frequency compensation



Academic Editor: Guanqiu Qi

Received: 27 May 2026

Revised: 8 July 2026

Accepted: 13 July 2026

Published: 16 July 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Image super-resolution reconstruction (SRR) aims to recover a high-resolution (HR) image from a low-resolution (LR) input through a series of transformations [1]. Since it was first proposed in the 1980s, SRR has attracted extensive attention from researchers and has become an important research topic in computer vision and image processing due to its broad practical applications and theoretical significance [2]. Existing super-resolution methods can generally be divided into interpolation-based methods [3,4], reconstruction-based methods [5], and deep learning-based methods [6,7]. Interpolation-based methods mainly estimate unknown pixel values from known neighboring pixels using mathematical functions in the spatial domain, thereby enlarging the image size. However, these methods

essentially generate pixel values rather than recover high-frequency details, which often leads to blurred edges, distorted textures, and ringing artifacts in the enlarged images. Reconstruction-based methods attempt to restore missing high-frequency details by exploiting the intrinsic structure and prior knowledge of images. Nevertheless, they usually require high computational costs, are difficult to apply to large-scale images, and still struggle to suppress artifacts and ringing effects. In recent years, convolutional neural network-based super-resolution methods have demonstrated remarkable performance.

Generative adversarial networks (GANs) [8] have brought revolutionary advances to the field of image generation through their unique adversarial training mechanism. In particular, they have shown significant advantages in producing visually realistic results, providing a new technical route for image reconstruction. Based on this idea, Ledig et al. proposed SRGAN, which was the first to introduce the GAN framework into single-image super-resolution reconstruction [9]. SRGAN innovatively designed a perceptual loss function that combines content loss with adversarial loss, successfully overcoming the limitations of traditional super-resolution methods in reconstructing high-frequency details. It not only significantly improves the visual quality of reconstructed images, but also generates more natural and realistic texture details, making the reconstructed images better aligned with human visual perception and laying an important foundation for subsequent GAN-based super-resolution studies.

Although SRGAN achieves favorable visual results in image super-resolution reconstruction, it still has several limitations. First, the feature extraction structure in its generator is relatively simple, which limits its ability to represent multi-level textures and edge details. Second, the upsampling process lacks an explicit high-frequency compensation mechanism, making the reconstructed images prone to blurred edges, texture loss, and ringing artifacts. Third, the original loss function provides insufficient constraints on local structural consistency, so structural distortion and artifacts may still occur in the reconstructed images. To address these issues, this paper proposes a novel GAN architecture based on SRGAN. Specifically, a Hierarchical Feature Aggregation Attention Block is introduced to improve the feature extraction capability of SRGAN, and a Laplacian Adaptive Upsampling module is designed to alleviate ringing artifacts and contour blurring during upsampling. Based on these improvements, a High-Frequency Compensation and Feature Aggregation Generative Adversarial Network (HFAGAN) is proposed. The main contributions of this paper are summarized as follows:

1. A Hierarchical Feature Aggregation Attention Block is proposed and used as the backbone of SRGAN. By progressively extracting image features at different levels through multiple convolutional layers and incorporating a high-frequency variance channel attention mechanism, the network enhances its ability to represent high-frequency details such as textures and edges, thereby improving the limited feature extraction capability of the original SRGAN.
2. A Laplacian high-frequency compensation upsampling module is designed. Based on sub-pixel upsampling, a Laplacian high-frequency branch is introduced to explicitly compensate for edge contours and texture details, thereby alleviating edge blurring and artifact problems in reconstructed images.
3. A dissimilarity structural similarity loss function is introduced to constrain the local structural consistency between reconstructed images and real images. This helps reduce edge distortion and structural artifacts, further improving the perceptual quality of reconstructed images.

2. Related Work

With the rise in deep learning, super-resolution methods based on convolutional neural networks (CNNs) have demonstrated stronger capability in feature representation than traditional methods and have gradually become the mainstream research direction in the field of image super-resolution.

Dong et al. [10] first applied convolutional neural networks to the super-resolution task and proposed the SRCNN model, which mainly consists of three stages: patch extraction and representation, nonlinear mapping, and image reconstruction. As subsequent studies revealed a close relationship between model performance and network depth, increasingly deeper super-resolution networks were proposed, such as VDSR [11], DRRN [12], and DBPN [13]. However, due to the effect of diminishing returns, simply increasing network depth became insufficient to continuously improve model performance, prompting researchers to explore other directions for model improvement. Shi et al. [14] found that sub-pixel convolution could improve image resolution and reconstruction efficiency, and proposed an efficient sub-pixel convolutional neural network, namely ESPCN. They introduced an efficient sub-pixel convolution layer into the neural network to perform image upsampling, thereby improving image resolution. Hu et al. [15] incorporated channel and attention mechanisms into their model to transform low-resolution features into more informative representations, achieving favorable reconstruction results.

Lim et al. [16] found that batch normalization layers tend to damage the contrast information of image features and have a negative impact on super-resolution tasks. They therefore proposed the Enhanced Deep Super-Resolution Network (EDSR). In this network, batch normalization layers were removed and a simplified residual block structure was adopted, which effectively reduced computational cost. This design made it possible to construct deeper and wider networks, significantly improving the accuracy and efficiency of image reconstruction. Zhang et al. [17] observed that different channel features in deep networks contribute unequally to reconstruction performance, and proposed the Residual Channel Attention Network (RCAN). They introduced a residual-in-residual structure and integrated a channel attention mechanism, which adaptively recalibrates the feature weights of each channel. As a result, the network can focus more on informative features, enabling the training of extremely deep networks with more than 400 layers and achieving excellent reconstruction quality.

With their unique adversarial mechanism, GAN models have attracted widespread attention in the field of image generation, leading to the development of models such as DCGAN [18], WGAN [19], and CGAN [20]. SRGAN was the first GAN-based model introduced into the field of super-resolution. Based on the core idea of GANs, SRGAN consists of a generator (G) and a discriminator (D). The generator extracts features from the input samples, learns the data distribution, and outputs generated samples. The real samples and generated samples are then paired as input data and fed into the discriminator. After processing by the discriminator, a classification result is obtained to determine whether the input sample is real or generated. This result is fed back to the generator, and the two networks are continuously trained and updated until they eventually reach a Nash equilibrium [21].

The generator continuously learns from the feedback provided by the discriminator and generates images that aim to deceive it, while the discriminator is responsible for identifying whether an image is generated by the generator. The two networks therefore form a competitive and adversarial relationship. The structure of the generator is shown in Figure 1.

The generator is composed of multiple residual blocks [22], which are used for feature extraction while helping stabilize the network. In super-resolution networks, especially when clear and realistic details need to be recovered from low-resolution images, it is necessary to capture deep semantic information. Shallow networks can only extract low-level features such as edges and colors and are unable to effectively capture deeper information. However, when the network depth reaches a certain level, the training error may increase instead of decreasing as expected [23]. The idea of residual blocks is to allow the network to learn the difference between the input and the output, making it easier to learn an identity mapping, preserve the original information flow, and avoid the problem above. By stacking multiple residual blocks, SRGAN forms a deeper network that can capture more complex and abstract features.

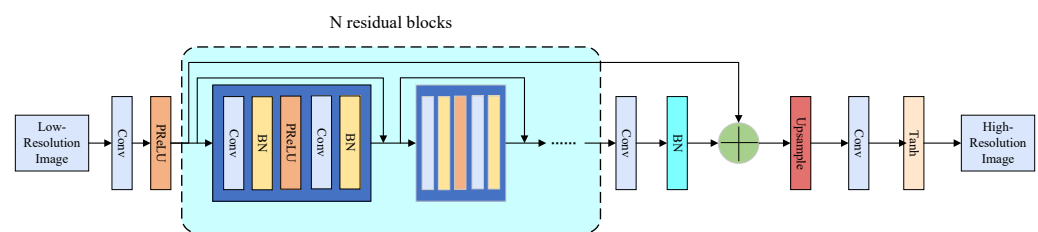


Figure 1. SRGAN generator.

The discriminator network extracts features from the input image through multiple convolutional layers, determines whether the image is an original high-resolution image or a reconstructed high-resolution (SR) image, and then feeds back the classification result [24]. Its structure is similar to that of classification models such as VGG [25], as it is also a purely convolutional neural network. However, unlike conventional classification networks, the discriminator uses strided convolutions for downsampling instead of pooling layers. From an overall perspective, the discriminator performs a feature extraction and compression process that transforms the input from high resolution with a small number of channels to low resolution with a large number of channels. In addition, unlike general classification models, the discriminator is designed for binary classification, that is, to distinguish between real and generated images.

The design of the discriminator has two main purposes. First, through adversarial training with the generator, it encourages both networks to improve together: as the discriminator becomes more capable of distinguishing real from generated images, the reconstruction ability of the generator is also enhanced. Second, the discriminator provides gradients for the generator. Its classification result is back-propagated as important feedback to guide the generator in further improving the reconstructed images. It is precisely because of the discriminator that SRGAN can overcome the overly smooth and blurry results often produced by traditional super-resolution methods. The discriminator is also one of the key components that enables the concept of perceptual loss in SRGAN. The structure of the discriminator is shown in Figure 2.

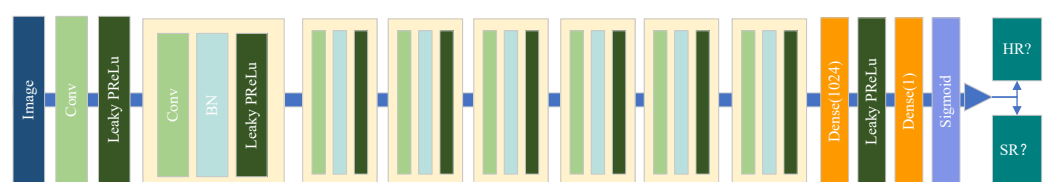


Figure 2. SRGAN discriminator.

3. Method

3.1. Network Architecture

Figure 3 shows the improved HFAGAN model. The HFAGAN model still follows the generator–discriminator framework of SRGAN. The generator consists of a Deep Feature Extraction Block (DFEB) and a Laplacian Adaptive Upsampling (LAU) module. In the DFEB, the proposed Hierarchical Feature Aggregation Attention Block (HFAB) is used as the backbone network to replace the original residual blocks. This design increases the network depth, improves the model’s multi-scale feature processing capability, and enables clearer image reconstruction. The LAU module introduces a Laplacian compensation path during upsampling to improve the sharpness of image edges and textures.

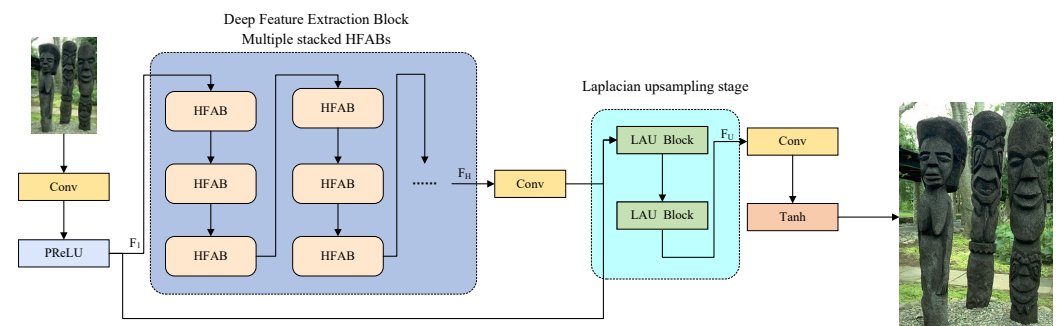


Figure 3. Structure of the HFAGAN Model.

The overall module expression is given in Equation (1):

$$F_u = M_{LAU}(F_1 + F_H) \quad (1)$$

where F_1 denotes the extracted shallow features, F_H denotes the features extracted by the stacked HFAB modules, and $M_{LAU}(\cdot)$ denotes the LAU module.

In the generator, the number of stacked HFAB modules is set to 23, and the basic feature channel number is set to 64. Each HFAB contains three consecutive 3×3 convolutional layers, followed by feature concatenation and channel fusion. For $\times 4$ super-resolution, two LAU modules are used, and each LAU performs $\times 2$ upsampling. The convolution kernels in the LAU module are mainly set to 3×3 , while 1×1 convolution is used for channel projection. No batch normalization is used in the generator, since normalization layers may weaken pixel-level details in super-resolution reconstruction.

3.2. Deep Feature Extraction Block

To address the problems of the overly simple structure and insufficient feature extraction capability of the original residual blocks, stacked HFAB modules are adopted in the Deep Feature Extraction Block. The core of HFAB consists of three consecutive 3×3 convolutional layers, which construct a multi-level feature extraction path. Features at different levels are concatenated and fused along the channel dimension, and a dynamic channel attention mechanism is introduced to adaptively recalibrate the channel weights. The module uses the PReLU activation function to enhance nonlinear representation capability. Meanwhile, a scaling factor is employed to control the strength of the residual connection, allowing the network to effectively capture multi-scale detail features while maintaining training stability. As a result, the proposed module significantly improves the image quality and detail restoration capability of super-resolution reconstruction.

3.2.1. HFAB Module

Batch normalization (BN) layers can accelerate network training and convergence in most deep learning tasks, such as image classification [26]. They also help control gradient explosion, alleviate gradient vanishing, and reduce overfitting. However, for pixel-level tasks such as image enhancement and super-resolution reconstruction, BN layers may weaken fine image details and adversely affect image generation quality [27]. Therefore, BN layers are removed in the HFAB module. Instead, three consecutive convolutional layers are adopted to process the input feature maps, forming a progressive process of dimensional expansion, refinement, and abstraction. The first convolutional layer projects the input features into a high-dimensional space. The subsequent convolutional layers further perform deep nonlinear refinement in the expanded feature space, while the activation function enhances the smoothness and representation capability of the features. This progressive feature extraction design not only maintains parameter efficiency, but also gradually expands the effective receptive field, enabling the network to capture broader contextual dependencies and provide rich multi-scale information for subsequent dense feature fusion. Its mathematical expression is shown in Equation (2):

$$\begin{aligned} F_1 &= \sigma(W_1 * F_{in} + b_1), \\ F_k &= \sigma(W_k * F_{k-1} + b_k), \quad k = 2, 3 \end{aligned} \quad (2)$$

where W_1 and b_1 denote the weight and bias of the first convolutional layer, respectively. W_k and b_k denote the weight and bias of the k -th convolutional layer for $k = 2, 3$. These layer-specific parameters are used to progressively transform and refine the feature representation at different convolutional stages. F_{in} denotes the input feature map, F_k denotes the feature map extracted by the k -th convolutional layer, and $\sigma(\cdot)$ denotes the activation function.

Inspired by DenseNet [28] and RDN [29], dense feature fusion is introduced into the HFAB module. Specifically, the intermediate features generated by each convolutional layer within the module are extracted and preserved, and then aggregated through channel-wise concatenation. This design enables the subsequent attention module to recalibrate features in a multi-scale feature space that contains low-level textures, mid-level structures, and high-level semantic information, as shown in Figure 4.

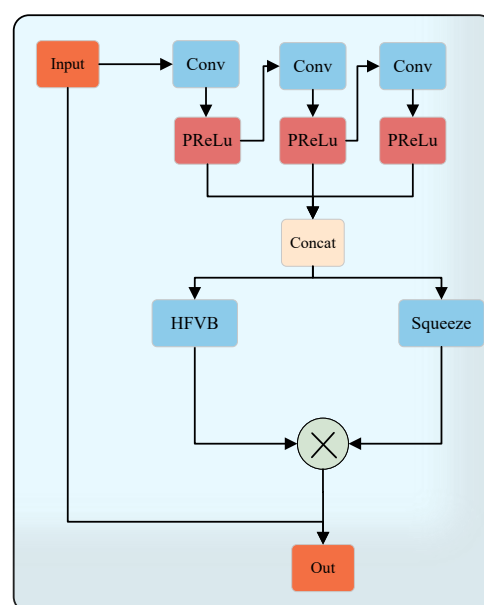


Figure 4. HFAB network architecture.

The effectiveness of HFAB lies mainly in two aspects: receptive field expansion and multi-level feature reuse. The consecutive convolutions do not simply increase the network depth; instead, they form a hierarchical feature extraction process from shallow to deep. The first convolutional layer mainly focuses on local edges, color variations, and fine textures. The second convolutional layer further aggregates contextual information and better extracts overall contours. The third convolutional layer captures higher-level semantic information and integrates the relationship between local details and global structures.

Meanwhile, HFAB concatenates the intermediate features generated by different convolutional layers along the channel dimension, allowing shallow texture information, intermediate structural information, and deep semantic features to jointly participate in the subsequent reconstruction process. For super-resolution reconstruction, edge textures and fine structural details are easily weakened during deep feature propagation. Therefore, compared with the relatively simple residual blocks in the original SRGAN, HFAB can enhance the model's ability to represent high-frequency textures and edge structures from the feature extraction stage.

Specifically, each HFAB takes a 64-channel feature map as input. The three internal convolutional layers use 128 output channels, and their outputs are concatenated into a 384-channel feature map, which is then reduced to 64 channels by a 1×1 fusion convolution.

The main difference between HFAB and conventional residual or dense blocks is that HFAB not only reuses intermediate features, but also organizes them into a hierarchical feature space for subsequent high-frequency-oriented channel recalibration.

3.2.2. High-Frequency Variance Channel Attention

Attention mechanisms have shown great potential in improving the performance of convolutional neural networks. Among them, the Convolutional Block Attention Module (CBAM) is one of the representative methods [30]. It adaptively recalibrates features by sequentially inferring a channel attention map and a spatial attention map. The structure of CBAM is shown in Figure 5.

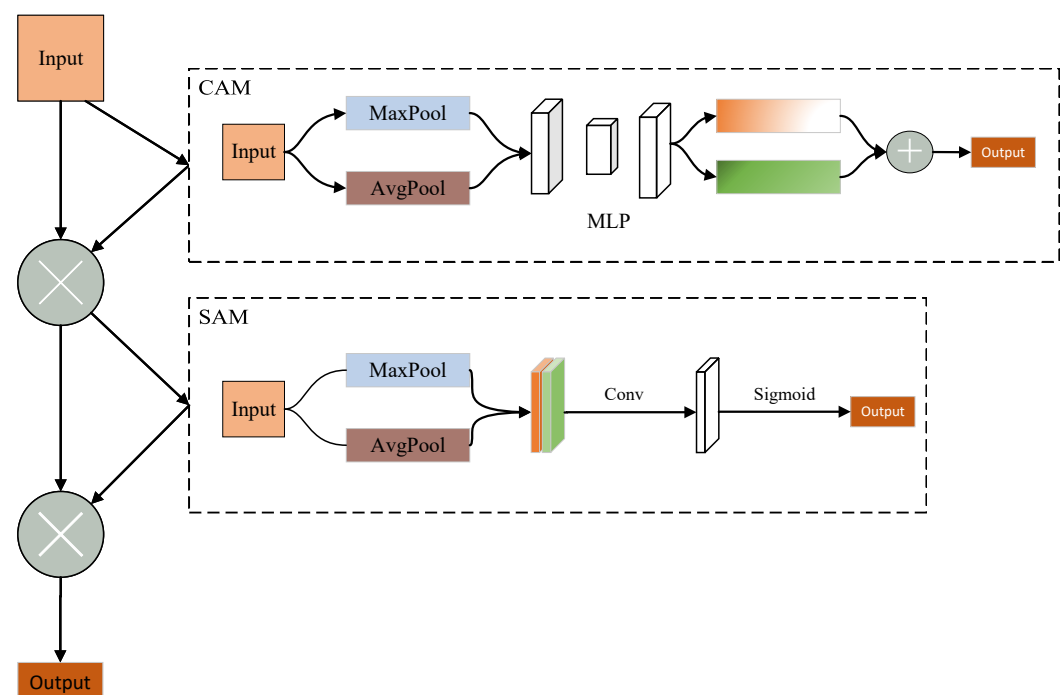


Figure 5. Structure of CBAM.

To adaptively select key features from dense features containing rich multi-scale information, a customized High-Frequency Variance Adaptive Channel Attention Block (HFVB) is integrated into HFAB. As shown in Figure 6, considering that the hierarchical convolutional feature extraction structure in HFAB already has strong capability for capturing spatial contextual information, directly applying the complete CBAM may introduce unnecessary computational redundancy. Therefore, the design of CBAM is adopted with several key adaptations and improvements. First, the spatial attention branch is removed, and only channel attention is retained to focus on the selection and suppression of feature channels. Unlike the original standard CBAM, which operates on a single input, the proposed attention mechanism can directly receive inputs from multiple scales.

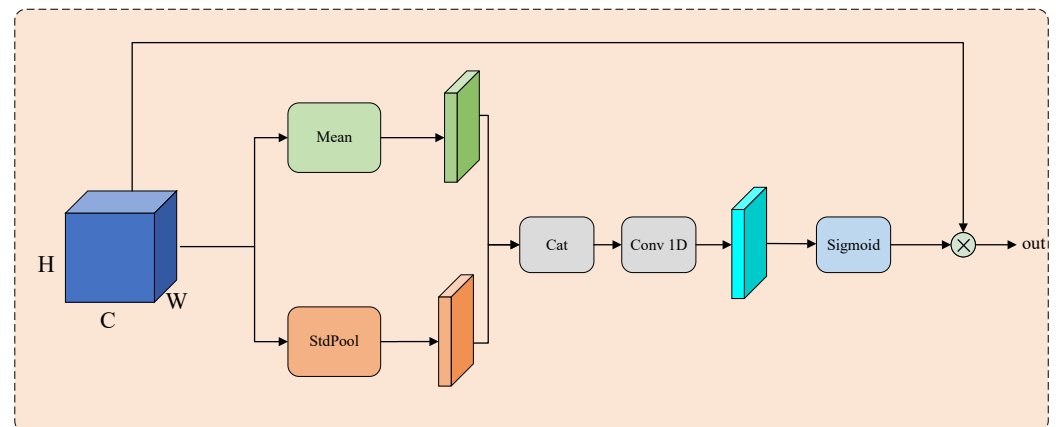


Figure 6. Architecture of the HFVB.

Unlike traditional channel attention methods, the proposed high-frequency variance channel attention does not use max pooling as the main descriptor of high-frequency responses. This is because although max pooling can highlight local strong-response regions, local maximum values in super-resolution reconstruction are often susceptible to noise points, artifacts, or abnormal activations. As a result, the attention weights may be biased toward a small number of prominent pixel responses while ignoring the overall texture distribution. Therefore, this paper combines average pooling with standard deviation pooling to model channel features from the perspectives of low-frequency structures and high-frequency variations.

Average pooling reflects the overall response intensity of each channel and is mainly used to describe low-frequency background information, color distribution, and basic structural information in an image. For super-resolution tasks, this information helps maintain the overall brightness, color consistency, and structural coherence of the image, thereby preventing large-scale color shifts or structural deformation in the generated results.

$$F_{std} = \sqrt{\frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (X_c(i, j) - F_{avg})^2 + \varepsilon} \quad (3)$$

where ε is a small constant used to ensure numerical stability. Since the gradient of the square-root operation may become large when the variance approaches zero, ε is added inside the square root to avoid division by an extremely small value during backpropagation. Therefore, the StdPool operation remains stable even in smooth regions with weak spatial response variations. F_{avg} denotes the global response of each channel after average pooling, H and W represent the height and width of the feature map, respectively, $X_c(i, j)$ denotes the feature value of the c -th channel at spatial position (i, j) . Standard deviation pooling can reflect the dispersion degree of pixel responses in the feature map. When a channel contains

abundant high-frequency details, such as edges, textures, and contours, its spatial responses usually exhibit stronger fluctuations, resulting in a larger standard deviation. In contrast, the responses in smooth background regions are relatively stable, leading to a smaller standard deviation. Therefore, by introducing standard deviation pooling, the network can explicitly perceive the variation intensity of high-frequency textures in each channel, thereby enhancing its attention to edge contours, fine textures, and complex structural regions.

Subsequently, to avoid the loss of high-frequency information caused by the dimensionality reduction operation in traditional attention mechanisms, a one-dimensional convolution without dimensionality reduction is adopted to achieve information interaction across channels. Finally, the average-response branch and the standard-deviation-response branch are fused, and the final channel attention weights are generated through the Sigmoid function.

Compared with the channel attention mechanism in the original CBAM, which uses “average pooling + max pooling,” the proposed high-frequency variance channel attention has two advantages. First, standard deviation pooling can better reflect the overall fluctuation degree of spatial responses than max pooling, thereby avoiding excessive interference from individual abnormal responses in attention allocation. Second, the one-dimensional convolution without dimensionality reduction avoids the possible loss of high-frequency details caused by channel compression in traditional multilayer perceptrons. As a result, weak textures, fine edges, and other easily overlooked information can still participate in the calculation of channel weights. Therefore, this module is more suitable for recovering high-frequency textures and edge details in image super-resolution reconstruction.

The key difference between HFVB and CBAM-like attention is that HFVB does not rely on max pooling or an additional spatial attention branch. Instead, it uses standard-deviation pooling to explicitly describe high-frequency spatial response variations, which is more suitable for edge and texture reconstruction.

3.3. Laplacian Upsampling Layer

Traditional upsampling operations often fail to accurately recover high-frequency textures, resulting in overly smooth reconstructed images and artifact problems. ESPCN demonstrates the efficiency of sub-pixel convolution, while LapSRN uses a Laplacian pyramid structure for residual prediction [31]. Inspired by these methods, this paper proposes a Laplacian Adaptive Upsampling module, as shown in Figure 7.

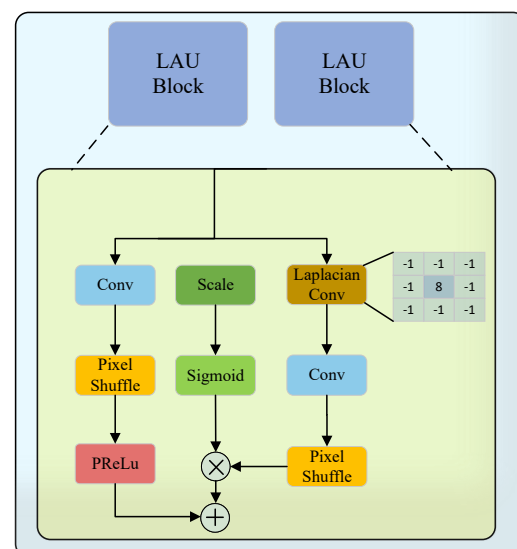


Figure 7. LAU Module.

The LAU module adopts a dual-branch parallel structure. The main branch uses a sub-pixel convolution layer to recover the basic semantic information and low-frequency content of the image. Let F_{in} denote the input feature. The output of this branch is given in Equation (4):

$$F_{base} = \rho_s(W_{base} * F_{in} + b) \quad (4)$$

where $\rho_s(\cdot)$ denotes the pixel shuffle operation, W and b denote the weight and bias of the convolutional layer, respectively, and F_{base} denotes the output of the main branch.

Meanwhile, to compensate for the lost high-frequency details, a high-frequency enhancement branch is introduced. This branch uses a fixed Laplacian operator K_{lap} as a non-trainable inductive bias to explicitly extract high-frequency information such as edges and textures from the input features. The extracted high-frequency information is then projected and upsampled to generate a high-frequency feature map F_{edge} . The calculation process is shown in Equation (5):

$$K_{lap} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

$$F_{edge} = \rho_s(w_{proj} * (K_{lap} \odot F_{in})) \quad (5)$$

Here, $\odot$ denotes the channel-wise application of the fixed Laplacian kernel to F_{in} , which is implemented as a depthwise convolution. In this operation, each input channel is filtered independently, and no cross-channel information mixing is performed. The learnable 1×1 projection convolution w_{proj} is then used as a pointwise convolution to map the extracted high-frequency responses to the target dimension required by the pixel shuffle operation. Therefore, the high-frequency branch follows a depthwise-separable-style design, but the Laplacian filtering itself is specifically a fixed depthwise convolution. F_{edge} denotes the output of the high-frequency branch.

Finally, a learnable scaling factor α is introduced to perform weighted fusion of the outputs from the two branches so as to balance structural information and texture details. The final upsampled output is obtained as shown in Equation (6):

$$F_{up} = F_{base} + \sigma(\alpha) \cdot F_{edge} \quad (6)$$

where $\sigma(\alpha)$ denotes the sigmoid function, which ensures that the weight is dynamically adjusted within the range of $(0, 1)$. In our implementation, the learnable scaling factor α is initialized to 0.3. This constraint prevents excessive high-frequency compensation during the early training stage and helps stabilize generator training. Through this explicit frequency compensation mechanism, LAU can effectively enhance image edge details.

Therefore, LAU differs from conventional sub-pixel upsampling and Laplacian-pyramid methods by directly embedding a Laplacian high-frequency compensation branch into the upsampling process and adaptively controlling its contribution through a learnable scaling factor.

3.4. Loss Function

SRGAN improves the visual realism of reconstructed images by introducing perceptual loss. Its loss function mainly consists of content loss and adversarial loss. Inspired by this idea, this paper jointly uses image loss, adversarial loss, and perceptual loss to constrain the training of the generator. The image loss is used to reduce the pixel-level error between the reconstructed image and the real high-resolution image, the adversarial loss is used to improve the visual realism of the generated image, and the perceptual loss is used to

constrain the difference between the generated image and the real image in the deep feature space. The basic loss function is given in Equation (7):

$$L_{base} = L_{image} + \lambda_{adv}L_{adv} + \lambda_{per}L_{per} \quad (7)$$

where L_{base} denotes the basic generator loss, L_{image} denotes the image loss, L_{adv} denotes the adversarial loss, L_{per} denotes the perceptual loss, and λ_{adv} and λ_{per} denote the weight coefficients of the adversarial loss and perceptual loss, respectively.

However, relying on the above loss functions only makes it difficult to sufficiently constrain the local structural consistency of reconstructed images, which may lead to edge distortion and structural artifacts. Therefore, this paper further introduces the Dissimilarity Structural Similarity Index Measure (DSSIM) loss function, which is defined in Equation (8):

$$L_{DSSIM} = 1 - SSIM(I_{SR}, I_{HR}) \quad (8)$$

where I_{SR} denotes the generated super-resolution image, I_{HR} denotes the real high-resolution image, and $SSIM(\cdot)$ denotes the structural similarity calculation. After introducing DSSIM, the final generator loss function is given in Equation (9):

$$L_G = L_{base} + \lambda_{dssim}L_{DSSIM} \quad (9)$$

By introducing the DSSIM loss, the model can further enhance the local structural consistency between the reconstructed image and the real image, thereby reducing edge distortion and structural artifacts.

To provide a clearer and more intuitive understanding of the proposed method, the overall flowchart of HFAGAN is shown in Figure 8. The LR image is first processed by shallow feature extraction and then fed into the Deep Feature Extraction Block, where stacked HFAB modules are used to extract and aggregate hierarchical features. Subsequently, the fused features are sent to the LAU module for upsampling and Laplacian high-frequency compensation, generating the final SR image. During training, the reconstructed SR image is jointly optimized using the HR image, discriminator feedback, and the combined loss function, including image loss, perceptual loss, adversarial loss, and DSSIM loss.

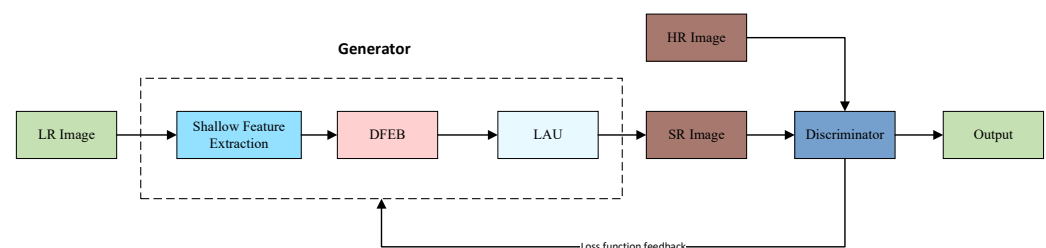


Figure 8. Overall flowchart of the proposed HFAGAN method.

4. Experimental Results and Analysis

4.1. Dataset

In this paper, the DIV2K dataset is used as the training and validation set, while four widely used public benchmark datasets, namely Set5, Set14, BSD100, and Urban100, are selected as the test sets. The DIV2K dataset contains a large number of high-quality natural images with 2K resolution and rich image content, providing sufficient texture, edge, and structural information for super-resolution models.

During the training stage, the DIV2K training set is used, and data augmentation operations such as horizontal flipping and vertical flipping are applied to the training

images to increase the number of samples and improve the generalization ability of the model. During the validation stage, the DIV2K validation set is used to observe the convergence behavior and performance changes of the model during training.

During the testing stage, experiments are conducted on Set5, Set14, BSD100, and Urban100, respectively. Among them, Set5 is mainly used to evaluate the model's ability to recover basic structures and edge contours; Set14 contains more complex texture and structural information; BSD100 includes a larger number of images with more diverse natural scenes, making it suitable for comprehensively evaluating the generalization performance of the model. In addition, Urban100 contains many urban scenes with dense edges, repeated building textures, and regular geometric structures, which places higher requirements on the recovery of high-frequency details and line structures. Therefore, introducing Urban100 can further verify the effectiveness of the proposed method in complex structural scenarios.

4.2. Experimental Environment and Training Parameter Settings

The hardware environment used in this experiment is as follows: the CPU is an Intel(R) Core(TM) i9-14900KF @ 3.20 GHz, and the GPU is an NVIDIA GeForce RTX 5070 Ti with 16 GB of video memory. The software environment includes the Python 3.10.16 programming language, Windows 11 operating system, PyTorch 2.8.0 deep learning framework, and CUDA 12.8 for accelerated computation. The Adam optimizer is used for both generator and discriminator optimization. The learning rate is set to 1×10^{-4} . The Adam parameters are configured as $\beta_1 = 0.9$ and $\beta_2 = 0.999$.

During the training stage, the images are cropped into smaller patches before being fed into the model. The generator is first pretrained for 50 epochs, followed by 200 epochs of adversarial training. The batch size is set to 8, and the Adam optimizer is used to improve the convergence speed.

4.3. Evaluation Metrics

This paper uses Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), and Feature Similarity Index Measure (FSIM) as evaluation metrics.

PSNR is one of the most common and widely used evaluation metrics in the field of super-resolution. It calculates the mean squared error (MSE) between two images at the pixel level and compares it with the maximum possible signal value of the image. Although SRGAN usually achieves lower PSNR than traditional super-resolution networks, its visual perceptual quality is superior to that of many conventional methods [32].

Different from PSNR, SSIM focuses more on the structural consistency between images. It evaluates image similarity from the perspectives of luminance, contrast, and structure, which makes it more consistent with the human visual perception of image quality.

LPIPS is introduced because PSNR and SSIM cannot fully represent the actual perceptual quality of reconstructed images [33]. The core idea of LPIPS is that human perception of image similarity relies more on high-level semantic features than on low-level pixels or structural information. It uses deep learning models to automatically learn human perceptual judgments and has become a new paradigm for image quality assessment.

To further measure the recovery of structural details and edge textures, FSIM is also used in this paper. FSIM mainly relies on phase congruency and gradient magnitude to evaluate feature similarity between images. Phase congruency reflects significant structural information such as edges, corners, and textures, while gradient magnitude describes local grayscale variations and edge strength.

4.4. Quantitative Analysis of the Model

To verify the effectiveness of the proposed HFAGAN model, comparative experiments are conducted on four public test datasets: Set5, Set14, BSD100, and Urban100. The proposed method is compared with several classical super-resolution models, including Bicubic, SRCNN [10], SRGAN [9], VDSR [11], and ESRGAN [34]. Four evaluation metrics, namely PSNR, SSIM, LPIPS, and FSIM, are adopted to comprehensively evaluate model performance from multiple perspectives, including reconstruction accuracy, structural similarity, and visual perceptual quality.

As shown in Table 1, the proposed method achieves favorable overall performance on the Set5 dataset. Specifically, the PSNR reaches 29.01 dB, which is the highest among all compared methods, and the LPIPS is reduced to 0.1097, also achieving the best result. Compared with SRGAN, the proposed method improves PSNR by 0.31 dB and reduces LPIPS by 0.0234, indicating that the proposed model improves both pixel-level reconstruction accuracy and visual perceptual quality. Compared with ESRGAN, HFAGAN shows a slight advantage in PSNR and LPIPS, while its SSIM and FSIM values are marginally lower. This indicates that the proposed method does not uniformly outperform ESRGAN on all metrics, but provides a modest improvement mainly in reconstruction accuracy and perceptual similarity.

Table 1. Performance metrics on the Set5 dataset.

Method	PSNR	SSIM	LPIPS	FSIM
Bicubic	26.89	0.7934	0.3450	0.8480
SRCNN	27.81	0.8120	0.2120	0.8836
SRGAN	28.70	0.8301	0.1331	0.9109
VDSR	28.87	0.8424	0.1556	0.9078
ESRGAN	28.99	0.8382	0.1248	0.9136
HFAGAN	29.01	0.8368	0.1097	0.9117

As shown in Table 2, the proposed method also performs well across all four evaluation metrics on the Set14 dataset. Specifically, the PSNR reaches 25.70 dB, the SSIM reaches 0.7332, the LPIPS is reduced to 0.2001, and the FSIM reaches 0.9121. Relative to SRGAN, HFAGAN improves all evaluation metrics, demonstrating enhanced reconstruction accuracy and perceptual quality. It also slightly outperforms ESRGAN in terms of PSNR, SSIM, LPIPS, and FSIM. Since the Set14 dataset contains more complex texture and structural information, it places higher requirements on the feature extraction and detail restoration capabilities of the model.

Table 2. Performance metrics on the Set14 dataset.

Method	PSNR	SSIM	LPIPS	FSIM
Bicubic	24.16	0.6927	0.4416	0.8372
SRCNN	24.80	0.7129	0.3013	0.8830
SRGAN	25.52	0.7287	0.2184	0.9082
VDSR	25.45	0.7303	0.2456	0.9027
ESRGAN	25.69	0.7325	0.2106	0.9103
HFAGAN	25.70	0.7332	0.2001	0.9121

As shown in Table 3, the proposed method also achieves favorable experimental results on the BSD100 dataset. Compared with SRGAN, the PSNR of the proposed method increases from 25.64 dB to 25.78 dB, the SSIM increases from 0.6906 to 0.6959, the LPIPS decreases from 0.3157 to 0.2935, and the FSIM increases from 0.8226 to 0.8296. The BSD100

dataset contains a larger number of images and more diverse scene content, which makes it more suitable for comprehensively evaluating the generalization performance of the model. The experimental results show that HFAGAN can still maintain stable reconstruction performance in complex natural image scenarios. This indicates that the proposed method not only improves perceptual quality, but also takes structural information preservation and detail restoration into account to a certain extent.

Table 3. Performance metrics on the BSD100 dataset.

Method	PSNR	SSIM	LPIPS	FSIM
Bicubic	24.75	0.6655	0.5314	0.7662
SRCNN	25.22	0.6841	0.3928	0.8029
SRGAN	25.64	0.6906	0.3157	0.8226
VDSR	25.67	0.6923	0.3449	0.8152
ESRGAN	25.76	0.6948	0.3035	0.8260
HFAGAN	25.78	0.6959	0.2935	0.8296

As shown in Table 4, the proposed method also achieves stable performance on the Urban100 dataset. Urban100 contains a large number of urban building images with dense edges, repeated window patterns, straight lines, and complex structural textures, making it more challenging than general natural image datasets. Compared with SRGAN, HFAGAN improves PSNR from 23.40 dB to 23.54 dB, increases SSIM from 0.7158 to 0.7212, reduces LPIPS from 0.2278 to 0.2056, and improves FSIM from 0.9744 to 0.9769. These results indicate that the proposed model can better preserve structural information and recover high-frequency details in complex urban scenes.

Table 4. Performance metrics on the Urban100 dataset.

Method	PSNR	SSIM	LPIPS	FSIM
Bicubic	20.43	0.5781	0.7564	0.4921
SRCNN	22.35	0.6718	0.3393	0.9631
SRGAN	23.40	0.7158	0.2278	0.9744
VDSR	23.22	0.7142	0.2597	0.9731
ESRGAN	23.51	0.7188	0.2132	0.9753
HFAGAN	23.54	0.7212	0.2056	0.9769

HFAGAN also achieves modest improvements over ESRGAN across all four evaluation metrics on Urban100. Specifically, PSNR and SSIM increase by 0.03 dB and 0.0024, respectively, while LPIPS decreases by 0.0076 and FSIM improves by 0.0016. Although these gains are numerically limited, they suggest that the proposed hierarchical feature aggregation and Laplacian high-frequency compensation modules can provide incremental benefits for reconstructing dense edges and regular geometric textures. Therefore, the results on Urban100 further demonstrate the generalization ability of the proposed method in more complex structural scenarios.

4.5. Ablation Study

To verify the contribution of each module to model performance, ablation experiments are conducted by progressively introducing the four variants into the baseline SRGAN: the Hierarchical Feature Aggregation Attention Block embedded with CBAM (+HFAB-C), the Hierarchical Feature Aggregation Attention Block integrated with HFVB (+HFAB-H), the LAU upsampling module (+LAU), and the improved loss function (+Loss). The experimental results are shown in Table 5.

As can be seen from the experimental results, after introducing the Hierarchical Feature Aggregation Attention Block, the PSNR, LPIPS, and FSIM metrics are all clearly improved on the three datasets. On the Set5 dataset, the PSNR increases from 28.70 dB to 28.83 dB, indicating that the hierarchical convolutional feature extraction and multi-scale feature fusion in HFAB can effectively enhance the feature extraction capability of the network and improve image reconstruction accuracy. Meanwhile, the significant decrease in LPIPS also shows that this module can better recover high-frequency details such as textures and edges.

Table 5. Ablation study results.

Dataset	Model	PSNR	SSIM	LPIPS	FSIM
Set5	SRGAN	28.70	0.8301	0.1331	0.9109
	+HFAB-C	28.76	0.8293	0.1291	0.9109
	+HFAB-H	28.83	0.8308	0.1285	0.9100
	+LAU	29.00	0.8363	0.1252	0.9125
	+Loss	29.01	0.8368	0.1097	0.9117
Set14	SRGAN	25.52	0.7287	0.2184	0.9082
	+HFAB-C	25.47	0.7229	0.2172	0.9069
	+HFAB-H	25.56	0.7253	0.2161	0.9084
	+LAU	25.67	0.7290	0.2112	0.9098
	+Loss	25.70	0.7332	0.2001	0.9121
BSD100	SRGAN	25.64	0.6906	0.3157	0.8226
	+HFAB-C	25.63	0.6873	0.3090	0.8230
	+HFAB-H	25.67	0.6886	0.3079	0.8239
	+LAU	25.74	0.6918	0.3038	0.8247
	+Loss	25.78	0.6959	0.2935	0.8296

After further introducing the LAU module, all metrics continue to improve. In particular, the improvement in FSIM is relatively evident, indicating that the Laplacian high-frequency compensation mechanism can effectively enhance the recovery of image edge contours and texture structures. This is consistent with the design idea proposed above, namely compensating for detailed information through a high-frequency branch. At the same time, the LPIPS value is further reduced, demonstrating that the LAU module plays a positive role in improving visual realism and reducing artifacts.

Finally, after adding the improved loss function, the model achieves the best LPIPS results on all three datasets. Specifically, the LPIPS decreases from 0.2112 to 0.2001 on Set14 and from 0.3038 to 0.2935 on BSD100, indicating that the improved loss function further enhances the model's ability to optimize perceptual quality. In addition, PSNR, SSIM, and FSIM also remain at relatively high levels, suggesting that the model can improve visual perceptual quality while still taking image structure and reconstruction accuracy into account.

Notably, the comparison between HFAB-C and HFAB-H further demonstrates the advantage of the proposed HFVB over CBAM. As shown in Table 5, HFAB-H achieves better PSNR, SSIM, and LPIPS than HFAB-C on Set5, Set14, and BSD100, indicating improvements in reconstruction accuracy, structural similarity, and perceptual quality. In addition, HFAB-H obtains higher FSIM on Set14 and BSD100, while only showing a very slight decrease on Set5. Overall, these results suggest that HFVB is more suitable for the proposed HFAB structure than CBAM. This is mainly because HFVB introduces standard-deviation pooling to explicitly capture high-frequency response variations, enabling the network to better focus on texture fluctuations, edge contours, and fine details in super-resolution reconstruction.

Overall, HFAB, LAU, and the improved loss function optimize the model from three aspects: feature extraction, high-frequency compensation, and structural constraint, respectively. The ablation results demonstrate that each module contributes positively to the improvement of model performance. When combined, the three components achieve the best or near-best overall performance, verifying the effectiveness of the proposed method.

4.6. Qualitative Analysis

To further verify the effectiveness of the proposed HFAGAN model in terms of visual perceptual quality, qualitative comparison experiments are conducted with several classical super-resolution models, including Bicubic, SRCNN, SRGAN, VDSR, and ESRGAN. The experimental results are shown in Figure 9.

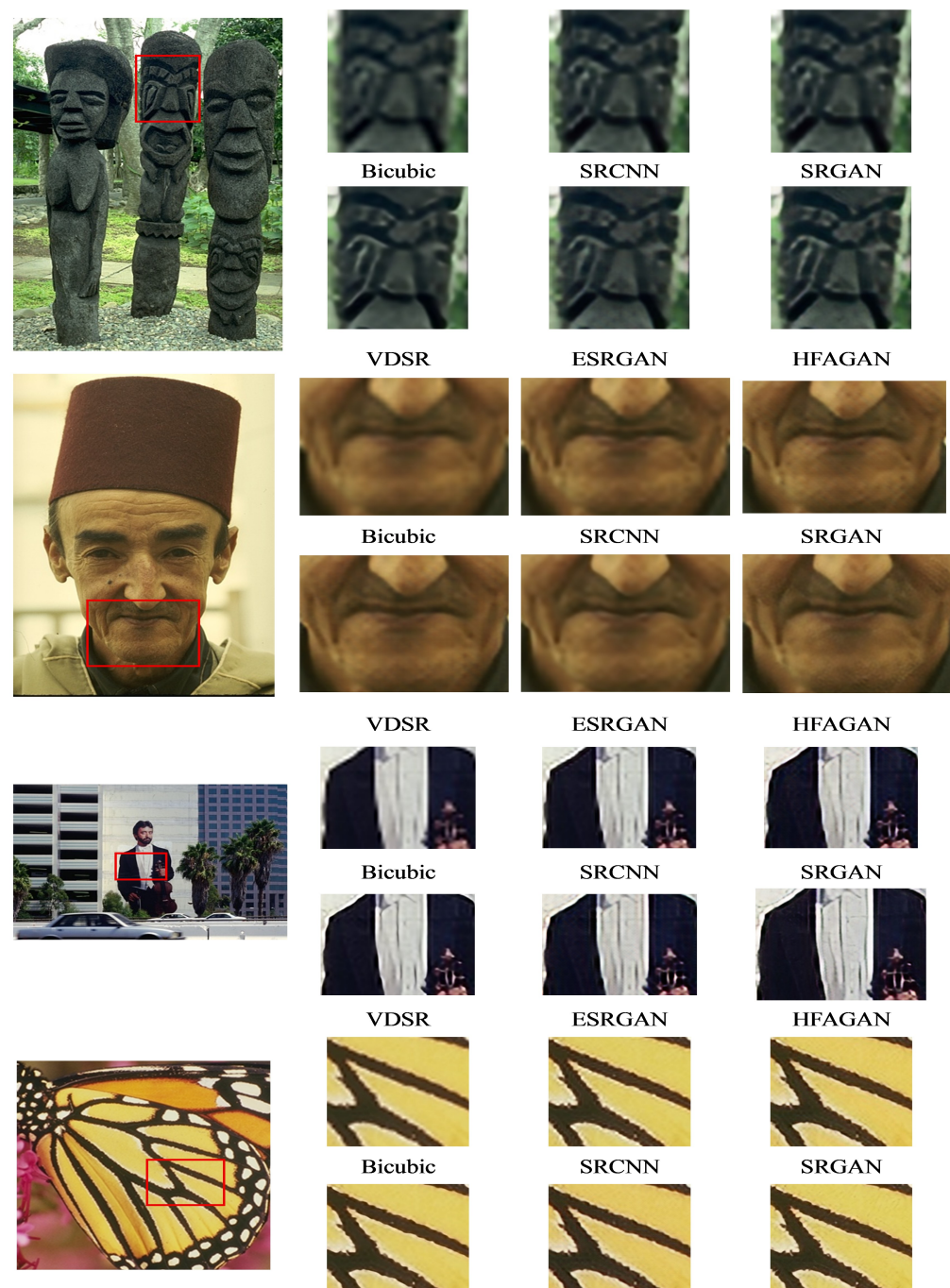


Figure 9. Qualitative comparison results.

In the first group of building scenes, HFAGAN tends to preserve clearer edge structures as well as the detailed textures of building edges. In particular, the line structures in local regions are clearer and more complete. Compared with the blurred edges produced by Bicubic and SRCNN, the proposed method achieves better line continuity and more natural local textures. Meanwhile, compared with the local artifacts and texture distortion observed in SRGAN and ESRGAN, the results generated by HFAGAN are closer to the real high-resolution images in terms of structural preservation and detail realism.

For the human image, HFAGAN shows more obvious restoration effects on the contour of the chin and the beard textures on both sides of the mouth. Compared with the blurred textures and missing details produced by other methods, the proposed method better preserves fine hair textures and edge transitions, making the facial structure clearer and more natural. As a result, its visual perceptual quality is closer to that of the real image.

For the line patterns on the clothing, the lines generated by HFAGAN are continuous, smooth, and sharp at the edges, while ringing effects are effectively avoided. HFAGAN can clearly restore local geometric shapes and fabric textures, achieving visual realism closer to the high-resolution reference image. This verifies the effectiveness of the proposed Laplacian upsampling layer in compensating for high-frequency details, as well as the advantage of the hierarchical feature aggregation module in preserving complex textures.

For the texture details of butterfly wings, HFAGAN shows more noticeable reconstruction effects on small spots and vein structures. Compared with the blurred color blocks or texture distortions produced by other methods, the proposed method better preserves the distribution of fine spots and the direction of veins, making the wing textures clearer and more natural. Its visual perceptual quality is therefore closer to that of the real high-resolution image.

5. Conclusions

Aiming at the limitations of existing super-resolution reconstruction methods in high-frequency texture recovery, edge detail preservation, and artifact suppression, this paper proposes HFAGAN, a generative adversarial network based on hierarchical feature aggregation and Laplacian high-frequency compensation. By introducing the HFAB module, the proposed method enhances the model's ability to extract multi-level features. Combined with the high-frequency variance channel attention mechanism, it further strengthens the representation of texture and edge information. In addition, the LAU module is used to compensate for high-frequency details during the upsampling process, thereby improving the clarity and visual naturalness of reconstructed images. Finally, the DSSIM loss function is introduced to constrain the local structural consistency of reconstructed images, reducing structural distortion and artifacts.

Experimental results show that the proposed method achieves favorable reconstruction performance on the Set5, Set14, BSD100, and Urban100 datasets. Compared with ESRGAN, HFAGAN also obtains modest improvements on several metrics, particularly in perceptual quality indicators such as LPIPS and FSIM. However, the numerical gains are relatively limited, indicating that the proposed method provides an incremental improvement rather than a substantial overall advantage. The qualitative experiments also demonstrate that the proposed method can effectively alleviate texture blurring and artifact problems.

Although the proposed method achieves good experimental results, there are still some limitations in restoring fine details in complex scenes and large-scale upsampling tasks. Meanwhile, the introduction of multi-scale feature fusion and attention mechanisms increases the computational complexity of the model to some extent. In future work, lightweight network structures, frequency-domain feature modeling, and super-resolution

reconstruction under real degradation scenarios will be further explored to improve the generalization ability and practical application value of the model.

Author Contributions: K.X.: conceptualization, methodology, validation, formal analysis, and writing—original draft; S.X.: data curation, formal analysis, and writing—review and editing; B.C.: formal analysis, resources, and writing—review and editing; Y.W.: writing—review and editing; R.u.H.M.: writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this article are all publicly available and can be found online.

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

References

1. Ye, L.; Zou, C. A survey of image super-resolution reconstruction based on deep learning. In *Proceedings of the International Conference on Wireless Communications, Networking and Applications*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 584–594.
2. Hu, C.-S.; Zhan, S.; Wu, C.-Z. Image super-resolution based on deep learning features. *Acta Autom. Sin.* **2017**, *43*, 814–821.
3. Wang, L.; Xiang, S.; Meng, G.; Wu, H.; Pan, C. Edge-directed single-image super-resolution via adaptive gradient magnitude self-interpolation. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *23*, 1289–1299. [[CrossRef](#)]
4. Ruangsang, W.; Aramvith, S. Efficient super-resolution algorithm using overlapping bicubic interpolation. In *Proceedings of the 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*; IEEE: New York, NY, USA, 2017; pp. 1–2.
5. Dun, Y.; Da, Z.; Yang, S.; Xue, Y.; Qian, X. Kernel-attended residual network for single image super-resolution. *Knowl.-Based Syst.* **2021**, *213*, 106663. [[CrossRef](#)]
6. Deng, Q.; Yu, L.; Hu, H. A review of single image super-resolution reconstruction based on deep learning. In *Proceedings of the 2021 International Workshop on Modern Science and Technology*; The International Center of National University Corporation Kitami Institute of Technology: Hokkaido, Japan, 2021; Volume 2021, pp. 251–256.
7. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: New York, NY, USA, 2021; pp. 1833–1844.
8. 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*, 2672–2680.
9. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2017; pp. 4681–4690.
10. 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: Berlin/Heidelberg, Germany, 2014; pp. 184–199.
11. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2016; pp. 1646–1654.
12. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2017; pp. 3147–3155.
13. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2018; pp. 1664–1673.
14. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2016; pp. 1874–1883.
15. Hu, Y.; Li, J.; Huang, Y.; Gao, X. Channel-wise and spatial feature modulation network for single image super-resolution. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 3911–3927. [[CrossRef](#)]
16. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*; IEEE: New York, NY, USA, 2017; pp. 136–144.
17. 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 (ECCV)*; Springer: Cham, Switzerland, 2018; pp. 286–301.
18. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
19. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein GAN. *arXiv* **2017**, arXiv:1701.07875. [[CrossRef](#)]

20. Gao, H.; Chen, Z.; Huang, B.; Chen, J.; Li, Z. Image super-resolution based on conditional generative adversarial network. *IET Image Process.* **2020**, *14*, 3006–3013. [[CrossRef](#)]
21. LIU, Y.; GE, G.; QI, Z. Research on embroidery image restoration based on improved deep convolutional generative adversarial network. *Laser Optoelectron. Prog.* **2023**, *60*, 68–78.
22. Li, X.; Hu, Z.; Huang, X. Combine relu with tanh. In *Proceedings of the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*; IEEE: New York, NY, USA, 2020; Volume 1, pp. 51–55.
23. Cherian, A.K.; Poovammal, E.; Rathi, Y. Improving image resolution on surveillance images using SRGAN. In *Inventive Systems and Control: Proceedings of ICISC 2021*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 61–76.
24. Peiyu, L.; Yali, Z. Face Image Super-Resolution Reconstruction Based on Improved SRGAN Model. *Comput. Eng.* **2023**, *49*, 199–205.
25. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, CA, USA, 7–9 May 2015.
26. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the International Conference on Machine Learning*; PMLR: Brookline, MA, USA, 2015; pp. 448–456.
27. Peng, Y.; Liu, L.; Wang, G.; Meng, X.; Li, Y. Single image super-resolution reconstruction based on split-attention networks. *Chin. J. Liq. Cryst. Disp.* **2024**, *39*, 950–960. [[CrossRef](#)]
28. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2017; pp. 4700–4708.
29. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2018; pp. 2472–2481.
30. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2018; pp. 3–19.
31. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2017; pp. 624–632.
32. Chen, H.; Lu, X.; Zhu, J. GAN-Based Super-Resolution with Enhanced Multi-Scale Laplacian Pyramid and Frequency Domain Loss. *IET Image Process.* **2025**, *19*, e70028. [[CrossRef](#)]
33. Yang, A.; Yu, C.; Wang, J.; Wei, Z.; Cao, J.; Liu, L. Cross-scale atomic feature enhanced network for high-fidelity single image super-resolution. *Multimed. Syst.* **2025**, *31*, 358. [[CrossRef](#)]
34. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*; Springer: Cham, Switzerland, 2018.

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.