Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 4, No. 1, 2025

FAFMNet: A Lightweight Super-Resolution Network via Frequency-Aware and Multi-Scope Feature Fusion

Rui Xu [1], Wei Fan [2]

[1] Chongqing University of Technology, Chongqing 400054, China

[2] Chongqing University of Technology, Chongqing Key Laboratory of Trusted Human-Machine Interaction and Security for Intelligent Connected Vehicles, Chongqing 400054, China

Abstract: Single Image Super-Resolution (SISR), a fundamental task in computer vision, aims to reconstruct high-quality images from their low-resolution counterparts. To address the limitations of existing methods in recovering high-frequency details and achieving efficient feature modeling, we propose a lightweight and effective Frequency-Aware and Fusion-Modulated Network (FAFMNet). The proposed network jointly enhances local texture representation and global structure modeling by incorporating a Frequency-Aware Modulator (FAM) and a Multi-Scope Fusion Block (MSFB), thereby enabling efficient cross-scale and cross-frequency feature extraction. Specifically, the FAM module models global dependencies in the frequency domain, leveraging low-frequency preservation and high-frequency enhancement strategies to improve sensitivity to structural regions. Meanwhile, the MSFB module aggregates contextual information across multiple receptive fields and utilizes a sparse channel separation mechanism to achieve lightweight multi-scale feature representation. Extensive experiments on several benchmark datasets demonstrate that FAFMNet achieves superior reconstruction performance compared to existing methods while maintaining low parameter count and fast inference speed, validating the effectiveness and practicality of the proposed design.

Keywords: Multi-Scale Feature Fusion, Frequency Modulation, Lightweight Network.

1. Introduction

Single Image Super-Resolution (SISR) aims to reconstruct a high-resolution (HR) image from a given low-resolution (LR) input, serving as a fundamental task in computer vision. It has been widely applied in various real-world scenarios such as video surveillance, medical imaging, and satellite remote sensing. In recent years, convolutional neural network (CNN)-based approaches [1] have achieved remarkable progress in SISR by constructing deep architectures to learn the complex mapping between LR and HR image domains, significantly improving reconstruction quality.

Despite the rapid advancement in reconstruction performance, two major challenges remain unresolved. First, conventional spatial-domain convolutions struggle to effectively capture long-range dependencies and high-frequency details, often resulting in incomplete recovery of image structures and textures. Second, the limited receptive field of standard convolutional operations hampers the integration of multi-scale contextual information, leading to suboptimal performance in complex scenarios.

To address these issues, we propose a novel Frequency-Aware and Fusion-Modulated Network (FAFMNet) that emphasizes both global perception and local precision under a lightweight design. The proposed FAFMNet is composed of three key components:

LiteDepthConv, designed to efficiently extract local edge-aware features with minimal computational cost;

Frequency-Aware Modulator (FAM), which leverages Fourier Transform to enhance structure-aware features by dynamically emphasizing salient regions in the frequency domain;

Multi-Scope Fusion Block (MSFB), which aggregates multi-scale contextual information in parallel and integrates it via frequency-domain attention to improve feature discriminability.

Extensive experiments demonstrate that FAFMNet achieves competitive reconstruction performance on multiple public benchmarks while maintaining a favorable balance between model complexity and inference cost, making it suitable for practical deployment. As shown in Figure 1, we compare several representative lightweight SR models on the Set5 dataset with a ×2 upscaling factor. Remarkably, FAFMNet achieves a PSNR of 38.10 dB with only ~270K parameters, demonstrating its strong capability to maintain high reconstruction quality under compact architecture constraints.

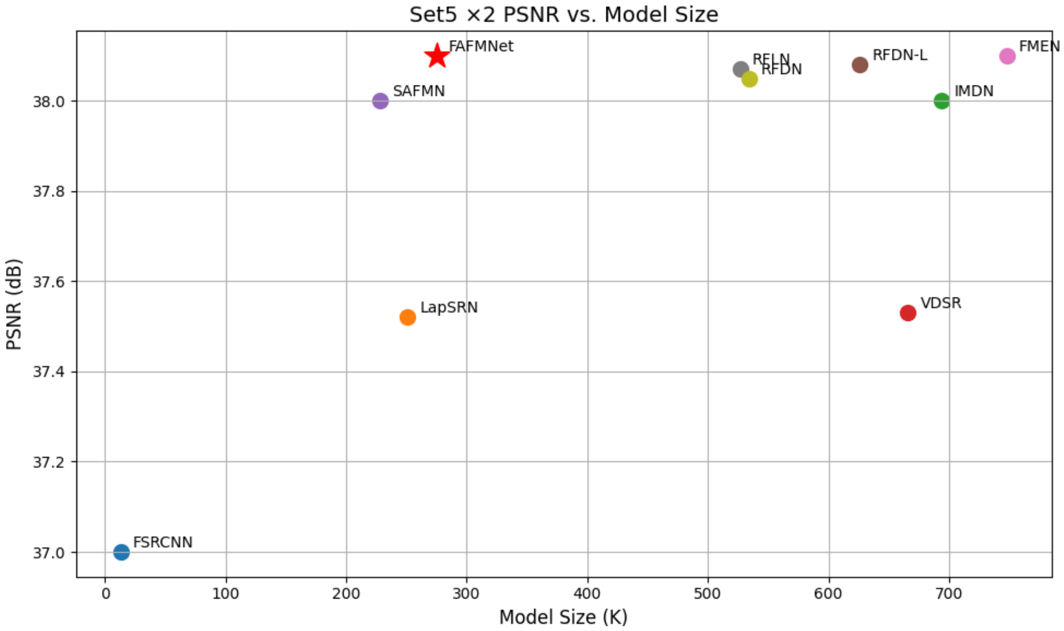

Figure 1. Comparative analysis of reconstruction performance and model complexity on the Set5 ×2 upscaling task, highlighting the efficiency of FAFMNet

2. Related Work

2.1. CNN-Based Super-Resolution Methods

In recent years, convolutional neural networks (CNNs) have made significant advancements in single image super-resolution (SISR). SRCNN [2] was one of the first approaches to apply deep learning to SISR, employing a three-layer convolutional structure to directly learn the mapping from low-resolution (LR) to high-resolution (HR) images. VDSR [3] further extended this by introducing deeper networks and residual learning, which effectively alleviated the gradient vanishing problem in deep architectures.

Following this, EDSR [4] removed unnecessary modules and adopted high-capacity residual blocks to achieve better performance with a simplified design. RDN [5] introduced dense connections to fully exploit hierarchical features, while RCAN [6] integrated channel attention mechanisms to adaptively emphasize important features across channels, thereby improving texture reconstruction quality.

Despite these improvements, CNN-based methods are inherently limited by the local nature of convolution operations, which restricts the receptive field and makes it difficult to model long-range dependencies. Furthermore, their ability to recover high-frequency structural information remains inadequate—especially in extremely low-resolution inputs or highly textured scenes—posing a major bottleneck for further performance gains.

2.2. Frequency Domain Modeling Methods

While spatial-domain convolutions excel at capturing local textures, they are inherently limited in modeling global dependencies and high-frequency responses. To address this issue, recent research has explored the integration of frequency-domain analysis into image reconstruction tasks. For instance, FDN (Frequency Domain Network) [7] designs network architectures directly in the frequency domain to process spectral information, thereby enhancing edge and detail reconstruction. FFTFormer combines frequency-domain representation with Transformer architectures by mapping input images into the frequency space, where self-attention mechanisms are applied to improve structural region recognition. Other approaches have also employed explicit frequency modeling using transforms such as Discrete Cosine Transform (DCT) or wavelet transforms.

However, many of these frequency-domain models are incorporated as auxiliary modules rather than being deeply integrated into the network backbone. Moreover, their frequency selection strategies are often coarse-grained, lacking fine-grained modulation mechanisms, which limits their effectiveness in complex scenarios.

To overcome these limitations, we propose the Frequency-Aware Modulator (FAM), which transforms fused features into the frequency domain to extract both low-frequency components and high-frequency residuals. These components are then enhanced in the frequency domain and transformed back to the spatial domain to modulate features. This enables dynamic spatial-frequency joint modeling, significantly improving the network's sensitivity to high-frequency structures.

2.3. Multi-Scale Feature Fusion Methods

Images contain semantic and structural information across multiple scales, especially in high-scale super-resolution tasks where features at different resolutions critically affect reconstruction quality. Consequently, effective multi-scale modeling and fusion strategies have become a focus of recent research.

MSRN (Multi-Scale Residual Network) [8] adopts multiple parallel convolutional branches to extract features at different scales, which are then fused via residual connections to enhance structural representation. MADNet (Multi-Attention Deep Network) [9] further integrates attention mechanisms and hierarchical convolutions to achieve multi-scale enhancement in salient regions.

Nonetheless, most of these methods rely on deep convolutional stacking or channel-wise fusion, resulting in large model sizes and high computational cost, which limits their applicability in lightweight deployment scenarios. Additionally, frequency-domain priors are often overlooked during feature fusion, weakening the model’s discriminative ability in preserving textures and edges.

To address these issues, we design the Multi-Scope Fusion Block (MSFB), which employs a sparse channel separation strategy to construct four parallel branches with 1×1, 3×3, 5×5 convolutions and lightweight depthwise convolutions for extracting multi-scale features. After concatenation, the features are fused using the frequency-aware modulation mechanism. This design balances local detail and global structure modeling, incorporating frequency priors into the fusion process. As a result, the output features exhibit enhanced structural sensitivity and stronger representational capacity, making the model particularly suitable for fine-grained texture reconstruction.

3. Methodology

This paper proposes FAFMNet (Frequency-Aware and Fusion-Modulated Network), a lightweight neural network architecture specifically designed for efficient single image super-resolution (SISR) tasks. The network is tailored to integrate spatial-domain structural information with frequency-domain global modeling capabilities, aiming to achieve superior reconstruction performance while maintaining low computational overhead.

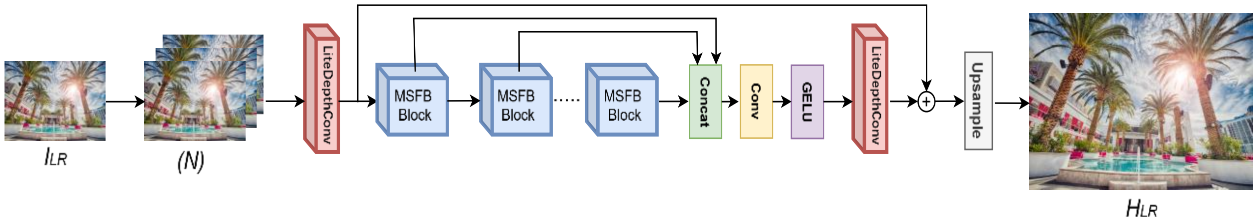

Figure 2. Overall architecture of FAFMNet

As illustrated in Figure 2, FAFMNet takes a low-resolution image as input and begins by enhancing the initial representation through channel-wise stacking. The image features are then fed into a shallow convolutional block named LiteDepthConv, which combines pointwise and depthwise convolutions to efficiently extract edge-aware features while maintaining low computational complexity.

The intermediate features are sequentially processed by multiple Multi-Scope Fusion Blocks (MSFBs). Each MSFB consists of several parallel branches with varying convolutional receptive fields to capture both local textures and contextual information. These features are then merged along the channel dimension. Furthermore, each MSFB integrates the proposed Frequency-Aware Modulator (FAM), which performs Fourier transform-based frequency decomposition and dynamically re-weights low- and high-frequency components to enhance texture and detail representation.

Within the main backbone, the output from each module is retained for subsequent fusion. All module outputs are concatenated and compressed using a 1×1 convolution, followed by a residual connection with the shallow features to improve stability and information flow. Finally, the fused features are passed through the upsampling module to restore the spatial resolution and generate the high-resolution output image.

3.1. LiteDepthConv Module

To enhance shallow feature representation while maintaining network efficiency, we introduce the LiteDepthConv module as the first stage of feature extraction. This module adopts a structural design that combines pointwise convolution and depthwise separable convolution, enabling effective capture of local information while significantly reducing the number of parameters and computational overhead.

Specifically, the input feature map $X \in R^{C_{in} \times H \times W}$ is first processed through a1×1 pointwise convolution for channel compression or projection, yielding the intermediate feature $X_1 \in R^{C_{out} \times H \times W}$:

$$X_1 = Conv_{1\times1}(X) \tag{1}$$

Next, X_1 is fed into a depthwise convolutional layer, where each channel is independently convolved using a $k \times k$ kernel (with k=3) to capture local spatial information. The operation is expressed as:

$$\mathrm{Y} = \mathrm{DWConv}_{\mathrm{k \times k}}(\mathrm{X}_1) \tag{2}$$

As a result, the output feature map is $Y \in R^{Cout \times H \times W}$. LiteDepthConv not only provides sufficient representational capacity but also demonstrates strong edge modeling and computational efficiency, making it particularly suitable for the initial feature extraction stage in lightweight SISR tasks.

3.2. Frequency-Aware Modulator (FAM)

In single image super-resolution tasks, traditional attention mechanisms typically operate solely in the spatial domain, where their ability to capture long-range dependencies and frequency components is inherently limited due to the constrained receptive field of local convolutional operators. This results in suboptimal performance, especially in the reconstruction of high-frequency details.

To address this limitation, we propose a Frequency-Aware Modulator (FAM), which enhances the network's capability in representing structural clarity and edge textures by incorporating frequency-domain modeling and modulation. As illustrated in Figure 3, the input features are first transformed into the frequency domain using the Fast Fourier Transform (FFT). Then, a frequency band selector is applied to modulate different frequency components, selectively enhancing the high-frequency response while preserving essential low-frequency information.

After this enhancement, the modified frequency-domain features are projected back to the spatial domain via the inverse FFT, and fused with the original input features. This process generates a frequency-enhanced attention map, which guides the network to better focus on critical structural regions and fine edge details, ultimately improving the quality of image reconstruction.

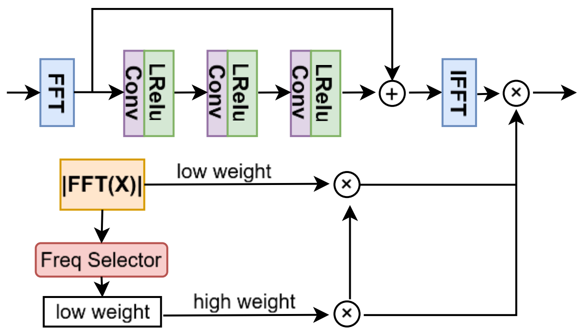

Figure 3. Structure of the FAM Module – Frequency-domain Enhancement and Modulation Pipeline

The core idea of this module is to project the input features from the spatial domain into the frequency domain, leveraging frequency magnitude information to guide the selective enhancement of high-frequency components and suppress low-frequency components via lightweight convolutional pathways for nonlinear modeling. Specifically, the input feature $X \in R^{C \times H \times W}$ is first transformed into the two-dimensional complex frequency spectrum through the real-valued Fast Fourier Transform (Real FFT), expressed as:

$$F = \mathcal{F}_{\mathrm{RF}}(X) \tag{3}$$

Where $F \in C^{C \times H \times W'}$ denotes the complex frequency representation, containing both real and imaginary components. Then, a frequency selector module is designed based on global spectral statistics. This module adaptively normalizes the spectrum and applies a 1×1 convolution layer to generate channel-wise high-frequency weights, thereby enabling frequency-aware selective modulation. The high-frequency components are passed through a series of lightweight convolutions (with LeakyReLU activation) to enhance their response, while the low-frequency components are either retained or fused accordingly.

Subsequently, the modulated frequency features are reconstructed back into the spatial domain as:

$$Y = X \odot \mathcal{F}^{-1}(F_{\mathrm{mod}}) \tag{4}$$

Here, F_{mod} represents the frequency features after enhancement and fusion, $\mathcal{F}^{-1}$ denotes the inverse Fast Fourier Transform (IRFFT), and $\odot$ indicates element-wise multiplication. This formulation yields a frequency-enhanced attention map that modulates the original features to better highlight key structural regions and achieve high-quality image restoration.

3.3. Multi-Scope Fusion Block

To effectively model multi-scale and multi-frequency semantic and structural features within an image, we propose the Multi-Scope Fusion Block (MSFB). This module integrates a multi-branch architecture with frequency-domain enhancement mechanisms, enabling the collaborative fusion of local detail modeling and global contextual awareness while maintaining a lightweight design.

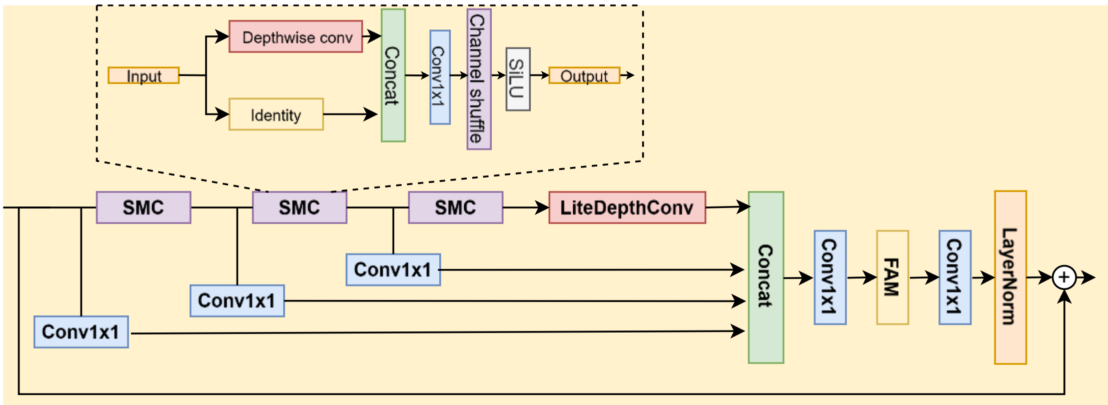

Figure 4. Structural diagram of the Multi-Scope Fusion Block (MSFB)

Figure 4 illustrates the overall structure and information flow of the proposed Multi-Scope Fusion Block (MSFB). This module comprises four parallel branches, each responsible for capturing feature information under different receptive fields: Branch 1: Applies a 1×1 convolution to the input feature map $X \in R^{C \times H \times W}$ for channel compression and extraction of low-level local features. Branch 2 and Branch 3: Employ SparseMixConv with 3×3 and 5×5 kernels, respectively, to extract mid-level and large-scale contextual features. Branch 4: Integrates LiteDepthConv to enhance the modeling of structural regions.

The output features from all four branches are concatenated along the channel dimension and then fused via a 1×1 convolution to obtain a unified feature representation:

$$\mathrm{F}_{\mathrm{fused}} = \mathrm{Conv}_{1\times1}(\mathrm{Concat}(\mathrm{B}_1, \mathrm{B}_2, \mathrm{B}_3, \mathrm{B}_4)) \tag{5}$$

To enhance the discriminability of the fused features, we introduce the Frequency-Aware Modulator (FAM) to perform frequency-domain modeling on $\mathrm{F}_{\mathrm{fused}}$ .Specifically, the feature map is first projected into the frequency domain, where the low- and high-frequency components are extracted and assigned different weight coefficients. The enhanced frequency response is then transformed back to the spatial domain to implement attention modulation:

$$F_{\mathrm{mod}} = \mathcal{F}^{-1}(\mathcal{F}(\mathrm{Ffused}) + \phi(\mathcal{F}(\mathrm{Ffused}))) \tag{6}$$

Here, $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fast Fourier Transform (FFT) and its inverse, respectively, while $\phi(\cdot)$ is the frequency enhancement mapping function. The modulated feature is further projected to the target number of channels and normalized on a per-pixel basis:

$$\mathrm{Y} = \mathrm{LayerNorm}\left(Conv_{1\times1}(F_{mod})\right) + \mathrm{X} \tag{7}$$

MSFB captures semantic features across various receptive fields through its multi-branch structure and significantly improves the model's responsiveness to textures and structural regions by incorporating frequency-domain modulation. With low computational requirements, this module is particularly suitable for lightweight image reconstruction tasks and demonstrates excellent performance and generalization in single image super-resolution (SISR).

4. Experimental Results and Analysis

This section aims to evaluate the performance of the proposed FAFMNet model in the task of single image super-resolution (SISR), validating its effectiveness in terms of image reconstruction quality, model parameters, and inference speed. We conduct comprehensive comparisons with existing methods on multiple benchmark datasets and perform ablation studies to verify the contribution of key modules.

4.1. Experimental Settings

Datasets: We adopt DIV2K [10], Set5 [11], Set14 [12], BSD100 [13], and Urban100 [14] as evaluation benchmarks. The model is trained on DIV2K and tested on the aforementioned datasets.

Scaling Factors: We evaluate the model under upscaling factors ×2, ×3, and ×4.

Evaluation Metrics: The reconstruction quality is assessed using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), both computed on the Y channel of the YCbCr color space.

Implementation Details: The model is implemented using the PyTorch framework. We employ the Adam optimizer with an initial learning rate of 1e-4. The training is conducted with a batch size of 64, input patch size of 48×48, and the total number of training epochs is set to 1000.

4.2. Comparison with Existing Methods

To validate the effectiveness of the proposed FAFMNet, we conduct systematic evaluations on five widely-used image super-resolution (SR) benchmark datasets: Set5, Set14, BSD100, Urban100, and Manga109. Experiments are carried out under upscaling factors of ×2, ×3, and ×4. We comprehensively compare our model with several classic and recent lightweight SR methods. The compared methods include the early lightweight network FSRCNN [15], the deep residual model VDSR [3], the multi-scale convolutional network LapSRN [16], the representative lightweight SR model IMDN [17], the attention-enhanced efficient network SAFMN [18], and the recent lightweight model RFDN [19], which achieves a good balance between performance and efficiency. Table 1 presents the performance comparison of these lightweight SR methods under different scaling factors in terms of PSNR and SSIM.

Table 1. Performance comparison (PSNR/SSIM) of various lightweight SR methods under different upscaling factors

Methods	Scale	Params	Set5	Set14	BSDS100	Urban100	Manga109
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
FSRCNN [15]	X2	13K	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9710
LapSRN [16]		251 K	37.52/0.9591	32.99/0.9124	31.80/0.8952	30.41/0.9103	37.27/0.9740
IMDN [17]		694K	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
VDSR [3]		666 K	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9750
SAFMN [18]		228 K	38.00/0.9605	33.54/0.9177	32.16/0.8995	31.84/0.9256	38.71/0.9771
RFDN [19]		534 K	38.05/0.9606	33.68/0.9184	32.16/0.8994	32.12/0.9278	38.88/0.9773
FAFMNet (ours)		275K	38.10/0.9611	33.81/0.9209	32.25/0.9009	32.69/0.9335	39.11/0.9779
FSRCNN [15]	X3	13 K	33.18/0.9140	29.37/0.8240	28.53/0.7910	26.43/0.8080	31.10/0.9210
IMDN [17]		703K	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
VDSR [3]		666 K	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9340
SAFMN [18]		233 K	34.34/0.9267	30.33/0.8418	29.08/0.8048	27.95/0.8474	33.52/0.9437
RFDN [19]		541 K	34.41/0.9273	30.34/0.8420	29.09/0.8050	28.21/0.8525	33.67/0.9449
FAFMNet (ours)		283K	34.58/0.9287	30.48/0.8457	29.14/0.8082	28.61/0.8623	34.12/0.9481
FSRCNN [15]	X4	13 K	30.72/0.8660	27.61/0.7550	26.98/0.7150	24.62/0.7280	27.90/0.8610
LapSRN [16]		813 K	31.54/0.8852	28.09/0.7700	27.32/0.7275	25.21/0.7562	29.09/0.8900
IMDN [17]		715K	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
VDSR [3]		666 K	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8870
SAFMN [18]		240 K	32.18/0.8948	28.60/0.7813	27.58/0.7359	25.97/0.7809	30.43/0.9063
RFDN [19]		550 K	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858	30.58/0.9089
FAFMNet (ours)		293K	32.38/0.8974	28.73/0.7851	27.61/0.7405	26.45/0.7975	30.86/0.9139

As shown in Table 1, FAFMNet consistently delivers top PSNR and SSIM results across all upscaling factors (×2, ×3, ×4) on five benchmark datasets: Set5, Set14, BSD100, Urban100, and Manga109. Its advantages are especially evident on Urban100 and Manga109, which contain complex structures and rich textures. For example, under ×2 upscaling, FAFMNet achieves 32.69 dB on Urban100, outperforming RFDN (32.12 dB) and IMDN (32.17 dB), highlighting its strength in recovering high-frequency details.

FAFMNet also demonstrates excellent parameter efficiency, with only 275K–293K parameters—significantly fewer than IMDN (~700K) and RFDN (~540K)—while maintaining superior reconstruction quality. This highlights a well-balanced design between compactness and performance. Compared to ultra-lightweight models like FSRCNN (~13K), FAFMNet achieves better feature modeling and reconstruction, making it ideal for resource-constrained yet performance-critical applications.

Beyond numerical metrics, qualitative comparisons further support FAFMNet's effectiveness. As shown in Figure 5, FAFMNet reconstructs sharper edges, clearer contours, and more realistic textures, especially in detailed or structurally complex regions. Competing methods tend to produce blurry edges, broken lines, or missing textures, whereas FAFMNet delivers higher visual fidelity and perceptual quality.

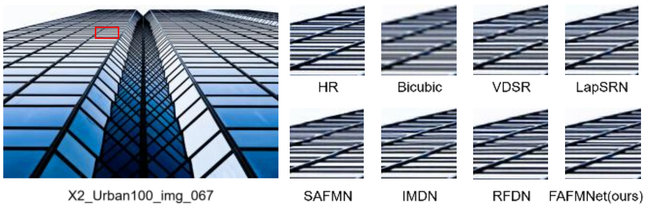

Figure 5. Visual comparison of ×2 SR results on representative images from Urban100 datasets.

This performance improvement is mainly attributed to the proposed Frequency-Aware Modulation module, which enhances the recovery of fine details by modeling and amplifying high-frequency components in the frequency domain. Meanwhile, the designed Multi-Scope Fusion Block integrates multi-scale structural information in the spatial domain, further boosting the model's contextual perception capability.

4.3. Ablation Study

To further validate the effectiveness of the key components in FAFMNet, we conducted two sets of ablation experiments focusing on the contributions of the Frequency-Aware Modulator (FAM) and the Multi-Scope Fusion Block (MSFB).

4.3.1. Effectiveness of the FAM Module

We first evaluate the impact of incorporating the frequency modeling module FAM on the overall network performance. Under the same network architecture, we remove the FAM module and retain only the basic convolutional flow. The results are shown in Table 2. It can be observed that excluding FAM leads to a notable drop in both PSNR and SSIM across multiple datasets, with particularly significant degradation on complex datasets such as Urban100 and Manga109. This indicates that FAM effectively enhances the model's ability to capture high-frequency details, thereby improving the reconstruction of structural and textural features.

Table 2. Impact of the FAM module on the super-resolution performance of FAFMNet (×2 Upscaling)

Model Configuration	Set5	Set14	BSD100	Urban100	Manga109
FAFMNet (w/o FAM)	37.92 / 0.9591	33.21 / 0.9150	32.01 / 0.8981	31.80 / 0.9244	38.55 / 0.9752
FAFMNet	38.10 / 0.9611	33.81 / 0.9209	32.25 / 0.9009	32.69 / 0.9335	39.11 / 0.9779

4.3.2. Impact of MSFB Module Quantity

To investigate the impact of the number of Multi-Scope Fusion Blocks (MSFB) on model performance, we varied the number of stacked MSFB modules to 5, 6, 7, 8, and 9 while keeping the rest of the architecture unchanged. The experimental results are shown in Table 3. From the results, we observe a consistent performance improvement as the number of MSFB modules increases from 5 to 8, indicating that deeper multi-scope fusion enhances the model's feature extraction capabilities. However, when the number increases to 9, the performance gain saturates and even slightly drops on certain datasets, likely due to redundancy and increased optimization difficulty caused by excessive depth. Considering both performance and computational efficiency, we adopt 8 MSFB modules as the default configuration in our final model.

Table 3. Impact of MSFB Module Quantity on the Super-Resolution Performance of FAFMNet (×2 Upscaling)

MSFB module count	Set5	Set14	BSD100	Urban100	Manga109
5	37.91 / 0.9596	33.29 / 0.9159	32.02 / 0.8986	31.94 / 0.9251	38.62 / 0.9757
6	38.00 / 0.9603	33.48 / 0.9183	32.13 / 0.8997	32.31 / 0.9293	38.85 / 0.9768
7	38.07 / 0.9609	33.67 / 0.9198	32.18 / 0.9004	32.55 / 0.9319	39.03 / 0.9773
8	38.10 / 0.9611	33.81 / 0.9209	32.25 / 0.9009	32.69 / 0.9335	39.11 / 0.9779
9	38.08 / 0.9610	33.78 / 0.9203	32.23 / 0.9007	32.64 / 0.9328	39.05 / 0.9775

5. Conclusion

This paper proposes a novel lightweight image super-resolution network, FAFMNet, which effectively enhances the reconstruction of image textures and structural details by integrating frequency modeling and multi-scale fusion strategies. By introducing the Frequency-Aware Modulator (FAM) into the network, FAFMNet captures global contextual information in the frequency domain and guides the reconstruction process to focus on crucial structural regions. Meanwhile, the Multi-Scope Fusion Block (MSFB) integrates multi-scale feature representations through a compact design, improving performance while maintaining low parameter complexity.

Experimental results demonstrate that FAFMNet outperforms existing lightweight methods across multiple benchmark super-resolution datasets, achieving new baselines in terms of PSNR and SSIM, with particularly notable performance on high-texture datasets such as Urban100. Furthermore, ablation studies confirm the critical role of the FAM and MSFB modules in enhancing performance. Future work will explore incorporating cross-modal learning mechanisms to further improve the generalization and adaptability of the model in real-world scenarios.

Acknowledgements

This research was supported by the project "Visual AI-Based Recognition Technology for Physical Defects in Microelectronic Components" (Project No. 2024ZYD0286). Focusing on the intelligent detection of fine-grained defects in electronic components, the project provided a valuable foundation that guided our investigation into image reconstruction and detail enhancement. Under this technical background, we proposed a lightweight super-resolution network, FAFMNet, which integrates frequency modeling and multi-scope feature fusion. The model significantly improves the reconstruction of high-frequency details and structural edges, making it particularly suitable for resource-constrained applications that require high-precision recovery. The research resources, experimental platforms, and theoretical foundations provided by the project played a crucial role in the design and validation of the proposed model. We would like to express our sincere gratitude to the project management office, academic supervisors, and all colleagues who offered their support throughout this study.

References

[1] Dong Li, Jia-Bin Huang, Yali Li, Shengjin Wang, and Ming HsuanYang. Weaklysupervisedobject localization with pro gressive domain adaptation. In IEEE Conference on Com puter Vision and Pattern Recognition, pages 3512–3520, 2016.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In European Conference on Computer Vi sion, pages 184–199. Springer, 2014.

[3] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional net works. In CVPR, pages 1646–1654, 2016. 7.

[4] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and KyoungMuLee. Enhanceddeepresidualnetworksforsingle image super-resolution. In CVPR Workshops, 2017. 1, 2, 5, 6, 7, 9.

[5] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature dis tillation network for lightweight image super-resolution. In ECCV, pages 41–55, 2020. 1, 6.

[6] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In European Confer ence on Computer Vision, pages 286–301, 2018.

[7] J. Li, S. You, and A. Robles-Kelly, “A Frequency Domain Neural Network for Fast Image Super-Resolution,” 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 2018, pp. 1–8.

[8] Y. Li, L. Fang, J. Liu, and Y. Zhang, “Multi-Scale Residual Network for Image Super-Resolution,” Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp.

[9] Y. Hui, X. Wang, Y. Gao, Y. Gao, and X. Chen, “MADNet: A Fast and Compact Network for Single Image Super-Resolution,” Proceedings of the 28th ACM International Conference on Multimedia (ACM MM), 2020, pp. 1237–1245.

[10] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In CVPRW, pages 126–135, 2017. 5.

[11] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In BMVC, pages 1–10, 2012. 5, 6, 7.

[12] Roman Zeyde, Michael Elad, and Matan Protter. On sin gle image scale-up using sparse-representations. In Interna tional conference on curves and surfaces, pages 711–730, 2010. 5, 6, 7.

[13] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In ICCV, pages 416–423, 2001. 5, 6, 7.

[14] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In CVPR, pages 5197–5206, 2015. 1, 5, 6, 7.

[15] Chao Dong, Chen Change Loy, and Xiaoou Tang. Acceler ating the super-resolution convolutional neural network. In ECCV, pages 391–407, 2016. 7.

[16] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In CVPR, pages 624–632, 2017. 2, 7.

[17] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi distillation network. In ACM MM, pages 2024–2032, 2019. 1, 2, 7.

[18] Long Sun, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Spatially-adaptive feature modulation for efficient image super-resolution. In ICCV, 2023. 3, 8.

[19] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature dis tillation network for lightweight image super-resolution. In ECCVW, pages 41–55, 2020. 1, 2, 7, 8.