ISSN: 2993-8643
Engineering: Open Access
Research Article

A Super-Resolution Reconstruction Method for Lightweight Building Images Based on An Expanding Feature Modulation Network

Yi Zhang[1*], Ruonan Lin[1], Ang Ping[2], Jincheng Li[3] and Xiao Yao[4]

[1]*School of Advanced Manufacturing, Fuzhou University, China*

[2]*Leicester International Institute, Dalian University of Technology, Panjin City, Liaoning, China*

[3]*Guangzhou Zili Information Consulting Co., Ltd. Guangzhou, China*

[4]*School of Information Science and Engineering, Shenyang Ligong University, China*

***Corresponding Author**
Yi Zhang, School of Advanced Manufacturing, Fuzhou University, China.

Submitted: 2025, Jul 10; **Accepted:** 2025, Aug 14; **Published:** 2025, Aug 25

Citation: Zhang, Y., Lin, R., Ping, A., Jincheng, L., Yao, X. (2025). A Super-Resolution Reconstruction Method for Lightweight Building Images Based on An Expanding Feature Modulation Network. *Eng OA, 3*(8), 01-09.

Abstract
This study proposes a lightweight method for building image super-resolution using a Dilated Contextual Feature Modulation Network (DCFMN). The process includes obtaining high-resolution images, down-sampling them to low-resolution, enhancing the low-resolution images, constructing and training a lightweight network model, and generating super-resolution outputs. To address challenges such as regular textures and long-range dependencies in building images, the DCFMN integrates an expansion separable modulation unit and a local feature enhancement module. The former employs multiple expansion convolutions equivalent to a large kernel to efficiently aggregate multi-scale features while leveraging a simple attention mechanism for adaptivity. The latter encodes local features, mixes channel information, and ensures no additional computational burden during inference through reparameterization. This approach effectively resolves the limitations of existing lightweight super-resolution networks in modeling long-range dependencies, achieving accurate and efficient global feature modeling without increasing computational costs, and significantly improving both reconstruction quality and lightweight efficiency for building image super-resolution models.

Index Terms: Building Images, Dilated Contextual Feature Modulation Network, Lightweight Network, Super-Resolution Reconstruction

1. Introduction

Structural health monitoring (SHM) plays a crucial role in ensuring the safety, reliability, and durability of engineering structures. Visual measurement methods, widely used in monitoring applications such as dynamic displacement, static displacement, and crack detection, are becoming increasingly important due to their non-contact a nature [1]. However, these methods often encounter limitations in image resolution caused by factors like measurement environment constraints and equipment costs. Super-resolution (SR) reconstruction techniques, which aim to generate high-resolution images from corresponding low-resolution images, offer a promising solution to this challenge [2].

In the context of building structures, which typically exhibit regular texture patterns, the SR reconstruction of images is particularly sensitive to noise. The quality of the reconstruction is heavily influenced by the network's ability to model long-range dependencies within the image [3]. Traditional methods often struggle to balance the need for high performance and computational efficiency, especially when dealing with architectural images that require modeling of distant spatial relationships. To address this issue, this paper proposes a lightweight SR reconstruction method based on an expanding feature modulation network, designed to optimize both performance and computational efficiency [4].

The proposed method introduces an adaptive inflated feature modulation network that uses multiple inflated convolutions stacked on a large convolution kernel. This approach achieves a large receptive field while minimizing computational cost. Additionally, a spatial attention mechanism, implemented via a simple element-wise multiplication operation, enables the model to focus adaptively on the most important features of the image. To further enhance efficiency, a local feature enhancement module, based on a reparameterization operation, reduces network inference time without compromising the reconstruction quality. The combination of these innovations leads to an effective and efficient SR reconstruction method for building images with regular textures, achieving a favorable trade-off between performance and computational efficiency.

Despite the advancements in SR techniques, no lightweight super-resolution method based on an expanding feature modulation network has been proposed for building images in the existing literature and patents [5]. The contributions of this invention fill this gap by providing a novel, resource-efficient solution for improving the SR quality of building images.

2. Related Work

2.1 Image Super-resolution

The ill-posed nature of SR tasks presents significant challenges in reconstructing high-quality images [6]. Numerous methods have been proposed for SR, which can be broadly classified into interpolation-based, prior-based, sparse representation-based, and learning-based approaches [7]. In recent years, deep learning methods have demonstrated impressive performance across various high-level visual tasks, leading many researchers to develop deep learning algorithms for image super-resolution. SRCNN, the pioneering work in this area, was the first to apply convolutional neural networks (CNNs) to image super-resolution [8].

Initially, researchers focused on improving the reconstruction performance of networks by increasing their depth and width. However, this approach often leads to a substantial increase in resource consumption, making model training more challenging[9]. Consequently, some researchers began incorporating various attention mechanisms into networks to enhance model reconstruction[10]. Specifically, RCAN (Residual Channel Attention Network) applies channel attention to the SR task, proposing a residual channel attention network to address the issue of unequal treatment of different feature information in the channel dimension[11].

In addition to channel attention, spatial attention has proven to be crucial for high-quality image reconstruction[12]. For instance, CSNLN (Cross-Scale Non-Local Network) introduces a cross-scale non-local attention module to extract all potential intrinsic prior knowledge within an image[13]. Similarly, NLSN (Non-Local Sparse Network) combines non-local attention operations with sparse representations, proposing a new non-local sparse attention mechanism with dynamic sparse attention patterns[14]. This integration allows the network to leverage long-range modeling through non-local attention while maintaining the robustness of sparse representations [15].

Moreover, self-attention mechanisms, particularly those in Vision Transformer (ViT)-based architectures, have also been successfully applied to SR tasks[16]. These methods have shown the potential to surpass the reconstruction performance of SR approaches based on traditional CNNs.

2.2 Lightweight Image Super-resolution

Due to the limited computational and memory resources of mobile devices, large-scale networks often do not perform well on these platforms [17]. As a result, the development of lightweight image super-resolution networks has garnered increasing attention from researchers [18].

For instance, BNN (Binarized Neural Networks) introduces a binarization method for image super-resolution, aiming to preserve network performance while significantly reducing computational resource consumption [19]. Recognizing the excessive redundancy in convolution operations, BSRN (Bilateral Separable Residual Network) optimizes convolutions by using blueprint separable convolutions and incorporates two effective attention modules to improve network efficiency [20]. VapSR, on the other hand, expands the effective receptive field of the network at a low cost by utilizing depth-wise convolution and depth-wise dilated convolution [21].

WDRN (Wavelet Domain Residual Network) applies wavelet transforms to image super-resolution, proposing wavelet feature mapping and wavelet coefficient reconstruction blocks to achieve efficient and accurate high-resolution reconstruction [22]. Class SR combines class and SR modules within a unified framework, utilizing different capacity networks for different image regions during reconstruction [23].

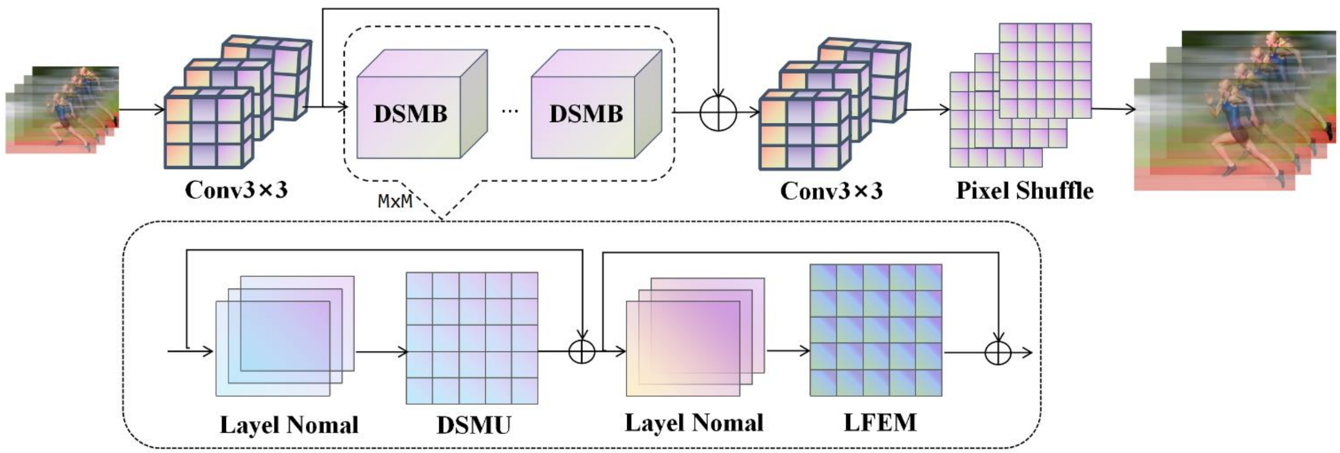
ABSR (Adaptive Binarized Super-Resolution) introduces an incentive selection mechanism that balances the inference of high-frequency information with the preservation of low-frequency information, thereby providing richer data for super-resolution reconstruction [24].

To address the limitations of traditional Vision Transformers, which typically model self-attention in a unidimensional fashion, OmniSR proposes an omni self-attention module that simultaneously models pixel interactions in both spatial and channel dimensions [25,26]. SAFMN (Spatially Adaptive Feature Modulation Network) introduces a spatially adaptive feature modulation mechanism to dynamically select representative feature representations, enhancing the network's ability to focus on the most relevant features [27].

3. Models

The proposed Dilated Contextual Feature Modulation Network (DCFMN) is composed of several key modules: a shallow feature extraction module, a deep feature extraction module that named

DSMB (Deep Feature Extraction Module), which consists of dilated separable modulation units (DSMU), a local feature enhancement module (LFEM), and an up-sampling feature reconstruction module.



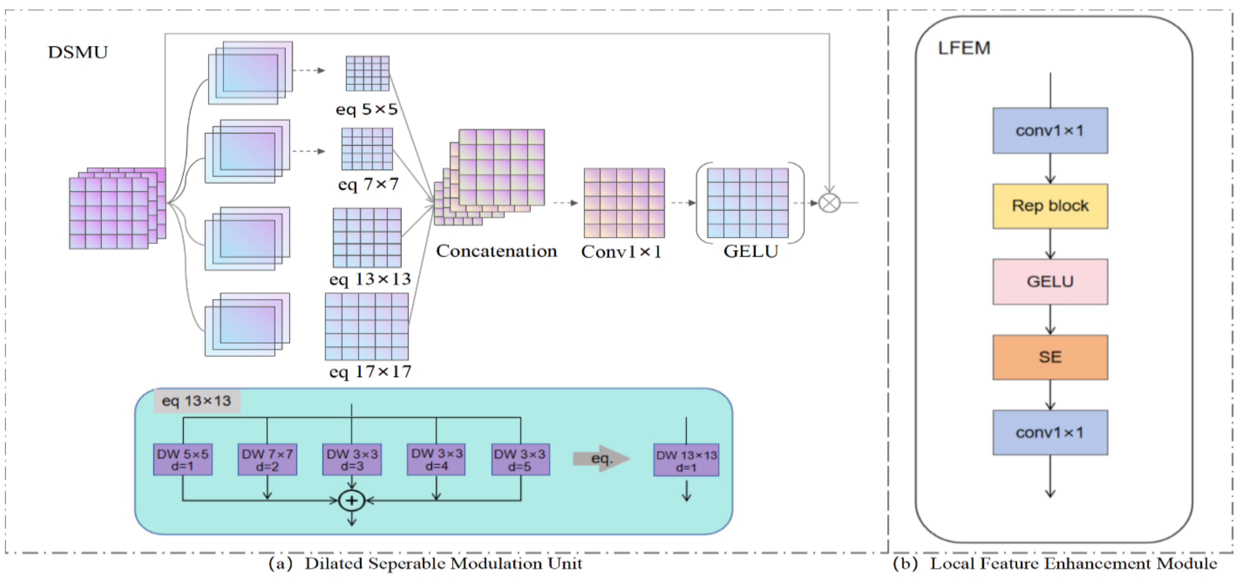


Figure 1: General Framework of DCFMN

Figure 2: Design Diagram of ESMU and LFEM based on Reparameterization Operations

The overall architecture is illustrated in Figure 1. Initially, low-resolution images are processed by the shallow feature extraction module to generate preliminary feature representations. Subsequently, the deep feature extraction module employs dilated convolutions and feature modulation mechanisms to extract and enhance long-range dependencies. Finally, the upsampling feature reconstruction module is used to restore the image, yielding the final high-resolution output.

3.1 Shallow Feature Extration

In the input stage of DCFMN, low-resolution images undergo preliminary feature extraction via a convolutional layer (with a 3x3 kernel). This layer transforms the low-resolution image into feature space, generating the shallow feature f_0:

$$f_0 = Conv_{3\times3}(I_{LR}) \quad (1)$$

where I_{LR} denotes the low-resolution image, and $Conv_{3\times3}$ represents a 3x3 convolutional layer.

3.2 DSMU

Long-term contextual information involves extensive feature interactions, which are essential for reconstructing finer details in dense pixel prediction tasks. To this end, we reconsidered the design of the core feature extraction units within the deep feature extraction module. Unlike self-attention mechanisms, traditional feature pyramids, and large kernel convolutions, we propose a more lightweight alternative: the DSMU. This unit learns long-range dependencies from features at various scales, thus facilitating a more efficient exploration of useful feature information for architectural image reconstruction, as depicted in Figure 2.

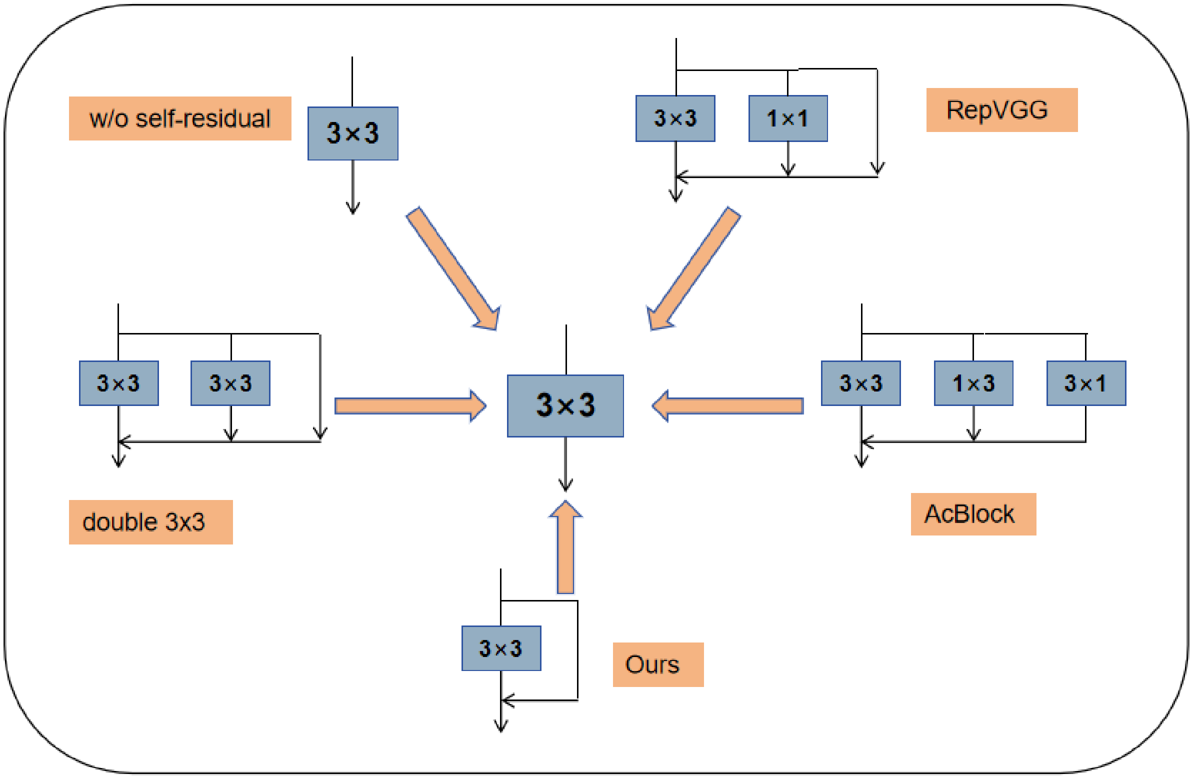


Figure 3: Reparametric Block Equivalence Diagram

In DCFMN, the DSMU processes the input features through a series of dilated convolutional layers to extract multi-scale feature information. Specifically, for the shallow feature f_0, we iteratively apply a series of dilated convolution units H_k, obtaining the deep features f_k:

$$f_k = H_k (f_{k-1}), k = 1, \cdots, n \quad (2)$$

This process enhances the model's ability to capture long-range dependencies by stacking dilated convolutional layers, thereby improving the precision of image reconstruction.

Next, we perform a channel chunking operation on the normalized input features to generate four separate feature maps, which are then passed through the DSMU for multi-scale feature extraction. Given input features X, this process can be expressed as:

$$[X_0, X_1, X_2, X_3] = \mathrm{CHUNK}(X) \quad (3)$$

$$\widetilde{X}_i = \mathrm{DW} - \mathrm{Conv}_{n\times n}(X_i), 0 \le i \le 3 \quad (4)$$

where CHUNK(·) denotes the channel chunking operation, and $\mathrm{DW\text{-}Conv}_{n\times n}$ (·) represents the deep convolutional kernels for multi-scale feature extraction, with n taking values from 5, 7, 13, to 17.

For a given pixel in the feature map, distant pixels may be more relevant than neighboring pixels. To reduce computational burden, we exploit the properties of standard convolutions to equivalently replace large convolution kernels with multiple smaller dilated convolution kernels, extracting features more efficiently. The hyperparameters include kernel size *k*, dilation rate *d*, and reconstruction kernel size K. As shown in Figure 3, with k = (5,7,3,3,3), d = (1,2,3,4,5), the kernel size K = 13 is achieved. The equivalent operations for the DSMU are as follows:

$$k = (3,3), d = (1,2) \to K = 5\#(5)$$

$$k = (5,3,3), d = (1,2,3) \to K = 7\#(6)$$

$$k = (5,9,3,3,3), d = (1,2,4,5,7) \to K = 17\#(7)$$

Results in Table 3 demonstrate that this approach reduces the number of parameters and floating-point operations (FLOPs) while improving performance.

Finally, we concatenate these multi-scale features and aggregate them using a 1×1 convolution. This process can be expressed as:

$$\dot{X} = \mathrm{Conv}_{1\times1}\left(\mathrm{Concat}\left(\left[\widetilde{X}_0, \widetilde{X}_1, \widetilde{X}_2, \widetilde{X}_3\right]\right)\right) \quad (8)$$

where Concat(·) denotes concatenation along the channel dimension, and $\mathrm{Conv}_{1\times1}$ (·) refers to a 1×1convolution.

After aggregating the multi-scale features, we normalize them using the GELU activation function, then add the result to the input features to generate the final output. This can be expressed as:

$$\overline{X} = \phi(\dot{X}) + X \quad (9)$$

where $\phi(\cdot)$ denotes the GELU function.

3.3 LFEM

Although dilated convolutions effectively capture long-range dependencies, the extraction of local contextual information remains crucial. Therefore, DCFMN incorporates a LFEM, which captures local feature information through multiple parallel 3×3 convolution structures.

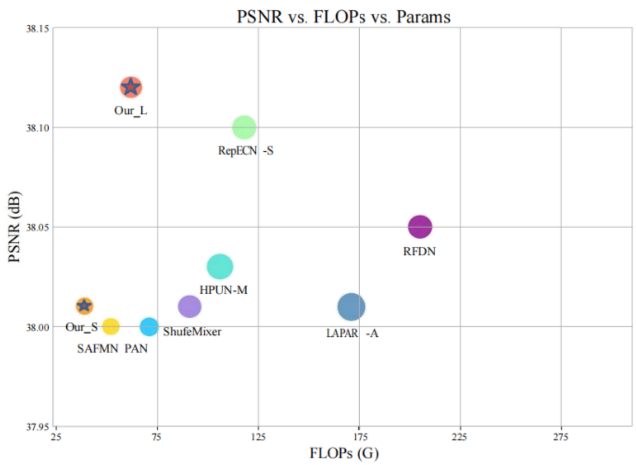


Figure 4: Schematic Comparison of Performance and Model Complexity for the SR Task

The design process of LFEM is as follows: Initially, input features undergo a 1×1 convolution to double the number of channels, enriching the feature representation:

$$X' = \mathrm{Conv}_{1\times1}(X_{in}) \quad (10)$$

Subsequently, the features are processed in parallel by multiple 3×3 convolutions to form a multi-branch structure that extracts local context. During this process, we introduce the GELU activation function to apply nonlinear mappings to the intermediate features and further enhance critical features using a spatial attention module (SE module). Finally, a 1×1 convolution reduces the channel dimensions to the original size, yielding the output features:

$$X_{out} = \mathrm{Conv}_{1\times1}(\mathrm{SE}(\phi(\mathrm{Rep}_{\mathrm{Conv}3\times3}(X')))) \quad (11)$$

where X_{in} and X_{out} represent the input and output features, respectively, $\mathrm{Rep}_{\mathrm{Conv}3\times3}$ denotes the equivalent 3×3 convolution kernel achieved using reparameterization techniques in the multi-branch structure.

3.4 Up-sampling Feature Reconstruction

After completing deep feature extraction, we use an up-sampling layer to reconstruct the high-resolution image. The up-sampling process employs a structure composed of a 3×3 convolutional layer and a subpixel convolutional layer to map the deep features to the high-resolution space. The final super-resolution image I_{SR} is obtained through the following equation:

$$I_{SR} = P_{\theta}(f_k + f_0) \quad (12)$$

The overall architecture incorporates the DSMB module, which consists of both the DSMU and LFEM modules. This process can be expressed as:

$$X' = \mathrm{DSMU}(\mathrm{LN}(X_{in})) + X \quad (13)$$

$$X_{out} = \mathrm{LFEM}(\mathrm{LN}(X')) + X' \quad (14)$$

4. Experiment

This chapter presents the experimental design, results, and comparative analysis of the building image super-resolution reconstruction method based on the DCFMN. We focus on evaluating the performance of the model, including reconstruction quality, computational complexity, and its ability to recover fine details in architectural images.

4.1 Loss Function Definition

Combining the traditional Mean Absolute Error (MAE) loss and the frequency loss based on Fast Fourier Transform (FFT), we define a composite loss function. This loss function aims to optimize the network's image reconstruction ability while enhancing the recovery of high-frequency details:

$$L = \lambda_1 \parallel I_{\mathrm{HR}} - I_{\mathrm{SR}} \parallel_1 + \lambda_2 \parallel F(I_{\mathrm{HR}}) - F(I_{\mathrm{SR}}) \parallel_2 \#(15)$$

where I_{HR} represents the high-resolution ground truth image, $\|\cdot\|_1$ denotes the L_1 norm, F refers to the Fast Fourier Transform, and γ are weighting factors.

Through this loss function, the model not only optimizes pixel-level reconstruction quality but also ensures that the frequency characteristics of the image are preserved, thereby improving the overall image quality.

4.2 Model Training Details

We trained two models of different sizes, denoted as DCFMN-S and DCFMN-L. DCFMN-S uses 10 modules, while DCFMN-L uses 16 modules. During training, each batch contains 32 low-resolution images with a size of 256×256. We utilized the Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.99$ to train the proposed models. The initial learning rate was set to 1×10^{-3}, with a minimum value of 1×10^{-6}, and the learning rate was updated using a cosine annealing algorithm. The Exponential Moving Average (EMA) weight was set to 0.999. The model was trained on an NVIDIA GeForce RTX 3080 GPU for a total of 10610^6106 iterations.

4.3 Testing and Performance Evaluation

To evaluate the performance of our method, we compared it against the most advanced lightweight SR methods, including SRCNN, VDSR, IMDN, EDSR-baseline, LAPAR, PAN, RFDN, ShuffleMixer, HPUN, HNCT, RepECN-S, and SAFMN. Table 1 presents the quantitative comparison results on benchmark datasets with scale factors of ×2, ×3, and ×4.

Table I: Quantitative evaluation of the super-resolution reconstruction performance in terms of PSNR/SSIM on four relevant datasets, along with a comparison to other existing methods (best and second-best performance highlighted in red and blue, respectively)

Methods	Scale	#Params[K]	Set5	Set14	B100	Urban100	Manga109
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic		-	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403	30.80/0.9339
SRCNN[]		57	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946	35.74/0.9661
VDSR		665	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9729
IMDN[]		694	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
EDSR-baseline[]		1370	37.99/0.9604	33.57/0.9175	32.16/0.8994	31.98/0.9272	38.54/0.9769
LAPAR-A[]		548	38.01/0.9605	33.62/0.9183	32.19/0.8999	32.10/0.9283	38.67/0.9772
PAN[]		261	38.00/0.9605	33.59/0.9181	32.18/0.8997	32.01/0.9273	38.70/0.9773
RFDN[]	× 2	417	38.05/0.9606	33.68/0.9184	32.16/0.8994	32.12/0.9278	38.88/0.9774
ShuffleMixer[]		394	38.01/0.9606	33.63/0.9180	32.17/0.8995	31.89/0.9257	38.83/0.9774
HPUN-M[]		492	38.03/0.9604	33.60/0.9185	32.20/0.9000	32.09/0.9282	38.83/0.9775
HNCT		365	38.08/0.9608	33.65/0.9182	32.22/0.9001	32.22/0.9294	38.87/0.9774
RepECN-S		411	38.10/0.9607	33.68/0.9187	32.24/0.9004	32.3/0.9301	38.76/0.9773
SAFMN[]		228	38.00/0.9605	33.54/0.9177	32.16/0.8995	31.84/0.9256	38.71/0.9771
DCFMN_S(Ours)		224	38.01/0.9605	33.55/0.9178	32.16/0.8996	31.97/0.9263	38.88/0.9776
DCFMN_L(Ours)		354	38.12/0.9609	33.75/0.9192	32.24/0.9005	32.26/0.9287	39.16/0.9782
Bicubic		-	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556
SRCNN[]		57	32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989	30.59/0.9107
VDSR		665	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9310
IDMN[]		703	34.36/0.9270	30.32/0.8417	29.09/0.8046	2817/0.8519	33.61/0.9445
EDSR-baseline[]		1555	34.37/0.9270	30.28/0.8417	29.09/0.8052	28.15/0.8527	33.45/0.9439
LAPAR-A[]		594	34.36/0.92676667	30.34/0.8421	29.11/0.8054	28.15/0.8523	33.51/0.9441
PAN[]		261	34.40/0.9271	30.36/0.8423	29.11/0.8050	28.11/0.8511	33.61/0.9448
RFDN[]	× 3	541	34.41/0.9273	30.34/0.8420	29.09/0.8050	28.21/0.8525	33.67/0.9449
ShuffleMixer[]		415	34.40/0,9272	30.37/0.8423	29.12/0.8051	28.08/0.8498	33.69/0.9448
HPUN-M[]		500	34.39/0.9269	30.33/0.8420	29.11/0.8052	28.06/0.8508	33.54/0.9441
HNCT		363	34.47/0.9275	30.44/0.8439	29.15/0.8067	28.28/0.8557	33.81/0.9459
RepECN-S		411	34.47/0.9277	30.41/0.8439	29.15/0.8064	28.30/0.8551	33.72/0.9456
SAFMN[]		233	34.34/0.9267	30.33/0.8418	29.08/0.8048	27.95/0.8474	33.52/0.9437
DCFMN_S(Ours)		230	34.37/0.9270	30.33/0.8417	29.12/0.8057	28.13/0.8505	33.80/0.9454
DCFMN_L(Ours)		360	34.52/0.9282	30.45/0.8440	29.17/0.8070	28.31/0.8542	34.04/0.9467
Bicubic		-	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
SRCNN[]		57	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.66/0.8505
VDSR		665	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8809
IDMN[]		715	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
EDSR-baseline[]		1518	32.09/0.8938	28.58/0.7813	27.57/0.7357	26.04/0.7849	30.35/0.9067
LAPAR-A[]	× 4	659	32.15/0.8944	28.61/0.7818	27.61/0.7366	26.14/0.7871	30.42/0.9074
PAN[]		261	32.13/0.8948	28.61/0.7822	27.59/0.7363	26.11/0.7854	30.51/0.9095
RFDN[]		550	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858	30.58/0.9089
ShuffleMixer[]		411	32.21/0.8953	28.66/0.7827	27.61/0.7366	26.08/0.7835	30.65/0.9093
HPUN-M[]		511	32.19/0.8946	28.61/0.7818	27.58/0.7364	26.04/0.7851	30.49/0.9078
HNCT		372	32.31/0.8957	28.71/0.7834	27.63/0.7381	26.20/0.7896	30.70/0.9112
RepECN-S		427	32.32/0.8964	28.69/0.7833	27.62/0.7375	26.19/0.7889	30.54/0.9099
SAFMN[]		240	32.18/0.8948	28.60/0.7813	27.58/0.7359	25.97/0.7809	30.43/0.9063
DCFMN_S(Ours)		239	32.25/0.8955	28.65/0.7824	27.61/0.7368	26.08/0.7833	30.60/0.9083
DCFMN_L(Ours)		369	32.33/0.8965	28.73/0.7843	27.65/0.7385	26.27/0.7888	30.83/0.9108

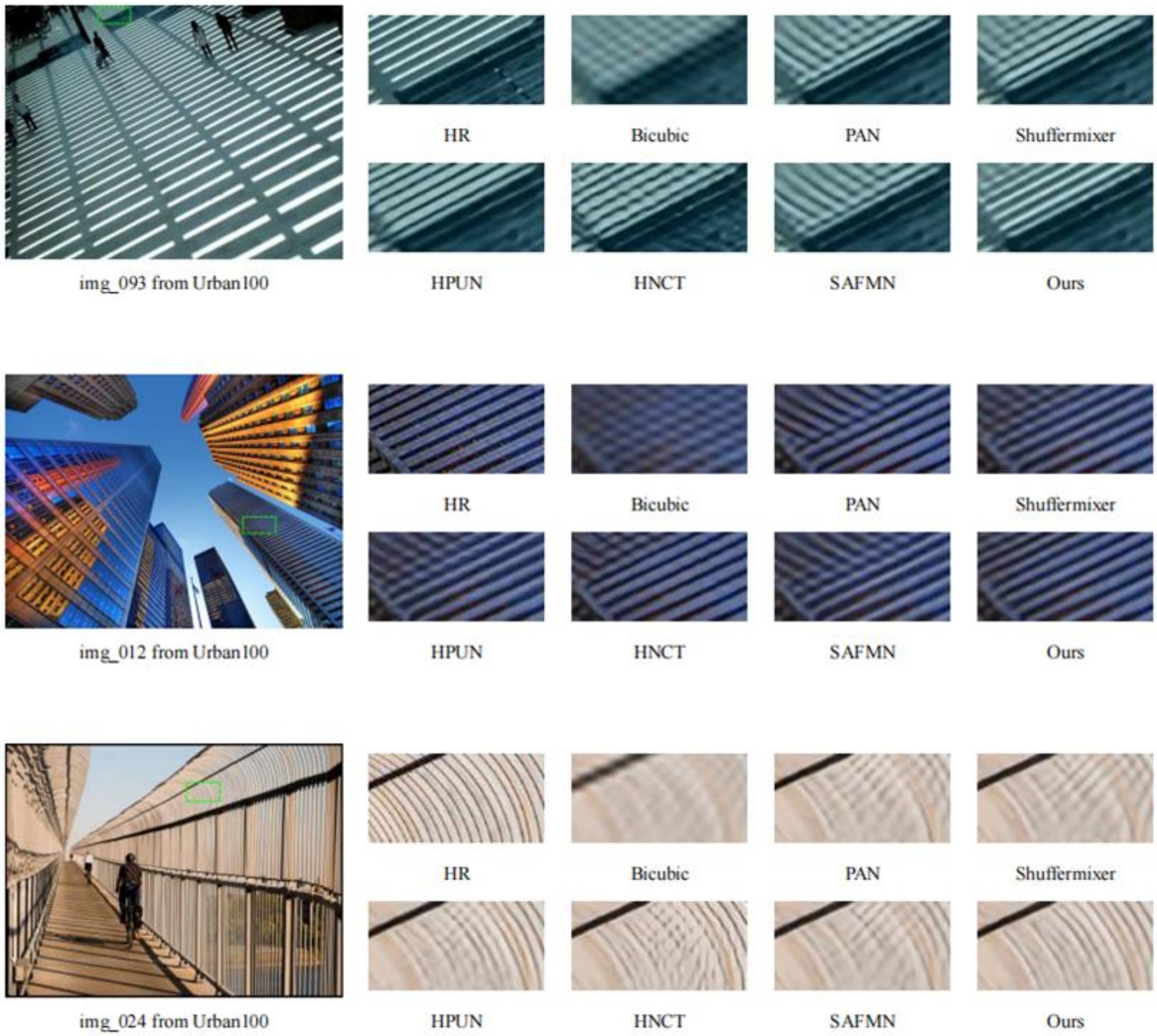


Figure 5: ×4 Comparison Plot of Detail Presentation on Architectural Images on the Super-resolution (SR) Task

Beyond PSNR/SSIM metrics, we also report the number of parameters (#params) and floating-point operations (#FLOPs). The #params relate to memory consumption, while #FLOPs are associated with energy consumption. #FLOPs were computed after super-resolving LR images to 1280×720 HR images using the fvcore library. Results in Table 1 and Figure 4 confirm that the proposed method based on the DCFMN achieves a favorable trade-off between model complexity and reconstruction performance.

To further demonstrate the efficacy of our method, we conducted a visual comparison on the Urban100 dataset with a scale factor of ×4. As shown in Figure 5, most of the compared lightweight super-resolution methods fail to accurately recover architectural image textures and details, exhibiting noticeable blurring artifacts and distortions. In contrast, our proposed lightweight building image super-resolution method based on the DCFMN achieves the best visual quality, recovering more architectural structural details.

Moreover, to validate the effectiveness of the DSMU and LFEM components in our design, we performed ablation experiments. The results in Tables 2 and 3 demonstrate that all components of DSMU and LFEM are crucial for enhancing performance

	Method	#Params	#FLOPs	Set5	Set14	Urban100
DSMU	Baseline	369K	16.29G	32.33/0.8965	28.73/0.7843	26.27/0.7888
	DSMU -> 3×3	382K	19.01G	32.27/0.8955	28.60/0.7816	26.06/0.7827
	DSMU -> Large kernel	407K	18.57G	32.26/0.8961	28.71/0.7839	26.21/0.7874
	DSMU -> 1×n+n×1	381K	15.80G	32.31/0.8964	28.69/0.7840	26.22/0.7879

Table 2: Ablation Study of The DSMU Module Performance

5. Conclusion

In this work, we have presented a lightweight building image super-resolution reconstruction method based on a DCFMN. Our approach overcomes the limitations of existing lightweight super-resolution models in capturing long-range dependencies by enabling precise and efficient global modeling without increasing the parameter count. This advancement substantially enhances both the robustness and reconstruction quality of lightweight super-resolution models.

	Method	#Params	#FLOPs	Set5	Set14	Urban100
LFEN	Baseline	239K	10.66G	32.25/0.8955	28.65/0.7824	26.08/0.7833
	w/o self-residual	239K	10.66G	32.18/0.8948	28.61/0.7816	26.05/0.7823
	LFEN -> RepVGG	239K	10.66G	32.24/0.8951	28.62/0.7822	26.07/0.7826
	LFEN -> AcBlock	239K	10.66G	32.19/0.8945	28.60/0.7815	26.07/0.7828
	LFEN->double 3x3	239K	10.66G	32.23/0.8952	28.63/0.7824	26.08/0.7835
	w/o SE	192K	10.66G	32.15/0.8943	28.60/0.7814	25.97/0.7800
	LFEN->Channal MLP	182K	10.17G	32.08/0.8935	28.56/0.7806	25.91/0.7783
	LFEN-> Inverted residual block	225K	12.54G	32.09/0.8936	28.58/0.7812	25.90/0.7782

Table 3: Ablation Study of the LFEM Module Performance

References

1. Zadeh, L. A., Klir, G. J., & Yuan, B. (1996). *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers* (Vol. 6). World scientific.
2. Bezdek, J. C., & Pal, S. K. (1993). *Fuzzy models for pattern recognition* (No. CONF-9410335--). USDOE Pittsburgh Energy Technology Center, PA (United States); Oregon State Univ., Corvallis, OR (United States). Dept. of Computer Science; Naval Research Lab., Washington, DC (United States); Electric Power Research Inst., Palo Alto, CA (United States); Bureau of Mines, Washington, DC (United States).
3. Mitra, S., & Pal, S. K. (2005). Fuzzy sets in pattern recognition and machine intelligence. *Fuzzy Sets and systems, 156*(3), 381-386.
4. Ishibuchi, H., Nozaki, K., & Tanaka, H. (1992). Distributed representation of fuzzy rules and its application to pattern classification. *Fuzzy sets and systems, 52*(1), 21-32.
5. Abe, S., & Lan, M. S. (1995). A method for fuzzy rules extraction directly from numerical data and its application to pattern classification. *IEEE transactions on fuzzy systems, 3*(1), 18-28.
6. Mille, S., Belz, A., Bohnet, B., Graham, Y., Pitler, E., & Wanner, L. (2018, July). The first multilingual surface realisation shared task (sr'18): Overview and evaluation results. In *Proceedings of the First Workshop on Multilingual Surface Realisation* (pp. 1-12).
7. Liu, D., Wen, B., Liu, X., Wang, Z., & Huang, T. S. (2017). When image denoising meets high-level vision tasks: A deep learning approach. *arXiv preprint arXiv:1706.04284*.
8. Ward, C. M., Harguess, J., Crabb, B., & Parameswaran, S. (2017, September). Image quality assessment for determining efficacy and limitations of Super-Resolution Convolutional Neural Network (SRCNN). In *Applications of Digital Image Processing XL* (Vol. 10396, pp. 19-30). SPIE.
9. Nguyen, T., Raghu, M., & Kornblith, S. (2020). Do wide and deep networks learn the same things? uncovering how neural network representations vary with width and depth. *arXiv preprint arXiv:2010.15327.*
10. Wu, Y., Ma, Y., Liu, J., Du, J., & Xing, L. (2019). Self-attention convolutional neural network for improved MR image reconstruction. *Information sciences, 490*, 317-328.
11. Lin, Z., Garg, P., Banerjee, A., Magid, S. A., Sun, D., Zhang, Y., ... & Pfister, H. (2022). Revisiting rcan: Improved training for image super-resolution. *arXiv preprint arXiv:2201.11279.*
12. Yingze Bao, S., Chandraker, M., Lin, Y., & Savarese, S. (2013). Dense object reconstruction with semantic priors. In Proceedings of the *IEEE conference on computer vision and pattern recognition* (pp. 1264-1271).
13. Mei, Y., Fan, Y., Zhou, Y., Huang, L., Huang, T. S., & Shi, H. (2020). Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 5690-5699).
14. Mei, Y., Fan, Y., & Zhou, Y. (2021). Image super-resolution with non-local sparse attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3517-3526).
15. Zhang, X., Zeng, H., Guo, S., & Zhang, L. (2022, October).

Efficient long-range attention network for image super-resolution. In *European conference on computer vision* (pp. 649-667). Cham: Springer Nature Switzerland.
16. Khan, A., Rauf, Z., Sohail, A., Khan, A. R., Asif, H., Asif, A., & Farooq, U. (2023). A survey of the vision transformers and their CNN-transformer based variants. *Artificial Intelligence Review, 56*(Suppl 3), 2917-2970.
17. Chen, C., Zhang, P., Zhang, H., Dai, J., Yi, Y., Zhang, H., & Zhang, Y. (2020). Deep learning on computational-resource-limited platforms: A survey. *Mobile Information Systems, 2020*(1), 8454327.
18. Hui, Z., Gao, X., Yang, Y., & Wang, X. (2019, October). Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia* (pp. 2024-2032).
19. Liang, S., Yin, S., Liu, L., Luk, W., & Wei, S. (2018). FP-BNN: Binarized neural network on FPGA. *Neurocomputing, 275,* 1072-1086.
20. Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., ... & König-Langlo, G. (2018). Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017). *Earth System Science Data, 10*(3), 1491-1501.
21. Zhou, L., Cai, H., Gu, J., Li, Z., Liu, Y., Chen, X., ... & Dong, C. (2022, October). Efficient image super-resolution using vast-receptive-field attention. In *European conference on computer vision* (pp. 256-272). Cham: Springer Nature Switzerland.
22. Puthussery, D., Panikkasseril Sethumadhavan, H., Kuriakose, M., & Charangatt Victor, J. (2020, August). Wdrn: A wavelet decomposed relightnet for image relighting. In *European Conference on Computer Vision* (pp. 519-534). Cham: Springer International Publishing.
23. Kong, X., Zhao, H., Qiao, Y., & Dong, C. (2021). Classsr: A general framework to accelerate super-resolution networks by data characteristic. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12016-12025).
24. Ma, C., Lin, X., Lv, H., & Wang, H. (2009, December). ABSR: An agent based self-recovery model for wireless sensor network. In *2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure* Computing (pp. 400-404). IEEE.
25. Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2022). Transformers in vision: A survey. *ACM computing surveys (CSUR), 54*(10s), 1-41.
26. Liu, T., Xu, C., Tang, L., Meng, Y., Xu, W., Wang, J., & Xu, J. (2024). OmniSR-M: A rock sheet with a multi-branch structure image super-resolution lightweight method. *Applied Sciences, 14*(7), 2779.
27. Sun, L., Dong, J., Tang, J., & Pan, J. (2023). Spatially-adaptive feature modulation for efficient image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 13190-13199).
28. Meneganti, M., Saviello, F. S., & Tagliaferri, R. (1998). Fuzzy neural networks for classification and detection of anomalies. *IEEE transactions on neural networks, 9*(5), 848-861.
29. Tagliaferri, R., Eleuteri, A., Meneganti, M., & Barone, F. (2001). Fuzzy min–max neural networks: from classification to regression. *Soft Computing, 5*(1), 69-76.
30. Kim, H. J., & Yang, H. S. (2005, August). A weighted fuzzy min-max neural network and its application to feature analysis. In *International Conference on Natural Computation* (pp. 1178-1181). Berlin, Heidelberg: Springer Berlin Heidelberg.
31. Gabrys, B., & Bargiela, A. (2000). General fuzzy min-max neural network for clustering and classification. *IEEE transactions on neural networks, 11*(3), 769-783.
32. Quteishat, A., & Lim, C. P. (2008). A modified fuzzy min–max neural network with rule extraction and its application to fault detection and classification. *Applied Soft Computing, 8*(2), 985-995.
33. Quteishat, A., Lim, C. P., & Tan, K. S. (2010). A modified fuzzy min–max neural network with a genetic-algorithm-based rule extractor for pattern classification. *IEEE transactions on systems, man, and cybernetics-part a: systems and humans, 40*(3), 641-650.
34. Nandedkar, A. V., & Biswas, P. K. (2007). A fuzzy min-max neural network classifier with compensatory neuron architecture. *IEEE transactions on neural networks, 18*(1), 42-54.
35. Bargiela, A., Pedrycz, W., & Tanaka, M. (2004). An inclusion/exclusion fuzzy hyperbox classifier. *International Journal of Knowledge-based and Intelligent Engineering Systems, 8*(2), 91-98.
36. Kohonen, T. (2012). *Self-organization and associative memory* (Vol. 8). Springer Science & Business Media.
37. https://cir.nii.ac.jp/crid/1574231874559408512.