

# Block-Based Multi-Scale Image Rescaling

Jian Li and Siwang Zhou\*

Hunan University, China  
{lijian543, swzhou}@hnu.edu.cn

## Abstract

Image rescaling (IR) seeks to determine the optimal low-resolution (LR) representation of a high-resolution (HR) image to reconstruct a high-quality super-resolution (SR) image. Typically, HR images with resolutions exceeding 2K possess rich information that is unevenly distributed across the image. Traditional image rescaling methods often fall short because they focus solely on the overall scaling rate, ignoring the varying amounts of information in different parts of the image. To address this limitation, we propose a Block-Based Multi-Scale Image Rescaling Framework (BBMR), tailored for IR tasks involving HR images of 2K resolution and higher. BBMR consists of two main components: the Downscaling Module and the Upscaling Module. In the Downscaling Module, the HR image is segmented into sub-blocks of equal size, with each sub-block receiving a dynamically allocated scaling rate while maintaining a constant overall scaling rate. For the Upscaling Module, we introduce the Joint Super-Resolution method (JointSR), which performs SR on these sub-blocks with varying scaling rates and effectively eliminates blocking artifacts. Experimental results demonstrate that BBMR significantly enhances the SR image quality on the of 2K and 4K test dataset compared to initial network image rescaling methods.

## Introduction

With the growing demand in practical applications, the technology of image rescaling (IR) has rapidly advanced. Specifically, it can significantly optimize storage utilization (Zhang and Wu 2023) and minimize the bandwidth required for images and videos transmission (Yeo et al. 2018, 2020; Yu et al. 2023). The availability of high-resolution displays and advancements in camera equipment have led to the widespread usage of images and videos with 2K resolution or higher. However, as image resolution increases, there is a significant exponential growth in the demand for storage space and network transmission bandwidth. Therefore, the development of IR techniques is crucial for effectively storing and transmitting high-resolution images and videos.

Typically, IR methods (Sun and Chen 2020; Xu et al. 2023; Xiao et al. 2023) demonstrate superior super-resolution (SR) image reconstruction performance com-

\*Corresponding author.

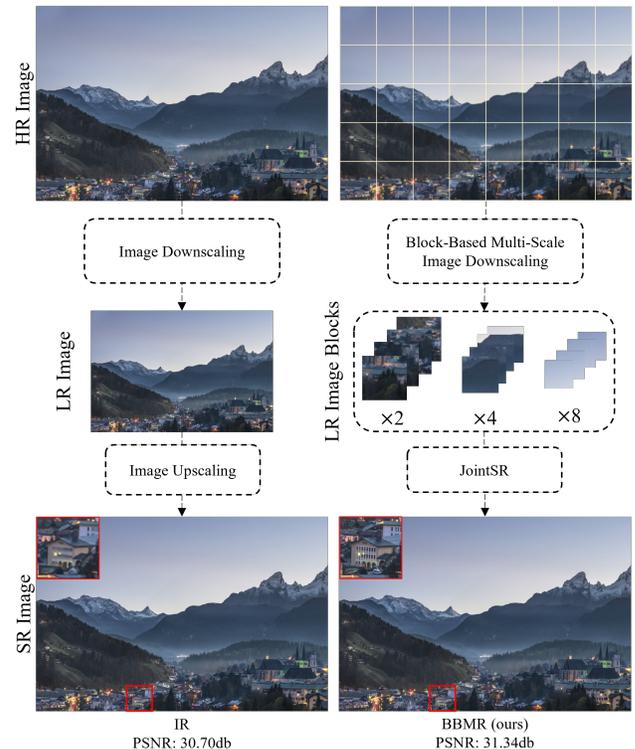


Figure 1: Comparison of traditional image rescaling methods (left) and our block-based multi-scale image rescaling framework (right) at  $\times 4$  scaling rate. Both methods use bicubic interpolation downscaling and OmniSR model upscaling. The improvement in the image quality of the red box is due to the smaller scaling rate.

pared to image super-resolution models (Wang et al. 2023; Chen et al. 2023a,b) trained with bicubic downscaling. This superiority is attributed to IR’s focus on designing well-crafted downscaling methods, which produce low-resolution (LR) images more suitable for SR reconstruction. In practice, from a holistic perspective, these methods emphasize overall image scaling. However, (Kong et al. 2021) suggests that the difficulty of super-resolution varies across different regions of an image, indicating that different parts contain varying amounts of information. Therefore, considering

only the overall scaling rate limits the reconstruction performance of IR methods.

In this paper, we consider an application scenario for image rescaling where LR images are utilized solely to save storage space and reduce transmission bandwidth without the need for visualization. Inspired by (Zhou et al. 2020), which improves overall reconstruction quality by dynamically adjusting the blocks' compression ratio, we adopt a similar approach for image rescaling. Specifically, we select different downscaling factors for different parts of the image, enhancing the overall super-resolution quality.

In this work, we introduce the BBMR framework for achieving multi-scale scaling, which comprises two components: the Downscaling Module and the Upscaling Module. Within the Client/Server Architecture, the Downscaling Module is typically managed on the server side, rendering the computational cost negligible. In this module, the HR image is initially divided into  $128 \times 128$  sub-blocks. Subsequently, we dynamically allocate the scaling rate for each sub-block while maintaining a consistent overall scaling rate, utilizing the assistance of the super-resolution model from the Upscaling Module during this process. For instance, regions of the image containing sky are assigned a higher scaling rate, whereas areas with buildings are allocated a lower scaling rate. As illustrated in Fig. 1, our block-based multi-scale downscaling method (right) contrasts with traditional downscaling methods (left). In the Upscaling Module, we utilize a joint super-resolution method (JointSR) to perform sub-block super-resolution and eliminate blocking artifacts. This method integrates a deblocking branch to address blocking effects at the feature level. Furthermore, we observe that deep learning-based SR models share certain similarities. To minimize server-side computation, a lightweight super-resolution model can be employed to assist in scaling rate allocation with minimal degradation of SR image quality.

Since what we propose is a general image rescaling framework that can be applied to various image rescaling methods, we used two different SR networks, OmniSR (Wang et al. 2023) and CRAFT (Li et al. 2023a), in our experiments. Additionally, we employed the neural network downscaling method described in (Fan, Lian, and Quan 2022) and bicubic interpolation to validate our BBMR framework.

Overall, our contributions can be summarized in three key aspects:

- We proposed the block-based multi-scale downscaling strategy, which can more reasonably allocate scaling rates for different parts of the image in the Downscaling Module, thereby significantly improving the image reconstruction quality.
- We propose the JointSR method, which adopts a transfer learning strategy and incorporates LR blocks of different resolutions for training. Additionally, it introduces a de-block branch to eliminate blocking artifacts at the feature level.
- We find that it can use a very lightweight super-resolution model to assist in scaling rate allocation in the Downscal-

ing Module, which can greatly reduce the computational complexity of the Downscaling Module.

## Related Work

### Single Image Super-Resolution

With the rapid development of deep learning, significant progress has been made in single image super-resolution. SRResNet (Ledig et al. 2017) utilized residual connections to preserve details from previous layers. Recent approaches, such as (Liang et al. 2021; Chen et al. 2023a,b), employ global receptive field attention mechanisms to further enhance SR quality. In addition, advancements in diffusion models (Gao et al. 2023; Yue, Wang, and Loy 2024) have been explored in the image SR process to generate clearer super-resolved images.

### Block-Base Super-Resolution

Block-based SR methods have also been extensively studied. (Li et al. 2023b) proposes a video block over-fitting super-resolution method to achieve real-time and high-quality video super-resolution. Other methods, such as (Kong et al. 2021; Wang et al. 2024; Zhang et al. 2024), adopt different types of super-resolution models according to the difficulty of super-resolving image sub-blocks to improve the overall super-resolution speed. (Luo et al. 2024) leverages the block characteristics of transformers, different categories of image sub-blocks to exit early in various Transformer layers to improve calculation speed. Unlike the above methods, which mainly focus on improving the super-resolution speed of the Upscaling Module, our proposed BBMR framework implements block multi-scaling in the Downscaling Module to enhance the image super-resolution quality of the Upscaling Module.

### Image Rescaling

Image rescaling (IR) and image super-resolution (SR) are distinct tasks. IR involves downscaling and upscaling, while SR only focuses on upscaling. In IR, a ground-truth HR image is downscaled for storage and transmission and recovered when necessary. Downscaling, which generates an LR version of an HR image, is the inverse of SR. Bicubic interpolation (Gao and Gruev 2011) is the most common downscaling method. Recently, more and more work has been treating downscaling and upscaling as a unified process. (Sun and Chen 2020) proposed a learning-based image reduction method using a Content-adaptive Resampler, which effectively improves SR reconstruction. (Kim et al. 2018) introduced a technique called task-aware image reduction to support SR tasks. Moreover, (Xiao et al. 2023) presented a reversible network that models bidirectional degradation and recovery from a new perspective. These works focus on the overall scaling of the image, but in many scenarios, LR images are only used as storage and transmission media (Yeo et al. 2018; Yu et al. 2023). Therefore, we propose our BBMR framework, whose LR images are stored as sub-blocks with different resolutions, and can be applied to most of the IR and SR methods.

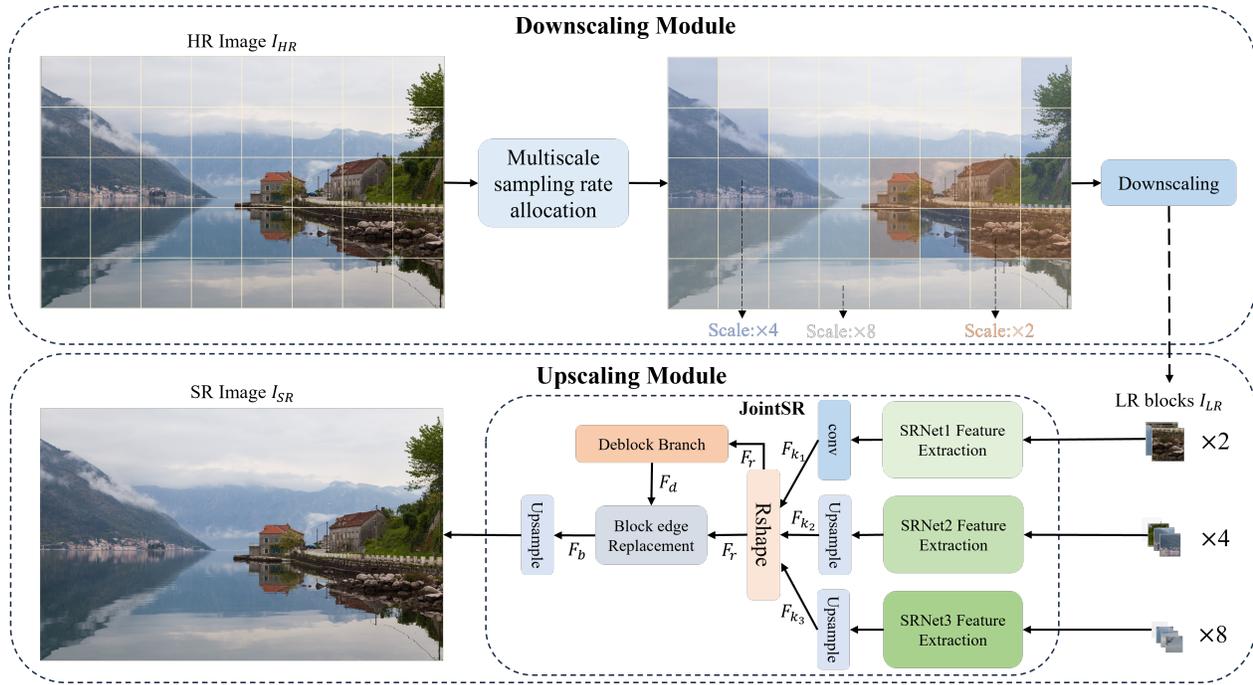


Figure 2: The overall framework structure of the proposed BBMR when the scaling rate is  $\times 2$ ,  $\times 4$  and  $\times 8$ . Downsampling Module: Outputs LR image sub-blocks with three different resolutions through block-based multi-scale downscaling strategy. Upscaling Module: Generates high-quality SR images using the JointSR method. Upsample contains a convolution layer and a pixelshuffle module.

## Methods

The overall architecture of our method is illustrated in Fig. 2, with the Block-Based Multi-Scale Image Rescaling Framework (BBMR) comprising two main components: the Downsampling Module and the Upscaling Module. The Downsampling Module divides the image into blocks and dynamically allocates the scaling rate for each block. The Upscaling Module then takes these blocks with varying scaling rates and processes them through our JointSR method, effectively performing super-resolution on the image blocks and removing blocking artifacts.

### Observation

In order to evaluate the difficulty of super-resolution for image sub-blocks with different resolutions, we averaged the HR images of the DIV2K validation dataset into  $128 \times 128$  sub-blocks and validated them using BICUBIC downscaling and OmniSR upscaling methods. As shown in Tab. 1, we found that approximately 20% of sub-blocks experienced only a 1.06 dB decrease when downscaled by  $\times 8$  compared to  $\times 4$ , which we term simple blocks. About 4.7% of the sub-blocks showed an 8.21 dB improvement when downscaled by  $\times 2$  compared to  $\times 4$ , which we term hard blocks. The remaining 75.3% showed no significant difference in PSNR changes between downscaling by  $\times 2$  and  $\times 8$  compared to  $\times 4$ , which we term medium blocks. This finding allows us to trade a small quality loss in simple sub-blocks for a significant quality gain in hard sub-blocks while maintaining

the overall scaling rate.

Scales	simple	medium	hard
$\times 2$	-	37.24	39.13
$\times 4$	39.02	31.62	30.92
$\times 8$	37.96	27.34	-

Table 1: PSNR at different super-resolution scales for different categories of image blocks in the DIV2K validation dataset. The tests use BICUBIC downscaling and OmniSR upscaling

### Block-Based Multi-Scale Downscaling Strategy

In Downsampling Module in order to realize adaptive allocation of image sub-block sample rate, we designed a sample rate allocation algorithm. The method first determines three downscaling factors  $k_1, k_2, k_3 (k_1 < k_2 < k_3)$  and takes  $k_2$  as the overall scaling rate. To ensure that the overall scaling rate is maintained at  $k_2$ , it is necessary to determine the values of  $a$  and  $c$ . The ratio  $a : c$  represents the proportion of image blocks with scaling factors  $k_1$  and  $k_3$ .  $a$  and  $c$  must satisfy the following formula.

$$\left\{ a \left( \frac{hw}{k_1^2} \right) + c \left( \frac{hw}{k_3^2} \right) \right\} \bmod k_2^2 = 0 \quad (1)$$

Where  $h$  and  $w$  is the height and width of the image blocks. The input HR image  $X^{HR}$  is divided into  $N$  sub-blocks  $\{x_i^{HR}\}_{i=1}^N$  of size  $(h, w)$ , and then all the sub-blocks

---

**Algorithm 1: Block Scaling Rate Allocation Algorithm**


---

**Input:**  $arr^{pay}, arr^{earn}, block\_max_{k_1}, a, c, t, N$ 
**Output:**  $blocks_{k_1}, blocks_{k_2}, blocks_{k_3}$ 

```

1: for  $i = 1 : block\_max_{k_1}$  do
2:    $earn \leftarrow Select(arr^{earn}, i * a, a)$ 
3:    $pay \leftarrow Select(arr^{pay}, i * c, c)$ 
4:   if  $pay + t > earn$  then
5:      $break$ 
6:   else
7:      $Append(blocks_{k_1}, arr^{earn}, i * a, a)$ 
8:      $Append(blocks_{k_3}, arr^{pay}, i * c, c)$ 
9:   end if
10: end for
11:  $blocks_{k_2} \leftarrow FindB.k2(N, blocks_{k_1}, blocks_{k_3})$ 

```

---

are downscaled into LR sub-blocks of three resolutions respectively, the process are as follows:

$$\{x_{i,s}^{LR}\}_{s=1}^3 = \{f_{Down}^s(x_i^{HR})\}_{s=1}^3 \quad (2)$$

Where  $s = 1, 2, 3$  correspond to  $k_1, k_2, k_3$ , respectively. Then all the LR sub-blocks are super-resolved into  $128 \times 128$  resolution SR blocks by the same SR method as the Upscaling Module, which is indicated below:

$$\{x_{i,s}^{SR}\}_{s=1}^3 = \{f_{up}^s(x_{i,s}^{LR})\}_{s=1}^3 \quad (3)$$

Calculate the PSNR values for all the SR and HR sub-blocks  $\{P_{i,s}\}_{i=1}^N = \{f_{PSNR}(x_i^{HR}, x_{i,s}^{SR})\}_{i=1}^N$ . Calculate the difference between all  $s = 1$  and  $s = 2$  sub-blocks in ascending order. Then, calculate the difference between all  $s = 2$  and  $s = 3$  sub-blocks in descending order sort. The process are as follows:

$$\{arr_k^{earn}\}_{k=1}^N = sort(\{P_{i,1} - P_{i,2}\}_{i=1}^N, asc) \quad (4)$$

$$\{arr_k^{pay}\}_{k=1}^N = sort(\{P_{i,2} - P_{i,3}\}_{i=1}^N, desc) \quad (5)$$

The specific sample rate allocation algorithm is shown in Alg. 1.

$block\_max_{k_1}$  is the maximum number of  $k_1$  scaling factor blocks that can be selected.  $arr^{pay}$  and  $arr^{earn}$  represent  $\{arr_k^{pay}\}_{k=1}^N$  and  $\{arr_k^{earn}\}_{k=1}^N$  respectively.  $t$  is a constant, set according to the super-resolution model and the image size.  $blocks_{k_1}, blocks_{k_2}, blocks_{k_3}$  hold the sub-block numbers corresponding to the three scaling rates.  $Select(arr^{earn}, i * a, a)$  means selecting  $a$  elements from  $arr^{earn}$  starting at position  $i * a$  and returning the sum of those  $a$  elements.  $Append(blocks_{k_1}, arr^{earn}, i * a, a)$  means selecting  $a$  elements from  $arr^{earn}$  starting at position  $i * a$  to append at the end of  $blocks_{k_1}$ .  $FindB.k2$  means to return all blocks in  $N$  that is not in  $blocks_{k_1}$  and  $blocks_{k_3}$ .

### Joint Super-Resolution Method

When stitching together super-resolved image blocks, block artifacts are inevitable. Typically, neural networks address block artifacts by processing the generated SR image, which significantly increases the computational load due to the input being in high resolution. To tackle this issue, we propose

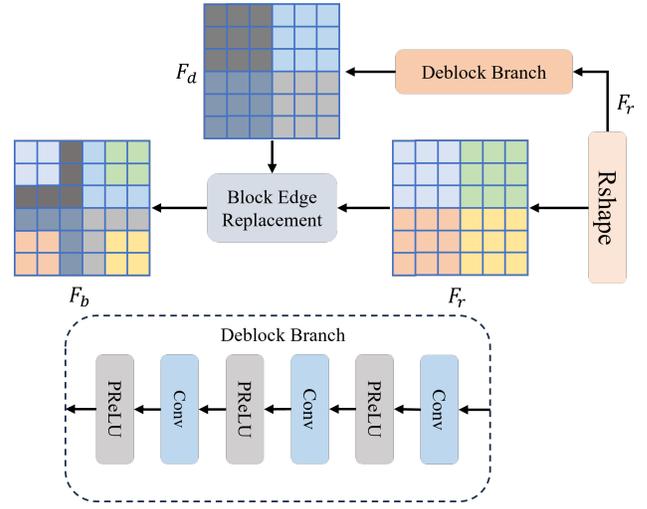


Figure 3: Illustration of the block edge replacement policy and deblock branch structure.

the joint super-resolution (JointSR) method, which performs deblocking while the features are still in the LR stage. As illustrated in Fig. 2, the SRNet Feature Extraction layer is derived from a pre-trained super-resolution model, where the final upscaling layer is removed. This part can be adapted depending on the super-resolution model used. Initially, LR image sub-blocks with different resolutions  $blocks_{k_1}, blocks_{k_2}, blocks_{k_3}$  pass through the SRNet Feature Extraction and Upsample modules sequentially to obtain feature blocks of the same size  $F_{k_1}, F_{k_2}, F_{k_3}$ .

Directly stitching these blocks into a full image feature to process would inevitably cause block artifacts, so we introduce the deblock branch. As shown in Fig. 3, Deblock branch consists of three layers of convolution and PReLU activation function. The features  $F_{k_1}, F_{k_2}, F_{k_3}$  are reshaped into a full feature image  $F_r$  according to their positions. The  $F_r$  passes through the deblock branch to obtain  $F_d$ , which is used to replace the block edges in  $F_r$  and resulting in  $F_b$ . Finally, an Upsample module outputs the SR image  $I_{SR}$ .

We train the JointSR model as an integrated unit. Three SRNet Feature Extraction layers use pre-trained weights.

### Very Lightweight Super-Resolution Model

We find that SR images obtained through deep learning-based super-resolution exhibit certain similarities. There are computational limitations in some cases. To reduce Downscaling Module computation, we propose using a fixed very lightweight super-resolution model instead of dynamically adjusting the Downscaling Module's SR model to match the SR model in Upscaling Module. This work not only significantly reduces the computational load on the server but also can reduce the overhead of model replacement.

To achieve this very lightweight super-resolution model, we streamline the architecture by reducing the number of OSAG modules from 5 to 1, based on the OmniSR (Wang et al. 2023). This reduction maintains the model's ability to

Evaluation index	SR method	BICUBIC		Down-Net		Mean
		Test2K	Test4K	Test2K	Test4K	
PI↓	OmniSR-O	6.3213	5.7104	5.4829	4.8712	5.5964
	BBMR-OmniSR	<b>5.5386</b>	<b>5.0819</b>	<b>4.6909</b>	<b>4.2477</b>	<b>4.8897</b>
	CRAFT-O	6.4482	5.8717	5.4919	4.8938	5.6764
	BBMR-CRAFT	<b>5.5865</b>	<b>5.1480</b>	<b>4.6519</b>	<b>4.2273</b>	<b>4.9034</b>
NIQE↓	OmniSR-O	6.6214	5.9650	5.6639	5.0877	5.8345
	BBMR-OmniSR	<b>5.5452</b>	<b>5.0814</b>	<b>4.5704</b>	<b>4.2640</b>	<b>4.8652</b>
	CRAFT-O	6.7598	6.1989	5.6746	5.1009	5.9335
	BBMR-CRAFT	<b>5.5626</b>	<b>5.2774</b>	<b>4.4778</b>	<b>4.2302</b>	<b>4.8870</b>
PSNR↑	OmniSR-O	36.50	33.15	39.60	36.75	36.50
	BBMR-OmniSR	<b>37.78</b>	<b>34.30</b>	<b>41.36</b>	<b>38.49</b>	<b>37.98</b>
	CRAFT-O	36.32	32.99	39.46	36.64	36.35
	BBMR-CRAFT	<b>37.53</b>	<b>34.05</b>	<b>41.43</b>	<b>38.46</b>	<b>37.86</b>

Table 2: PI(Blau et al. 2018), NIQE(Mittal, Soundararajan, and Bovik 2012) and PSNR values on Test2K, Test4K. The best results are remarked in bold font. O: the original networks with overall image scaling. BBMR: Block-Based Multi-Scale Image Rescaling Framework. OmniSR and CRAFT denote upscaling models. BICUBIC and Down-Net denote downscaling methods.

SR method	Parameters	Test2k	FLOPs	Test4k	FLOPs
OmniSR-O	0.77M	36.50	<b>243.46G(100%)</b>	33.15	<b>327.87G(100%)</b>
BBMR-OmniSR	2.3M	<b>37.78</b>	250.15G(103%)	<b>34.30</b>	337.22G(103%)
CRAFT-O	0.72M	36.32	<b>242.20G(100%)</b>	32.99	<b>326.17G(100%)</b>
BBMR-OmniSR	2.20M	<b>37.53</b>	249.04G(103%)	<b>34.05</b>	335.61G(103%)

Table 3: Comparison of Parameters, PSNR and FLOPs using bicubic downscaling method and different upscaling models. Parameters and FLOPs only computes the Upscaling Module. The best results are remarked in bold font.

effectively upscale images while dramatically cutting down the computational requirements.

## Experiment

### Setting

Our proposed BBMR is a general image rescaling framework that can be applied to various image rescaling methods. For verification, we selected OmniSR (Wang et al. 2023) and CRAFT (Li et al. 2023a) as super-resolution models and employed both BICUBIC downscaling and a neural network-based downscaling method proposed by (Fan, Lian, and Quan 2022). This allows us to effectively demonstrate the effect of our framework. For ease of experimentation, we select  $\times 2$ ,  $\times 4$ , and  $\times 8$  as our multi-scale factors, with the overall scaling factor set to  $\times 4$ .

**Training Data** We use the DIV2k and Flickr2k (Timofte et al. 2017) dataset for training. With BICUBIC downscaling, LR image sizes are  $64 \times 64$  for  $\times 2$  and  $\times 4$  scaling factors, and  $32 \times 32$  for  $\times 8$ . Down-Net employs end-to-end training with  $256 \times 256$  HR input and SR output. For JointSR, we randomly crop  $768 \times 768$  regions from DIV2k and Flickr2k as HR images and divide them into 36 HR sub-blocks with the size of  $128 \times 128$ . Subsequently, we use three different scaling rates to randomly select 12 sub-blocks for downscaling as the input of JointSR. Data augmentation includes random horizontal flips and  $90/270$  degree rotations.

**Testing data** Test2K and Test4K each contain 100 images, which were selected from the DIV8K (Gu et al. 2019) dataset and downsampled using bicubic interpolation. PSNR comparisons are performed on the Y channel in YCbCr.

**Training details** The batch size for OmniSR and CRAFT at all upscaling rates is set to 32, trained for 1000 epochs. JointSR uses a batch size of 4 and trained for 100 epochs. All networks start with a learning rate of 0.0005, for SR networks halved every 250 epochs, and for JointSR halved every 20 epochs. Training utilizes the AdamW optimizer with L1 loss. All models are built with PyTorch and trained on NVIDIA GeForce RTX 4090 GPUs. The loss of end-to-end training is shown below, where  $k$  is the scaling factor.

$$Loss_{p2p} = \left(\frac{1}{2^k}\right)L1Loss(LR', LR) + L1Loss(SR, HR) \quad (6)$$

$LR'$  represents the output obtained through Down-Net downscaling, while  $LR$  is derived via bicubic downscaling.  $SR$  is the output after super-resolution, and  $HR$  is the original high-resolution image.

### Evaluation Index Of BBMR

We select PSNR, NIQE, and PI as the evaluation metrics for our experiments, where NIQE and PI are no-reference visual evaluation metrics. As shown in Tab. 2, no matter which image rescaling method is chosen, our BBMR framework

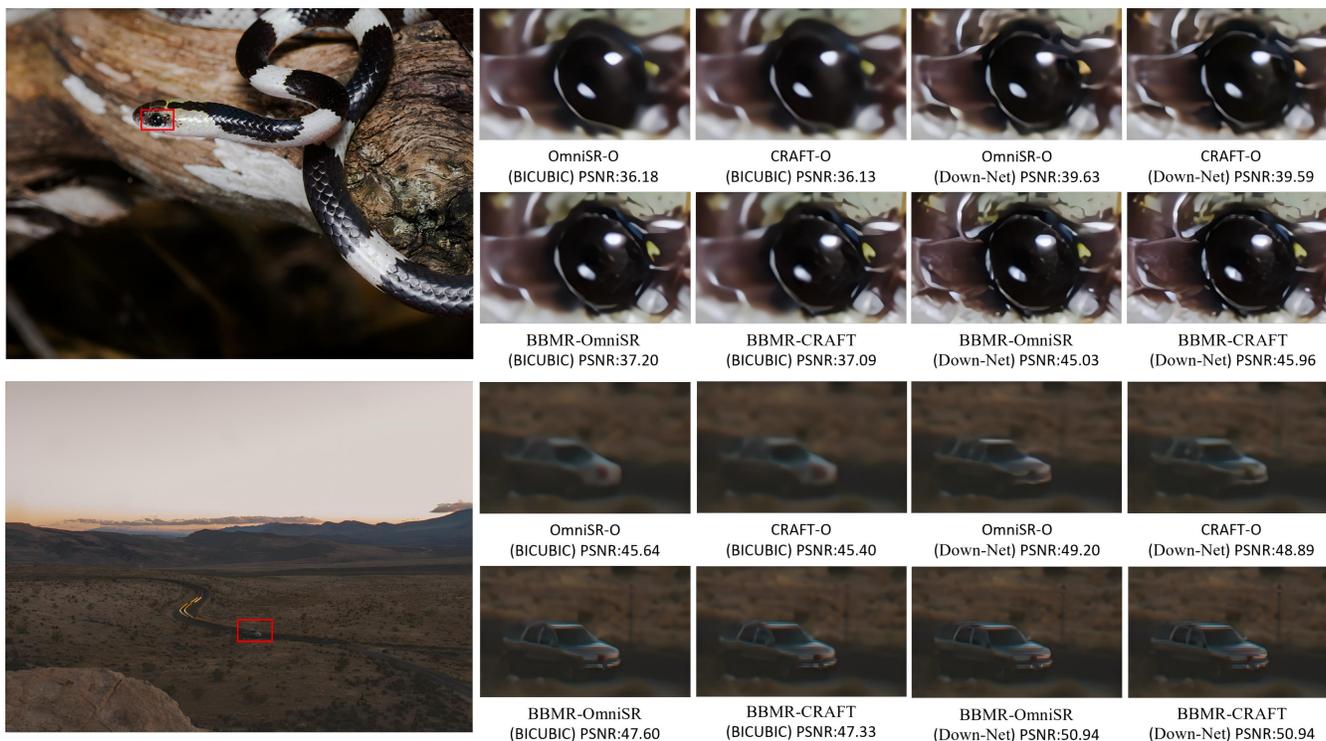


Figure 4: Comparison of visual quality between the BBMR method and traditional image rescaling method using different downscaling and upscaling approaches.

SR method	Test2k	FLOPs	Test4k	FLOPs
BBMR-OmniSR w/o light	<b>37.78</b>	1258.53G(100%)	<b>34.30</b>	1694.85G(100%)
BBMR-OmniSR w/ light	37.57	<b>324.01G(26%)</b>	34.11	<b>436.34(26%)</b>
BBMR-CRAFT w/o light	<b>37.53</b>	1256.80G(100%)	<b>34.05</b>	1692.52G(100%)
BBMR-CRAFT w/ light	37.39	<b>324.01G(26%)</b>	33.93	<b>436.34G(26%)</b>

Table 4: Comparison of the computational load of the Downscaling Module using a very lightweight super-resolution model and the original super-resolution model. FLOPs only computes the Downscaling Module. The best results are remarked in bold font.

demonstrates significant improvements on the Test2k and Test4k datasets compared to the original network method. The PSNR value is improved by approximately 1.5dB on average. Notably, when using the neural network downscaling method, the PSNR value can be enhanced by up to 1.97dB compared to the original network method. Furthermore, the substantial improvement in NIQE and PI scores suggests that our BBMR framework can enhance the overall visual quality of image super-resolution.

### Visual Quality Comparison Of BBMR

Fig. 4 shows the visual quality comparison of the BBMR framework using different image rescaling methods compared to the original network methods. It is evident that for primary subjects like cars or animals, BBMR assigns them a lower scaling rate, resulting in much clearer images.

### Comparison Of The Calculation Amount Of Upscaling Module

As shown in Tab. 3, our BBMR method achieves a significant improvement in quality while only increasing the computational load in the Upscaling Module by 3% compared to the original network method, which is almost negligible. BBMR ensures that the overall scaling rate size of the LR image remains unchanged, which indicates that our strategy can significantly enhance image super-resolution quality while keeping the transmission data size and the computational load in the Upscaling Module almost constant.

### The Effect Of Very Lightweight Super-Resolution Model

We used the very lightweight super-resolution models to assist in allocating the scaling rate of image sub-blocks instead of using the super-resolution model employed on the Upscaling Module. As shown in Tab. 4, using a very lightweight



Figure 5: Visual quality comparison of block effects using bicubic downscaling. The boundary of the image block is located at the cross in the middle of the image. The first row shows the upscaling results of OmniSR, while the second row shows the upscaling results of CRAFT.

super-resolution model in the Downscaling Module reduces the computation by approximately three-quarters compared to the previous method, with a PSNR decrease ranging from 0.21db to 0.12db. This is very useful under conditions of limited server-side computational capacity.

### Ablation Study

Case	Block	BMD	Joint	DB	PSNR	FLOPs
1	×	×	×	×	33.15	327.87G
2	✓	×	×	×	33.08	327.65G
3	✓	✓	×	×	34.26	328.30G
4	✓	✓	✓	×	34.32	330.53G
5	✓	✓	✓	✓	34.30	337.22G

Table 5: Ablation study of the proposed BBMR on Test4K dataset with OmniSR upscaling and BICUBIC downscaling for 4× SR. Block: image super-resolution with average block. BMD: using the block-based multi-scale downscaling strategy. Joint: Using the JointSR method. DB: Using the JointSR method with deblock branch. FLOPs only computes the Upscaling Module.

**The role of block-based multi-scale downscaling strategy** As shown in Tab. 5, simply dividing the image into blocks and using the same scaling rate will degrade the quality of the super-resolved image. This is because each image block loses information at the edges and creates severe block artifacts. In contrast, using our block-based multi-scale downscaling strategy shows significant improvement on the Test4k dataset, indicating that our block-based

multi-scale downscaling strategy can reduce the quality loss caused by blocking and enhance the overall quality of super-resolved images.

**The role of deblock branch In JointSR** As shown in Tab. 5, although the PSNR of the JointSR method w/o deblock branch is 0.02db higher on Test4k compared to w/ deblock branch, the block artifacts w/o deblock branch are very obvious, as seen in Fig. 5. In contrast, the block artifacts are almost completely eliminated after adding the deblock branch.

**The role of JointSR** JointSR is our proposed method that integrates super-resolution with deblocking at the feature level to reduce computational cost. As shown in Tab. 5, the JointSR w/ deblock branch increases the PSNR by 0.04db on the Test4k dataset and JointSR w/o deblock branch increases the PSNR by 0.06db with almost unchanged computation. Furthermore, as illustrated in Fig. 5, it is evident that JointSR can effectively remove blocking artifacts, no matter which SR model is used.

### Conclusion

This paper proposes a Block-Based Multi-Scale Image Rescaling (BBMR) framework. Additionally, within the Upscaling Module, we propose a joint super-resolution method (JointSR) to remove image blocking artifacts. The key idea is to selectively allocate the scaling rate for each image block based on the difficulty of its super-resolution, while keeping the overall scaling rate constant. Extensive experiments show that the framework improves the PSNR by 1.06 to 1.96 dB on the Test2K and Test4K test sets.

## Acknowledgments

This work was supported in part by the National Science Foundation of China (62172153) and Hunan Provincial Key Research and Development Program of China (2024AQ2032).

## References

- Blau, Y.; Mechrez, R.; Timofte, R.; Michaeli, T.; and Zelnik-Manor, L. 2018. The 2018 PIRM challenge on perceptual image super-resolution. In *Proceedings of the European conference on computer vision (ECCV) workshops*, 0–0.
- Chen, X.; Wang, X.; Zhou, J.; Qiao, Y.; and Dong, C. 2023a. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 22367–22377.
- Chen, Z.; Zhang, Y.; Gu, J.; Kong, L.; Yang, X.; and Yu, F. 2023b. Dual aggregation transformer for image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, 12312–12321.
- Fan, Z.-E.; Lian, F.; and Quan, J.-N. 2022. Global sensing and measurements reuse for image compressed sensing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8954–8963.
- Gao, S.; and Gruev, V. 2011. Bilinear and bicubic interpolation methods for division of focal plane polarimeters. *Optics express*, 19(27): 26161–26173.
- Gao, S.; Liu, X.; Zeng, B.; Xu, S.; Li, Y.; Luo, X.; Liu, J.; Zhen, X.; and Zhang, B. 2023. Implicit diffusion models for continuous super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10021–10030.
- Gu, S.; Lugmayr, A.; Danelljan, M.; Fritsche, M.; Lamm, J.; and Timofte, R. 2019. Div8k: Diverse 8k resolution image dataset. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, 3512–3516. IEEE.
- Kim, H.; Choi, M.; Lim, B.; and Lee, K. M. 2018. Task-aware image downscaling. In *Proceedings of the European conference on computer vision (ECCV)*, 399–414.
- Kong, X.; Zhao, H.; Qiao, Y.; and 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*, 12016–12025.
- Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4681–4690.
- Li, A.; Zhang, L.; Liu, Y.; and Zhu, C. 2023a. Feature modulation transformer: Cross-refinement of global representation via high-frequency prior for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12514–12524.
- Li, G.; Ji, J.; Qin, M.; Niu, W.; Ren, B.; Afghah, F.; Guo, L.; and Ma, X. 2023b. Towards high-quality and efficient video super-resolution via spatial-temporal data overfitting. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 10259–10269. IEEE.
- Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; and Timofte, R. 2021. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1833–1844.
- Luo, X.; Ai, Z.; Liang, Q.; Liu, D.; Xie, Y.; Qu, Y.; and Fu, Y. 2024. AdaFormer: Efficient Transformer with Adaptive Token Sparsification for Image Super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 4009–4016.
- Mittal, A.; Soundararajan, R.; and Bovik, A. C. 2012. Making a “completely blind” image quality analyzer. *IEEE Signal processing letters*, 20(3): 209–212.
- Sun, W.; and Chen, Z. 2020. Learned image downscaling for upscaling using content adaptive resampler. *IEEE Transactions on Image Processing*, 29: 4027–4040.
- Timofte, R.; Agustsson, E.; Van Gool, L.; Yang, M.-H.; and Zhang, L. 2017. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 114–125.
- Wang, H.; Chen, X.; Ni, B.; Liu, Y.; and Liu, J. 2023. Omni aggregation networks for lightweight image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22378–22387.
- Wang, Y.; Liu, Y.; Zhao, S.; Li, J.; and Zhang, L. 2024. CAMixerSR: Only Details Need More” Attention”. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 25837–25846.
- Xiao, M.; Zheng, S.; Liu, C.; Lin, Z.; and Liu, T.-Y. 2023. Invertible rescaling network and its extensions. *International Journal of Computer Vision*, 131(1): 134–159.
- Xu, B.; Guo, Y.; Jiang, L.; Yu, M.; and Chen, J. 2023. Downscaled representation matters: Improving image rescaling with collaborative downscaled images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 12237–12247.
- Yeo, H.; Chong, C. J.; Jung, Y.; Ye, J.; and Han, D. 2020. Nemo: enabling neural-enhanced video streaming on commodity mobile devices. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 1–14.
- Yeo, H.; Jung, Y.; Kim, J.; Shin, J.; and Han, D. 2018. Neural adaptive content-aware internet video delivery. In *13th USENIX Symposium on Operating Systems Design and Implementation (OSDI 18)*, 645–661.
- Yu, Q.; Li, Q.; He, R.; Tyson, G.; Shi, W.; Lv, J.; Yuan, Z.; Zhang, P.; Lan, Y.; and Li, Z. 2023. BiSR: Bidirectionally Optimized Super-Resolution for Mobile Video Streaming. In *Proceedings of the ACM Web Conference 2023*, 3121–3131.
- Yue, Z.; Wang, J.; and Loy, C. C. 2024. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in Neural Information Processing Systems*, 36.

Zhang, T.; Kasichainula, K.; Zhuo, Y.; Li, B.; Seo, J.-S.; and Cao, Y. 2024. Transformer-Based Selective Super-resolution for Efficient Image Refinement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 7305–7313.

Zhang, X.; and Wu, X. 2023. Dual-layer image compression via adaptive downsampling and spatially varying upconversion. *arXiv preprint arXiv:2302.06096*.

Zhou, S.; He, Y.; Liu, Y.; Li, C.; and Zhang, J. 2020. Multi-channel deep networks for block-based image compressive sensing. *IEEE Transactions on Multimedia*, 23: 2627–2640.