2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Dual Path Attention and Re-parameterization Network for efficient image super-resolution

Yao Li (✉)[1][0009-0001-0146-3467] and Junjie Huang[1] and Ka Chen (✉)[2] and Guoqiang Zhao[1] and Aodie Cui[1] and Yongzhuo Zhu[1] and Hanyang Pan[1] and Chuang Li [1]

[1] Wuhan Textile University, Wuhan, 430200, Hubei, China
[2] Guangxi Zhongjin Lingnan Mining Co., Ltd, Laibin, 546100, Guangxi, China

Abstract. Efficient deep learning methods have achieved significant performance in single image super-resolution. Recent research on efficient super-resolution has mainly focused on reducing the number of parameters and computational complexity through various network de- signs, enabling models to be better deployed on resource constrained devices. In this work, we propose a novel and effective super-resolution model based on attention mechanism and structural re-parameterization, called Dual Path Attention and Re-parameterization Network (DPARN), which uses a hybrid attention mechanism to balance the running speed and reconstruction quality of the model. Specifically, we utilize grouped convolution to introduce both parameter free and enhanced spatial attention to improve the feature extraction capability of student networks. Meanwhile, we adopted a novel lightweight network training strategy that first uses knowledge distillation for initial training, during which structured knowledge from the teacher network is transmitted to the student network. Then, multiple loss functions are combined to fine- tuning the student network, in order to preserve high-frequency details and avoid excessive smoothing caused by pixel loss. Finally, extensive experiments conducted on four benchmark datasets demonstrated the effectiveness and efficiency of proposed DPARN. Our method achieves PSNR/SSIM performance comparable to state-of-the-art efficient super- resolution models, with faster inference speed and fewer network parameters.

Keywords: Super-resolution, Balance, Attention, Re-parameterization.

1 Introduction

Single image super-resolution (SISR) is a mature task in low-level computer vision, aimed at reconstructing high resolution (HR) images from a single low resolution (LR) image. This task has broad applicability in enhancing image quality in various fields [1,2,3,4,6,7], such as in the field of intelligent mining, image super resolution mainly solves the problems of low image quality, blurry details, and difficult extraction of effective information in coal mine images[51]. The emergence of deep learning has driven significant progress in this field [8,9,10,11,12,13,14,16,17]. The latest research in super-resolution tasks is largely driven by attention mechanisms. Many state-of-the-art networks adopt it and even use larger visual transformers (ViT) as model

architectures [18,19,20,21,22,31,23,24]. These networks emphasize long-range dependencies between key features and image blocks through attention mechanisms, capturing broader contextual information to ensure continuity of details and accuracy of edge textures. However, the computational requirements of attention mechanisms (involving complex network structures and a large number of additional parameters) lead to challenges such as large model size and slow inference speed. These challenges limit the applicability of these models and hinder their use in efficient and high-speed computing scenarios, such as SISR tasks on resource constrained mobile devices.

Many existing efficient super-resolution (ESR) techniques have achieved notable success in enhancing model efficiency. Certain models primarily concentrate on minimizing FLOPs and parameters, accomplishing this through techniques like Group Convolution and Depth-wise Separable Convolution [25,26]. Nevertheless, merely decreasing FLOPs or parameters occasionally fails to notably boost the model's inference speed and may even compromise its accuracy. Other models [27] diminish the size of model parameters by sharing feature information and pruning non-attention branches. However, despite their complexity, these models still encompass numerous parameters, leading to prolonged runtimes. Preserving a straightforward network topology is essential to guarantee rapid inference speeds. However, conventional attention mechanisms frequently result in more intricate network architectures. To address this issue, inspired by SPAN [42], we introduced a parameter free attention mechanism to balance model performance and speed. At the same time, to preserve the advantages of both attention mechanisms, we designed a dual path structure using feature split operations, which can also reduce the number of parameters in the model.

Our proposed method introduces two different attention mechanisms through a dual-path structure, ensuring the running speed of the model and solving the time-consuming problem caused by traditional methods and the performance challenge brought by parameter free theory. At the same time, we use re-parameterization and knowledge distillation techniques to improve the performance. Pixel loss forces the generated image to align with the real image at the pixel level, but tends to output averaged results, resulting in high-frequency detail loss, blurry or lack of sharpness in the generated image. To solve this problem, we use multiple loss functions to fine-tuning the lightweight model, including balancing robustness and optimization stability against outliers with Charbonnier loss, perceptual loss making the generated image closer to the real image in texture, structure, and semantics, edge loss can constrain the differences between the generated image and the real image in the edge region. By enhancing the sharpness, coherence, and positional accuracy of the edges to improve the quality of high-frequency detail restoration.

In this article, our contributions can be summarized as follows:

Design a novel dual-path structure, which uses Enhanced Spatial Attention and Swift Parameter-free Attention to balance the speed and accuracy of the model.

Utilizing knowledge distillation and re-parameterization to optimize the student networks, the performance can be improved without increasing the computational complexity.

Combining multiple loss functions to compensate for the high-frequency information mislaying caused by pixel loss, making the generated super-resolution image more in line with human senses.

2 Related Work

2.1 Efficient Image Super-resolution

In the past few years, deep neural networks (DNNs) have demonstrated outstanding capabilities in improving SISR performance. The groundbreaking work is SRCNN [28], which applies bicubic down-sampling to HR images to construct data pairs and uses a simple Convolutional Neural Network (CNN) to learn end-to-end mapping from LR to HR images. Subsequently, a large number of CNN based methods have been proposed to achieve better performance [29,30,31,28,32]. For example, Kim et al. [12] proposed a 20 layers network with residual learning, which inspired the development of SISR deeper and wider networks. EDSR [27] follows the idea of residual learning, modifying residual blocks by removing batch normalization layers to construct a very deep and wide network. In addition, some studies use advanced losses such as VGG loss [24], perceptual loss [33], and GAN loss [34] to learn real image details. Recently, Transformer based super-resolution methods [13,41] have become increasingly popular, achieving high performance. However, most of these methods require a large amount of computing resources and have a large number of parameters, FLOP, and inference time, which is not conducive to practical deployment and application in edge devices.

Efficient image super-resolution aims to reduce the computational workload and parameter count of SR networks, while achieving faster inference time and maintaining high performance. In the deployment of SR models in the real world, the computing power of deployed devices is usually limited, such as edge devices. In this case, the efficiency of the SR network becomes an important aspect. In order to meet the growing demand for deploying SR models with limited computing resources, many works have refocused their attention on efficient image SR techniques [21,18]. At the same time, many competitions, such as NTIRE [37] and AIM [36], have launched efficient image super-resolution entries to promote the development of related research. In recent related studies, CARN [8] proposed local and global level linkage mechanisms to implement lightweight SR networks. IMDN [11] designed an information multiple distillation network by constructing cascaded information multi-distillation blocks to extract hierarchical features. The following RFDN [14] work further improved the network by introducing feature distillation blocks, which use 1×1 convolutional layers to achieve dimensional changes. Based on RFDN, RLFN [20] studied its speed bottleneck and improved its speed by removing layered distillation connections. In addition, RLFN proposes a feature extractor to extract more edge and texture information. With these advancements, they achieved first place in the NTIRE 2022 Efficient Super Resolution Challenge [35].

2.2 Attention Mechanism

The attention mechanism is widely used in computer vision tasks. By dynamically re-weighting features, computing resources are directed to the most prominent parts of the input, thereby improving the efficiency and effectiveness of various tasks [44,45,46,47,48]. Attention-based super-resolution networks generally require larger receptive fields to capture local and global information in order to improve performance, however, using parameterized attention maps may slow down inference. For ESR tasks, the application of lightweight attention mechanisms plays an important role in improving model performance without significantly increasing complexity [15]. Inspired by SPAN [42], attention maps can be generated without additional training and parameters while still having a positive impact on performance. The key to this lightweight attention method is to maximize the representation power of the super-resolution network within a limited model budget. This method makes up for the speed disadvantage of parametric attention. By combining both mechanisms, we can develop a dual-path architecture that balances speed and performance. Our method combines the advantages of parameterized and parameter-free attention, and uses grouped convolution and re-parameterization to further reduce the model size while ensuring model performance. This not only reduces computational complexity but also enhances its ability to localize weak objects, which is crucial for improving super-resolution technology.

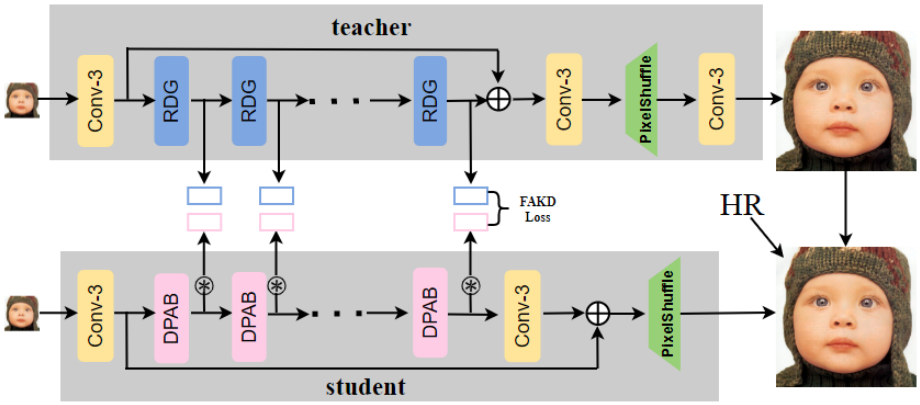

Fig. 1. The training pipeline of our method.

3 Method

In this section, we will first introduce our proposed Dual-path Attention Re-parameter Network (DPARN) in Section 3.1. In Section 3.2, we describe a feature extractor designed by combining two attention mechanisms, called Dual-path Attention Block (DPAB). In Section 3.3, we designed a structure that is friendly to image super-resolution and can improve model performance without increasing computational costs, named Re-parameterization Residual Block (ReRB). In Section 3.4, we use multiple

loss functions to jointly fine-tune the student network, which can retain more high-frequency information of the generated image.

3.1 Network Architecture

The network architecture of our proposed DPARN is shown in Fig. 1. Our DPARN mainly consists of two parts: a large teacher network and a lightweight student network. The teacher network first conduct LR feature extraction implemented by single 3×3 convolution. Then, the key component of our teacher network utilizes multiple stacked Residual-Dense Group (RDG) blocks from the DRCT [49] for deep feature extraction. The final pixel reconstruction module consists of two learnable layers and a non-parametric operation (sub-pixel convolution) for saving parameters as much as possible. Given an input LR image I^{LR} the corresponding super-resolution image I_t^{SR} can be generated by:

$$I_t^{SR} = N_T(I^{LR}) \tag{1}$$

where $N_T(\bullet)$ is our teacher network, the loss function of it can be expressed by:

$$L_T = \| I_t^{SR} - I^{HR} \|_1 \tag{2}$$

where I^{HR} indicates the target HR image and $\|\|_1$ is l_1 norm. After the teacher network training converges, save its weights as part of the next stage of the lightweight network training. In the first module of our proposed student network, we use one 3×3 convolution to extract the shallow features:

$$F_{shallow} = m_{conv}^1(I^{LR}) \tag{3}$$

where $m_{conv}^1(\bullet)$ denotes the first 3×3 convolution operation and $F_{shallow}$ is the extracted feature maps. Then we use multiple DPABs for deep feature extraction. This stage can be expressed by:

$$F_n = m_{DPAB}^n(m_{DPAB}^{n-1}(\ldots m_{DPAB}^0(F_{shallow})\ldots)) \tag{4}$$

where m_{DPAB}^n denotes the n-th DPAB function, and F_n is the n-th output feature maps. Last, the reconstruction module aims to generate the final output I_S^{SR}:

$$I_S^{SR} = m_{rec}(m_{conv}^2(F_n) + F_{shallow}) \tag{5}$$

where $m_{rec}(\bullet)$ denotes sub-pixel operation and $m_{conv}^2(\bullet)$ represents the second 3×3 convolution. When training the student network, we use the I_s^{SR} and HR image for supervised learning, we compute the following losses:

$$L_{SL} = \alpha \| I_s^{SR} - I_t^{SR} \|_1 + \beta \| I_s^{SR} - I^{HR} \|_1 \tag{6}$$

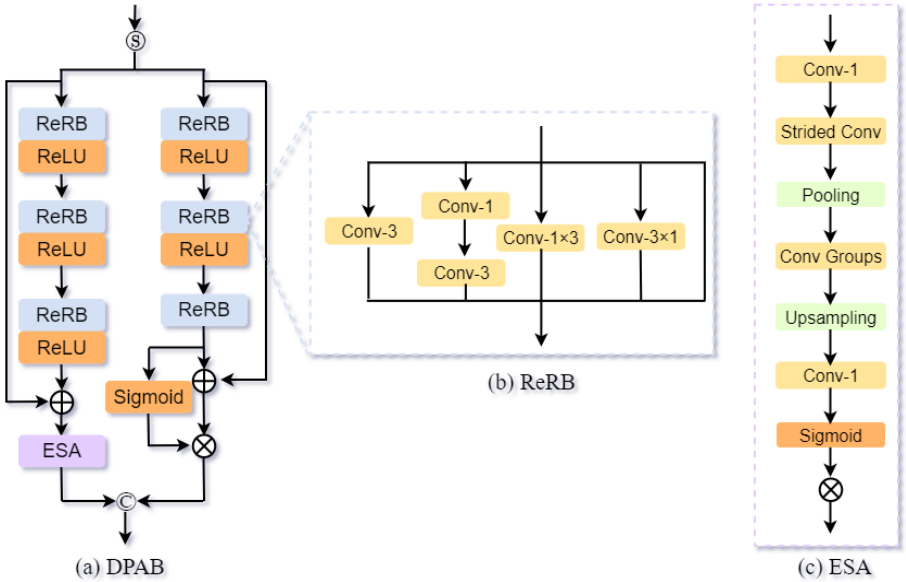

Fig. 2. (a) Structure of DPAB. (b) The structure of ReRB. (c) The structure of ESA

where α and β are penalty coefficients to balance different aspect of loss. In addition to knowledge distillation, feature distillation is also a commonly used method to improve model performance in the field of image super-resolution. Thus, we introduce Feature Affinity-based Distillation (FAKD) [5] as another loss function for student network training, it can be expressed by:

$$L_{fa} = \frac{1}{|M|}\sum_{i=1}^{\hat{i}} \| M_i^T - \uparrow (M_i^S) \|_1 \tag{7}$$

where M_i^T and M_i^S represent the affinity matrices of the teacher and student network extracted from the feature map of the i-th layer, $\hat{i}$ is the number of layers that we choose to extract. $|M|$ represents the number of elements in the affinity matrix, $\uparrow$ denotes using 1×1 convolution to expand the channel of features which extracted from student network. The final loss during training the student network is:

$$L_S = L_{SL} + \gamma L_{fa} \tag{8}$$

where γ is weight of L_{fa}. Using this loss, the student networks can better learn the feature distribution of the teacher, thereby reducing overfitting of training data.

3.2 Attention Mechanism

In this section, we introduced the Dual Path Attention Module (DPAB), which can significantly reduce inference time while maintaining model performance. As shown in Fig. 2a, our proposed DPAB abandons the idea of a single attention mechanism and instead uses grouped convolution to introduce both parameter free and enhance spatial attention to balance speed and effectiveness. Both the left and right branches of DPAB use several stacked convolutions and ReLU activation function for local feature extraction. Differently, in the left structure, the features generated by the third ReLU is added to the left input and then processed by ESA. In the right structure, the last convolution output is fed into the Sigmoid activation function to generate attention feature maps, and is added to the right input to generate fused features. The fused features are multiplied with the attention feature map to obtain the right final features. Finally, concatenate the outputs on the left and right sides to obtain the final output of the entire module. Given the input Fin, the structure can be described as:

$$
\begin{aligned}
&F_{left}^{in}, F_{right}^{in} = Split(F_{in}) \\
&F_{left}^{out} = E(CR_3^{left}(CR_2^{left}\left(CR_1^{left}(F_{left}^{in})\right)) + F_{left}^{in}) \\
&F_{right}^{3} = CR_3^{right}(CR_2^{right}(CR_1^{right}(F_{right}^{in}))) \\
&F_{right}^{out} = \sigma(F_{right}^{3}) \otimes (F_{right}^{3} + F_{right}^{in}) \\
&F_{out} = Concat(F_{left}^{out}, F_{right}^{out})
\end{aligned}
\tag{9}
$$

where $Split$ denotes channel split layer of DPAB, F_{left}^{in} and F_{right}^{in} represent the input features of the left and right branches, CR_j^{left} and CR_j^{right} indicate the j -th CONV+RELU layer of the left and right, E is Enhanced Spatial Attention, F_{left}^{out} and F_{right}^{out} represent the out features of the left and right, CR_3^{right} is the third convolution layer of right, σ is Sigmoid activation function, $\otimes$ is element-wise product, $Concat$ is concatenate layer.

3.3 Re-parameterization Residual Block

To further improve the performance of our student network, we used structural re-parameterization to enhance its feature extraction capability. Re-parameterization has achieved good results in super-resolution tasks and can improve the model's expressiveness without increasing complexity [17]. Therefore, we designed a novel module suitable for image super-resolution, namely Re-parameterization Residual Block (ReRB), which can effectively extract edge features in the horizontal and vertical directions of the image. As shown in the Fig. 2b, the first component is a 3×3 convolution, which is used to ensure the baseline performance of the model. The second component is composed of a 1×1 and a 3×3 convolution, where 1×1 is used for channel expansion because wider features can improve the model's expressiveness. 3×3 is

used for channel compression, converting the channel to the original input number. The third component is a 1×3 convolution, which is used to extract horizontal edges in the image. The fourth component is a 3×1 convolution, which can detect vertical textures in the image. In order to solve the gradient dissipation problem of deep networks, we added residual connection to the structure. The mathematical expression of the entire module is as follow:

$$\tilde{F}_{out} = C_{3\times3}(\tilde{F}_{in}) + C_{3\times3}(C_{1\times1}(\tilde{F}_{in})) + C_{1\times3}(\tilde{F}_{in}) + C_{3\times1}(\tilde{F}_{in}) + \tilde{F}_{in} \tag{10}$$

where $C_{i\times j}$ denotes a convolution with kernel size (i, j), $\tilde{F}_{in}$ and $\tilde{F}_{in}$ are input and output feature of ReRB. In the inference stage, the second component can be merged into one single normal convolution. According to [53], the first, third and fourth component also can be incorporated into a standard 3×3 convolution. Finally, the two parallel convolutions are transformed into one after re-parameterization.

3.4 Multiple Loss Fine-tuning Strategy

L_1 loss focuses solely on pixel-level absolute differences, without considering human visual perception. As a result, it often struggles to effectively restore textures, edges, and high-frequency details, leading to noticeable information loss in fine structures. The generated textures tend to appear overly smooth, making the results less visually satisfying. In order to address this issue, some approaches combine perceptual loss with L1 loss during training. While this improves perceptual quality, it still lacks sufficient texture-related information. Inspired by MPRNet [52], we introduce edge loss to compensate for this limitation. By integrating Charbonnier loss, perceptual loss, and edge loss, we fine-tuned the student network through joint optimization. The fine-tuning loss function is formulated as follows:

$$\begin{aligned}
L_{ft} &= L_{char}(I_s^{SR}, I^{HR}) + \lambda_1 L_{percep}(I_s^{SR}, I^{HR}) + \lambda_2 L_{edge}(I_s^{SR}, I^{HR}) \\
L_{char} &= \sqrt{\| I_s^{SR} - I^{HR} \|^2 + \varepsilon^2} \\
L_{percep} &= \frac{1}{C_j H_j W_j} \| \Phi_j(I_s^{SR}) - \Phi_j(I^{HR}) \|_2^2 \\
L_{edge} &= \sqrt{\| \Delta(I_s^{SR}) - \Delta(I^{HR}) \|^2 + \varepsilon^2}
\end{aligned} \tag{11}$$

where λ_1 and λ_2 are balance factor of three loss functions, which are set to 0.005 and 0.05. ε denotes a constant set to 0.001. $\Phi_j(x)$ is the output feature maps of j-th layer of network Φ when processing the image x, $C_j H_j W_j$ are channel, height and width of $\Phi_j(x)$, Δ denotes the Laplacian operator. $\|\|_2^2$ means the Euclidean distance between two feature representations.

4 Experiments

4.1 Setup and Details

We adopt the widely used high-quality (2K resolution) DIV2K dataset, which contains 800 training samples, following some previous studies [20,40]. We test the performance of our method on five benchmark datasets: Set5, Set14, BSD100, and Urban100.

Table 1. Comparison with state-of-the-art methods on benchmark datasets. The best performance in red colors. The second-best results are marked in blue. MAF means MAFFSRN and Shuffle denotes ShuffleMixer

Scale	Model	Params	Runtime	Set5	Set14	B100	U100
		K	ms	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
×2	SRCNN	24	6.92	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946
	VDSR	666	35.37	37.53/0.9587	33.05/0.9127	31.90/0.8960	30.77/0.9141
	LapSRN	251	53.98	37.52/0.9591	32.99/0.9124	31.80/0.8952	30.41/0.9103
	CARN	1592	159.10	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256
	IMDN	694	77.34	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283
	RFDN	534	74.51	38.05/0.9606	33.68/0.9184	32.16/0.8994	32.12/0.9278
	MAF	402	152.91	37.97/0.9603	33.49/0.9170	32.14/0.8994	31.96/0.9268
	ECBSR	596	39.96	37.90/0.9615	33.34/0.9178	32.10/0.9018	31.71/0.9250
	RLFN	527	60.39	38.07/0.9607	33.72/0.9187	32.22/0.9000	32.33/0.9299
	Shuffle	411	144.22	38.01/0.9606	33.63/0.9180	32.17/0.8995	31.89/0.9257
	SAFMN	228	118.07	38.00/0.9605	33.54/0.9177	32.16/0.8995	31.84/0.9256
	SPAN	481	50.39	38.08/0.9608	33.71/0.9183	32.22/0.9002	32.24/0.9294
	Ours	462	47.45	38.06/0.9603	33.72/0.9184	33.20/0.9001	32.25/0.9296
×4	SRCNN	57	1.9	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221
	VDSR	666	8.95	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524
	LapSRN	502	66.81	31.54/0.8852	28.09/0.7700	27.32/0.7275	25.21/0.7562
	CARN	1592	39.96	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837
	IMDN	715	20.56	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838
	RFDN	550	20.40	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858
	MAF	441	39.69	32.18/0.8948	28.58/0.7812	27.57/0.7361	26.04/0.7848
	ECBSR	603	10.21	31.92/0.8946	28.34/0.7817	27.48/0.7393	25.81/0.7773
	RLFN	543	16.41	32.24/0.8952	28.62/0.7813	27.60/0.7364	26.17/0.7877
	Shuffle	411	144.22	32.21/0.8953	28.66/0.7827	27.61/0.7366	26.08/0.7835
	SAFMN	240	72.06	32.18/0.8948	28.60/0.7813	27.58/0.7359	25.97/0.7809
	SPAN	498	13.67	32.20/0.8953	28.66/0.7834	27.62/0.7374	26.18/0.7879
	Ours	482	12.32	32.21/0.8952	28.65/0.7829	27.63/0.7376	26.18/0.7877

We evaluate our method and the comparison methods on the Y channel of YCbCr space using two common metrics, namely Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). In addition, to verify the efficiency of the method, we count the number of parameters and inference time of the network structure of each method in a standard way as validation metrics, which are statistically obtained in the same computing environment.

Four DPAB modules with 52 channeled feature maps are deployed for training procedure. All experiments were completed on NVIDIA2080Ti. In the training phase, we use random flipping, rotation augmentation and select Adam as the optimizer. When

training the teacher network, we set the initial learning rate to 1e-4, halve it every 100,000 iterations, and then use L1 loss for supervision. When training the

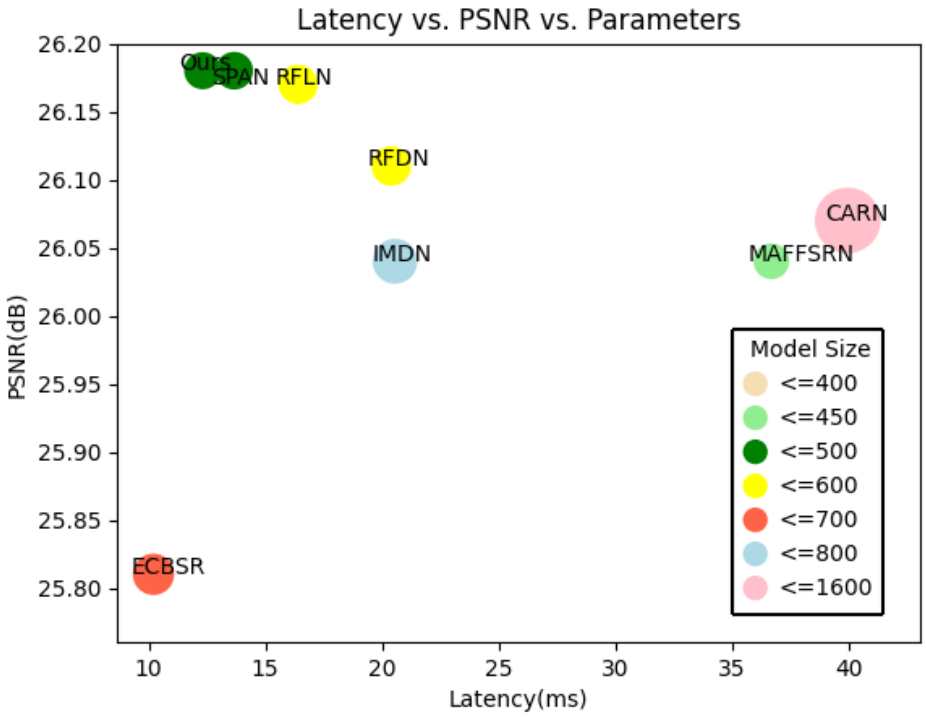

Fig. 3. Latency, PSNR and complexity of model comparison on Urban100 dataset in $\times 4$ scale factor task.

student network using the progressive strategy, the initial learning rate is 2e-5 and halved every 20,000 iterations. In this process, the training patch size is gradually increased to improve the performance, and is selected from [64, 128, 256, 384]. In the fine-tuning process, the initial learning rate is 1e-5 and halved every 40,000 iterations, fine-tuning is performed using 384×384 patches as input.

4.2 Quantitative Results

In this study, we scaled up the DPARN model by $2\times$ and $4\times$ in various benchmark tests, and compared its detailed test results with the current state-of-the-art and efficient super-resolution models [11,12,14,16,17,20,21,25,38,39,42,54]. The detailed results refer to Table 1. In multiple benchmark tests, DPARN has shown better performance than most models in terms of PSNR and SSIM, especially in terms of inference time. Compared with SPAN, DPARN can achieve very similar PSNR and SSIM performance indicators with faster inference speed. As shown in Fig. 3, by visualizing the relationship between image quality, inference time, and model size, we observed that under

equivalent PSNR, inference speed of DPARN was significantly faster than other models; When the model parameter counts are similar, it not only performs better, but also runs faster. Therefore, DPARN achieves the best balance in imaging quality, parameter quantity, and inference speed.

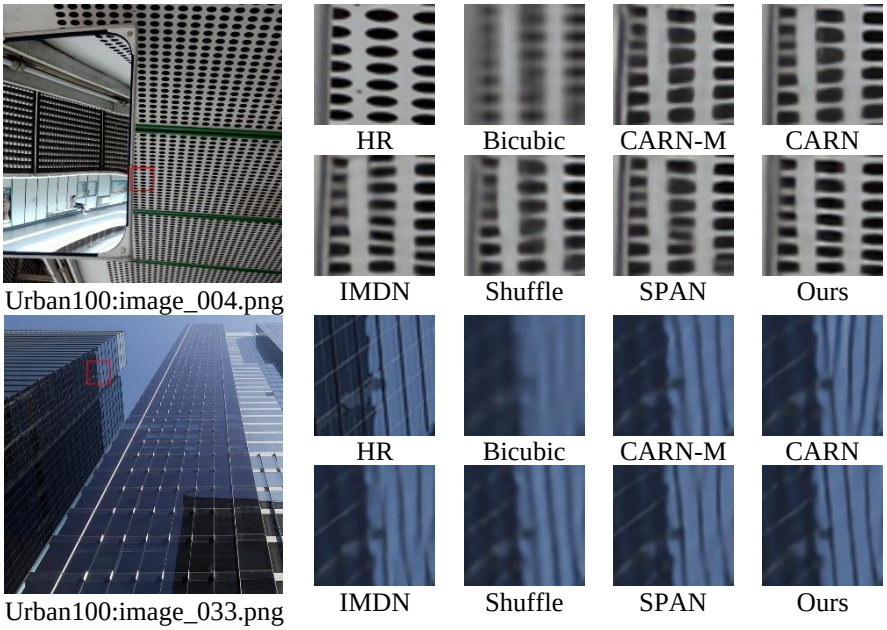

Fig. 4. Visual comparison of the results of ours and other methods on the validation set of Urban100 for ×4 upscaling task. Shuffle means ShuffleMixer method

4.3 Quanlitative Results

As shown in Fig. 4, we compared our method with some recent efficient super-resolution methods. From the figure, it can be seen that although our model is small, we can still achieve good super-resolution performance. Compared with other larger methods, the images after super-resolution have clearer edges. This is because we used multiple loss functions to fine tune the student network, which makes our method more inclined to retain high-frequency objects during the training process.

Table 2. Effect of using different degrees of attention on PSNR and SSIM, esa means for use only Enhance Spatial Attention, spa denotes Swift Parameter-free Attention alone

Model	Runtime ms	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	U100 PSNR/SSIM
DPAB_esa	15.8	32.23/0.8952	28.67/0.7830	27.64/0.7378	26.19/0.7878
DPAB_spa	11.9	32.19/0.8949	28.63/0.7825	27.59/0.7372	26.16/0.7874
DPAB (ours)	12.3	32.21/0.8952	28.65/0.7829	27.63/0.7376	26.18/0.7877

4.4 Ablation Study

Effectiveness of different attention. The results in Table 2 show the advantages of our dual-path method. Compared with the network that directly adopts a single attention, our method has a good balance between the speed and accuracy of model inference. Compared with only using enhanced spatial attention, our method is faster, but the accuracy is only slightly reduced. Compared with only using swift parameter-free attention, our method has a decrease in accuracy, but the speed is only slightly increased. However, compared to the slight decrease in speed, the significant improvement in PSNR and SSIM performance indicators is worthwhile.

Table 3. Effect of using different degrees of distillation on PSNR and SSIM. Teacher means teacher network, Shallow means the features of the shallow blocks, and Deep means the features of the deep blocks.

Teacher	Shallow	Deep	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	U100 PSNR/SSIM
			32.15/0.8944	28.58/0.7823	27.58/0.7369	26.13/0.7869
✓			32.17/0.8948	28.60/0.7824	27.60/0.7371	26.14/0.7873
✓	✓		32.17/0.8949	28.61/0.7826	27.61/0.7373	26.16/0.7874
✓		✓	32.19/0.8950	28.62/0.7828	27.63/0.7374	26.17/0.7876
✓	✓	✓	32.21/0.8952	28.65/0.7829	27.63/0.7376	26.18/0.7877

Effectiveness of different distillation. The results in Table 3 show the advantages of our multi-stage feature distillation. Compared to directly training a small model end-to-end, our method can significantly improve model performance, benefit from the powerful representation of the teacher's model capabilities. Meanwhile, our student model will not incur additional computational costs. From the results in the table, we found that using only deep features is better than using only shallow features, and the best effect is achieved when both are used simultaneously.

Table 4. Performance comparison of using different loss functions evaluated on four benchmark datasets. CL denotes chain loss, PL means perceptual loss and EL denotes edge loss.

CL	PL	EL	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	U100 PSNR/SSIM
			32.18/0.8947	28.61/0.7826	27.58/0.7371	26.15/0.7874
✓			32.19/0.8949	28.63/0.7826	27.59/0.7372	26.16/0.7875
✓	✓		32.20/0.8950	28.64/0.7828	27.60/0.7375	26.17/0.7875
✓		✓	32.18/0.8949	28.63/0.7827	27.61/0.7376	26.16/0.7876
✓	✓	✓	32.21/0.8952	28.65/0.7829	27.63/0.7376	26.18/0.7877

Effectiveness of different loss. We show the impact of different loss function combinations on PSNR and SSIM during the fine-tuning process in Table 4. In previous work, there have been no examples of training solely using perceptual loss or edge loss, so we will not make a comparison here. From the table, it can be seen that compared to initial training, using only chain loss for fine-tuning can achieve improvement on

PSNR. By combining two loss functions for fine-tuning, the chain loss and perceptual loss showed the most significant improvement on PSNR. But overall, the maximum PSNR can only be achieved when all three loss functions are simultaneously applied to fine-tuning..

Effectiveness of re-parameterization. As Table 5 showed, we compared the impact of using re-parameterization on performance, and the experimental results showed that the use of re-parameterization technology can comprehensively improve the performance of the model, verifying the effectiveness of the ReRB module.

Table 5. Comparison of our model with re-parameterization and no_rep without using the re-parameterization technology for training process

Model	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	U100 PSNR/SSIM
no_rep	32.16/0.8948	28.62/0.7825	27.58/0.7372	26.15/0.7873
with_rep	32.21/0.8952	28.65/0.7829	27.63/0.7376	26.18/0.7877

Generalization. We also investigate the generalization of our proposed multiple loss fine-tuning strategy. We apply this strategy to IMDN [11] and the quantitative comparison results are shown in Table 6, indicating that our proposed method has generality and can be applied to other existing SISR models.

Table 6. Effect of multiple loss for ×4 SR. IMDN_ML employs multiple loss in fine-tuning.

Model	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	U100 PSNR/SSIM
IMDN	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838
IMDN_ML	32.22/0.8950	28.62/0.7813	27.58/0.7355	26.07/0.7839

5 Conclusion

This article proposes a dual path attention fusion method. Specifically, we use grouped convolution to adopt Enhanced Spatial and Swift Parameter-free Attention mechanism, and employ lightweight model training strategies such as structural re-parameterization, knowledge distillation, and multi loss joint finetuning to further improve the performance of the network. Our experiments have demonstrated the effectiveness of our method and achieved advanced performance in terms of time consumption and model size on common open-source datasets. This makes our proposed method very suitable for real-world applications, especially in resource constrained scenarios such as mobile devices. Future research may apply the dual path attention mechanism to other computer vision tasks and further optimize the network to improve efficiency

6 References

1. C. H. Genitha and K. Vani, “Super resolution mapping of satellite images using hopfield neural networks,” in IEEE, 2010, pp. 114–118.Author, F., Author, S.: Title of a proceedings paper. In: Editor, F., Editor, S. (eds.) CONFERENCE 2016, LNCS, vol. 9999, pp. 1–13. Springer, Heidelberg (2016)
2. W. Liu, D. Lin, and X. Tang, “Hallucinating faces: TensorPatch super-resolution and coupled residue compensation,” in IEEE Computer Society Conference on Computer Vision & Pattern Recognition, 2005.Author, F.: Contribution title. In: 9th International Proceedings on Proceedings, pp. 1–2. Publisher, Location (2010)
3. R. C. Patel and M. V. Joshi, “Super-resolution of hyperspectral images using com pressive sensing based approach,” 2012.
4. R. R. Peeters, P. Kornprobst, M. Nikolova, S. Sunaert, T. Viéville, G. Malandain, R. Deriche, O. D. Faugeras, M. K. P. Ng, and P. V. Hecke, “The use of super resolution techniques to reduce slice thickness in functional mri,” Wiley Subscription Services, Inc., A Wiley Company, no. 3, 2004.
5. Z. He, T. Dai, J. Lu, Y. Jiang, and S.-T. Xia, “Fakd: Feature-affinity based knowledge distillation for efficient image super-resolution,” in 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020, pp. 518–522.
6. A. Singh and J. Singh, “Super resolution applications in modern digital image processing,” International Journal of Computer Applications, vol. 150, no. 2, pp. 6–8, 2016.
7. Z. Yuan, J. Wu, S. I. Kamata, A. Ahrary, and P. Yan, “Fingerprint image enhancement by super resolution with early stopping,” IEEE.
8. N. Ahn, B. Kang, and K. A. Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” 2018.
9. G. Rui, S. Xiao-Ping, and J. Dian-Kun, “Learning a deep convolutional network for image super-resolution reconstruction,” Journal of Engineering of Heilongjiang University, 2018.
10. C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” Springer, Cham, 2016.
11. Z. Hui, X. Gao, Y. Yang, and X. Wang, “Lightweight image super-resolution with information multi-distillation network,” ACM, 2019.
12. J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” IEEE, 2016.
13. J. Liang, J. Cao, G. Sun, K. Zhang, and R. Timofte, “Swinir: Image restoration using swin transformer,” IEEE, 2021.
14. J. Liu, J. Tang, and G. Wu, Residual Feature Distillation Network for Lightweight Image Super-Resolution. Computer Vision – ECCV 2020 Workshops, Glasgow, UK, August 23–28, 2020, Proceedings, Part III, 2021.
15. J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, “Residual feature aggregation network for image super-resolution,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
16. L. Sun, J. Pan, and J. Tang, “Shufflemixer: An efficient convnet for image super resolution,” Advances in Neural Information Processing Systems, vol. 35, pp. 17 314–17 326, 2022.
17. X. Zhang, H. Zeng, and L. Zhang, “Edge-oriented convolution block for real-time super resolution on mobile devices,” in Proceedings of the 29th ACM International Conference on Multimedia, 2021, pp. 4034–4043.
18. J. Cao, Q. Wang, Y. Xian, Y. Li, B. Ni, Z. Pi, K. Zhang, Y. Zhang, R. Timofte, and L. Van Gool, “Ciaosr: Continuous implicit attention-in-attention network for arbitrary-scale image

super-resolution," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1796–1807.
19. H. Choi, J. Lee, and J. Yang, "N-gram in swin transformers for efficient lightweight image super-resolution," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 2071–2081.
20. F. Kong, M. Li, S. Liu, D. Liu, J. He, Y. Bai, F. Chen, and L. Fu, "Residual local feature network for efficient super-resolution," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 766–776.
21. A. Muqeet, J. Hwang, S. Yang, J. Kang, Y. Kim, and S.-H. Bae, "Multi-attention based ultra lightweight image super-resolution," in Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16. Springer, 2020, pp. 103–118.
22. H. Wang, X. Chen, B. Ni, Y. Liu, and J. Liu, "Omni aggregation networks for lightweight image super-resolution," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 22 378–22 387.
23. Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3147–3155.
24. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.
25. W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 624–632.
26. W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 1874–1883.
27. B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 136–144.
28. Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 2472–2481.
29. M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 1664–1673.
30. Z. Luo, L. Yu, X. Mo, Y. Li, L. Jia, H. Fan, J. Sun, and S. Liu, "Ebsr: Feature enhanced burst super-resolution with deformable alignment," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 471–478.
31. Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution us ing very deep residual channel attention networks," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 286–301.
32. Z. Luo, Y. Li, S. Cheng, L. Yu, Q. Wu, Z. Wen, H. Fan, J. Sun, and S. Liu, "Bsrt: Improving burst super-resolution with swin transformer and flow-guided deformable alignment," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 998–1008.
33. J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer, 2016, pp. 694–711.

34. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Communications of the ACM, vol. 63, no. 11, pp. 139–144, 2020.
35. F. S. Khan and S. Khan, "Ntire 2022 challenge on efficient super-resolution: Methods and results," in Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2022, pp. 1061–1101.
36. A. Ignatov, R. Timofte, M. Denna, A. Younes, G. Gankhuyag, J. Huh, M. K. Kim, K. Yoon, H.-C. Moon, S. Lee et al., "Efficient and accurate quantized image super resolution on mobile npus, mobile ai & aim 2022 challenge: report," in European conference on computer vision. Springer, 2022, pp. 92–129.
37. Y. Li, Y. Zhang, R. Timofte, L. Van Gool, L. Yu, Y. Li, X. Li, T. Jiang, Q. Wu, M. Han et al., "Ntire 2023 challenge on efficient super-resolution: Methods and results," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1922–1960.
38. C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13. Springer, 2014, pp. 184–199.
39. L. Sun, J. Dong, J. Tang, and J. Pan, "Spatially-adaptive feature modulation for efficient image super-resolution," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 13 190–13 199.
40. L. Yu, X. Li, Y. Li, T. Jiang, Q. Wu, H. Fan, and S. Liu, "Dipnet: Efficiency distillation and iterative pruning for image super-resolution," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1692–1701.
41. X. Chen, X. Wang, J. Zhou, Y. Qiao, and C. Dong, "Activating more pixels in image super-resolution transformer," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 22 367–22 377.
42. C. Wan, H. Yu, Z. Li, Y. Chen, Y. Zou, Y. Liu, X. Yin, and K. Zuo, "Swift parameter-free attention network for efficient super-resolution," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 6246–6256.
43. Y. Li, K. Zhang, J. Liang, J. Cao, C. Liu, R. Gong, Y. Zhang, H. Tang, Y. Liu, D. Demandolx et al., "Lsdir: A large scale dataset for image restoration," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1775–1787.
44. C. Cao, Q. Dong, and Y. Fu, "Learning prior feature and attention enhanced image inpainting," in European Conference on Computer Vision, 2022.
45. S. Gao, C. Zhou, C. Ma, X. Wang, and J. Yuan, "Aiatrack: Attention in attention for transformer visual tracking," Springer, Cham, 2022.
46. Y. Jang, W. Shin, J. Kim, S. Woo, and S. H. Bae, "Glamd: Global and local attention mask distillation for object detectors," in European Conference on Computer Vision, 2022.
47. T.-D. Truong, Q.-H. Bui, C. N. Duong, H.-S. Seo, S. L. Phung, X. Li, and K. Luu, "Direcformer: A directed attention in transformer approach to robust action recognition," in Proceedings of the IEEE/CVF conference on computer vision and pat tern recognition, 2022, pp. 20 030–20 040.
48. H. Lee, H. Choi, K. Sohn, and D. Min, "Knn local attention for image restoration," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 2139–2149.
49. C.-C. Hsu, C.-M. Lee, and Y.-S. Chou, "Drct: Saving image super-resolution away from information bottleneck," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2024, pp. 6133–6142.

50. S. Ioffe, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” arXiv preprint arXiv:1502.03167, 2015.
51. Z. Su, Y. Wang, X. Ma, M. Sun, D. Cheng, C. Li, and H. Jiang, “Single image super-resolution via wide-activation feature distillation network,” Sensors, vol. 24, no. 14, p. 4597, 2024.
52. S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao,“Multi-stage progressive image restoration,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 14 821–14 831.
53. X. Ding, Y. Guo, G. Ding, and J. Han, “Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks,” in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1911–1920.
54. N. Ahn, B. Kang, and K.-A. Sohn, “Fast, accurate, and lightweight super-resolution with cascading residual network,” in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 252–268.