Computing and Informatics, Vol. 43, 2024, 393–413, doi: 10.31577/cai_2024_2_393

EFFECTIVE LIGHTWEIGHT DUAL-PATH SHIFT COMPENSATION NETWORK FOR IMAGE SUPER-RESOLUTION

Yu Yang, Pan Wang, Yajuan Wu*

School of Computer Science
China West Normal University
Nanchong, China
e-mail: 838759385@qq.com, scwuyajuan@163.com

Abstract. In this paper, we propose a lightweight dual-path convolutional neural network for image super-resolution (SR). We introduce shift convolution and propose a shift-channel attention (shift-ca) mechanism to build an effective network. Shift-ca produces an attentional map with a larger field of view, and its formulation is similar to channel attention and spatial attention. In addition, we propose the Local Shift-Channel Attention Feature Extraction (LCFE) module as the main part of the Dual Path Shift Attention Block (DPSAB). Using the dual-path structure allows us to reduce the network depth and retain more original features for the subsequent up-sampling compensation operation. In the final HR reconstruction module, we combine the nearest neighbor upsampling layer, convolutional layer, and activation layer to form the compensated nearest neighbor upsampling module (C-NUM) to improve the reconstruction quality with a small parameter cost. Our final model is the Dual Path Shift Attention Network (DPSAN), and it achieves similar performance to the lightweight network WMRN (36.38 % for WMRN) with only 195 k parameters. Applying our module to the EDSR-baseline also yielded good results. The effectiveness of each proposed component was verified by an ablation study.

Keywords: Deep learning, super resolution, shift convolution, dual-path, compensation operation

Mathematics Subject Classification 2010: 68U10

* Corresponding author

1 INTRODUCTION

Image super-resolution reconstruction (SRR) is a computer vision technique that uses single or multiple low-resolution images (LR) and algorithms to generate high-resolution images (HR) without changing imaging hardware conditions. It has many applications, such as biometric recognition, image analysis, and monitoring. However, SRR is an ill-posed problem due to the countless corresponding high-resolution images for a low-resolution image. To address this issue, researchers have proposed deep learning-based image SRR algorithms.

In 2014, Dong et al. proposed SRCNN [1], the first super-resolution convolutional neural network which used only three convolutional layers to extract internal image features and significantly improves reconstruction performance. In 2016, Shi et al. built on the basic model of SRCNN [1] to propose ESPCN [2] which extracted features directly from the low-resolution image size and improved efficiency. In 2017, Lim et al. introduced enhanced deep residual networks (EDSR) [3] by optimizing the residual structure, achieving a network depth of 160 layers but with a parameter size of 43 MB. In 2018, Zhang et al. proposed a very deep residual channel attention network (RCAN) [4] with an attention mechanism, reducing the parameter size to 16 MB. However, these methods increased network depth to achieve satisfactory results which increased the computational cost and made them unsuitable for portable devices such as mobile phones and cameras.

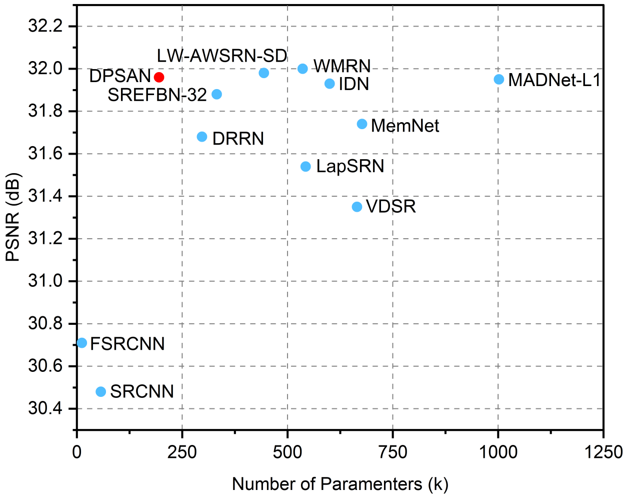


Figure 1. The performance and parameter comparison of our DPSAN with other lightweight networks on set5 dataset for upscaling factor × 4

To reduce computational burden and memory consumption, various methods have been proposed for image super-resolution. Cascading residual network (CARN) [5] introduced a cascading network architecture but with poor hypersegmentation performance. Information distillation network (IDN) [6] and its successor, information multi-distillation network (IMDN) [7], improved performance through information distillation. Residual feature distillation network (RFDN) [8] proposed a lighter residual distillation network. However, these methods are not lightweight enough, and performance can be further improved. In [9], the authors propose a parameter-free, FLOP-free shift operation as an alternative to spatial convolution. This kind of convolution achieves good results in the non-image super-resolution domain. Meanwhile, DANv2 [10] and DCLS [11] have been proposed by using different two-way networks. One path is introduced as an additional path to the features of the estimation kernel to achieve excellent results in the field of blind super-resolution.

The aim of this article is to enhance the existing dual-path attention module DPCB by proposing a new dual-path shift attention block (DPSAB) that utilizes displacement convolution to reduce model complexity. The DPSAB consists of two shift-conv operations and a special attention mechanism module (LFE) to extract local structural information effectively. Unlike previous approaches that use the dual-path structure to connect the stretched kernel with blurred features, we propose to use basic blurry features as an additional path to compensate for artefacts and errors introduced by the estimated kernel, resulting in a lightweight super-resolution model called DPSAN. The DPSAN comprises multiple groups of DPSAB that receive auxiliary and original features, and the feature is directly magnified, convolved, and activated to obtain weights, which are then enhanced with the up-sampled LR image for feature refinement before being combined with the main feature. Our proposed method achieves a good balance between model complexity and performance, resulting in a better-performing lightweight dual-path shift attention network (DPSAN) for fast and accurate image SR. Our contribution can be summarized as follows:

1. We propose a simple and basic attention scheme shift-ca based on shift-conv and design, a compensated upsampling dual-path network.
2. We have integrated shift-ca with dual-path conditional block to propose DPSAB, which is efficient and constructive.
3. We use the mechanism of upsampling followed by dual-path compensation in the high-resolution reconstruction process, which greatly reduces the parameters, and few people study the subsequent operation of upsampling.

2 RELATED WORK

2.1 CNN-Based SR Methods

Deep neural networks have greatly improved the results of image reconstruction [12, 4, 13], but their high computational cost and a large number of parameters limit their practical application. To address this issue, some researchers [14] have used the original low-resolution (LR) image as input instead of the upsampled image, and others have proposed techniques such as group convolution [15], depth-wise separable convolution [16], and self-attention convolution [17] to accelerate deep models [5, 18]. These techniques have also been used in super-resolution (SR) models with promising results. For example, CARN-M [5] used group convolution to achieve efficient SR, while IMDN [7] extracted hierarchical features step-by-step and aggregated them using 1×1 convolution. In this work, we introduce a novel approach using a shift-conv scheme and a dual-path attention network in our DPSAN model to achieve efficient and concise SR.

2.2 Attention Scheme

The attention mechanism has become a popular technique in computer vision tasks such as object detection, classification, and image segmentation. Channel attention was first introduced by Hu et al. [19]. In image classification tasks, which enhances network representational ability by modelling the relationship between channels. Non-local methods [20], such as those used in RCAN [4] and Liu et al. [21], capture long-range dependencies by calculating the response of pixel positions as a weighted sum of features of all pixel positions in the image. Other attention mechanisms, such as second-order attention [22] and cross-scale non-local attention [23], have also been proposed for image super-resolution. However, most of these methods require complex attention modules to achieve better performance. In contrast, our approach aims to learn effective attention with lower computational complexity and generate 3D attention features with 1×1 convolution layers. We incorporate attention mechanisms into our proposed framework to improve high-level feature representation.

2.3 Reconstruction Methods in SR Networks

In the early stages of super-resolution networks [24, 25], interpolation-based upsampling methods were used, while learning-based reconstruction methods, such as pixel-shuffle [2], were typically implemented at the end of the network. However, in recent works, interpolation-based upsampling methods have also been used at the end of the network to achieve good performance [26]. Therefore, the reconstruction module now essentially consists of upsampling (interpolation-based or learning-based) and convolution layers. In the reconstruction method of DPSAN, we use interpolation-based nearest neighbour upsampling, convolution layers, and activation functions.

Moreover, previous works have shown that the compensation of reconstructed images can be performed during the reconstruction process, but few researchers have studied it during the reconstruction stage. Therefore, in this work, we use a compensation operation similar to a dual path at the reconstruction stage to achieve better reconstruction.

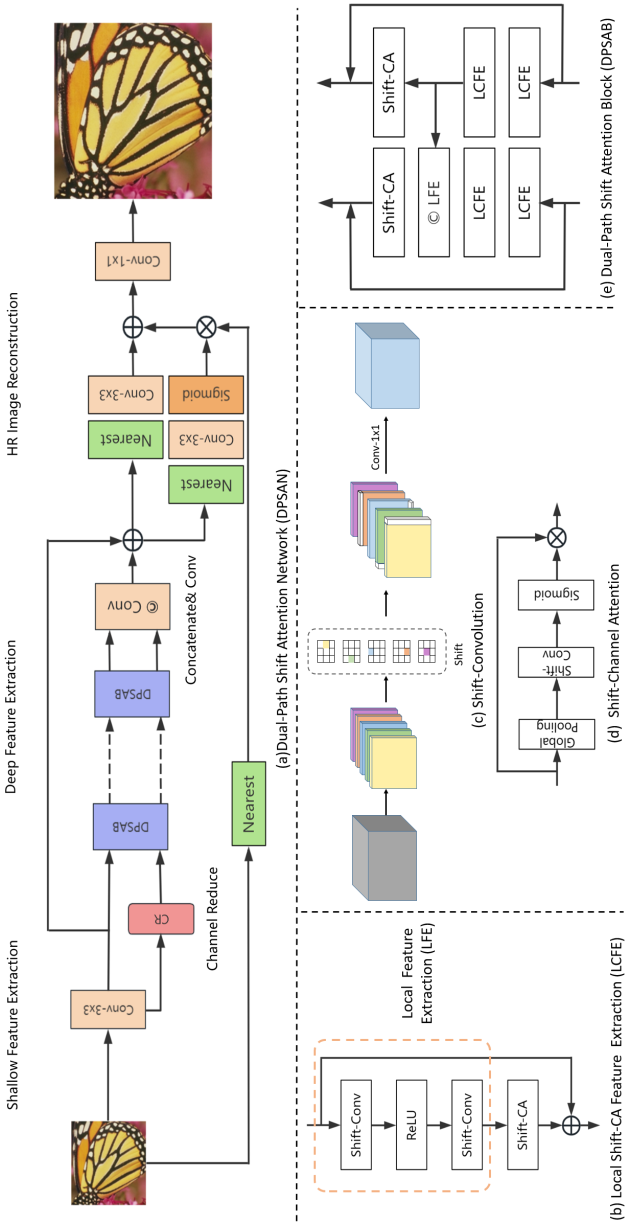


Figure 2. The network architecture of the proposed DPSAN. a) The overall pipeline of DPSAN, which contains the shallow range feature extraction module, depth feature extraction module and high score image reconstruction module. b) The LCFE architecture. c) Description of shift-conv, which consists of a shift operation and a 1 × 1 convolution. d) Description of shift-ca. e) Description of DPSAB.

3 METHOD

In this section, we first introduce the flow of the dual-path shift attention network (DPSAN) for SR tasks and then discuss in detail its key components, namely the dual-path shift attention block (DPSAB) and the compensated nearest upsampling module(C-NUM).

3.1 Network Architecture

As shown in Figure 2, our DPSAN network architecture consists of three modules, namely the shallow feature extraction module, the DPSAB-based deep feature extraction module, and the HR image reconstruction module. Prior to input into the HR reconstruction module, there are multi-branch global shortcut connections from the output of shallow feature extraction module to deep feature extraction module. Specifically, for a given degraded LR image $X_l \in \mathbb{R}^{3\times H\times W}$, where H and W are the height and width of the LR image, respectively, we first apply the shallow feature extraction module denoted as $H_{SF}(\cdot)$, which contains only a 3×3 convolution and a channel halving operation, to extract the local bilateral features $X_s \in \mathbb{R}^{C\times H\times W}$and $X_{s^r} \in \mathbb{R}^{\frac{C}{2}\times H\times W}$:

$$(X_s, X_{s^r}) = H_{SF}(X_l), \tag{1}$$

where C is the channel number of the intermediate features. X_s and X_{s^r} enter the depth feature extraction module, respectively, denoted by $H_{DF}(\cdot)$ which consists of M cascaded DPSABs. That is:

$$X_d = H_{DF}(X_s, X_{s^r}), \tag{2}$$

where $X_d \in \mathbb{R}^{C\times H\times W}$ denotes the output. Using X_d and X_s as input, the HR image X_h is reconstructed as:

$$X_h = H_{RC}(X_s + X_d), \tag{3}$$

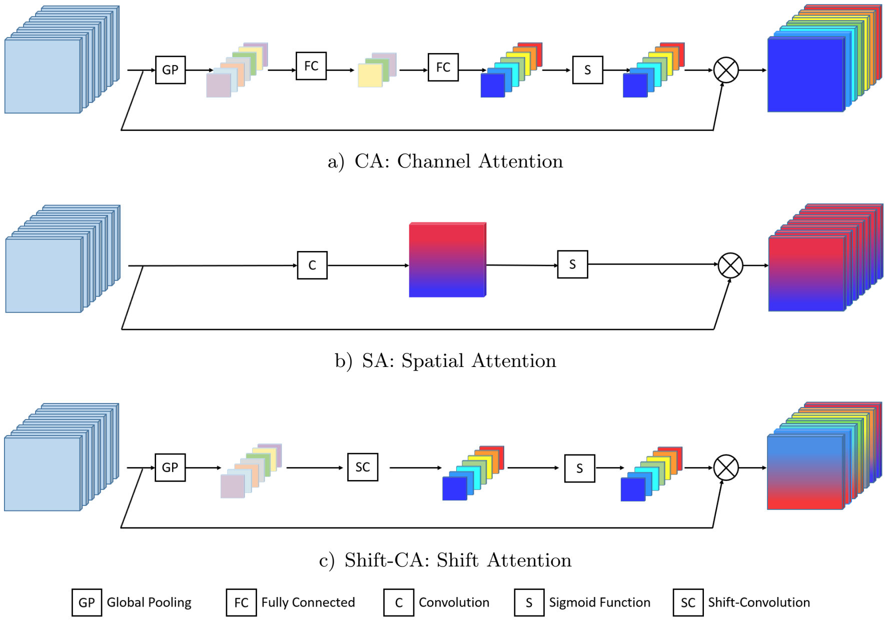
where H_{RC} is the reconstruction module. There are several design options for the reconstruction module [24, 14, 2, 3]. To achieve high efficiency, we also use a dual-path reconstruction structure to build it. DPSAN can be optimized using commonly used SR loss functions such as L_2 [25], L_1 [12], and perceptual loss [27, 28]. For simplicity, given N ground truth HR images $\{X_{t,i}\}_{i=1}^{N}$, we optimize the parameters of DPSAN by minimizing the pixel-level L_1 loss:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} ||X_{h,i} - X_{t,i}||_1. \tag{4}$$

Adam optimizer is used to optimize our DPSAN because it has good performance in low-level vision tasks.

3.2 Shift Attention Scheme

Firstly, in order to adapt to our network model, we re-examine channel attention [19] and spatial attention [29]. As shown in Figure 3, channel attention aims to obtain a one-dimensional ($C \times 1 \times 1$) attention feature vector, while spatial attention obtains a two-dimensional ($1 \times H \times W$) attention map. In contrast, our shift attention is able to generate a 3D ($C \times H \times W$) matrix as an attention feature and using shift-conv can provide a larger receptive field with almost the same computational complexity as a 1×1 convolution.



a) CA: Channel Attention

b) SA: Spatial Attention

c) Shift-CA: Shift Attention

Figure 3. Comparison of three different attention mechanisms

3.3 Deep Feature Extraction Scheme

As shown in Figure 2 a), we use five DPSABs and a final fusion convolution to construct the Deep Feature Extraction module. Unlike previous works [30, 10], for DPSAB, we use a reduced channel path as an additional path to control the parameters of our lightweight super-resolution network. This additional path compensates for the introduced artefacts and errors during network training using the original blurry features. The specific DPSAB is shown in Figure 2 e), where the left side is the main path consisting of two groups of LCFE and one group of fusion LFE that will merge the features from the compensatory path. It receives the shallow feature extraction module's feature: $X_s \in \mathbb{R}^{C\times H\times W}$. The compensatory path on the right side consists of only two groups of LCFE and receives the feature $X_{s^r} \in \mathbb{R}^{\frac{C}{2}\times H\times W}$ from the shallow feature extraction module. The specific LCFE is shown in Figure 2 b), consisting of two shift-conv and one shift-ca.

3.4 HR Image Reconstruction Module

Our reconstruction module incorporates a compensation-like mechanism, as illustrated in Figure 2 a). In previous super-resolution networks, the reconstruction module primarily comprised upsampling and convolution layers, with little attention paid to compensation mechanisms during upsampling. In our approach, the input features to the reconstruction module are split into two paths and upsampled using nearest neighbour interpolation. The compensatory upsampling path involves introducing the blurred HR image, obtained after nearest neighbour sampling, and multiplying it with the compensatory path sigmoid to generate the compensatory feature. The compensatory feature is then fused with the main feature to yield enhanced super-resolution reconstruction outcomes. We used nearest neighbour interpolation multiple times in the reconstruction module to reduce the parameter cost. Our experimental findings reveal that incorporating compensatory upsampling substantially boosts performance with minimal additional parameters.

4 EXPERIMENTAL RESULTS

In this section, we conduct extensive experiments to quantitatively and qualitatively validate the superior performance of our DPSAN for light-weight and classic SR tasks on five SR benchmark datasets. We also present comprehensive ablation studies to evaluate the design of our proposed DPSAN.

4.1 Datasets and Metrics

We use the DIV2K dataset as our training dataset. LR images are obtained by bicubic downsampling of HR images. In the testing phase, five standard benchmark datasets, Set5, Set14, B100, Urban100, and Manga109 are used for evaluation. The widely used Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) on the Y channel are used as evaluation metrics. Additionally, all MultiAdds are calculated by assuming that the resolution of the HR image is 720 p.

4.2 Implementation Details

During the training process, we use the DIV2K dataset to train our DPSAN. The HR patch size is set to 256 × 256, and the batch size was 32. We use random rotations of 90°, 180°, 270°, and horizontal flips to augment the data. We use the L1 loss function and Adam optimizer for model training. We train the model using the Adam optimizer with an initial maximum learning rate of 1e−4 and a minimum learning rate of 1e−7 for a total of 250 epochs. The learning rate is multiplied by 0.5 at the 50th, 100th, 150th, and 200th epochs. The proposed algorithm has been implemented in the PyTorch framework on a computer equipped with an NVIDIA GTX 1080Ti GPU.

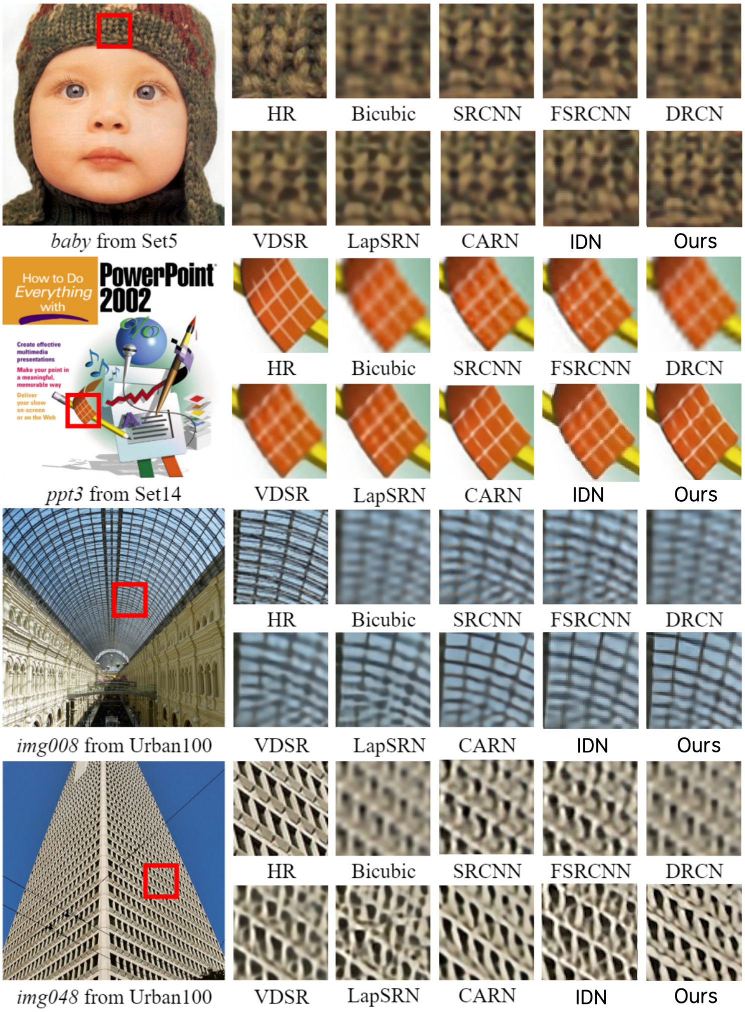


Figure 4. Visual comparison for upscaling factor × 4

Scale	Model	Params	Mult-Adds	Set5	Set14	B100	Urban100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
2	SRCNN [1]	57 k	52.7 G	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946	35.60/0.9663
	FSRCNN [14]	12 k	6.0 G	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9710
	VDSR [25]	665 k	612.6 G	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9750
	DRRN [31]	298 k	6,796.9 G	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.88/0.9749
	LapSRN [32]	251 k	29.9 G	37.52/0.9590	33.08/0.9130	31.80/0.8950	30.41/0.9100	37.27/0.9740
	MemNet [33]	677 k	623.9 G	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195	37.72/0.9740
	IDN [6]	553 k	180.9 G	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	38.01/0.9749
	SREFBN-32 [34]	310 k	–	37.80/0.9591	33.43/0.9158	32.09/0.8981	31.76/0.9242	38.27/0.9755
	LW-AWSRN-SD [35]	348 k	79.6 G	37.86/0.9600	33.41/0.9161	32.07/0.8984	31.67/0.9237	38.20/0.9762
	MADNet-L1 [36]	878 k	187.1 G	37.85/0.9600	33.38/0.9161	32.04/0.8979	31.62/0.9233	–
	WMRN [37]	452 k	103 G	37.83/0.9599	33.41/0.9162	32.08/0.8984	31.68/0.9241	38.27/0.9763
	DPSAN	**195 k**	**65.32 G**	37.87/0.9600	33.41/0.9162	32.09/0.8986	31.70/0.9244	38.25/0.9763
3	SRCNN [1]	57 k	52.7 G	32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989	30.48/0.9117
	FSRCNN [14]	12 k	5.0 G	33.16/0.9140	29.43/0.8242	28.53/0.7910	26.43/0.8080	31.10/0.9210
	VDSR [25]	665 k	612.6 G	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9340
	DRRN [31]	297 k	6,796.9 G	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.71/0.9379
	LapSRN [32]	290 k	115.0 G	33.81/0.9220	29.79/0.8325	28.82/0.7980	27.07/0.8275	32.21/0.9350
	MemNet [33]	677 k	623.9 G	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376	32.51/0.9369
	IDN [6]	588 k	60.1 G	34.14/0.9259	30.13/0.8383	28.98/0.8026	27.86/0.8463	33.11/0.9416
	SREFBN-32 [34]	319 k	–	34.14/0.9245	30.17/0.8386	28.99/0.8021	27.80/0.8444	33.09/0.9406
	LW-AWSRN-SD [35]	388 k	39.5 G	34.18/0.9273	30.21/0.8398	28.99/0.8027	27.80/0.8444	33.13/0.9416
	MADNet-L1 [36]	930 k	88.4 G	34.16/0.9253	30.21/0.8398	28.98/0.8023	27.77/0.8439	–
	WMRN [37]	556 k	57.0 G	34.11/0.9251	30.17/0.8390	28.98/0.8021	27.80/0.8448	33.07/0.9413
	DPSAN	**195 k**	**29.8 G**	34.19/0.9256	30.18/0.8393	28.99/0.8028	27.75/0.8438	32.99/0.9409

Scale	Model	Params	Mult-Adds	Set5	Set14	B100	Urban100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
4	SRCNN [1]	57 k	52.7 G	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.58/0.8555
	FSRCNN [14]	12 k	4.6 G	30.71/0.8657	27.59/0.7535	26.98/0.7150	24.62/0.7280	27.90/0.8610
	VDSR [25]	665 k	612.6 G	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8870
	DRRN [31]	297 k	6,796.9 G	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.45/0.8946
	LapSRN [32]	543 k	139.3 G	31.54/0.8852	28.09/0.7700	27.32/0.7275	25.21/0.7562	29.09/0.8900
	MemNet [33]	677 k	623.9 G	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630	29.42/0.8942
	IDN [6]	600 k	34.5 G	31.93/0.8923	28.45/0.7781	27.48/0.7326	25.81/0.7766	30.04/0.9026
	SREFBN-32 [34]	332 k	–	31.88/0.8905	28.45/0.7781	27.47/0.7322	25.74/0.7746	30.04/0.9017
	LW-AWSRN-SD [35]	444 k	25.4 G	31.98/0.8921	28.46/0.7786	27.48/0.7368	25.74/0.7746	30.09/0.9024
	MADNet-L1 [36]	1 002 k	54.1 G	31.95/0.8917	28.44/0.7780	27.47/0.7327	25.76/0.7746	–
	WMRN [37]	536 k	45.7 G	32.00/0.8952	28.47/0.7786	27.49/0.7328	25.89/0.7789	30.11/0.9040
	DPSAN	**195 k**	**17.34 G**	31.96/0.8917	28.49/0.7785	27.49/0.7329	25.89/0.7780	30.05/0.9028
8	SRCNN [1]	57 k	52.7 G	25.34/0.6471	23.86/0.5443	24.14/0.5043	21.29/0.5133	22.46/0.6606
	FSRCNN [14]	12 k	4.6 G	25.42/0.6440	23.94/0.5482	24.21/0.5112	21.32/0.5090	22.39/0.6357
	VDSR [25]	665 k	612.6 G	25.73/0.6743	23.20/0.5110	24.34/0.5169	21.48/0.5289	22.73/0.6688
	DRCN [38]	1 774 k	17,974 G	25.93/0.6743	24.25/0.5510	24.49/0.5168	21.71/0.5289	23.20/0.6686
	LapSRN [32]	813 k	–	26.15/0.7028	24.45/0.5792	24.54/0.5293	21.81/0.5555	23.39/0.7068
	MemNet [33]	677 k	623.9 G	26.16/0.7414	24.38/0.6199	24.58/0.5842	21.89/0.5825	23.56/0.7387
	MSRN [39]	6 226 k	89.6 G	26.59/0.7254	24.88/0.5961	24.70/0.5410	22.37/0.5977	24.28/0.7517
	DPSAN	**230 k**	**6.01 G**	26.69/0.7619	24.71/0.6308	24.67/0.5893	22.13/0.5982	23.98/0.7561

Table 1. Quantitative results of the advanced SR method for all amplification factors × 2, ×3, ×4 and ×8. Red/blue text: the first/second best of all methods. Overstriking: our methods.

Module Type	Params	Mult-Adds	PSNR/SSIM Set5	PSNR/SSIM Se14	PSNR/SSIM BSD100	PSNR/SSIM Urban100	PSNR/SSIM Manga109
EDSR-baseline	1 370 k	126.55 G	37.91/0.9602	33.53/0.9172	32.15/0.8995	31.99/0.9270	38.40/0.9765
EDSR-baseline-C	1 222 k	112.96 G	37.99/0.9605	33.60/0.9175	32.18/0.8999	32.08/0.9285	38.51/0.9769
RCAN	15 444 k	5 785.60 G	38.27/0.9614	34.12/0.9216	32.41/0.9027	33.34/0.9384	39.44/0.9786
RCAN-C	15 296 k	5 724.16 G	38.10/0.9609	33.67/0.9185	32.24/0.9005	33.24/0.9380	39.28/0.9754

Table 2. Results of Substitution into C-NUM in different networks. We use the PSNR values obtained on the five datasets as a scaling factor × 2. We record the results in 1×10^6 iterations. Red/green text: the rise/fall of method.

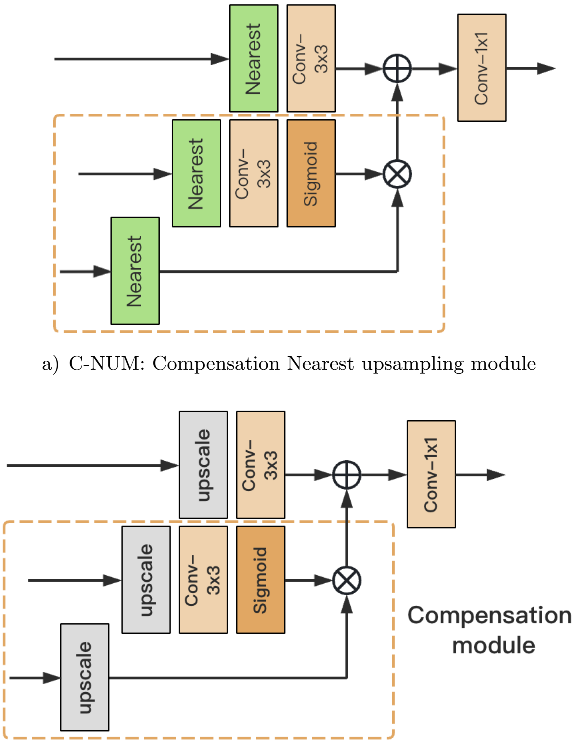
4.3 Comparison with Lightweight SR Model

We compare the proposed DPSAN with commonly used lightweight SR models for amplification factors × 2, ×3, ×4 and ×8, including SRCNN [1], FSRCNN [14], VDSR [25], LapSRN [32], DRRN [31], MemNet [33], IDN [6], SREFBN-32 [34], LW-AWSRN-SD [35], MADNet-L1 [36] and WMRN [37]. Table 1 shows the quantitative results in terms of PSNR and SSIM for the five benchmark datasets obtained by different algorithms. In addition, the number of parameters and Mult-Add of the compared models are given. From Table 1, we find that our DPSAN has less than 200 k parameters, but outperforms most state-of-the-art methods. Specifically, WMRN [37] partially achieves similar performance to ours, but with nearly 556 k parameters, which is about two and a half times more than ours. Compared to SREFBN-32 [34], we can achieve higher PSNR on most datasets. On the amplification factor × 8 we also get great performance with very small parameters.

Qualitative comparison. Then, we qualitatively compare the SR quality of different lightweight models. The ×4 SR results for the four example images are shown in Figure 6. It can be seen that our model is able to reconstruct the stripe and line patterns more accurately. For the image "ppt3", we observe that most of the compared methods produce significant artifacts and blurring effects, while our method produces more accurate lines. For the building details in "img008" and "img048", DPSAN enables a reconstruction with fewer artifacts.

Module Type	Params	Mult-Adds	PSNR	
RB	222 k	81.16 G	34.575 dB	
LFE	194 k	71.98 G	34.577 dB	(↑ 0.002 dB)
RB-CA	228 k	81.16 G	34.579 dB	(↑ 0.004 dB)
RB-CBAM	223 k	82.05 G	34.574 dB	(↓ 0.001 dB)
LCFE	195 k	65.32 G	34.604 dB	(↑ 0.029 dB)

Table 3. Comparison of the number of parameters and average values of PSNR obtained on five data sets for basic RB, LFE, RB-CA, RB-CBAM and LCFE with a magnification factor × 2. We recorded the results for 5×10^5 iterations.



a) C-NUM: Compensation Nearest upsampling module

b) C-PSUM: Compensation Pixel shuffle upsampling module

Figure 5. Comparison of using the same compensation mechanism but with different upsampling methods

4.4 Comparison of Runtime and FLOPs

In this section we also compare the more specific parameters Runtime and FLOPs. From Table 4, we can see that the FLOPs of DPSAN are only 33.4 G and the Runtime is only 37 ms, while the other networks are much larger than the parameters of our network, thus better reflecting the effectiveness of our proposed method.

4.5 Ablation Study

4.5.1 Comparison of Different Attention Schemes

To demonstrate the effectiveness of our shift-ca layer, we use DPSAN as the base network with DPSAB as the basic module and replace the LCFE module with residual blocks (RB), residual blocks with channel attention (RB-CA), and five

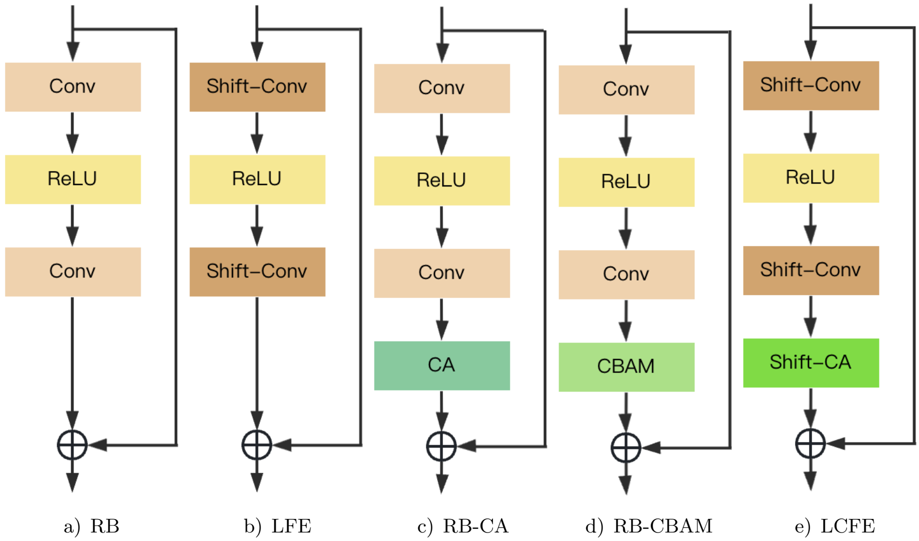


Figure 6. Comparison of five different feature extraction modules. a) RB: Basic residual block; b) LFE: Basic residual block with shift; c) RB-CA: Basic residual block with channel attention; d) RB-CBAM: Basic residual block with spatial attention and channel attention; e) LCFE: Basic residual attention block with shift attention.

residual blocks with spatial attention and channel attention (RB-CBAM). We also compare the effects of shift convolution. The results are shown in Figure 4.

In Table 3, we compare the number of parameters, Multi-Adds and PSNR performance for all methods. Note that all results are the average of PSNR calculated from 328 images on 5 benchmark datasets. It is observed that LCFE can improve 0.029 dB on average but with much fewer parameters and Multi-Adds,

Model	Params	FLOPs	Runtime	Set5 PSNR/SSIM
Bicubic	–	–	–	33.66/0.9299
FSRCNN	12 k	34 G	13 ms	37.00/0.9558
SRCNN	57 k	144 G	60 ms	36.66/0.9542
IND	553 k	393 G	138 ms	37.83/0.9600
CARN	1 592 k	503 G	199 ms	37.76/0.9590
SREFBN-32	310 k	–	–	37.80/0.9591
LW-AWSRN-SD	348 k	87 G	76 ms	37.86/0.9600
MADNet-L1	878 k	352.1 G	156 ms	37.85/0.9600
WMRN	452 k	102 G	88 ms	37.83/0.9599
DPSAN	**195 k**	**33.4 G**	**37 ms**	37.87/0.9600

Table 4. Comparison of Runtime and FLOPs for several different models on top of Set5×2

while RB-CBAM is slightly worse than RB. This indicates that relying on spatial attention and channel attention do not work well under my network model, but shift-ca feature extraction is more effective than channel attention and spatial attention.

4.5.2 Efficiency of the Compensation Upsampling Scheme

Our results are averaged over five data sets, and the validity of our method is verified by conducting four experiments on two variables. Specifically, as shown in Figure 5, we use the Pixel shuffle upsampling module (PSUM), which is used by most people, as a baseline to compare the results after removing the Nearest upsampling module (NUM) of the compensation module and replacing the upsampling part. As shown in Table 5, we find that compared to PSUM, the NUM with only the upsampling part replaced can greatly reduce the parameters, but the effect is also worse, while both C-PSUM and C-NUM can improve the PSNR. In addition, C-PSUM is 0.02 dB better than C-NUM compared to PSUM, but the parameters of C-NUM are only 72.5 % of C-PSUM. This shows that C-NUM can achieve more significant improvements compared to the conventional Pixel shuffle.

Basic Module	Params Diff		Mult-Adds		PSNR	
PSUM	248 k		31.25 G		28.684 dB	
NUM	192 k	(↓ 56 k)	17.02 G	(↓ 14.23 G)	28.682 dB	(↓ 0.002 dB)
C-PSUM	269 k	(↑ 21 k)	34.33 G	(↑ 3.08 G)	28.697 dB	(↑ 0.013 dB)
C-NUM	195 k	(↓ 53 k)	17.34 G	(↓ 13.91 G)	28.695 dB	(↑ 0.011 dB)

Table 5. Pixel shuffle upsampling module (PSUM) is used as the basic comparison model where we compare C-PSUM with C-NUM and Nearest upsampling module (NUM) with a magnification factor × 4. We recorded the results for 5×10^5 iterations.

4.5.3 Comprehensive Comparison

We selected one of the LCFE and RB-CA modules and one of the PSUM and C-NUM modules to combine to produce four results. We used RB-CA + PSUM as a control group. Table 6 shows that the combined best results were achieved using the combination of LCFE and C-NUM. The other two combinations LCFE + PSUM and RB-CA + C-NUM were combinations that changed one variable each, and they both improved the parameters and the results for the control group. All in all whatever combination was chosen showed some enhancement over the control group. It can be said that both our proposed components LCFE and C-NUM are effective.

4.5.4 Role of C-NUM in Other Models

We also investigated the effectiveness of C-NUM in networks with different model sizes. For comparison, we chose two networks, EDSR-baseline and RCAN, which

LCFE	✓	✗	✗	✓
RB-CA	✗	✓	✓	✗
PSUM	✓	✓	✗	✗
C-NUM	✗	✗	✓	✓
PSNR	34.592 dB	34.575 dB	34.579 dB	34.604 dB
Params	248 k	281 k	228 k	195 k

Table 6. Comparison of the number of parameters and average values of PSNR obtained on five data sets for any two opposing modules with a magnification factor ×2. We recorded the results for 5×10^5 iterations.

have 1 370 k and 15 592 k parameters, respectively. We then replaced the upsampling with C-NUM in each of the two networks and named them EDSR-baseline-C and RCAN-C, respectively. Since training larger networks takes more time, here we record the results for 1×10^6 iterations. As shown in Table 2, for the lighter network EDSR-baseline, replacing it with C-NUM brings an improvement in PSNR (around 0.1 dB) and a reduction in parameters. However, C-NUM seems to degrade the performance of the larger network (RCAN). For example, RCAN-C is worse than RCAN, with a drop of about 0.2 dB in PSNR for various data sets. The experimental results show that C-NUM can make the lightweight model EDSR-baseline improve the effect and reduce the parameter, but it produces a bad effect for the larger model RCAN. This may be due to the complexity of the neural network structure. Changing a module in a large network may increase the complexity of the network and make training more difficult. Large networks are already complex in themselves, and adding a module may increase the number of parameters, introduce more nonlinear relationships, and make it more difficult for the network to learn, leading to performance degradation. On the other hand, for small networks, replacing a new module may provide additional information and features that help the model learn and generalize better. A relatively small network structure may be easier to train and be able to utilize the information provided by the added module more efficiently, so adding a module to a small network may lead to performance improvement.

5 CONCLUSIONS

The present study puts forward a novel convolutional neural network designed for image super-resolution that is lightweight yet efficient. Our proposed network incorporates a new shift-attention scheme, or shift-ca, which features only a few parameters but yields improved reconstruction outcomes. We also introduce a DPSAB module based on the shift-attention scheme. To further enhance the performance of the SR, we introduce a compensation module in the reconstruction branch consisting of nearest neighbour upsampling, convolutional layers, and activation functions. The framework not only enhances SR performance at a low parameter cost but also improves the performance of other lightweight networks. The experimental results demonstrate that our final model, DPSAN, achieves comparable results to advanced lightweight networks while utilizing significantly fewer parameters.

REFERENCES

[1] Dong, C.—Loy, C. C.—He, K.—Tang, X.: Learning a Deep Convolutional Network for Image Super-Resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.): Computer Vision – ECCV 2014. Springer, Cham, Lecture Notes in Computer Science, Vol. 8692, 2014, pp. 184–199, doi: 10.1007/978-3-319-10593-2_13.

[2] Shi, W.—Caballero, J.—Huszár, F.—Totz, J.—Aitken, A. P.—Bishop, R.—Rueckert, D.—Wang, Z.: Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1874–1883, doi: 10.1109/CVPR.2016.207.

[3] Lim, B.—Son, S.—Kim, H.—Nah, S.—Mu Lee, K.: Enhanced Deep Residual Networks for Single Image Super-Resolution. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1132–1140, doi: 10.1109/CVPRW.2017.151.

[4] Zhang, Y.—Li, K.—Li, K.—Wang, L.—Zhong, B.—Fu, Y.: Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.): Computer Vision – ECCV 2018. Springer, Cham, Lecture Notes in Computer Science, Vol. 11211, 2018, pp. 294–310, doi: 10.1007/978-3-030-01234-2_18.

[5] Ahn, N.—Kang, B.—Sohn, K. A.: Fast, Accurate, and Lightweight Super-Resolution with Cascading Residual Network. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.): Computer Vision – ECCV 2018. Springer, Cham, Lecture Notes in Computer Science, Vol. 11214, 2018, pp. 256–272, doi: 10.1007/978-3-030-01249-6_16.

[6] Hui, Z.—Wang, X.—Gao, X.: Fast and Accurate Single Image Super-Resolution via Information Distillation Network. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 723–731, doi: 10.1109/CVPR.2018.00082.

[7] Hui, Z.—Gao, X.—Yang, Y.—Wang, X.: Lightweight Image Super-Resolution with Information Multi-Distillation Network. Proceedings of the 27th ACM International Conference on Multimedia (MM '19), 2019, pp. 2024–2032, doi: 10.1145/3343031.3351084.

[8] Liu, J.—Tang, J.—Wu, G.: Residual Feature Distillation Network for Lightweight Image Super-Resolution. In: Bartoli, A., Fusiello, A. (Eds.): Computer Vision – ECCV 2020 Workshops. Springer, Cham, Lecture Notes in Computer Science, Vol. 12537, 2020, pp. 41–55, doi: 10.1007/978-3-030-67070-2_2.

[9] Wu, B.—Wan, A.—Yue, X.—Jin, P.—Zhao, S.—Golmant, N.—Gholaminejad, A.—Gonzalez, J.—Keutzer, K.: Shift: A Zero FLOP, Zero Parameter Alternative to Spatial Convolutions. 2018 IEEE/CVF Conference-

ence on Computer Vision and Pattern Recognition, 2018, pp. 9127–9135, doi: 10.1109/CVPR.2018.00951.

[10] Luo, Z.—Huang, Y.—Li, S.—Wang, L.—Tan, T.: End-to-End Alternating Optimization for Blind Super Resolution. CoRR, 2021, doi: 10.48550/arXiv.2105.06878.

[11] Luo, Z.—Huang, H.—Yu, L.—Li, Y.—Fan, H.—Liu, S.: Deep Constrained Least Squares for Blind Image Super-Resolution. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 17621–17631, doi: 10.1109/CVPR52688.2022.01712.

[12] Zhang, Y.—Tian, Y.—Kong, Y.—Zhong, B.—Fu, Y.: Residual Dense Network for Image Super-Resolution. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 2472–2481, doi: 10.1109/CVPR.2018.00262.

[13] Zhang, W.—Liu, Y.—Dong, C.—Qiao, Y.: RankSRGAN: Generative Adversarial Networks with Ranker for Image Super-Resolution. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 3096–3105, doi: 10.1109/ICCV.2019.00319.

[14] Dong, C.—Loy, C. C.—Tang, X.: Accelerating the Super-Resolution Convolutional Neural Network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.): Computer Vision – ECCV 2016. Springer, Cham, Lecture Notes in Computer Science, Vol. 9906, 2016, pp. 391–407, doi: 10.1007/978-3-319-46475-6_25.

[15] He, J.—Dong, C.—Qiao, Y.: Modulating Image Restoration with Continual Levels via Adaptive Feature Modification Layers. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11048–11056, doi: 10.1109/CVPR.2019.01131.

[16] Agustsson, E.—Timofte, R.: NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 1122–1131, doi: 10.1109/CVPRW.2017.150.

[17] Liu, J. J.—Hou, Q.—Cheng, M. M.—Wang, C.—Feng, J.: Improving Convolutional Networks with Self-Calibrated Convolutions. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10093–10102, doi: 10.1109/CVPR42600.2020.01011.

[18] Wang, Z.—Chen, J.—Hoi, S. C. H.: Deep Learning for Image Super-Resolution: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43, 2021, No. 10, pp. 3365–3387, doi: 10.1109/TPAMI.2020.2982166.

[19] Hu, J.—Shen, L.—Sun, G.: Squeeze-and-Excitation Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.

[20] Wang, X.—Girshick, R.—Gupta, A.—He, K.: Non-Local Neural Networks. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7794–7803, doi: 10.1109/CVPR.2018.00813.

[21] Liu, D.—Wen, B.—Fan, Y.—Loy, C. C.—Huang, T. S.: Non-Local Recurrent Network for Image Restoration. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.): Advances in Neural Information Processing Systems 31 (NeurIPS 2018). Curran Associates, Inc., 2018, pp. 1673–1682,

doi: 10.48550/arXiv.1806.02919.

[22] Dai, T.—Cai, J.—Zhang, Y.—Xia, S. T.—Zhang, L.: Second-Order Attention Network for Single Image Super-Resolution. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11057–11066, doi: 10.1109/CVPR.2019.01132.

[23] Mei, Y.—Fan, Y.—Zhou, Y.—Huang, L.—Huang, T. S.—Shi, H.: Image Super-Resolution with Cross-Scale Non-Local Attention and Exhaustive Self-Exemplars Mining. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5689–5698, doi: 10.1109/CVPR42600.2020.00573.

[24] Dong, C.—Loy, C. C.—He, K.—Tang, X.: Image Super-Resolution Using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 38, 2016, No. 2, pp. 295–307, doi: 10.1109/TPAMI.2015.2439281.

[25] Kim, J.—Lee, J. K.—Lee, K. M.: Accurate Image Super-Resolution Using Very Deep Convolutional Networks. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1646–1654, doi: 10.1109/CVPR.2016.182.

[26] Wang, X.—Yu, K.—Wu, S.—Gu, J.—Liu, Y.—Dong, C.—Qiao, Y.—Change Loy, C.: ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In: Leal-Taixé, L., Roth, S. (Eds.): Computer Vision – ECCV 2018 Workshops. Springer, Cham, Lecture Notes in Computer Science, Vol. 11133, 2018, pp. 63–79, doi: 10.1007/978-3-030-11021-5_5.

[27] Huang, G.—Liu, Z.—Van Der Maaten, L.—Weinberger, K. Q.: Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[28] Sajjadi, M. S. M.—Schölkopf, B.—Hirsch, M.: EnhanceNet: Single Image Super-Resolution Through Automated Texture Synthesis. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4501–4510, doi: 10.1109/ICCV.2017.481.

[29] Woo, S.—Park, J.—Lee, J. Y.—Kweon, I. S.: CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.): Computer Vision – ECCV 2018. Springer, Cham, Lecture Notes in Computer Science, Vol. 11211, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.

[30] Gu, J.—Lu, H.—Zuo, W.—Dong, C.: Blind Super-Resolution with Iterative Kernel Correction. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1604–1613, doi: 10.1109/CVPR.2019.00170.

[31] Tai, Y.—Yang, J.—Liu, X.: Image Super-Resolution via Deep Recursive Residual Network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2790–2798, doi: 10.1109/CVPR.2017.298.

[32] Lai, W. S.—Huang, J. B.—Ahuja, N.—Yang, M. H.: Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5835–5843, doi: 10.1109/CVPR.2017.618.

[33] Tai, Y.—Yang, J.—Liu, X.—Xu, C.: MemNet: A Persistent Memory Network for Image Restoration. 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 4549–4557, doi: 10.1109/ICCV.2017.486.

[34] Ketsoi, V.—Raza, M.—Chen, H.—Yang, X.: SREFBN: Enhanced Feature Block Network for Single-Image Super-Resolution. IET Image Processing, Vol. 16, 2022, No. 12, pp. 3143–3154, doi: 10.1049/ipr2.12546.

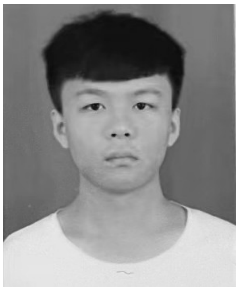
[35] Li, Z.—Wang, C.—Wang, J.—Ying, S.—Shi, J.: Lightweight Adaptive Weighted Network for Single Image Super-Resolution. Computer Vision and Image Understanding, Vol. 211, 2021, Art. No. 103254, doi: 10.1016/j.cviu.2021.103254.

[36] Lan, R.—Sun, L.—Liu, Z.—Lu, H.—Pang, C.—Luo, X.: MADNet: A Fast and Lightweight Network for Single-Image Super Resolution. IEEE Transactions on Cybernetics, Vol. 51, 2021, No. 3, pp. 1443–1453, doi: 10.1109/TCYB.2020.2970104.

[37] Sun, L.—Liu, Z.—Sun, X.—Liu, L.—Lan, R.—Luo, X.: Lightweight Image Super-Resolution via Weighted Multi-Scale Residual Network. IEEE/CAA Journal of Automatica Sinica, Vol. 8, 2021, No. 7, pp. 1271–1280, doi: 10.1109/JAS.2021.1004009.

[38] Kim, J.—Lee, J. K.—Lee, K. M.: Deeply-Recursive Convolutional Network for Image Super-Resolution. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1637–1645, doi: 10.1109/CVPR.2016.181.

[39] Li, J.—Fang, F.—Mei, K.—Zhang, G.: Multi-Scale Residual Network for Image Super-Resolution. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.): Computer Vision – ECCV 2018. Springer, Cham, Lecture Notes in Computer Science, Vol. 11212, 2018, pp. 527–542, doi: 10.1007/978-3-030-01237-3_32.



Yu Yang is a postgraduate in the School of Computer Science, at the China West Normal University. His main research interest is image super-resolution.

Pan Wang is a postgraduate in the School of Computer Science, at the China West Normal University. His main research interest is image super-resolution.

Yajuan Wu graduated from the Sichuan University and obtained her Ph.D. from the Sichuan University. She is currently Associate Professor in the School of Computer Science at China West Normal University. Her research interests are digital image processing and mathematical modeling methods.