

# DARN: Distance Attention Residual Network for Lightweight Remote-Sensing Image Superresolution

Qingjian Wang , Sen Wang , Mingfang Chen , and Yang Zhu 

**Abstract**—The application of single-image superresolution (SISR) in remote sensing is of great significance. Although the state-of-the-art convolution neural network (CNN)-based SISR methods have achieved excellent results, the large model and slow speed make it difficult to deploy in real remote sensing tasks. In this article, we propose a compact and efficient distance attention residual network (DARN) to achieve a better compromise between model accuracy and complexity. The distance attention residual connection block (DARCB), the core component of the DARN, uses multistage feature aggregation to learn more accurate feature representations. The main branch of the DARCB adopts a shallow residual block (SRB) to flexibly learn residual information to ensure the robustness of the model. We also propose a distance attention block (DAB) as a bridge between the main branch and the branch of the DARCB; the DAB can effectively alleviate the loss of detail features in the deep CNN extraction process. Experimental results on two remote sensing and five super-resolution benchmark datasets demonstrate that the DARN achieves a better compromise than existing methods in terms of performance and model complexity. In addition, the DARN achieves the optimal solution compared with the state-of-the-art lightweight remote sensing SISR method in terms of parameter amount, computation amount, and inference speed. Our code will be available at <https://github.com/candygogogogo/DARN>.

**Index Terms**—Convolution neural network, lightweight, remote sensing, single image superresolution (SISR).

## I. INTRODUCTION

SUPERRESOLUTION(SR) reconstruction refers to constructing a nonlinear mapping relationship between a pair of high-resolution and low-resolution images. Single image superresolution (SISR), as the most representative low-level vision task, has been intensively studied. The SISR algorithm can obtain high-definition images, which makes it a great contribution to various fields such as military, industry, aerospace, and remote sensing [1], [2], [3], [4]. However, the data collection in the remote sensing field is affected by long distances, wide viewing angles, and optical hardware equipment. Even with advanced collection equipment, it is difficult to obtain high-definition images that meet the needs of the mission. Therefore, the research

Manuscript received 19 August 2022; revised 27 October 2022; accepted 30 November 2022. Date of publication 8 December 2022; date of current version 27 December 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 52065035 and Grant 51965029, and in part by the Science and Technology Program of Yunnan Province under Grant 202002AC080001. (*Corresponding author: Sen Wang.*)

The authors are with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming 650500, China (e-mail: wqj3031@163.com; wangsen0401@126.com; mfchen111@sina.com; zhuyang1023@foxmail.com).

Digital Object Identifier 10.1109/JSTARS.2022.3227509

on the SISR algorithm that can recover the high-frequency information of remote sensing images is very significant.

Since the first pioneering SISR method SRCNN [5] was proposed, the end-to-end mapping of low-resolution images to high-resolution images has been incorporated into SR reconstruction methods, but the new solution inevitably leaves some flaws that must be addressed. For example, the unreasonable design of convolution kernels and nonlinear mapping results in very slow network inference. The FSRCNN [6] derived some optimization methods that depend on different algorithms based on the SRCNN framework. The ESPCN [7] improves reconstruction accuracy by replacing traditional interpolation upsampling with subpixel convolution functions. Deeper convolutional neural network (CNN) architectures [8] were shown to boost model performance. The dense residual block [9], [10] proposed by [11] can alleviate the problem of local information loss caused by long-distance residuals.

Based on predecessors' exploration of the CNN, the theory of SISR began to mature gradually. However, these methods [12], [13], [14], [15] that always try to improve the model accuracy by increasing the model capacity are difficult to show the application value in practical tasks. In particular, the large resolution characteristics of remote sensing images often lead to extremely slow inference speed of common SISR algorithms. Therefore, how to design a lightweight, efficient, and accurate SISR model has become an insurmountable problem in the field of SR.

From the perspective of trying to reduce model parameters, DRCN [16] and DRRN [17] adopted a recursive way to increase model parameter sharing, but the recursive loss requires a deeper CNN for information compensation, which makes it forced to reduce the inference speed of the model. CAUN-M [18] attempts to remove or reduce feature redundancy and speed up model inference through parameter pruning, but PSNR suffers a large loss of accuracy. Obviously, in order to obtain an effective trade-off between the accuracy, speed, and parameter quantity of the model, the expressive ability of the feature must be strengthened within the limited parameter quantity. Therefore, the information distillation method proposed by the information distillation network (IDN) [19] can achieve a moderate tradeoff by splitting feature channels to reduce feature redundancy. However, the simple channel splitting method limits the performance of the model to extract effective features due to the loss of part of the deep feature information. The heterogeneous structure adopted by LESRCNN [20] can improve the model reconstruction ability by flexibly combining low-frequency features and high-frequency features. MADNet [21] accelerates the inference of the model

through a multilateral residual module similar to inception. However, this multibranch inference solution also suffers from the disadvantage of a large proportion of model parameters. In the field of SR, FeNet [22] was inspired by the idea of the IDN channel splitting and constructed a lightweight LLB module. To achieve the purpose of enhancing the channel feature expression, the LLB module adopts the channel attention mechanism to construct the information communication of the upper and lower branches while channel splitting reduces the model parameters. However, another problem arises in the process of information distillation. That is to say when splitting channels, the side branches generated during distillation always leave the defect that deep features cannot be effectively extracted.

This article constructs a novel lightweight SR network to solve the aforementioned problems, namely distance attention residual network (DARN). The DARN improves the reconstruction ability of the network by strengthening the representation of feature channel information and introducing an effective attention module. The distance attention residual connection block (DARCB) we constructed is the core component of the DARN. The DARCB explicitly divides the input features into two branches by feature thinning convolution. Among them, the features after branch refinement are retained, and the main branch is further refined and extracted by the CNN module. The CNN used by the DARCB is a shallow residual block (SRB) based on a convolutional layer, a spanning connection, and an end activation unit. The most effective solution to enhance feature representation is to reduce channel feature redundancy and alleviate the loss of shallow features caused by deep CNNs. This is the specific reason why we construct the distance attention block (DAB). Distance attention means that the shallow features can remotely control the feature extraction of each SRB in the main branch through the means of attention. The DAB uses the shallow features of the branch as prior information, which can effectively alleviate the phenomenon of shallow information loss in the feature extraction of the main branch deep CNN, thereby enhancing the feature expression of the module. Finally, we adopt a multistage fusion mechanism to fuse the refined features of each stage to reduce the influence of redundant features of the main branch on the module output. Properly introducing attention modules [23], [24], [25] has effectively improved SR networks' performance. Therefore, we introduce an enhanced spatial attention (ESA) module to strengthen the capabilities of the model constructed in this article. As shown in Fig. 1, the proposed DARN method achieves state-of-the-art reconstruction performance when compared with existing lightweight SR networks. In particular, our model DARN-S still has a considerable competitive advantage under the premise of being sufficiently lightweight. This benefits from the powerful feature expression capability of the DARCB module proposed in this article.

The contributions of this article are as follows.

- 1) Designing the DARCB component to use multistage feature aggregation to enhance feature representation achieves superior progress over simple CNN concatenation modules.
- 2) The constructed DAB module can effectively apply shallow features to suppress the phenomenon of shallow feature loss in the process of deep CNN feature extraction.

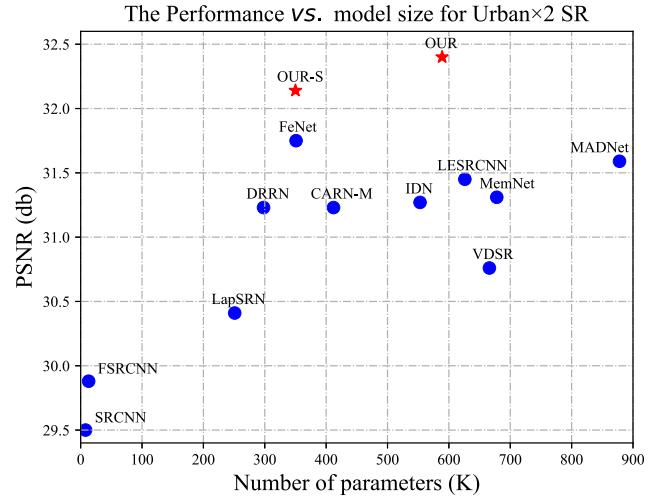


Fig. 1. Model parameters and accuracy tradeoff with the state-of-the-art lightweight methods on Urban100 for 2x SR. Our proposed DARN achieves superior performance, and our DARN-baseline also maintains competitive performance.

- 3) In this article, a lightweight image SR reconstruction model DARN is proposed, which achieves a good compromise between accuracy and efficiency.

## II. RELATED WORK

### A. Deep Network for SR

The SR task has developed rapidly since Dong et al. [5] proposed the seminal work SRCNN, which significantly outperformed the traditional methods. With the in-depth research of relevant researchers on SR tasks, the effectiveness of optimization strategies such as large models, deep convolution, and globalization of feature information on model performance has gradually been reflected. VDSR as proposed by Kim et al. [8] has achieved a dramatic improvement in SR tasks, which proves that depthwise convolution can improve the model performance. Lim et al. [12] adopt a wider model structure to increase model parameters to achieve a better performance. The EDPN [26] replicates the input image into a sequence and employs deformable convolutions to learn image internal self-similarity. Liu et al. [27] introduced window transformers [28] into the SR domain, which can strengthen the correlation of image globalization information. Chen et al. [29] combined the Swin-Transformer and channel attention mechanism, and the proposed HAT model refreshed the state-of-the-art SR performance. These methods have achieved great progress in performance, but large model parameters and high computational costs make it difficult to deploy to practical applications.

### B. Lightweight SR

The strict requirement of lightweight models for practical tasks has prompted researchers to focus on the development of more efficient SR models [1], [19], [20], [21], [22], [23], [24], [30], [31], [32]. The IDN proposed in [19] can separately extract the result of feature segmentation through two channels. The authors in [30] relies on the pyramid structure to gradually reconstruct the high-frequency residual features

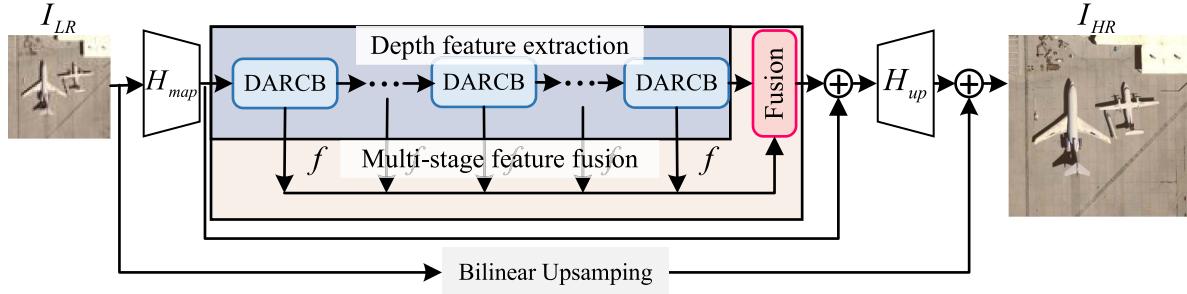


Fig. 2. Architecture of the DARN.

of the input image, and the use of deconvolution instead of the bilinear interpolation algorithm can also greatly reduce the computational complexity. The authors in [31] chose to abandon the direct learning of the mapping between a pair of high- and low-resolution images, and instead achieve accelerated inference of the model by transforming the SR task into linear regression for multiple basis filters. The pixel attention network (PAN) proposed by the authors in [32] adopts a dual-branch architectural approach, which improves the final reconstruction quality with little parameter cost. The multistage information distillation refinement structure designed in [23] can achieve multi-stage feature reuse. Li et al. [24] introduced separable convolutions to achieve more competitive performance with less parameter usage. Unlike other deep-learning-based SR models, LAPAR [31] reduces the SR task to a linear regression task with multiple base filters.

The remote sensing images collected from long distances are of poor quality. Therefore, the SR reconstruction of remote sensing images is a very meaningful work. LGCNet [33] is the first CNN-based remote sensing image SR model that utilizes local and global representations to learn image residuals between HR images and upscaled LR images. SCViT [34] proposes a spatial channel feature preservation model that considers the detailed geometric information of the high-spatial-resolution imagery. TransENet [35] employs a multiscale transformer to aggregate multidimensional spatial features while focusing on image spatial self-similarity. The LLB module proposed by FeNeT [22] adopts the channel attention mechanism to construct the information communication of the upper and lower branches while ensuring the model lightweight by channel splitting.

### III. NETWORK ARCHITECTURE

#### A. Framework View

For the input low-quality satellite remote sensing image  $I_{LR}$ , our method is to reconstruct a high-quality image  $I_{HR}$ , which should be close to the ground truth  $I_{GT}$ . As shown in Fig. 2, our DARN is mainly composed of four parts: a shallow feature extraction module  $H_{map}$ , a deep feature extraction module formed by concatenating  $N$  DARCBs, a multistage feature fusion block  $H_{fusion}$  and a reconstruction module  $H_{up}$ .

The input image first goes through the shallow feature extraction block  $H_{map}$  to map the low-dimensional image to the high-dimensional space, which can enrich the representation of image details. Then, the deep feature extraction module

composed of multiple DARCBs gradually refines the extracted features. It can be expressed as

$$F_0 = H_{map}(I_{LR}) \quad (1)$$

$$F_K = H_{DARCB_K}(F_{K-1}), K = 1, \dots, N \quad (2)$$

where  $K$  represents the  $K$ th DARCB.  $F_K$  represents both the output feature of the  $K$ th DARCB and the input feature of the  $(K+1)$ th DARCB.

The optimization of lightweight models refers to improving the performance and speed of the model under smaller model parameter constraints. Therefore, fusing features from different depths of the model is an effective way to improve the performance of lightweight models. As shown in Fig. 2, the fusion module improves the feature reuse rate by fusing multistage features. We use the fused feature  $F_{final}$  to restore high-quality remote sensing images through the reconstruction module  $H_{up}$ . Furthermore, rational application of residual learning is an effective way to improve the model performance. The aforementioned process can be expressed as

$$F_{final} = H_{Fusion}(f(F_0, F_1, \dots, F_N)) + F_0 \quad (3)$$

$$I_{HR} = H_{up}(F_{final}) + I_{LR} \quad (4)$$

where  $f$  represents connecting multistage features in the spatial dimension. Based on the relevant work [13], [23], [24], [26], [27] experience of other researchers, the DARN is optimized using the  $L_1$  generalized loss function. It can be expressed as

$$L_1 = \|I_{GT} - I_{HR}\|_1. \quad (5)$$

#### B. Thinking of Lightweight Structure

The determination of the network model architecture is the first challenge during the model design phase. Model lightweighting is to improve model accuracy as much as possible under the constraints of low parameter quantity, low computational complexity, and high inference speed. Common deep learning architectures mainly include convolution, transformer, and MLP. Transformer and MLP have a high computational complexity, so they are not suitable for SR light tasks based on pixel-level computation. The inference speed of convolution is fast and the computational complexity is low. Reasonable structural design can make the convolution show excellent performance under the limited parameters.

Thinking of lightweight structure are shown in Fig. 3. To limit the number of parameters of the model, we only use three  $3 \times 3$

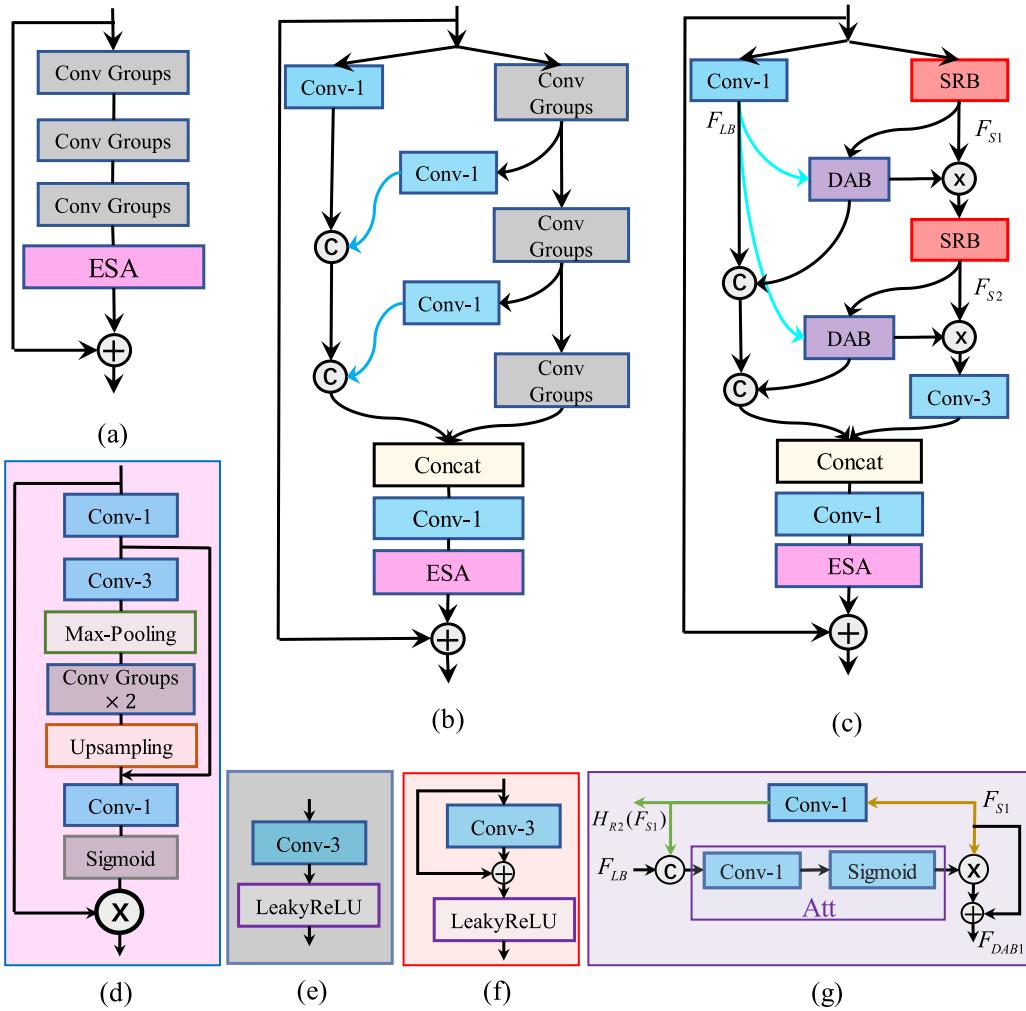


Fig. 3. (a) Simple baseline composed of three convolutions. (b) Feature distillation splicing (FDC) block formed by the simple model baseline transformation. (c) Distance attention residual connection block (DARCB). (d) Enhanced spatial attention (ESA). (e) Conv Groups. (f) Shallow residual block (SRB). (g) Distance attention block (DAB) for guiding CNN depth feature extraction. (a) Basic. (b) FDC. (c) DARCB. (d) ESA. (e) Conv Groups. (f) SRB. (g) DAB.

ordinary convolutions to compose the deep feature extraction component. The concatenation of three Conv Groups and an ESA module constitutes the baseline for our model design. However, simple convolution concatenation is bound to bring a lot of feature redundancy to the lightweight model, which inevitably limits the efficiency of component feature extraction. Therefore, we design a feature distillation connection (FDC) module by combining knowledge distillation and multistage feature fusion. FDC has the following two significant advantages over baseline:

- 1) the way that FDC fuses multistage features within a limited amount of parameters improves the feature utilization;
- 2) the distillation operation adopted by FDC can refine the feature information of the channel.

In the fusion stage, the influence of the deep features extracted by the main branch CNN block on the output of the FDC module is reduced. The influence of feature redundancy and random errors generated by the main branch on the performance of the module can be effectively reduced by assigning the output weights of each stage.

Feature redundancy and accumulated random errors formed by simple concatenation are common problems of current CNN models. To improve the feature extraction efficiency of the main branch, this article designs a DAB module, which uses the input refined features to suppress the loss of detail features of the entire main branch CNN module. Shallow features contain all the original image information. CNNs have problems such as feature loss, feature redundancy, and error accumulation when extracting deep features. Therefore, the shallow feature-supervised CNN can effectively reduce the error accumulation in the extraction process. In addition, the attention mechanism in the DAB module can compensate for the loss of information, enhance the proportion of effective features and reduce redundant information. An SRB is introduced as the main building block of the main branch to keep the network light enough. In addition, replacing the Conv Groups module with SRB can enable the model to flexibly learn residual information and make the model more robust. From formula (2), it can be known that the input feature of the  $K$ th DARCB is  $F_{K-1}$ , and the output feature is  $F_K$ . The DARCB first divides the input feature  $F_{K-1}$  into two paths of the main branch and branch for feature extraction, where

TABLE I  
QUANTITATIVE COMPARISON OF DIFFERENT CORE COMPONENTS

Methods	Params	Set5		Set10		BSD100		Urban100		Manga109	
		PSNR	SSIM								
Basic	550K	32.10	0.8944	28.53	0.7802	27.54	0.7363	25.96	0.7818	30.27	0.9053
FDC	590K	32.15	0.8950	28.57	0.7814	27.56	0.7369	26.04	0.7840	30.36	0.9068
DARCB	606K	<b>32.19</b>	<b>0.8952</b>	<b>28.58</b>	<b>0.7814</b>	<b>27.58</b>	<b>0.7373</b>	<b>26.07</b>	<b>0.7851</b>	<b>30.44</b>	<b>0.9078</b>

DARCB is the core component proposed in this article.

The bold values represents the optimal performance under the corresponding indicator.

TABLE II  
ABLATION STUDY OF ESA AND DAB

Methods	Params	Set5		Set10		BSD100		Urban100		Manga109	
		PSNR	SSIM								
DARN-woESA	559K	32.06	0.8943	28.51	0.7804	27.53	0.7357	25.94	0.7807	30.27	0.9056
DARN-woDAB	590K	32.14	0.8952	28.56	0.7814	27.56	0.7368	26.02	0.7837	30.40	0.9073
DARN	606K	<b>32.19</b>	<b>0.8952</b>	<b>28.58</b>	<b>0.7814</b>	<b>27.58</b>	<b>0.7373</b>	<b>26.07</b>	<b>0.7851</b>	<b>30.44</b>	<b>0.9078</b>

The bold values represents the optimal performance under the corresponding indicator.

TABLE III  
FEATURE REFINEMENT RATE OPTIMAL SOLUTION IS INVESTIGATED ON THE BENCHMARK DATASET WITH A SCALE FACTOR OF 4×

Refinement rate	Params	Set5		Set10		BSD100		Urban100		Manga109	
		PSNR	SSIM								
0.25	606K	<b>32.19</b>	<b>0.8952</b>	28.58	0.7814	<b>27.58</b>	0.7373	26.07	0.7851	30.44	0.9078
0.5	646K	32.12	0.8946	28.57	0.7813	27.58	0.7370	26.05	0.7839	30.41	0.9072
0.75	686K	<b>32.19</b>	0.8951	<b>28.59</b>	<b>0.7817</b>	<b>27.58</b>	<b>0.7375</b>	<b>26.10</b>	<b>0.7858</b>	<b>30.48</b>	<b>0.9082</b>

The bold values represents the optimal performance under the corresponding indicator.

the branch adopts feature refinement convolution to preserve the original information  $F_{LB}$  of the input features

$$F_{LB} = H_{R1}(F_{K-1}). \quad (6)$$

Among them,  $H_R$  is the feature refinement convolution. Second, two RSB modules are used to extract deep features, and two DAB modules are used to enhance the efficiency of deep feature extraction. Finally, a convolutional layer is employed to refine the depth feature  $F_{DAB2}$ . The specific process can be expressed as

$$F_{S1} = H_{S1}(F_{K-1}) \quad (7)$$

$$F_{DAB1} = H_{Att1}(f(F_{LB}, H_{R2}(F_{S1}))) * F_{S1} + F_{S1} \quad (8)$$

$$F_{S2} = H_{S2}(F_{DAB1}) \quad (9)$$

$$F_{DAB2} = g(H_{Att2}(f(F_{LB}, H_{R3}(F_{S2}))) * F_{S2} + F_{S2}). \quad (10)$$

Among them,  $H_S$  indicates the RSB feature extraction module,  $H_{Att}$  represents the attention mechanism in the DAB, and  $g$  represents several convolutional layers. Finally, the DARCB fuses multistage features

$$F_f = g(F_{LB}, H_{S2}(F_{S1}), H_{S3}(F_{S2}), F_{DAB2}). \quad (11)$$

To further enhance the representational power of the model while maintaining the efficiency of the model, we introduce a lightweight ESA block. Therefore, it can be concluded that the output characteristics of the  $k$ th DARN are

$$F_K = H_{ESA}(F_f) + F_f. \quad (12)$$

In general, the DARCB proposed in this article only uses three  $3 \times 3$  convolutional layers to extract deep features, which ensures the lightness of the model. The proposal of DAB strengthens feature extraction so that the model can achieve high accuracy while maintaining lightness.

## IV. EXPERIMENT ANALYSIS

### A. Datasets and Metrics

Building on previous work [22], [23], the extensive SR dataset DIV2K [36] is selected to train our model. The DIV2K dataset contains 800 high-quality RGB training images and 100 validation images. The reconstruction performance of the model is tested using two remote sensing data RS-T1 and RS-T2 proposed by FeNet [22]. Both RS-1 and RS-2 consist of 120 images covering 21 complex ground-truth remote sensing scenarios. To further comprehensively test the robustness of the model, five widely used SR benchmarks are used: Set5 [37], Set14 [38], BSD100 [39], Urban100 [40], and Manga109 [41]. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) were used to measure the quality of image reconstruction.

### B. Implementation Details

To obtain training data, the datasets introduced in the previous subsection are downsampled by  $2 \times$ ,  $3 \times$ , and  $4 \times$  bicubic interpolation. Note that all data that is not divisible by  $2 \times 3 \times 4 = 12$  is clipped to multiples of 12 before downsampling. The minibatch size is set to 64 and the patch size of each LR input is set to  $64 \times 64$ . To achieve data augmentation, horizontal flips and random rotations of  $90^\circ$  employed to augment the training set. ADAM trains the model as an optimizer and sets  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 1 \times 10^{-8}$ . The model is trained in  $1 \times 10^7$  batches, the initial learning rate is set to  $5 \times 10^{-4}$ , and the learning rate is halved every  $2 \times 10^5$  minibatches. The  $2 \times$  reconstruction model is trained from scratch, and when the model converges, the  $2 \times$  reconstruction model weights will be used as the pretrained models for the  $3 \times$  and  $4 \times$  models. DARN-S

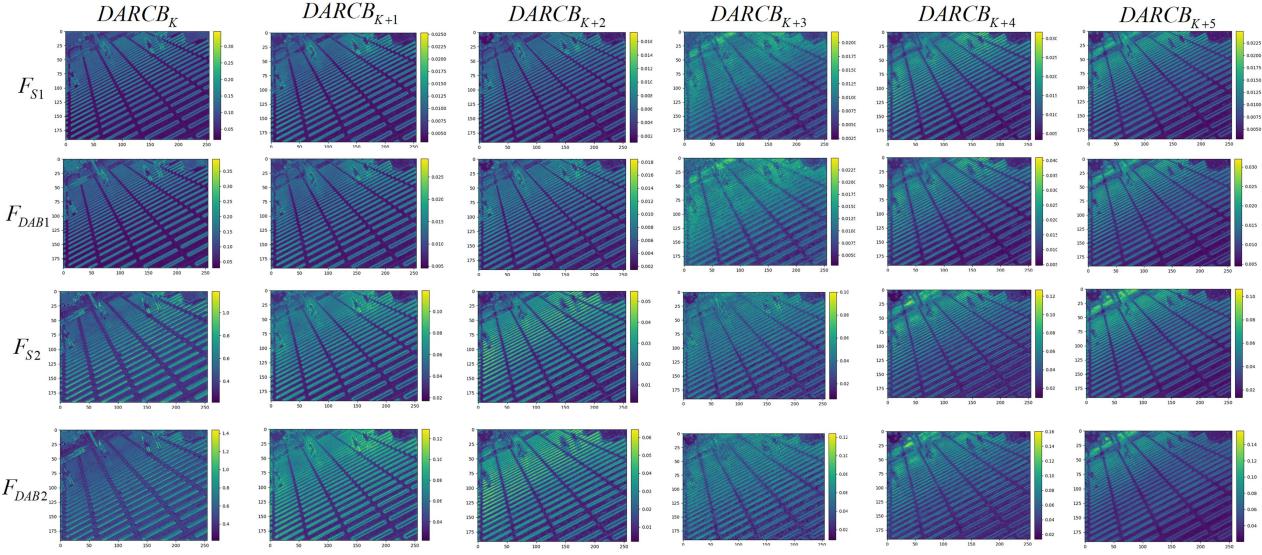


Fig. 4. Average feature maps at various stages of the DARCB.

uses 40 feature channels to ensure further light-weight, while the DARN uses 52 channels for better reconstruction quality. Both DARN-S and DARN models use six DARCB modules as the backbone for deep feature extraction. To keep the model lightweight, each DARCB contains two SRBs. Our model is implemented in the pytorch framework on an NVIDIA 3090 GPU, where DARN-S takes about 36 h to train for one million batches.

### C. Model Analysis

Experiment statement in this subsection. In order to facilitate the completion of a large number of comparative experiments and ablation experiments, the experiments in this section are all trained from scratch. The number of training steps is  $3 \times 10^5$ , the minibatch size is 16, and the number of channels is 52.

Structural design of core components. To verify the theoretical analysis in Section III-B, we show in Table I the performance of the models with Basic, FDC, and DARCB as the core components on five test sets. It can be found that FDC far outperforms Basic on the five test sets. This is because the feature redundancy and accumulated random errors of simple CNN concatenations limit the model performance.

1) *Ablation Study*: We conduct ablation experiments to demonstrate the effectiveness of the DAB and ESA module. The experimental results are shown in Table II. It can be found that the addition of DAB only brings 16 K additional parameters. The introduction of a small amount of parameters makes the model achieve a comprehensive lead on the five SR benchmark datasets. This is mainly due to the fact that the DAB module uses shallow features to guide the deep feature extraction of the main branch deep CNN, thus effectively suppressing the loss of detail features. Although the addition of the ESA module brings a large amount of parameters, the accuracy gains brought by it are very considerable.

2) *Feature Refinement Rate*: Feature refinement is an operation that refines the channel information of the current feature and reduces feature redundancy. The feature refinement rate is

TABLE IV  
QUANTIFY HOW LIGHTWEIGHT THE MODEL IS ON A BENCHMARK DATASET  
WITH A SCALING FACTOR OF 4 $\times$

Methods	Params	Mutil-adds	Times	Manga109	
				PSNR	SSIM
LESRCNN [20]	774K	241.6G	33.27ms	29.94	0.9002
FeNet [22]	352K	20.4G	13.49ms	29.85	0.8992
DARN-S	350K	19.7G	9.86ms	30.48	0.9084
DARN	605K	32.9G	10.70ms	30.65	0.9105

The best and next best results are red and blue, respectively. Computing multiple addition and inference speed corresponds to a 1280  $\times$  720 HR image.

the degree of compression of the channel information. In order to obtain the optimal refinement rate, it was set to 0.25, 0.5, and 0.75 for comparative experiments. The experimental results are shown in Table III. It can be found that when the refinement rate is set to 0.75, the optimal solution is taken by the model but the model parameters are increased by  $80^{\sim}k$ . However, with a refinement rate of 0.25, the model achieves still decent accuracy and has the smallest number of parameters. For a more lightweight model, we set the feature refinement rate to 0.25.

3) *Model Efficiency*: A lightweight model refers to a low number of parameters, low computational complexity, and high inference speed. The amount of model parameters (Params) and computational complexity (Mutil-adds) affect the difficulty of actual model deployment. Speed affects model work efficiency. We show the parameters, computational complexity, and inference time of the four models of LESRCNN, FeNet, DARN-S, and DARN in Table IV, respectively. It can be found that our DARN and DARN-S models cover the optimal solutions of the four lightweight indicators. It is worth noting that although the amount of parameters and computational complexity of the DARN is larger than that of FeNet, the reasonable model design of the DARN reduces the inference time by 27% and improves the PSNR score by 0.8 db compared to FeNet. In general, the rationality, efficiency, and lightness of the proposed model are proved by comparing with the existing excellent methods.

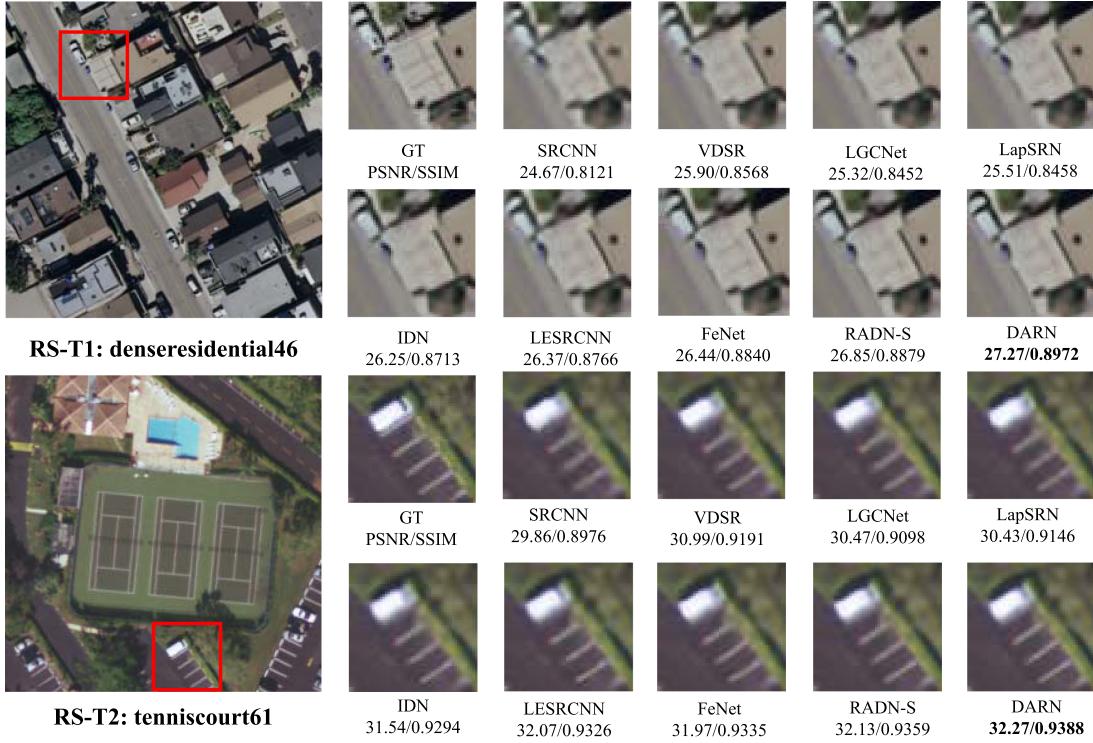


Fig. 5. Visual comparison of DARN with other SR methods on RS-T1 and RS-T2 remote sensing datasets. Please zoom in for better visualization.

4) *Feature Visual Analysis:* In order to more intuitively prove that the DAB module proposed in this article can alleviate the loss of details in the feature extraction process of deep CNN, this article visualizes the feature information of each stage in the DARCB inference process. It can be seen from Fig. 4 that in the same DARCB module, the output feature  $F_{\text{DAB}}$  of DAB obviously retains more detailed features than the input feature  $F_S$ . This proves that the DAB module uses the guidance of shallow features to make the model retain more detailed features. In addition, with the progress of the DARCB module, the features have a clear trend of refinement, which intuitively proves the rational design of the DARN.

#### D. Results Analysis of Remote Sensing Datasets

To verify the effectiveness of the proposed algorithm on remote sensing datasets, the DARN is compared with existing lightweight models, which including SRCNN [5], VDSR [8], LGCNet [33], LapSRN [30], IDN [19] and LESRCNN [20], CARN-M [42], and FeNet [22]. Among them, FeNet is a work on lightweight models for remote sensing data released in 2022. Comparing with advanced algorithms can better prove the effectiveness of the model in this article. All the aforementioned methods are directly tested on remote sensing data using pretrained models provided by relevant workers. In addition, the selected comparison model training sets are all DIV2K, and the same training set ensures the fairness of the comparison results. The quantitative comparison results of different methods are shown in Table V. It can be found that our model (DARN) and its lightweight version (DARN-S) achieve optimal and suboptimal

TABLE V  
QUANTITATIVE COMPARISON RESULTS OF REMOTE SENSING TEST IMAGES  
RS-T1 AND RS-T2

Methods	Scale	Params	RS-T1	RS-T2
			PSNR/SSIM	PSNR/SSIM
Bicubic	$\times 2$	-	33.25/0.8934	30.64/0.8837
SRCNN [5]		57K	35.18/0.9243	32.87/0.9209
VDSR [8]		666K	35.85/0.9312	33.86/0.9312
LGCNet [33]		193K	35.65/0.9298	33.47/0.9281
LapSRN [30]		251K	35.69/0.9304	33.57/0.9286
IDN [19]		553K	36.13/0.9339	34.07/0.9329
LESRCNN [20]		626K	36.04/0.9328	34.00/0.9320
CARN-M [42]		412K	35.77/0.9314	33.84/0.9315
FeNet [22]		351K	36.23/0.9341	34.22/0.9337
DARN-S		350K	<b>36.31/0.9347</b>	<b>34.35/0.9348</b>
DARN		589K	<b>36.38/0.9348</b>	<b>34.42/0.9357</b>
Bicubic	$\times 3$	-	29.73/0.7818	27.23/0.7697
SRCNN [5]		57K	30.95/0.8228	28.59/0.8180
VDSR [8]		666K	31.55/0.9352	29.40/0.8391
LGCNet [33]		193K	31.30/0.8314	29.03/0.8312
LapSRN [30]		290K	31.47/0.8338	29.22/0.8352
IDN [19]		553K	31.73/0.8430	29.59/0.8450
LESRCNN [20]		810K	31.68/0.8398	29.65/0.8444
CARN-M [42]		412K	31.72/0.8426	29.62/0.8452
FeNet [22]		357K	31.89/0.8432	29.80/0.8481
DARN-S		350K	<b>32.00/0.8483</b>	<b>29.98/0.8518</b>
DARN		589K	<b>32.08/0.8470</b>	<b>30.05/0.8537</b>
Bicubic	$\times 4$	-	27.91/0.6968	25.40/0.6770
SRCNN [5]		57K	28.87/0.7382	26.46/0.7296
VDSR [8]		666K	29.33/0.7546	27.03/0.7525
LGCNet [33]		193K	29.13/0.7481	26.76/0.7426
LapSRN [30]		543K	29.51/0.7614	27.24/0.7600
IDN [19]		553K	29.56/0.7623	27.31/0.7627
LESRCNN [20]		774K	29.62/0.7625	27.41/0.7646
CARN-M [42]		412K	29.57/0.7624	27.37/0.7647
FeNet [22]		366K	29.70/0.7688	27.45/0.7672
DARN-S		350K	<b>29.78/0.7682</b>	<b>27.59/0.7732</b>
DARN		589K	<b>29.85/0.7710</b>	<b>27.67/0.7758</b>

“-” indicates that the result is unknown. The best and next best results are red and blue, respectively.

TABLE VI  
QUANTITATIVE COMPARISON WITH STATE-OF-THE-ART METHODS ON SR BENCHMARK DATASETS

Mothods	Scale	Params	Multi-Add	Set5		Set14		BSD100		Urban100		Manga109	
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM							
Bicubic	$\times 2$	-	-	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403	30.80/0.9339	35.60/0.9663	36.67/0.9710	37.22/0.9750	-	-
SRCNN [5]		8K	52.7G	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663	36.67/0.9710	37.22/0.9750	-	-	-
FSRCNN [6]		13K	6.0G	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	30.76/0.9140	37.22/0.9750	-	-	-	-
VDSR [8]		666K	612.6G	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	30.53/0.9112	37.22/0.9750	-	-	-	-
LGCNet [33]		193K	178.1G	37.31/0.9580	32.94/0.9120	31.74/0.8939	30.53/0.9112	37.22/0.9750	-	-	-	-	-
DRCN [16]		1774K	-	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.55/0.9732	-	-	-	-	-
LapSRN [30]		251K	29.9G	37.52/0.9591	32.99/0.9124	31.80/0.8952	30.41/0.9103	37.27/0.9740	-	-	-	-	-
DRRN [17]		298K	6796.9G	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.88/0.9749	-	-	-	-	-
MenNet [43]		678K	2662.4G	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195	37.72/0.9740	-	-	-	-	-
IDN [19]		553K	124.6G	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	38.01/0.9749	-	-	-	-	-
LESRCNN [20]		626K	281.5G	37.65/0.9586	33.32/0.9148	31.95/0.8964	31.45/0.9206	37.89/0.9746	-	-	-	-	-
MADNet [21]		878K	187.1G	37.85/0.9600	33.39/0.9161	32.05/0.8981	31.59/0.9234	37.58/0.9741	-	-	-	-	-
CARN-M [42]		412K	91.2G	37.53/0.9583	33.26/0.9141	31.92/0.8960	31.23/0.9193	38.18/0.9752	-	-	-	-	-
FeNet [22]		351K	77.9G	37.90/0.9602	33.45/0.9162	32.09/0.8985	31.75/0.9245	38.18/0.9752	-	-	-	-	-
DARN-S		350K	78.9G	37.97/0.9609	33.54/0.9172	32.19/0.9005	32.14/0.9284	38.74/0.9773	-	-	-	-	-
DARN		589K	131.6G	38.04/0.9610	33.63/0.9186	32.25/0.9012	32.40/0.9305	38.87/0.9776	-	-	-	-	-
Bicubic	$\times 3$	-	-	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556	-	-	-	-	-
SRCNN [5]		8K	52.7G	32.75/0.9090	29.30/0.8215	28.41/0.7863	26.43/0.7989	30.48/0.9117	-	-	-	-	-
FSRCNN [6]		13K	5.0G	33.18/0.9140	29.37/0.8240	28.53/0.7910	26.43/0.8080	31.10/0.9210	-	-	-	-	-
VDSR [8]		666K	612.6G	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9340	-	-	-	-	-
LGCNet [33]		193K	79.0G	33.32/0.9172	29.67/0.8289	28.63/0.7923	26.77/0.8180	32.24/0.9343	-	-	-	-	-
DRCN [16]		1774K	-	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.21/0.9350	-	-	-	-	-
LapSRN [30]		502K	115.2G	33.81/0.9220	29.79/0.8325	28.82/0.7980	27.07/0.8275	32.71/0.9379	-	-	-	-	-
DRRN [17]		298K	6796.9G	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.51/0.9369	-	-	-	-	-
MenNet [43]		678K	2662.4G	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376	32.78/0.9384	-	-	-	-	-
IDN [19]		553K	124.6G	34.11/0.9253	29.99/0.8354	28.95/0.8013	27.42/0.8359	32.71/0.9381	-	-	-	-	-
LESRCNN [20]		810K	238.9G	33.93/0.9231	30.12/0.8380	28.91/0.8005	27.70/0.8415	32.76/0.9389	-	-	-	-	-
MADNet [21]		930K	88.4G	34.14/0.9251	30.20/0.8395	28.98/0.8023	27.78/0.8439	32.99/0.9394	-	-	-	-	-
CARN-M [42]		412K	46.1G	33.99/0.9236	30.08/0.8367	28.91/0.8000	27.55/0.8385	33.59/0.9446	-	-	-	-	-
FeNet [22]		357K	35.2G	34.21/0.9256	30.15/0.8383	28.98/0.8020	27.82/0.8447	33.76/0.9457	-	-	-	-	-
DARN-S		355K	35.0	34.35/0.9274	30.34/0.8428	29.09/0.8065	28.17/0.8528	33.76/0.9457	-	-	-	-	-
DARN		596K	58.4G	34.48/0.9286	30.41/0.8443	29.15/0.8076	28.38/0.8570	33.76/0.9457	-	-	-	-	-
Bicubic	$\times 4$	-	-	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866	-	-	-	-	-
SRCNN [5]		8K	52.7G	30.48/0.8626	27.50/0.7513	26.90/0.7101	24.52/0.7221	27.58/0.8555	-	-	-	-	-
FSRCNN [6]		13K	4.6G	30.72/0.8660	27.61/0.7550	26.98/0.7150	24.62/0.7280	27.90/0.8610	-	-	-	-	-
VDSR [8]		666K	612.6	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8870	-	-	-	-	-
LGCNet [33]		193K	44.5G	30.87/0.8746	27.82/0.7630	27.08/0.7186	24.82/0.7399	28.93/0.8854	-	-	-	-	-
DRCN [16]		1774K	-	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	29.09/0.8900	-	-	-	-	-
LapSRN [30]		502K	149.4G	31.54/0.8852	28.09/0.7700	27.32/0.7275	25.21/0.7562	29.45/0.8946	-	-	-	-	-
DRRN [17]		298K	6796.9	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.42/0.8942	-	-	-	-	-
MenNet [43]		678K	2662.4G	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630	29.41/0.8942	-	-	-	-	-
IDN [19]		553K	32.3G	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632	29.41/0.8942	-	-	-	-	-
LESRCNN [20]		774K	241.6G	31.88/0.8903	28.44/0.7772	27.45/0.7313	25.77/0.7732	29.94/0.9002	-	-	-	-	-
MADNet [21]		1002K	54.1G	32.01/0.8925	28.45/0.7781	27.47/0.7327	25.77/0.7751	29.80/0.8989	-	-	-	-	-
CARN-M [42]		412K	32.5G	31.92/0.8903	28.42/0.7762	27.44/0.7304	25.63/0.7688	29.85/0.8992	-	-	-	-	-
FeNet [22]		366K	20.4G	32.02/0.8919	28.38/0.7764	27.47/0.7319	25.75/0.7747	29.85/0.8992	-	-	-	-	-
DARN-S		363K	19.7	32.16/0.8951	28.58/0.7817	27.57/0.7374	26.08/0.7859	30.48/0.9084	-	-	-	-	-
DARN		606K	32.9G	32.24/0.8963	28.64/0.7830	27.61/0.7390	26.25/0.7913	30.65/0.9105	-	-	-	-	-

“—” indicates that the result is unknown. The best and next best results are red and blue, respectively. Computing multiple addition corresponds to a 1280 × 720 HR image.

scores on all magnifications of the remote sensing datasets RS-T1 and RS-T2. Specifically, the DARN achieves 0.15–0.25 dB PSNR score improvement compared with the state-of-the-art lightweight remote sensing image SR method FeNet. It is worth noting that DARN-S, which has the same parameters as FeNet, still improves the PSNR score by 0.08–0.13 dB at 2× sampling rate. The comparison results show that our DARN can effectively enhance the feature expression ability to improve the effect of remote sensing image restoration.

To evaluate the perceptual quality, we present the restoration results of two remote sensing images in the RS-T1 and RS-T2 datasets in Fig. 5, respectively. In addition, the patches compared in Fig. 5 are all evaluated quantitatively in order to distinguish

the recovery differences between different methods. It can be found that the DARN shows a better recovery effect than other methods on the texture of the outline and details of the object. To further demonstrate the stability of the model, the restoration results of two real remote sensing satellite images are shown in Fig. 6. We use five different algorithms (Bicubic, CARN-M [42], LESRCNN [20], FeNet [22], and DARN) to enlarge the real satellite remote sensing images by four times for visual perception comparison. It can be found that our method has better visual experience than other methods in both overall and detailed texture. In general, the DARN as a lightweight model achieves better results both quantitatively and qualitatively with existing models.

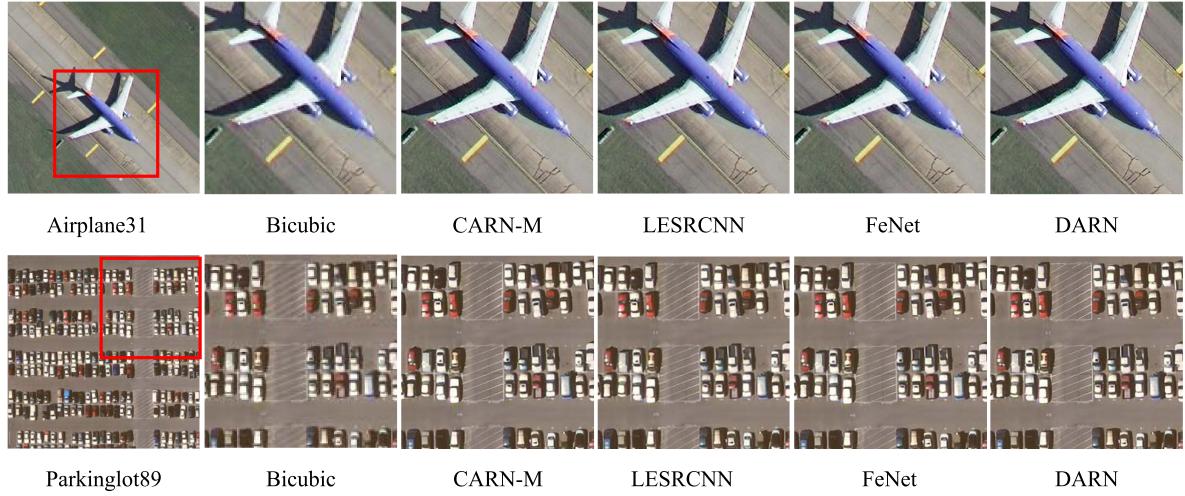


Fig. 6. Visual comparison of the DARN with other SR methods on real remote sensing images. Please zoom in for better visualization.

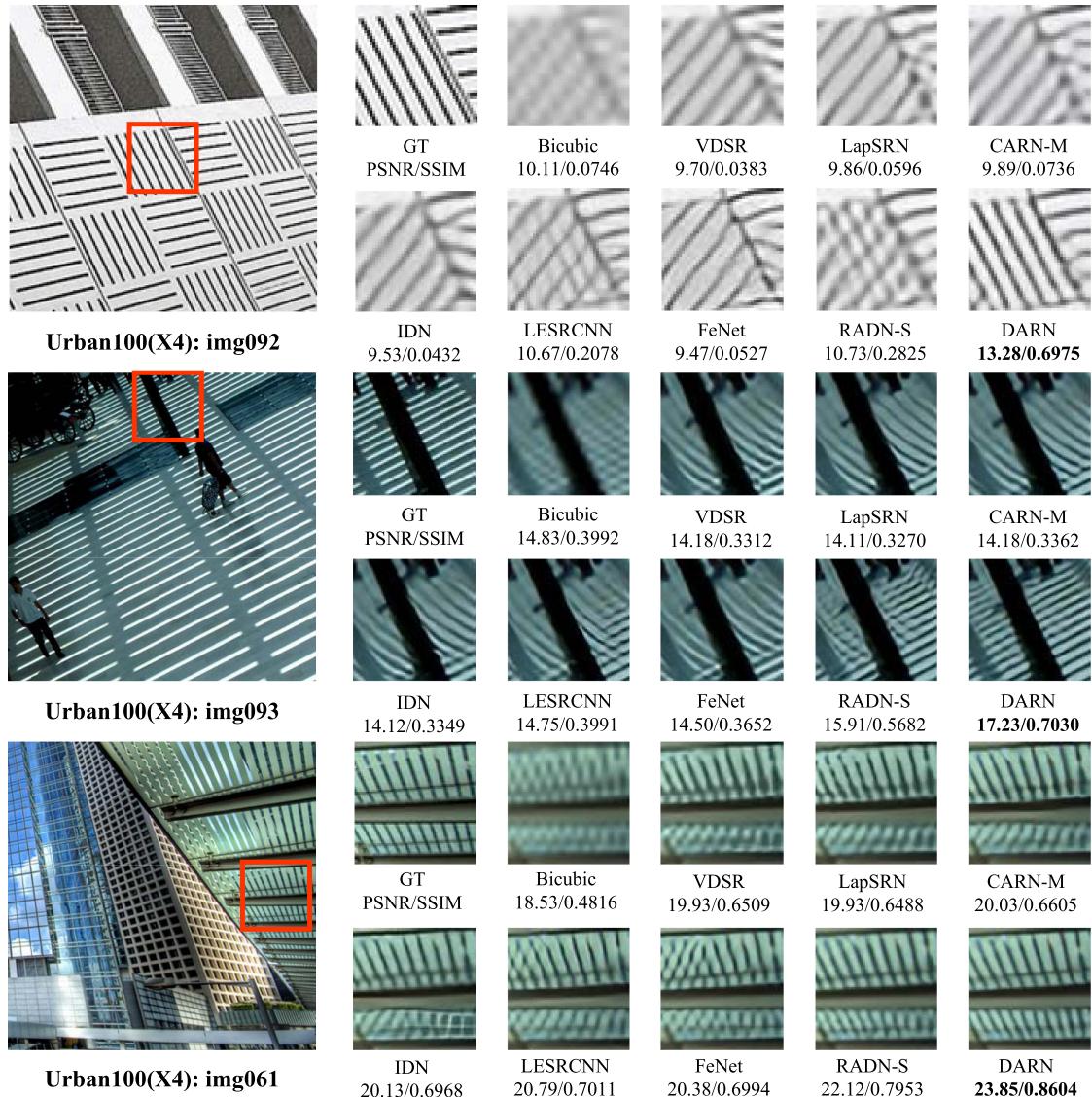


Fig. 7. Visual comparisons of the DARN with other SR methods on Urban100( $\times 4$ ) datasets. Please zoom in for better visualization.

### E. Analysis of SR Benchmark Dataset Results

The dataset composed of Set5 [37], Set14 [38], Urban100 [39], BSD100 [40], and Manga109 [41] is the most impartial test set in the field of image SR reconstruction, which covers urban architecture, ecological environment, animals and plants, and animation and other scene images. In order to further verify the generalization performance of the model, the PSNR and SSIM scores of 12 existing models (SRCNN [5], FSRCNN [6], VDSR [8], DRCN [16], LapSRN [30], DRRN [17], MenNet [41], IDN [19], LESRCNN [20], MADNet [21], CARN-M [42], and FeNet [22]) and the model (DARN-S and DARN) in this article are tested on the aforementioned five test sets, respectively. The models used in the test are all pretrained models provided by relevant workers. The test results are shown in Table VI. It can be seen that DARN and the lighter version DARN-S achieve the best scores on all test sets of  $\times 2$ ,  $\times 3$ , and  $\times 4$  sampling factors. Compared with the advanced FeNet method, the DARN leads the PSNR score by 0.14–0.69 db at  $2\times$  sampling factor, 0.17–0.77 db at  $3\times$  sampling factor, and 0.14–0.8 db at  $4\times$  sampling factor PSNR score. Even the DARN-S with comparable model size still outperforms the advanced FeNet algorithm in PSNR and SSIM scores on all sampling factors of the five test sets. Under the condition of the same parameter quantity, the performance of the two models mainly depends on the ability of the model framework to express the features. The overall lead of DARN-S is enough to prove the correctness of the feature enhancement idea of the DARN module using shallow features as prior information to guide CNN deep feature extraction. Even DARN-S still achieves a comprehensive lead in all test set metrics compared to MADNet, which has nearly three times the number of parameters. It is worth mentioning that DARN-S is still the optimal solution in terms of speed and computational complexity (see Table IV).

To evaluate the perceptual quality, we show in Fig. 7 three recovery results of some models on the Urban100 test set. The images in the Urban100 test set have more line textures for the human eye to distinguish the differences more intuitively. It can be clearly seen that the DARN achieves the best visual experience in terms of overall image block clarity and detailed line texture. Bicubic interpolates estimates directly between image pixels without changing the original pixel values. By observing GT and Bicubic, it can be found that when the image is downsampled by a factor of 4, the image appears extremely blurred and even some areas have false artifacts. For example, the actual line of img092 is extended to the lower right, and some artifacts that are biased to the lower left appear during downsampling. The generation of these artifacts makes the models VDSR [8], LapSRN [30], CARM-M [42], IDN [19], LESRCNN [20], and FeNet [22] obtain additional wrong information, which leads to serious errors in the final model recovery results. However, the DARN recovered accurate results and clear lines undisturbed by misinformation. Among them, some error messages also appeared in the DARN-S recovery results. This is because the parameter amount of DARN-S is smaller than that of the DARN, which has a weaker antiinterference ability. Nevertheless, the recovery effect of DARN-S is still better than other models.

### V. CONCLUSION

In this article, a lightweight single-image SR reconstruction model DARN is proposed to solve the problem of high-frequency information reconstruction of low-quality images collected by remote sensing satellites. The model is divided into 350 k light version DARN-S and 589 K normal version DARN for the choice of different tasks. Extensive experiments show that the proposed two versions of the model both show the best results compared with the existing lightweight models. The DARN exhibits excellent grades mainly depending on the core component DARN. Specifically, the DARN divides the module into two branches by refinement of shallow features and preserves the refinement features. The refinement features are then used as prior information to guide the deep feature extraction of the main branch CNN module through DAB. Prior information can suppress the loss of detailed feature information in the process of deep extraction. The feature visualization of the DARCB further proves that the use of prior information in our DAB can suppress the loss of detailed features of deep CNN. Furthermore, in order to fully utilize the extracted features under a limited amount of parameters, the backbone of both DARN and DARN-S adopts a multistage feature fusion operation. Finally, the experimental results on two remote sensing and five SR benchmark datasets show that our network can achieve a better balance between performance and model complexity.

### REFERENCES

- [1] P. Xu, H. Tang, J. Ge, and L. Feng, “ESPC-NASUnet: An end-to-end super-resolution semantic segmentation network for mapping buildings from remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, no. 3, pp. 5421–5435, May 2021, doi: [10.1109/JSTARS.2021.3079459](https://doi.org/10.1109/JSTARS.2021.3079459).
- [2] L. Chen, H. Liu, M. Yang, Y. Qian, Z. Xiao, and X. Zhong, “Remote sensing image super-resolution via residual aggregation and split attentional fusion network,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9546–9556, Sep. 2021, doi: [10.1109/JSTARS.2021.3113658](https://doi.org/10.1109/JSTARS.2021.3113658).
- [3] J. Tu, G. Mei, Z. Ma, and F. Piccialli, “SWGAN: Generative adversarial network combining swin transformer and CNN for remote sensing image super-resolution,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5662–5673, Jul. 2022, doi: [10.1109/JSTARS.2022.3190322](https://doi.org/10.1109/JSTARS.2022.3190322).
- [4] H. Chen, H. Zhang, J. Du, and B. Luo, “Unified framework for the joint super-resolution and registration of multiangle multi/hyperspectral remote sensing images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2369–2384, May 2020, doi: [10.1109/JSTARS.2020.2993629](https://doi.org/10.1109/JSTARS.2020.2993629).
- [5] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2014, pp. 184–199.
- [6] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 391–407.
- [7] W. Shi et al., “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1874–1883.
- [8] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1646–1654.
- [9] T. Tong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4799–4807.
- [10] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.
- [11] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, “Enhanced deep residual networks for single image super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1132–1140.

- [12] B. Lim, S. Son, H. Kim, and S. Nah, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1132–1140.
- [13] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.
- [14] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11057–11066.
- [15] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2472–2481.
- [16] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1637–1645.
- [17] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2790–2798.
- [18] V. Lebedev and V. Lempitsky, "Fast convnets using group-wise brain damage," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 2554–2564.
- [19] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 723–731.
- [20] C. Tian et al., "Lightweight image super-resolution with enhanced CNN," *Knowl.-Based Syst.*, vol. 205, 2020, Art. no. 106235.
- [21] R. Lan, L. Sun, Z. Liu, H. Lu, C. Pang, and X. Luo, "MADNeT: A fast and lightweight network for single-image super resolution," *IEEE Trans. Cybern.*, vol. 51, no. 3, pp. 1443–1453, Mar. 2021.
- [22] Z. Wang et al., "FeNet: Feature enhancement network for lightweight remote-sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Apr. 2022.
- [23] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proc. 27th ACM Int. Conf. Multimedia*, 2019, pp. 2024–2032.
- [24] Z. Li et al., "Blueprint separable residual network for efficient image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 832–842.
- [25] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2356–2365.
- [26] R. Xu, Z. Xiao, J. Huang, Y. Zhang, and Z. Xiong, "EDPN: Enhanced deep pyramid network for blurry image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 414–423.
- [27] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1833–1844.
- [28] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 9992–10002.
- [29] X. Chen, X. Wang, J. Zhou, and C. Dong, "Activating more pixels in image super-resolution transformer," May 2022, *arXiv:2205.04437*. [Online]. Available: <https://ui.adsabs.harvard.edu/abs/2022arXiv220504437C>
- [30] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5835–5843.
- [31] W. Li, K. Zhou, L. Qi, N. Jiang, J. Lu, and J. Jia, "LAPAR: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 20343–20355, 2020.
- [32] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong, "Efficient image super-resolution using pixel attention," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2020, pp. 56–72.
- [33] S. Lei, Z. Shi, and Z. Zou, "Super-resolution for remote sensing images via local-global combined network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1243–1247, Aug. 2017.
- [34] P. Lv, W. Wu, Y. Zhong, F. Du, and L. Zhang, "SCViT: A spatial-channel feature preserving vision transformer for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 3, pp. 1–12, Mar. 2022.
- [35] S. Lei, Z. Shi, and W. Mo, "Transformer-based multistage enhancement for remote sensing image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, Dec. 2021.
- [36] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang, "Ntire 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 1110–1121.
- [37] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2012, pp. 135.1–135.10, doi: [10.5244/C.26.135](https://doi.org/10.5244/C.26.135).
- [38] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. Curves Surfaces*, 2010, pp. 711–730.
- [39] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th IEEE Int. Conf. Comput. Vis.*, 2001, vol. 2, pp. 416–423.
- [40] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1110–1121.
- [41] Y. Matsui et al., "Sketch-based manga retrieval using manga109 dataset," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21811–21838, 2017.
- [42] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 252–268.
- [43] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4539–4547.



**Qingjian Wang** received the B.Eng. degree in mechanical design and automation from Anhui University of Science and Technology, Huainan, China, in 2020. He is currently working toward the M.Eng. degree with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, China.

His research interests include image super-resolution, object detection, and deep learning.



**Sen Wang** received the Ph.D. degree in mechanical and electronic engineering from the Kunming University of Science and Technology, Kunming, China, in 2017.

He is currently an Associate Professor with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology. His research interests include computer vision, visual fault diagnosis, and visual vibration measurement.



**Mingfang Chen** received the Ph.D. degree in mechanical manufacturing and automation from Northeastern University, Shenyang, China, in 2015.

He is currently a Professor with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, China. His research interests include complex electromechanical system integration and application, circuit analysis and engineering applications, intelligent control theory and applications.



**Yang Zhu** received the B.Eng. degree in mechanical engineering from the Wenzhou University of Technology, Wenzhou, China, in 2021. He is currently working toward the M.Eng. degree with the Faculty of Mechanical and Electrical Engineering, Kunming University of Science and Technology, Kunming, China.

His main research interests include image super-resolution and deep learning.