Multi-level Dispersion Residual Network for Efficient Image Super-Resolution

Yanyu Mao[1*] Nihao Zhang[1*] Qian Wang[1,2†] Bendu Bai[1,2] Wanying Bai[1]
Haonan Fang[1] Peng Liu[1] Mingyue Li[1] Shengbo Yan[1]
[1] Xian University of Posts and Telecommunications, Xi'an, China
[2] National Engineering Laboratory for Cyber Event Warning and Control Technologies
{bolttt, zwyczhang}@stu.xupt.edu.cn, {wqabby, baibendu}@xupt.edu.cn, 18791473467@163.com
{fanglolo, 2101200069, lmy}@xupt.edu.cn, xupt_yanshengbo@163.com

Abstract

Recently, single image super-resolution (SISR) has made great progress, especially through the combination of convolutional neural network (CNN) and Transformer, but the huge model complexity is not desirable for the efficient image super-resolution (EISR), nor is it affordable for edge devices. As a result, many lightweight methods have been investigated for EISR, such as distillation and pruning. However, investigating more powerful attention mechanisms is also a promising solution to improve network efficiency. In this paper, we propose a multi-level dispersion residual network (MDRN) for EISR. As the basic block of MDRN, enhanced attention distillation block (EADB) includes the proposed multi-level dispersion spatial attention (MDSA) and enhanced contrast-aware channel attention (ECCA), respectively. MDSA introduces multi-scale and variance information to obtain more accurate spatial attention distribution. ECCA effectively combines lightweight convolution layers and residual connections to improve the efficiency of channel attention. The experimental results show that the proposed methods are effective and our MDRN achieves a better balance of performance and complexity than the SOTA models. In addition, we won the first place in the model complexity track of the NTIRE 2023 Efficient SR Challenge. The code is available at https://github.com/bbbolt/MDRN.

1. Introduction

Single image super resolution (SR) aims to reconstruct high resolution (HR) images from corresponding low resolution (LR) images. In recent years, a lot of powerful SR networks [5,8,18,27,45] have achieved high-quality recovery of images. However, to pursue better performance, most networks use larger models with huge computational complexity, which is divorced from the fact that mobile devices need to be deployed. Therefore, many lightweight methods [3,10,32,46] have been proposed to solve the efficiency problem of SR. In the early stage, the recursive neural network [26] and group convolution strategy [2] was used to reduce the model parameters, but the amount of computation is still huge and the performance decreased. Recently, more and more efficient methods have been proposed. IDN [17] has introduced feature distillation which uses channel segmentation to reduce the cost of convolution computation. IMDN [16], RFDN [29] and BSRN [25] further improved distillation mechanisms and achieved more efficient performance. In addition to distillation mechanisms, many model compression techniques have also been proposed (*e.g.*, pruning [13], kernel decomposition [36] and reparameter [7, 43]). However, the improvement brought by efficient attention mechanisms cannot be ignored in EISR. Powerful attention mechanisms can better guide the network to select key information, which enables the network to better pursue the balance between the complexity and performance of the model.

In this paper, we rethink the two commonly used lightweight attention mechanisms, enhanced spatial attention (ESA) [30] and contrast-aware channel attention (CCA) [16], to further investigate the more efficient attention mechanisms for achieving a good trade-off between performance and model complexity. Based on existing advanced methods, we propose an efficient SR model multi-level dispersion residual network (MDRN) which is equipped with both spatial and channel attention mechanisms to enhance network representation capabilities. On the spatial dimension, ESA, which uses a single large-size compression/dispersion process (*i.e.*, maximum pooling and bilinear interpolation), over-compresses spatial information, resulting in the key regions being easily ignored. Therefore, we extend the original single-level compression and dispersion process to multiple levels to get the improved spatial attention mechanism multi-level dispersion spatial

[*] indicates contribute equally. [†] Corresponding author

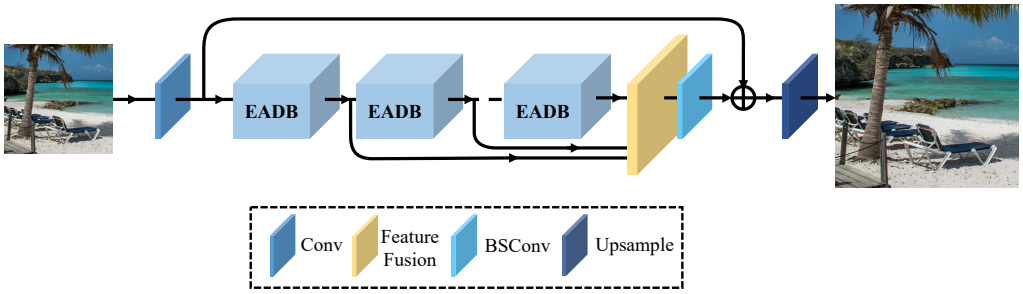


Figure 1. The framework of Multi-level Dispersion Residual Network (MDRN)

attention (MDSA). Except for multi-scale information, the local variance (L-var) is added into MDSA to pay more attention to structural information. In addition, we propose enhanced contrast-aware channel attention (ECCA), which can further improve the input and output features of CCA, as well as improve the flow of feature information through the residual connection. The introduction of ECCA and MDSA enables MDRN to obtain more powerful performance with less model complexity. Compared with other efficient SR models, MDRN obtains SOTA performance. As a variant of MDRN, MDRN-S was used to participate in the NTIRE2023 ESR competition [24].

2. Related work

2.1. Lightweight SR models

To deploy SR models on edge devices, some lightweight networks are proposed. DRCN [19] proposed a deep recursive convolutional network to increase the depth of the network, ensuring effectiveness while reducing the burden of too many parameters. DRRN [38] introduced residual structure based on DRCN and proposed a deep recursive residual network. Although the recursive layer reduces the number of parameters to a certain extent, the amount of computation is still unsatisfactory. MemNet [39] used a gating mechanism to fuse the features of different layers. IDN [17] proposed an information distillation network that splits features in the channel dimension and then processes them individually. IMDN [16] further improved IDN and proposed an information multiple distillation block. RFDN [29] introduced 1×1 convolution into the distillation operation, and proposed the shallow residual block (SRB). BSRN [25] used BSConv [11] to replace traditional convolution for reducing the number of network parameters. In addition, re-parameterization [9, 43] is also used to reduce the number of parameters and inference time.

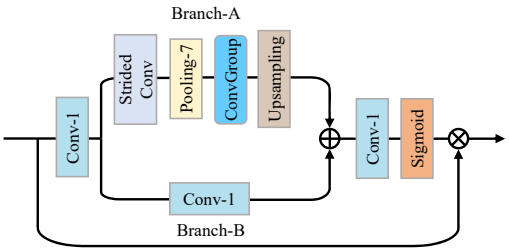
2.2. Attention Mechanism in SR

Attention Mechanism has been widely used in the field of vision recently. It can guide the network to focus on important information and suppress unnecessary information. SENet [14] proposed channel attention (CA) and achieved significant performance improvements in image classification tasks. RCAN [44] introduced CA into the residual block (RB) to model interdependencies across feature channels. Subsequently, SAN [6] proposed a novel second-order channel attention (SOCA) module to enhance the discriminative ability of the network. IMDN [16] proposed contrast-aware channel attention (CCA) to enhance image details (related to SSIM). RFANet [30] proposed an effective enhanced spatial attention (ESA) block. Since then, some efficient image super-resolution (EISR) methods [25, 29] introduced ESA to model the spatial position relationships. Recently, due to the success of Transformer [40] in the field of vision, Transformer-based methods have also been introduced into SR tasks. SwinIR [27] introduced Swin Transformer [31] into SR task and HAT [5] introduced CA into SwinIR to obtain significant performance improvements.

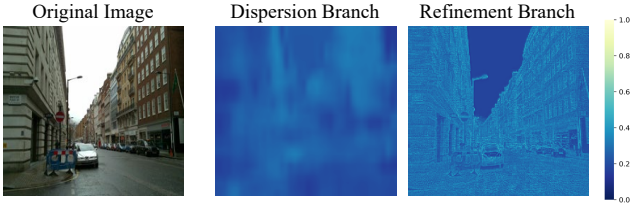
3. Method

3.1. Rethinking the Enhanced Spatial Attention

The original enhanced spatial attention (ESA) module in RFDN [29] and BSRN [25] has greatly improved the performance of the networks with a few parameters. It uses one large-size strided max-pooling to squeeze spatial information, calculates the attention weight on the down-sampled feature, and finally uses bilinear interpolation to disperse the attention map to the original spatial size. Then, the high-resolution features before spatial compression were directly mapped to the end of the block by a 1×1 convolution. ESA with a larger receptive field can well consider the prior knowledge of spatial information redundancy to simplify the mapping of the model. As shown in Fig. 2 (a), ESA consists of two parts: Branch-A and Branch-B. We activate both branches separately by the sigmoid function and visualize the results as shown in Fig. 2 (b). Interestingly, we found that Branch-A was mainly used to generate fine attention map in original resolution space, whereas Branch-B was used to generate dispersion attention map in lower resolution space because of feature spatial compression and attention weight dispersion (*i.e.*, maximum pooling with bi-



(a) Architecture of ehanced spatial attention (ESA).



(b) Visualization results of different branches in ESA.

Figure 2. Enhanced spatial attention (ESA) architecture and the visualization results. (a) The specific architecture of ESA. (b) Visualization results of different branches in ESA.

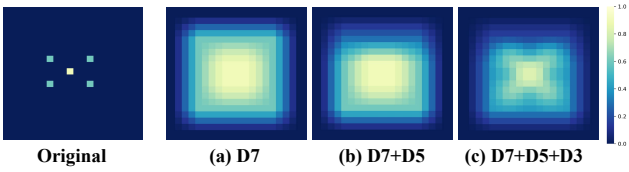


Figure 3. Rethinking the dispersion branch. Branch DN means the original image is processed by N × N max-pooling to squeeze spatial information and interpolated to the original spatial size (*i.e.*, N × N dispersion branch). (a) The original image is processed by branch D7. (b) Processing the original image by different branches D7 and D5 respectively, and adding them. (c) Processing by different branches D7, D5 and D3 respectively, and adding them.

linear interpolation). For convenience in description, the two branches are referred to as the refinement branch and the dispersion branch, respectively.

However, because ESA only uses a single large-size pooling kernel in the dispersion branch, as shown in Fig. 3 (a), surrounding unimportant information can easily be given high weights, and areas that contain true key information cannot be paid attention to.

Based on the thought of coarse-to-fine, we extend the single-level dispersion branch to multiple levels to focus the dispersion attention scores on key areas, making such areas gradually more prominent, as shown in Fig. 3 (b) and (c).

3.2. Multi-level Dispersion Spatial Attention (MDSA)

By rethinking ESA, we introduce the multi-level dispersion spatial attention (MDSA), which extends the single-level dispersion branch to a multiple-level fashion, as shown in Fig. 4. Compared to ESA, MDSA can mine the areas where key information is distributed. To better focus on the image patches that contain more structural information, we introduce local variance to represent structure information. The specific process of MDSA is shown in Fig. 6 (c). It is worth noting that local variance calculation is added only once, and the kernel size and stride remain consistent with the max-pooling of 7×7. In addition, we also reduce the depth of the original Conv Group. Therefore, MDSA does not introduce too much model complexity compared to ESA. Specifically, for an input F_{s_in}, the first step is to reduce the channel dimensions of feature by one 1×1 convolution layer $H^{cr}_{conv\_1}(\cdot)$ to ensure lightweight. This process can be formulated as

$$F_{cr} = H^{reduction}_{conv\_1}(F_{s\_in}), \tag{1}$$

and then the channel reduction feature F_{cr} is used to generate the spatial attention maps, including refinement attention map and dispersion attention map generated by different branches, which can be formulated as

$$\begin{aligned} F^i_{scr} &= H_{pool\_i}(H^{stride}_{conv\_3}(F_{cr})), i = 1, 2, 3, \\ Amap^i_{LR} &= H^i_g(F^i_{scr}), \\ Amap^i_{HR} &= H_{inter}(Amap^i_{LR}), \\ Amap_{HR} &= \sum_{i=1}^{3} Amap^i_{HR} + H_{conv\_1}(F_{cr}), \end{aligned} \tag{2}$$

where i means the i-th branch of dispersion. F^i_{scr} is the output of F_{cr} after spatial squeeze in the i-th dispersion branch. $H^{stride}_{conv\_3}(\cdot)$ and $H_{conv\_1}(\cdot)$ represent 3×3 convolution with stride of 2 and 1×1 convolution, $H^i_g(\cdot)$ denotes convolution group composed of two BSConv [11], and $H_{inter}(\cdot)$ represents bilinear interpolation which is used to upsample the low-resolution attention map $Amap^i_{LR}$ into the high-resolution space $Amap^i_{HR}$. Finally, we expand the channel number of the attention map $Amap_{HR}$ that combines multi-level information to be consistent with the input feature, and then use the combined attention map $Amap^{\uparrow}_{HR}$ to process the input feature F_{s_in}. This process can be formulated as

$$\begin{aligned} Amap^{\uparrow}_{HR} &= H^{expansion}_{conv\_1}(Amap_{HR}), \\ F_{s\_out} &= F_{s\_in} \otimes Sigmoid(Amap^{\uparrow}_{HR}) + F_{s\_in}, \end{aligned} \tag{3}$$

where symbol $\otimes$ denotes element-wise multiplication operation.

3.3. Enhanced Contrast-aware Channel Attention (ECCA)

RCAN [44] combines residual structure with channel attention and achieves impressive performance improve-

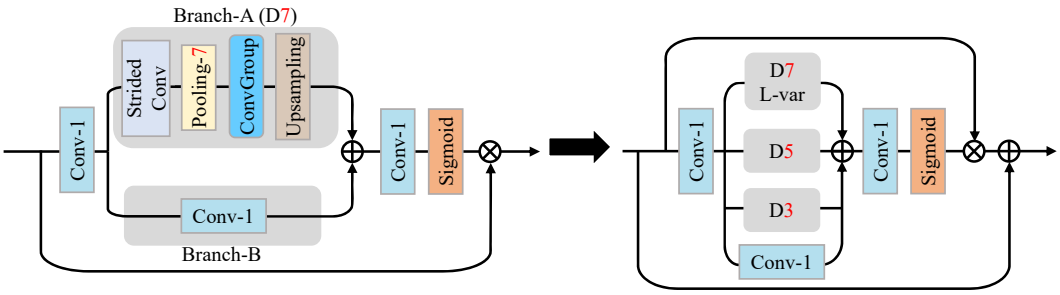


Figure 4. The evolutionary design from ESA to MDSA. (a) ESA with a single level dispersion branch. (b) MDSA extends a single level dispersion branch to multiple levels and finally adds them, where 'L-var' denotes the local variance calculation.

ment. IMDN [16] proposes contrast-aware channel attention (CCA) by taking both mean and standard deviation into account to select channel information and won the championship in the 2019 AIM Constrained Image Super-Resolution Challenge [42]. Inspired by these methods, we try to combine residual structure with CCA. As shown in Fig. 5 (b), it is worth noting that to avoid introducing too many parameters, we use the blueprint shallow residual block (BSRB) [25] to replace the vanilla convolution. Subsequently, inspired by the design of the switch spatial attention module in VapSR [47], we further remove the residual connection in the BSRB and insert the CCA module between the point-wise and the depth-wise convolution.

Finally, the enhanced contrast-aware channel attention (ECCA) module is constructed, as shown in Fig. 5 (c). This process can be formulated as

$$\begin{aligned} F_c &= GELU(H_{conv\_1}(F_{c\_in})), \\ F_c &= H_{CCA}(F_c), \\ F_{c\_out} &= H_{dwconv\_3}(F_c) + F_{c\_in}, \end{aligned} \tag{4}$$

where $H_{conv\_1}(\cdot)$ and $H_{dwconv\_3}(\cdot)$ represent the point-wise and depth-wise convolution with a kernel size of 3. F_{c_in}, F_{c_out} are the input and output features of ECCA respectively.

3.4. Enhanced Attention Distillation Block (EADB)

The specific process is shown in Fig. 6. Inspired by ESDB in BSRN [25], we design the enhanced attention distillation block (EADB), which is more efficient and powerful due to the introduction of two proposed attention mechanisms MDSA and ECCA. Specifically, given input F_{in}, the feature distillation can be formulated as

$$\begin{aligned} F_{d\_1}, F_{r\_1} &= D_1(F_{in}), R_1(F_{in}), \\ F_{d\_2}, F_{r\_2} &= D_2(F_{r\_1}), R_2(F_{r\_1}), \\ F_{d\_3}, F_{r\_3} &= D_3(F_{r\_2}), R_3(F_{r\_2}), \\ F_{d\_4} &= D_4(F_{r\_3}), \end{aligned} \tag{5}$$

where $D_i(\cdot)$ and $R_i(\cdot)$ represent the i-th distillation and refinement layers, respectively. F_{d_i}, F_{r_i} are i-th distilled and

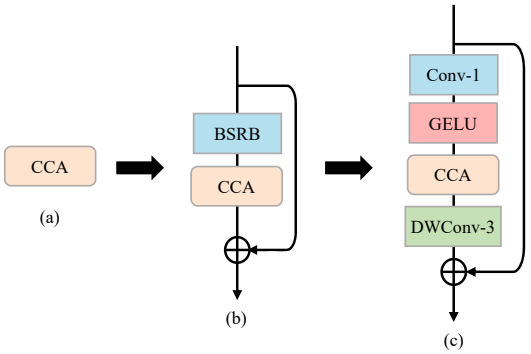


Figure 5. The evolutionary design from CCA to ECCA. (a) Base module CCA. (b) Combining CCA and BSRB in the way of residual structure. (c) Switching the CCA module to the middle of convolution layers and remove residual connection in BSRB.

refined features. Subsequently, the distilled features from different distillation layers are concatenated and fused as

$$F_{fused} = H_{conv\_1}(Concat(F_{d\_1}, F_{d\_2}, F_{d\_3}, F_{d\_4})), \tag{6}$$

where F_{fused} is the fused feature. Finally, the improved spatial attention MDSA and channel attention ECCA are used to further enhance the representation ability of the network more effectively as

$$F_{enhanced} = H_{ECCA}(H_{MDSA}(F_{fused})) \tag{7}$$

where $H_{MDSA}(\cdot)$, $H_{ECCA}(\cdot)$ represent spatial and channel attention modules MDSA and ECCA, and $F_{enhanced}$ is the enhanced feature.

3.5. Network Structure

The overall structure of the network is shown in Fig. 1. Given an input I_{LR} , we use BSConv to perform shallow feature extraction. This process can be described as follows

$$F_0 = H_{SF}(I_{LR}) \tag{8}$$

where $H_{SF}(\cdot)$ and F_0 represent the shallow feature extraction module and its output respectively. Then we feed F_0

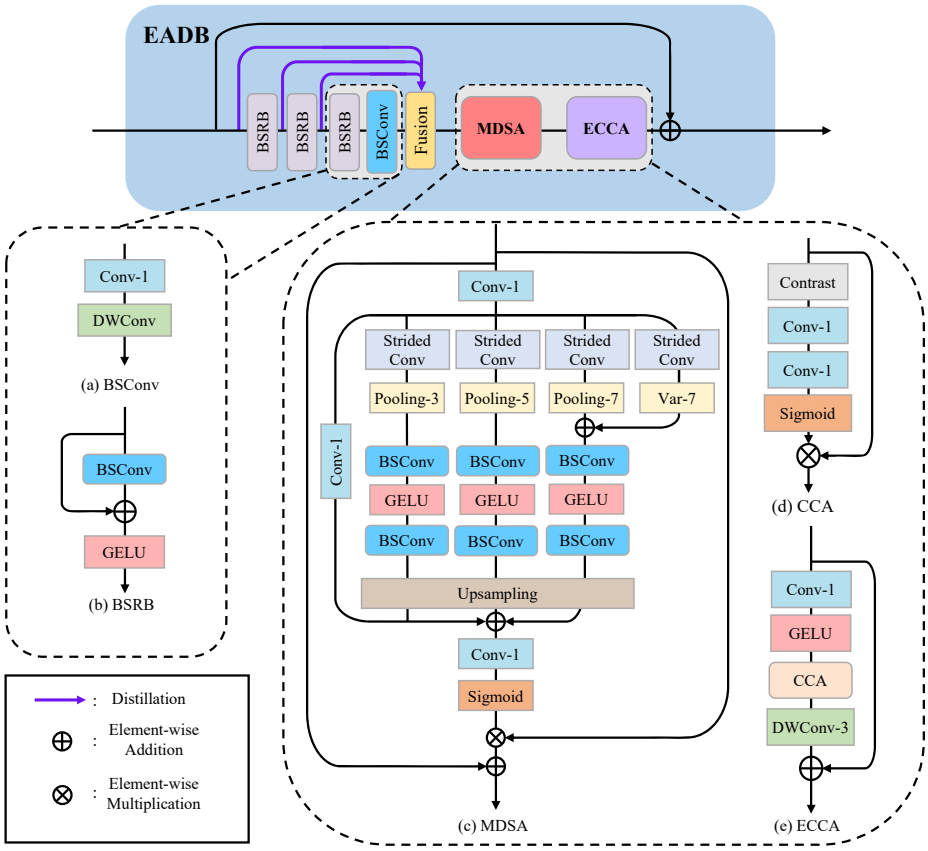


Figure 6. The specific architecture of the proposed enhanced attention distillation block (EADB). (a) BSConv: blueprint separable convolution. 'DWConv' means depth-wise convolution. (b) BSRB: Blueprint Shallow Residual Block. (c) MDSA: multi-level dispersion spatial attention. (d) CCA: contrast-aware channel attention. (e) ECCA: Enhanced contrast-aware channel attention.

into a stack of multi-level dispersion attention blocks (MDABs) to extract deep features. This process can be described as follows

$$F_m = H_m(F_{m-1}),\ m = 1, 2, \dots n, \tag{9}$$

where $H_m(\cdot)$ represents m-th MDAB. F_{m-1} and F_m denote the input feature and the output feature of m-th MDAB. To take advantage of the hierarchical features, we then use fusion module containing 1×1 convolution and GELU [12] to fuse the features of different layers, as:

$$F_{fused} = GELU(H_{conv\_1}(Concat(F_1, \dots F_n))) \tag{10}$$

where F_{fused} represents the fused features. Finally, in the reconstruction phase, the output I_{SR} of the model is generated by the following process

$$I_{SR} = H_{rec}(F_{fused} + F_0) \tag{11}$$

where $H_{rec}(\cdot)$ represents the reconstruction function containing a 3×3 convolution and sub-pixel operation [37]. L_1 loss function is used to optimize our model, which is expressed as follows

$$L_1 = \| I_{SR} - I_{HR} \|_1 \tag{12}$$

4. Experiments

4.1. Experimental Setup

Datasets and Metrics. The training dataset consists of 800 images from DIV2K [1] and the first 10k images from LSDIR [22]. Standard benchmark datasets used for evaluation include Set5 [4], Set14 [41], BSD100 [34], Urban100 [15], and Manga109 [35], PSNR and SSIM on the Y channel (*i.e.*, luminance) are used as the evaluation metrics.

Implementation details. MDRN consists of 8 EADBs and the number of channels is set to 56. As a small variant of MDRN, MDRN-S is used for the challenge and the channel number is set to 28. During training, 64×64 patches are randomly cropped from LR images as input. Training dataset was further employed data augmentation by horizontal flipping and 90-degree rotations. The model

Table 1. The performance of MDRN with the different attention modules.

Method	Params[K]	DIV2K_val (RGB) PSNR/SSIM	Set5 PSNR/SSIM	Set14 PSNR/SSIM	Urban100 PSNR/SSIM	BSDS100 PSNR/SSIM	Manga109 PSNR/SSIM
ESA+CCA	287	29.07/0.8210	32.26/0.8957	28.70/0.7841	26.31/0.7926	27.62/0.7377	30.71/0.9113
MDSA+ECCA	322	29.15/0.8230	32.33/0.8964	28.75/0.7848	26.43/0.7958	27.66/0.7391	30.88/0.9129

Table 2. Ablation study on the proposed MDSA. 'Base-D7' means the base model with ESA reduces the depth of ConvGroup and introduces a residual connection. 'DN' represents the dispersion branch that uses the max-pooling of N $\times$ N size. 'L-var' denotes the local variance calculation.

Method	Params[K]	DIV2K_val (RGB) PSNR/SSIM	Set5 PSNR/SSIM	Set14 PSNR/SSIM	Urban100 PSNR/SSIM	BSDS100 PSNR/SSIM	Manga109 PSNR/SSIM
wo ESA	280	28.97/0.8186	32.14/0.8950	28.60/0.7822	26.08/0.7848	27.59/0.7364	30.51/0.9086
Base (w/ ESA)	317	29.09/0.8214	32.30/0.8962	28.70/0.7839	26.37/0.7934	27.63/0.7380	30.77/0.9115
Base-D7	303	29.08/0.8212	32.23/0.8958	28.72/0.7845	26.33/0.7921	27.64/0.7382	30.78/0.9118
+D5	311	29.12/0.8221	32.32/0.8962	28.73/0.7842	26.39/0.7942	27.64/0.7385	30.84/0.9126
++D3	319	29.13/0.8226	32.33/0.8966	28.73/0.7847	26.42/0.7953	27.65/0.7390	30.89/0.9128
+++L-var (MDSA)	322	29.15/0.8230	32.33/0.8964	28.75/0.7848	26.43/0.7958	27.66/0.7391	30.88/0.9129

is trained by using the Adam optimizer [20] with β_1=0.9, β_2=0.999. The total training iterations are set 1000k with mini-batch size 64. The initial learning rate is initialized as 2e-3 and halved at [100k, 500k, 800k, 900k, 950k]-step. The model training is implemented by Pytorch framework on two NVIDIA RTX 3090 GPUs.

4.2. Ablation Study

In this section, we first verify the effect of different attention modules for MDRN. Subsequently, we implement comprehensive experiments to study the impact of different designs of MDSA and ECCA respectively.

Effectiveness of Attention Components. We take the model equipped with ESA and CCA as the baseline to verify the effectiveness of the proposed attention module. The results are shown in Tab. 1. Compared with ESA and CCA, the performance of MDRN equipped with MDSA and ECCA has been greatly improved with only 12.19% increase of parameters. Specifically, it has been improved by 0.08dB on the DIV2K validation set (100 images) and 0.45dB on other common benchmarks.

Study of Design in MDSA. We further conduct more detailed experiments to analyze the impacts of different designs step by step. First, we verify the impact of the spatial attention (SA) mechanism on the performance of MDRN. Comparing the first two rows in Tab. 2, we can see that the base model has a significant improvement at the cost of only 37K (11.67%) more parameters compared to the model without ESA. In particular, the DIV2K validation set is improved by 0.12dB. So, the improvement brought by the spatial attention mechanism to the SR model cannot be ignored. After that, we reduce the depth of the convolution group (ConvGroup) in MDSA. The performance decreased by only 0.01dB on the DIV2K validation set, and the other benchmark datasets have an average drop of 0.01dB, which

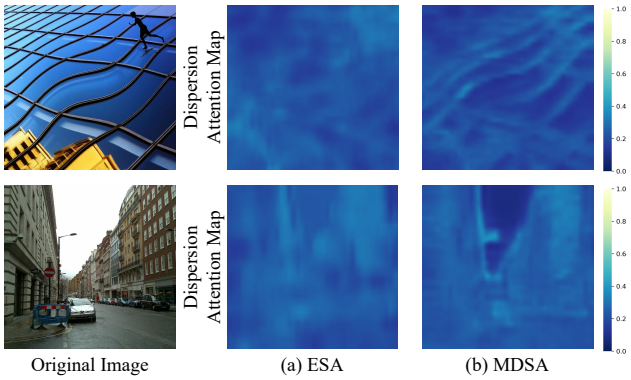


Figure 7. Visulization results of the dispersion attention map from dispersion branch. (a) The attention map of dispersion branch in ESA. (b) The attention map of dispersion branch in MDSA.

is consistent with the conclusion of RLFN that the original ConvGroup was redundant for ESA. Then, based on the model with shallow ConvGroup, we gradually introduce multi-scale spatial compression and dispersion processes (*i.e.*, D5 and D3 in Tab. 2.) and clearly observe the gradual improvement of performance. Subsequently, we introduce local variance calculation to improve the ability of the network to capture texture information, which further improves the performance as shown in the last row of Tab. 2. Finally, the proposed MDSA was successfully constructed. Compared to the model without ESA, MDRN brought a 0.18dB improvement on the DIV2K validation set with an increase of 42K (15%) parameters, as well as a 1.13dB improvement on the other datasets. Compared to the model with ESA, MDRN brings a 0.06dB improvement on the DIV2K validation set and a 0.28dB improvement on the other benchmark datasets with only a 5K (1.58%) increase in the number of parameters. We found MDSA can better distribute

Table 3. Ablation study on the proposed ECCA. 'LR' means the base model with CCA introduces lightweight convolution layers and a residual connection. 'Switch' denotes CCA is switched to the middle of the lightweight convolution layers.

Method	Params[K]	DIV2K_val (RGB)	Set5	Set14	Urban100	BSDS100	Manga109
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
wo CCA	289	29.10/0.8217	32.28/0.8960	28.72/0.7843	26.34/0.7934	27.64/0.7383	30.79/0.9121
Base (w/ CCA)	292	29.09/0.8216	32.28/0.8963	28.70/0.7841	26.36/0.7935	27.63/0.7384	30.80/0.9121
+LR	322	29.13/0.8226	32.30/0.8963	28.74/0.7848	26.40/0.7948	27.66/0.7392	30.85/0.9128
++Switch (ECCA)	322	29.15/0.8230	32.33/0.8964	28.75/0.7848	26.43/0.7958	27.66/0.7391	30.88/0.9129

Table 4. Ablation study on different branches of MDSA. 'Refine' represents refinement branch, and 'Dispersion' dispersion branch.

Method	#Params	DIV2K_val (RGB)	Set5	Set14	Urban100	BSDS100	Manga109
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
wo Dispersion	295K	29.01/0.8192	32.24/0.8957	28.66/0.7832	26.24/0.7892	27.60/0.7371	30.65/0.9103
wo Refine	320K	29.11/0.8221	32.33/0.8965	28.75/0.7848	26.38/0.7941	27.65/0.7389	30.84/0.9126
Dispersion+Refine	322K	29.15/0.8230	32.33/0.8964	28.75/0.7848	26.43/0.7958	27.66/0.7391	30.88/0.9129

Table 5. Quantitative comparison (average PSNR/SSIM) with state-of-the-art methods, and multiply-accumulate operations is evaluated on a 1280 $\times$ 720 HQ image. The best and second-best performance are in red and blue colors, respectively.

Method	Scale	Params	Multi-Adds	Set5	Set14	BSD100	Urban100	Manga109
				PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
EDSR-Baseline [28]	×2	1370K	316.3G	37.99/0.9604	33.57/0.9175	32.16/0.8994	31.98/0.9272	38.54/0.9769
IMDN [16]		694K	158.8G	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
RFDN [29]		534K	123.0G	38.05/0.9606	33.68/0.9184	32.16/0.8994	32.12/0.9278	38.88/0.9773
LatticeNet [33]		756K	169.5G	38.15/0.9610	33.78/0.9193	32.25/0.9005	32.43/0.9302	—
RLFN [21]		527K	115.4G	38.07/0.9607	33.72/0.9187	32.22/0.9000	32.33/0.9299	—
FMEN [9]		748K	172.0G	38.10/0.9609	33.75/0.9192	32.26/0.9007	32.41/0.9311	38.95/0.9778
BSRN [25]		332K	73.0G	38.10/0.9610	33.74/0.9193	32.24/0.9006	32.34/0.9303	39.14/0.9782
MDRN (ours)		304K	65.0G	38.11/0.9610	33.84/0.9205	32.32/0.9016	32.84/0.9350	39.14/0.9782
EDSR-Baseline [28]	×3	1,555K	160.2G	34.37/0.9270	30.28/0.8417	29.09/0.8052	28.15/0.8527	33.45/0.9439
IMDN [16]		703K	71.5G	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
RFDN [29]		541K	55.4G	34.41/0.9273	30.34/0.8420	29.09/0.8050	28.21/0.8525	33.67/0.9449
LatticeNet [33]		765K	76.3G	34.53/0.9281	30.39/0.8424	29.15/0.8059	28.33/0.8538	—
FMEN [9]		757K	77.2G	34.45/0.9275	30.40/0.8435	29.17/0.8063	28.33/0.8562	33.86/0.9462
BSRN [25]		340K	33.3G	34.46/0.9277	30.47/0.8449	29.18/0.8068	28.39/0.8567	34.05/0.9471
MDRN (ours)		311K	29.6G	34.58/0.9286	30.51/0.8453	29.21/0.8081	28.70/0.8627	34.07/0.9476
EDSR-Baseline [28]	×4	1518K	114.0G	32.09/0.8938	28.58/0.7813	27.57/0.7357	26.04/0.7894	30.35/0.9067
IMDN [16]		715K	40.9G	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
RFDN [29]		550K	31.6G	32.24/0.8952	28.61/0.7819	27.57/0.7360	26.11/0.7858	30.58/0.9089
LatticeNet [33]		777K	43.6G	32.30/0.8962	28.68/0.7830	27.62/0.7367	26.25/0.7873	—
RLFN [21]		543K	29.8G	32.24/0.8952	28.62/0.7813	27.60/0.7364	26.17/0.7877	—
FMEN [9]		769K	44.2G	32.24/0.8955	28.70/0.7839	27.63/0.7379	26.28/0.7908	30.70/0.9107
BSRN [25]		352K	19.4G	32.35/0.8966	28.73/0.7847	27.65/0.7387	26.27/0.7908	30.84/0.9123
MDRN (ours)		322K	17.3G	32.35/0.8970	28.80/0.7861	27.69/0.7404	26.60/0.8005	31.02/0.9146

the attention weight to surrounding important regions than ESA, which is consistent with the analysis in the method part. In addition, to better reflect the influence of multi-level dispersion, we visualize the attention map of the dispersion branch, as shown in Fig. 7, we can see that the way of multi-level dispersion can better locate the important area.

Study of Design in ECCA. We conduct detailed experiments to verify the effectiveness of our proposed ECCA. First, we verified the impact of CCA on MDRN. As shown in Tab. 3, from the first two rows of the table, we can observe that CCA only brings about a weak performance improvement compared with the model without CCA. Subsequently, we combine BSRB [25] with CCA in the way of residual structure, and then further remove the residual connection in BSRB and insert the CCA module between point-wise convolution and depth-wise convolution, as shown in Fig. 5 (b) and (c). From the last two rows of Tab. 3, we can observe that ECCA brings significant performance improvement, compared with the base model with CCA.

Study of Different Branches. In this section, we verify the effect of different branches on the performance of MDRN. As shown in Tab. 4, we can observe that the performance of MDRN with both dispersion and refinement branches is higher than the two models using only one single branch. Specifically, MDRN achieves a 0.19dB and 0.05dB improvement on Urban100 with parameter increases of 9.2% and 0.6% respectively. This demonstrates that both dispersion and refinement branches are beneficial for reconstruction.

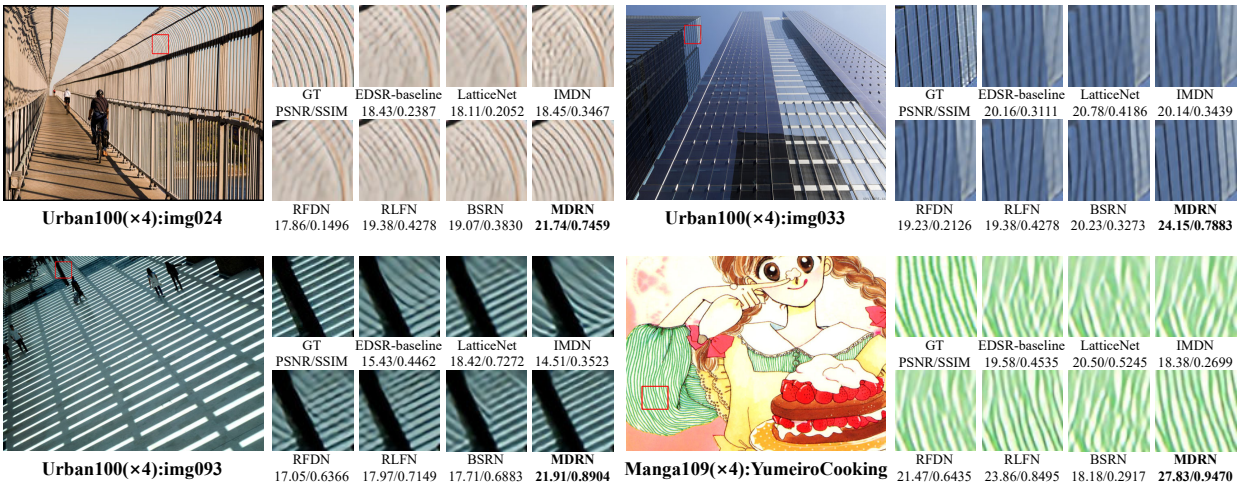


Figure 8. Visual comparison for ×4 SR. The patches for comparison are marked with red boxes in the original images. PSNR/SSIM is calculated based on the patches to better reflect the performance difference.

Table 6. Results of NTIRE 2023 Efficient Super-Resolution Sub-Track 1: Model Complexity. ‡ denotes the results in NTIRE 2022 Efficient Super-Resolution [23] challenge.

Team	Val PSNR	Test PSNR	Params[M]	FLOPs[G]	Acts[M]	Mem[M]	Runtime[ms]
TelunXupt (ours)	29.00	27.09	0.095	5.58	220.88	517.14	75.89
FRL_Team0	29.01	26.98	0.115	7.38	170.26	2028.66	196.64
Dase-DEALab	29.00	27.07	0.118	9.06	332.39	1114.77	130.73
Set5_Baby	29.01	27.08	0.129	8.29	202.70	652.41	99.79
FRL_Team4	28.95	27.02	0.173	10.60	187.32	1266.92	124.13
XPixel‡	29.01	—	0.156	9.50	65.76	729.94	—
NJUST_ESR‡	28.96	—	0.176	8.73	160.43	1346.74	—

4.3. Comparison with State-of-the-art Methods

We compare the proposed MDRN with other EISR works, including EDSR-Baseline [28], IMDN [16], RFDN [29], LatticeNet [33], RLFN [21], FMEN [9], BSRN [25], as shown in Tab. 5. Our MDRN achieves the best performance on all datasets with fewer parameters and Multi-Adds compared with other methods. Specifically, on the Urban100 dataset, the PSNR is 0.43dB and 0.33dB higher than RLFN and BSRN on ×4 SR, respectively. To demonstrate the restoration performance, visualization comparisons are shown in Fig. 8.

4.4. MDRN-S for NTIRE2023 Challenge

As a variant of MDRN, our MDRN-S won the 1st place in the NTIRE2023 Efficient Super-Resolution Challenge [24] Sub-Track 1: Model Complexity. The results are shown in Tab. 6. Specifically, 'Val PSNR' is the PSNR result tested on the validation set of 100 images from DIV2K and 'Test PSNR' is performed on a test set consisting of 100 LR test images from DIV2K and 100 LR test images from LSDIR. Compared to other competing solutions, our method has the fewest number of parameters and FLOPs and the best performance on the test set.

5. Conclusion

In this paper, we propose a multi-level dispersion residual network (MDRN) for efficient image super-resolution (EISR). The design of MDRN is inspired by blueprint separable residual network (BSRN). We adopt the similar architecture of BSRN but introduce a more efficient enhanced attention distillation block (EADB) by replacing original spatial and channel attention mechanisms with the proposed multi-level dispersion spatial attention (MDSA) and enhanced contrast-aware channel attention (ECCA). Specifically, MDSA divides the original attention map calculation into two branches: the refinement branch and the dispersion branch. For the dispersion branch, MDSA introduces multi-scale information to improve the original dispersion branch (*i.e.*, the branch using a single large-size pooling and interpolation operations) in the enhanced spatial attention (ESA) in BSRN, based on the thought of coarse-to-fine, which allows attention weight to better focus on the areas where important information is distributed. ECCA effectively combines CCA and lightweight convolution layers to optimize the input and output feature information of CCA. Extensive experiments show the effectiveness of the proposed methods. Our method achieves the best performance with lower model complexity compared to the state-of-the-art efficient SR methods. Besides, as a variant of MDRN, MDRN-S won the first place in the model complexity track of the NTIRE 2023 efficient super-resolution challenge.

Acknowledgements. This work was supported in part by the Shaanxi International Science and Technology Cooperation Program (2021KW-06) and the National Natural Science Foundation of China under Grant 41831072 and Grant 41874173.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 5

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European conference on computer vision (ECCV)*, pages 252–268, 2018. 1

[3] Mustafa Ayazoglu. Imdeception: Grouped information distilling super-resolution network. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 755–764, 2022. 1

[4] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 5

[5] Xiangyu Chen, Xintao Wang, Jiantao Zhou, and Chao Dong. Activating more pixels in image super-resolution transformer. *arXiv preprint arXiv:2205.04437*, 2022. 1, 2

[6] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11065–11074, 2019. 2

[7] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13733–13742, 2021. 1

[8] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014. 1

[9] Zongcai Du, Ding Liu, Jie Liu, Jie Tang, Gangshan Wu, and Lean Fu. Fast and memory-efficient network towards efficient image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 853–862, 2022. 2, 7, 8

[10] Jinsheng Fang, H. Lin, Xinyu Chen, and Kun Zeng. A hybrid network of cnn and transformer for lightweight image super-resolution. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1102–1111, 2022. 1

[11] Daniel Haase and Manuel Amthor. Rethinking depthwise separable convolutions: How intra-kernel correlations lead to improved mobilenets. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14600–14609, 2020. 2, 3

[12] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. 5

[13] Zejiang Hou and Sun-Yuan Kung. Efficient image super resolution via channel discriminative deep neural network pruning. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3647–3651. IEEE, 2020. 1

[14] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018. 2

[15] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015. 5

[16] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia*, pages 2024–2032, 2019. 1, 2, 4, 7, 8

[17] Zheng Hui, Xiumei Wang, and Xinbo Gao. Fast and accurate single image super-resolution via information distillation network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 723–731, 2018. 1, 2

[18] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016. 1

[19] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016. 2

[20] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014. 6

[21] Fangyuan Kong, Mingxi Li, Songwei Liu, Ding Liu, Jingwen He, Yang Bai, Fangmin Chen, and Lean Fu. Residual local feature network for efficient super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 766–776, 2022. 7, 8

[22] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 5

[23] Yawei Li, Kai Zhang, Radu Timofte, Luc Van Gool, Fangyuan Kong, Mingxi Li, Songwei Liu, Zongcai Du, Ding Liu, Chenhui Zhou, et al. Ntire 2022 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1062–1102, 2022. 8

[24] Yawei Li, Yulun Zhang, Luc Van Gool, Radu Timofte, et al. Ntire 2023 challenge on efficient super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2, 8

[25] Zheyuan Li, Yingqi Liu, Xiangyu Chen, Haoming Cai, Jinjin Gu, Yu Qiao, and Chao Dong. Blueprint separable residual network for efficient image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 833–843, 2022. 1, 2, 4, 7, 8

[26] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3867–3876, 2019. 1

[27] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021. 1, 2

[28] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017. 7, 8

[29] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 41–55. Springer, 2020. 1, 2, 7, 8

[30] Jie Liu, Wenjie Zhang, Yuting Tang, Jie Tang, and Gangshan Wu. Residual feature aggregation network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2359–2368, 2020. 1, 2

[31] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 2

[32] Zhisheng Lu, Juncheng Li, Hong Liu, Chao Huang, Linlin Zhang, and Tieyong Zeng. Transformer for single image super-resolution. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 456–465, 2021. 1

[33] Xiaotong Luo, Yuan Xie, Yulun Zhang, Yanyun Qu, Cuihua Li, and Yun Fu. Latticenet: Towards lightweight image super-resolution with lattice block. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*, pages 272–289. Springer, 2020. 7, 8

[34] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001. 5

[35] Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76:21811–21838, 2017. 5

[36] Qian Ning, Weisheng Dong, Guangming Shi, Leida Li, and Xin Li. Accurate and lightweight image super-resolution with model-guided deep unfolding network. *IEEE Journal of Selected Topics in Signal Processing*, 15(2):240–252, 2020. 1

[37] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. 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)*, pages 1874–1883, 2016. 5

[38] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017. 2

[39] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE international conference on computer vision*, pages 4539–4547, 2017. 2

[40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2

[41] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012. 5

[42] Kai Zhang, Shuhang Gu, Radu Timofte, Zheng Hui, Xiumei Wang, Xinbo Gao, Dongliang Xiong, Shuai Liu, Ruipeng Gang, Nan Nan, LI, Xueyi Zou, Ning Kang, Zhan-Han Wang, Hang Xu, Chaofeng Wang, Zheng Li, Linlin Wang, Jun Shi, Wenyu Sun, Zhiqiang Lang, Jiangtao Nie, Wei Wei, Lei Zhang, Yazhe Niu, Peijin Zhuo, Xiangzhen Kong, Long Sun, and Wenhao Wang. Aim 2019 challenge on constrained super-resolution: Methods and results. *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 3565–3574, 2019. 4

[43] Xindong Zhang, Hui Zeng, and Lei Zhang. Edge-oriented convolution block for real-time super resolution on mobile devices. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 4034–4043, 2021. 1, 2

[44] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 2, 3

[45] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018. 1

[46] Hengyuan Zhao, Xiangtao Kong, Jingwen He, Y. Qiao, and Chao Dong. Efficient image super-resolution using pixel attention. In *ECCV Workshops*, 2020. 1

[47] Lin Zhou, Haoming Cai, Jinjin Gu, Zheyuan Li, Yingqi Liu, Xiangyu Chen, Yu Qiao, and Chao Dong. Efficient image super-resolution using vast-receptive-field attention. In *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, pages 256–272. Springer, 2023. 4