

# Asymmetric Information Distillation Network for Lightweight Super Resolution

Zhikai Zong, Lin Zha, Jiande Jiang and Xiaoxiao Liu

Qingdao Hi-image Technologies Co.,Ltd (Hisense Visual Technology Co.,Ltd.)

zzksdu@163.com, {zhalin, jiangjiande, liuxiaoxiao3}@hisense.com

## Abstract

The purpose of this paper is to design a lightweight network to achieve image super resolution performance equivalent to SRResNet. We design an asymmetric information distillation block (AIDB) with distillation information multiplexing and asymmetric information extraction capabilities to better achieve this goal. Distillation information multiplexing refers to the repeated processing of distilled information to supplement the ability of key information extraction. Asymmetric information enhancement block (AIEB) refers to identify different features in the image by the horizontal and vertical feature extraction. AIEB greatly reduces the number of parameters, and distillation information multiplexing works as a supplement to the lost high dimensional information. A large number of experiments show that our asymmetric information distillation network (AIDN) achieves a better balance of performance and complexity than SOTA model. Moreover, Our proposed AIDN ranked second in the model complexity track of NTIRE2022 efficient super resolution challenge. Compared with the first place in this track, we achieves higher PSNR performance on testset with a slight disadvantage in the number of parameters. The code is available at <https://github.com/zzksdu/AIDN>.

## 1. Introduction

Single image super resolution (SISR) is a classic low-level computer vision task, which aims to reconstruct a super resolution (SR) image from a single low resolution (LR) image. Since SRCNN [3] was proposed, many deep learning methods [5, 10, 16, 17, 20, 23, 28, 31, 32] have been proposed to improve the performance of image super resolution. SRCNN [3] is the first attempt to reconstruct SR from LR using convolution neural network (CNN). Subsequently, Kim et al. [12] achieved better super resolution performance by increasing the network depth of SRCNN [3], which also shows that the network depth is closely related to the super resolution performance. Bee et al. [16] made greater progress in super resolution by introducing residual struc-

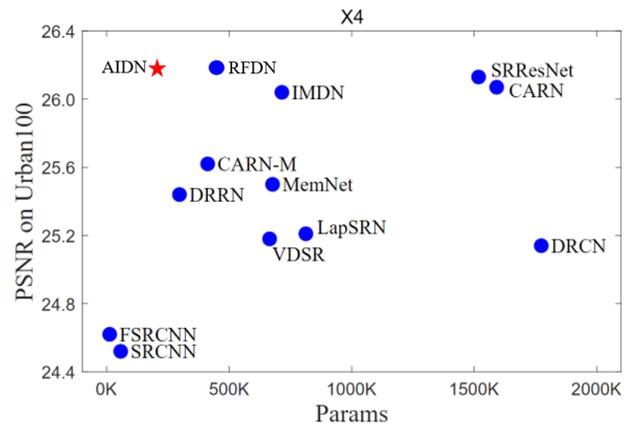


Figure 1. performance and parameters comparison between our AIDN and other State-of-the-art light weight network on urban100 dataset for upscaling factor  $\times 4$

tures that further increased the depth of the network.

In this paper, our purpose is to design a lightweight network structure, which can achieve the image super resolution performance equivalent to SRResNet [15]. Before that, lots of works [11, 23, 32] have been proposed to reduce the number of parameters in super resolution network, and achieved good performance. What impresses us is that IMDN [8] improved IDN [9] by introducing information multiple distillation block (IMDB), and got the first place in AIM2019 efficient super resolution challenge. RFDN [20] on the basis of IMDN [8] by replacing split operation and proposing shallow residual block (SRB) further reduce the runtime of SR while maintaining the super resolution performance. Although there are so many methods try to lightweight super resolution network, many works still have large amount of parameters and huge computational complexity. To solve this problem, NTIRE2022 held the efficient super resolution challenge [18], which aims to require participants to use the least computing resources to achieve performance equivalent to SRResNet [15]. Moreover, PSNR requires a minimum of 29dB, and any evaluation of parameter, flops, runtime, memory, and activation layer are lower than SRResNet [15]. This challenge will

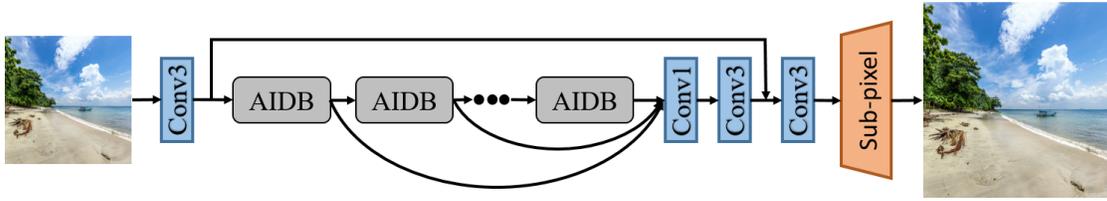


Figure 2. The architecture of asymmetric information distillation network (AIDN)

greatly promote the development of lightweight super resolution network. In order to minimize the parameters, we propose a very simple and effective super resolution network (AIDN) with less computational complexity. Moreover, in model complexity track we achieved the second place among all the effective submissions. And compared with the first place in this track, we achieved higher PSNR performance on testset with a slight disadvantage in the number of parameters. In the rest of this paper, we will introduce our method. First we rethink the network structure of RFDN [20]. RFDN [20] is composed of four residual feature distillation blocks (RFDB). In RFDB, the multistage distillation structure is used to extract the key information of SR reconstruction layer by layer. The input features are divided into two parts in RFDB. One part of the information is retained through simple CNN calculation, and the other part enhances the nonlinear ability by SRB module, which further improves the information extraction ability. We found that SRB in each RFDB can be completely equivalent to a convolution calculation in the final inference stage. In other words, there are many redundant calculations in RFDB. In order to further reduce the channel dimension of RFDB, we reduce the channel dimension at the beginning of each block and reduce redundant calculations. Moreover, we use asymmetric convolution in different directions to replace the original traditional convolution operation, which reduces the amount of parameters by 50% compared with RFDB. However by reducing the amount of parameters directly, the performance of super resolution will inevitably weaken with the reduction of network parameters. Therefore, we multiplex the distilled information obtained previously to enhance the distillation information behind in every block, so as to achieve better super resolution performance. The main contributions of this paper can be summarized as follows:

1. We propose an asymmetric information enhancement block (AIEB) to achieve faster and better image super resolution performance. Moreover, we reach the second place in the model complexity track of efficient super resolution challenge [18].

2. We designed an AIDB module. It has the same super resolution performance as the baseline network (SRResNet [15]). However, the number of parameters is reduced by

84% and the number of flops is reduced by 91.5%.

3. We systematically summarized the design of RFDN [20]. Based on our understanding of RFDN [20], we redesigned our model structure to realize the lightweight super resolution network.

## 2. Related work

In recent years, methods based on deep learning have been widely used in the field of super resolution (SR) and have made great progress. The SRCNN [3] network proposed by Dong et al. was the first attempt to which utilize convolution neural networks to address image super resolution. SRCNN [3] was a three layer convolution neural network to achieve image reconstruction in an end to end manner. Compared with SRCNN [3], VDSR [12] further improved the performance of SR by stacking 20 convolutional layers, which to some extent indicates that the depth of the network has a positive effect on the result of SR reconstruction. Kim et al. proposed the DRCN [13] structure by recursively using the feature extraction layer, which reduces the computational complexity of the SR network to a certain extent. DRRN [8] achieves better results by combining recursive and residual network schemes on the basis of DRCN [13], and reduces the amount of parameters. The laplacian pyramid super-resolution network (LapSRN [14]) takes the original LR image as input and reconstructs the subband residuals of the HR image step by step to improve speed and accuracy. For model acceleration, Zhang et al. [29] proposed a generalized singular value decomposition method to asymmetric reconstruction and improved the running speed of the model. In order to further improve the operation speed of the SR network, Shi et al. [24] designed an efficient sub-pixel convolution to upscale the resolution of the feature maps at the end of the SR model, which can make a lot of calculations in low-dimensional feature space. Similarly, fast SRCNN (FSRCNN [4]) use transposed convolutions as upsampling layers to accomplish resolution upscaling. Lim et al. [19] proposed EDSR and MDSR by removing unnecessary modules in traditional residual networks, which achieved significant improvements in the SR performance. On the basis of EDSR [19], Zhang et al. [31] proposed residual dense network (RDN) by introducing dense connections in residual

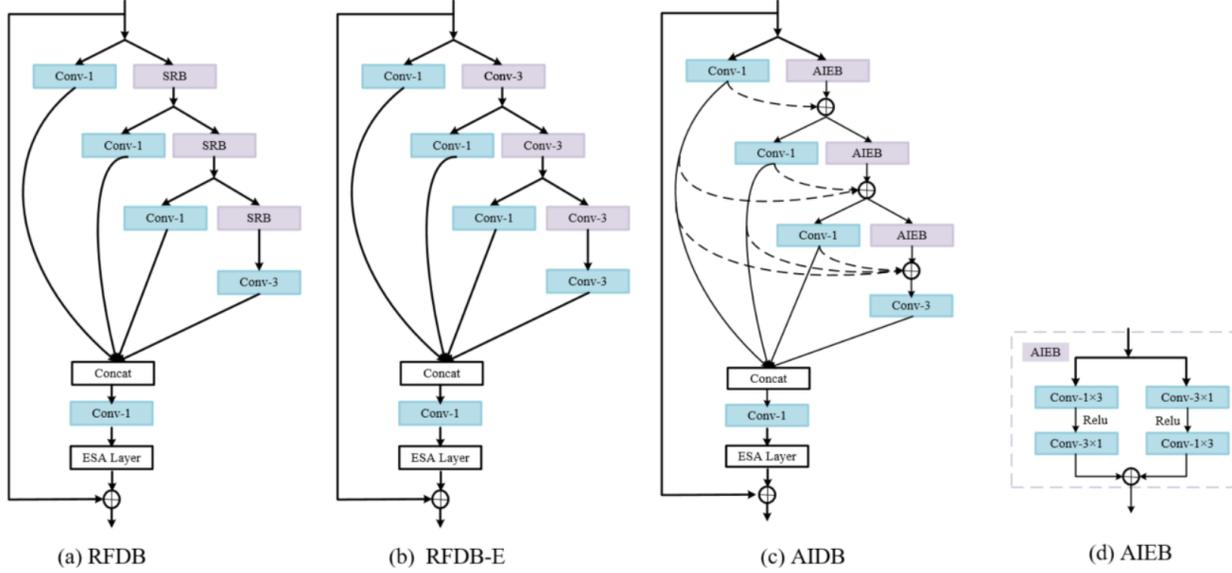


Figure 3. (a) RFDB: residual feature distillation block. (b) RFDB-E: the equivalent of RFDB. (c) AIDB: asymmetric information distillation block. (d) Asymmetric information enhancement block (AIEB)

blocks. They also introduced the channel attention mechanism into the improved SR residual block, and proposed the Residual Channel Attention Network (RCAN [30]), which utilizes global adaptive pooling to extract channel statistics. It can rescale channel-wise features to help train very deep networks by taking into account the interdependencies between channels. CAUN-M [26] improved the efficiency of super resolution by using group convolutions and achieves comparable results to models of computational complexity. Dai et al. [2] proposed SAN, which uses a second order attention mechanism to adaptively adjust the scale of features by considering feature statistics higher than first-order. Guo et al. [6] proposed a dual regression scheme, which introduced a dual branch based on the traditional SR network, which completed the process from SR to LR, thereby forming a closed-loop system and improving the performance of SR model. Despite the great success of CNN-based SR methods, most of them are not suitable for mobile devices. To address this issue, Ahn et al. [1] proposed a CARN-M model for mobile scenarios through a cascaded network architecture. Hui et al. [9] proposed an Information Distillation Network (IDN) to explicitly divide the extracted features into two parts. Building on the IDN [9], they also propose a fast and lightweight Information Multiple Distillation Network (IMDN [8]), which is the winning solution of the AIM 2019 Constrained Image Super-Resolution Challenge. On the basis of IMDN [8], RFDN [20] replaced the split operation of IMDN [8] with the traditional  $1 \times 1$  convolution, and proposed the SRB module, which won the championship of AIM2020-efficient super resolution challenge [18].

### 3. Method

#### 3.1. Residual feature distillation block

As shown in Figure3 (a), RFDN [20] proposes a more effective information distillation block (RFDB) based on IMDN, and it is still a progressive refining network. In the process of information distillation, one branch uses a single convolution for information distillation, and the other branch further enhances the input feature through a simple SRB. In the information distillation process, every stage will distill some information. Then all the distilled information is fused in the tail of block. Finally, the distilled information is then enhanced by enhanced spatial attention (ESA) layer [20]. The whole process can be expressed as equation (1).

$$\begin{aligned}
 F_{distilld.1}, F_{coarse.1} &= Conv_{1 \times 1}(F_{in}), SRB(F_{in}) \\
 F_{distilld.2}, F_{coarse.2} &= Conv_{1 \times 1}(F_{coarse.1}), SRB(F_{coarse.1}) \\
 F_{distilld.3}, F_{coarse.3} &= Conv_{1 \times 1}(F_{coarse.2}), SRB(F_{coarse.2}) \\
 F_{distilld.A} &= SRB(F_{coarse.3})
 \end{aligned} \quad (1)$$

Where  $F_{in}$  represents the input feature.  $Conv_{11}$  is the convolution operation with kernel size of  $1 \times 1$ .  $F_{distilled_i}$ ,  $F_{coarse_i}$  is the  $i$ -th distillation information and coarse information respectively. Equation 2 indicates concat the distilled features along the channel dimension.

$$F_{distilld} = Concat(F_{distilld.1}, F_{distilld.2}, F_{distilld.3}, F_{distilld.A}) \quad (2)$$

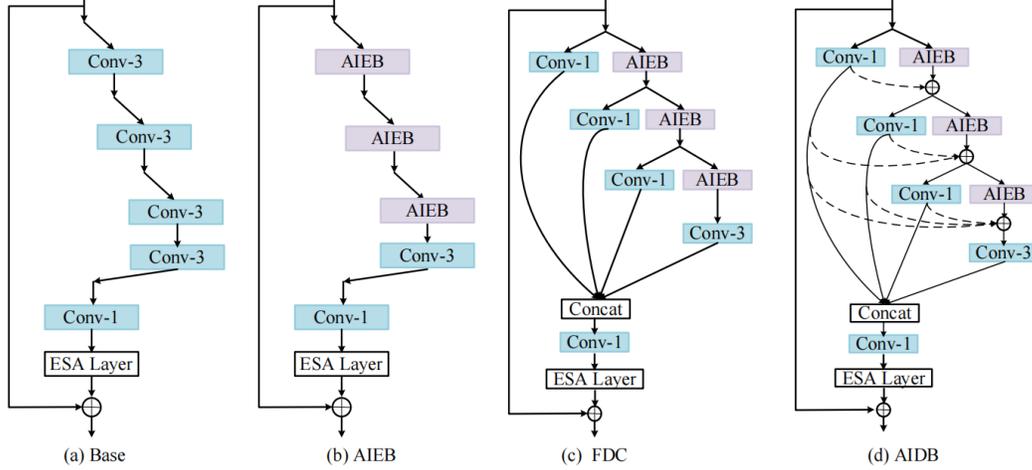


Figure 4. The Base block, AIEB block, FDC block and AIDB used in ablation study.

### 3.2. Rethinking the RFDB

Compared with IMDB, RFDB has made great progress in parameter and flops, but there are still some redundant calculations. Firstly, SRB in RFDB can be completely equivalent to an ordinary convolution operation. In other words, RFDB can be completely equivalent to Fig3 (b) with sufficient training. Then, we found that in the information distillation of the following layers within each block, the dimension of channel can be appropriately reduced, and there is no need to maintain a high dimension. As described in Fig. 3 (b), the structure of the whole block can be described as equation 3:

$$\begin{aligned}
 F_{distilld.1}, F_{coarse.1} &= Conv_{1 \times 1}(F_{in}), Conv_{3 \times 3}(F_{in}) \\
 F_{distilld.2}, F_{coarse.2} &= Conv_{1 \times 1}(F_{coarse.1}), Conv_{3 \times 3}(F_{coarse.1}) \\
 F_{distilld.3}, F_{coarse.3} &= Conv_{1 \times 1}(F_{coarse.2}), Conv_{3 \times 3}(F_{coarse.2}) \\
 F_{distilld.4} &= Conv_{3 \times 3}(F_{coarse.3})
 \end{aligned} \quad (3)$$

We call this new structure as RFDB-E, which removes obvious redundant computing. In addition, we believe that in the process of high-dimensional information extraction, a single convolution cannot provide better information distillation capability, and there should be some more effective ways to complete information extraction.

### 3.3. Asymmetric information distillation block

In this section, we will introduce our proposed AIDB, which is a lighter and more effective information extraction module. The overall structure of AIDB is shown in figure Figure3(c). You can observe that the information distillation operation is completed through a  $1 \times 1$  convolution operation, which can greatly reduce the amount of parameters. In the deeper information distillation module, we propose an AIEB, which improves the ability of information extraction. The AIEB is composed of two pairs of asymmetric convolutions in different directions. The sizes of these two asymmetric convolutions in different directions are  $1 \times 3$

and  $3 \times 1$  respectively, and they are used in different order on each branch. And a ReLU activation layer is added between every two asymmetric convolutions, which also improves the nonlinear ability of AIDB, so as to improve the ability of information extraction. Due to the powerful information extraction ability of AIEB, the input dimension can be reduced to half of the original dimension at the beginning of each block.

In the process of information distillation at the later levels within each block, some additional features are often needed to make up for the lost information caused by more nonlinear layers. By multiplexing the distilled information previously to the distillation information of the later levels, better distilled information can be completed with less parameters. Next, we will focus on the proposed AIDN in the next section.

### 3.4. Overall framework

We use the same structure as RFDN [20] and IMDN [8] to implement the overall network framework, as shown in Figure 2. The overall network structure consists of four parts: the head block, information extraction module, feature fusion module and upsampling module. (a) The head block. Head block is the process of extracting rough features from LR images through a  $3 \times 3$  convolution operation. On the other hand, the features extracted from the head block can be used as a residual information of the later block to supplement the distilled features. (b) Information extraction module. The information extraction module is formed by four AIDB stacks. With the enhancement of AIDB's ability to distill information, the output feature of each block are more accurate and more expressive. Moreover, we verified that with the increase of the number of AIDB, the performance of super resolution will significantly improved, with small increase in the number

Table 1. Comparison of CARN [1], SRResNet [15], RFDN [20] and AIDN for upscaling factor  $\times 2$ ,  $\times 3$  and  $\times 4$ . Red/Blue text: best/second-best

Scale	Method	Params	Set5 / PSNR	Set14 / PSNR	B100 / PSNR	Urban100 / PSNR	Manga109 / PSNR
$\times 2$	CARN [1]	1592K	37.76	33.52	32.09	31.92	38.36
$\times 2$	SRResNet [15]	1370K	<b>38.05</b>	33.64	<b>32.22</b>	<b>32.23</b>	38.05
$\times 2$	RFDN [20]	534K	<b>38.05</b>	<b>33.68</b>	32.16	32.12	<b>38.88</b>
$\times 2$	AIDN(ours)	323K	<b>38.07</b>	<b>33.72</b>	<b>32.18</b>	<b>32.24</b>	<b>38.89</b>
$\times 3$	CARN [1]	1592K	34.29	30.29	29.06	28.06	33.50
$\times 3$	SRResNet [15]	1554K	<b>34.41</b>	<b>30.36</b>	<b>29.11</b>	28.20	33.54
$\times 3$	RFDN [20]	541K	<b>34.41</b>	30.34	<b>29.09</b>	<b>28.21</b>	<b>33.67</b>
$\times 3$	AIDN(ours)	330K	<b>34.43</b>	<b>30.35</b>	<b>29.11</b>	<b>28.25</b>	<b>33.69</b>
$\times 4$	CARN [1]	1592K	32.13	<b>28.60</b>	<b>27.58</b>	26.07	30.47
$\times 4$	SRResNet [15]	1518K	32.17	<b>28.61</b>	<b>27.59</b>	26.12	30.48
$\times 4$	RFDN [20]	550K	<b>32.24</b>	<b>28.61</b>	27.57	<b>26.11</b>	<b>30.58</b>
$\times 4$	AIDN(ours)	339K	<b>32.26</b>	<b>28.60</b>	<b>27.58</b>	<b>26.16</b>	<b>30.59</b>

Table 2. Investigations of AIEB and FDC on the benchmark datasets with scale factor of  $\times 4$ . The best results are highlighted.

Method	Params	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Managa109 PSNR/SSIM
Base	352K	32.10/0.8933	28.51/0.7801	27.42/0.7341	26.01/0.7840	30.42/0.9072
AIEB	352K	32.19/0.8949	28.57/0.7809	27.49/0.7349	26.07/0.7849	30.50/0.9078
FDC	339K	32.23/0.8951	28.57/0.7810	27.54/0.7354	26.12/0.7857	30.56/0.9083
AIDB	339K	<b>32.26/0.8953</b>	<b>28.60/0.7818</b>	<b>27.58/0.7360</b>	<b>26.16/0.7864</b>	<b>30.59/0.9089</b>

of parameters. but the number of parameters does not increase significantly. (c) Feature fusion module. Similar to RFDN [20] and IMDN [8], we fuse the output features of each AIDB through concat and a  $3 \times 3$  convolution operation. The formula can be expressed as equation (2). (d) Upsampling module. In the feature fusion module, we got a more accurate and effective feature, then fintune the feature through a  $3 \times 3$  convolution operation, and then upsample the image through a pixelshuffle module to get the final SR image. The loss function of our model can be expressed as follows:

$$L(\theta) = \frac{1}{N} \sum_{i=0}^N \|H_{AIDN}(I_i^{LR} - I_i^{HR})\|_1 \quad (4)$$

Where  $\|\cdot\|_1$  represents L1 norm, LR, HR represents the low resolution image and the corresponding high resolution image. AIDN represents the final network structure. N represents batch size and  $\theta$  represents the parameters of AIDN.

## 4. Experiment

In this section, we will systematically compare our AIDN method with SOTA method on the benchmark

dataset. In addition, we conducted a large number of ablation experiments to verify the effectiveness of each component we proposed.

### 4.1. Datasets and metrics

We used DIV2K, Flickr2K, OST datasets as our training data. Among them, LR is obtained by HR through bicubic down sampling. We have a total of 13774 high resolution images, including 800 images in DIV2K, 2650 images in Flickr2K and 10324 images in OST. In the test phase, we use five widely used benchmark datasets, namely Set5 [4], SET14 [27], BSD100 [21], Urban100 [7] and Manga109 [22]. Moreover, PSNR and SSIM, two commonly used metrics in image restoration, are used to evaluate the image quality on the Y channel after super resolution.

### 4.2. Implementation details

In the training phase, we use three commonly used datasets (DIV2K, Flickr2K, OST) to train our AIDN. LR images ( $\times 2$ ,  $\times 3$ , and  $\times 4$ ) are obtained by downsampling HR image using bicubic in MATLAB. The methods of data augmentation include random flip and random rotation of  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$ . The image size of HR is set to  $256 \times 256$ , which

Table 3. Comparison of CARN [1], SRResNet [15], RFDN [20] and AIDN for upscaling factor  $\times 2$ ,  $\times 3$  and  $\times 4$ . Red/Blue text: best/second-best

Method	Params	Flops	Set5 PSNR/SSIM	Set14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM	Managa109 PSNR/SSIM
Scale			$\times 2$				
SRCNN [3]	57K	15.10G	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663
FSRCNN [4]	13K	1.72G	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9710
VDSR [12]	666K	175.53G	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9750
DRCN [13]	1774K	5150.23G	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.55/0.9732
LapSRN [14]	251K	8.57G	37.52/0.9591	32.99/0.9124	31.80/0.8952	30.41/0.9103	37.27/0.9740
DRRN [8]	298K	1947.54G	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.88/0.9749
MemNet [25]	678K	762.87G	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195	37.72/0.9740
CARN [1]	1592K	63.84G	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765
SRResNet [15]	1370K	101.44G	<b>38.05/0.9607</b>	33.64/0.9178	<b>32.22/0.9002</b>	<b>32.23/0.9295</b>	38.05/0.9607
IMDN [8]	694K	45.23G	38.00/0.9605	33.63/0.9177	<b>32.19/0.8996</b>	32.17/0.9283	<b>38.88/0.9774</b>
RFDN [20]	534K	32.07G	<b>38.05/0.9606</b>	<b>33.68/0.9184</b>	32.16/0.8994	32.12/0.9278	<b>38.88/0.9773</b>
AIDN(ours)	323K	19.21G	<b>38.07/0.9607</b>	<b>33.72/0.9192</b>	32.18/0.8995	<b>32.24/0.9289</b>	<b>38.89/0.9774</b>
RDN [31]	22123K	-	38.24/0.9614	34.01/0.9212	32.34/0.9017	32.89/0.9353	39.18/0.9780
RCAN [30]	15444K	-	38.27/0.9614	34.12/0.9216	32.41/0.9027	33.34/0.9384	39.44/0.9786
SAN [2]	15674K	-	38.31/0.9620	34.07/0.9213	32.42/0.9028	33.10/0.9370	39.32/0.9792
Scale			$\times 3$				
SRCNN [3]	57K	33.78G	32.75/0.9090	29.30/0.8215	28.41/0.7863	26.24/0.7989	30.48/0.9117
FSRCNN [4]	13K	3.21G	33.18/0.9140	29.37/0.8240	28.53/0.7910	26.43/0.8080	31.10/0.9210
VDSR [12]	666K	392.69G	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9340
DRCN [13]	1774K	11521.99G	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.24/0.9343
LapSRN [14]	502K	199.36G	33.81/0.9220	29.79/0.8325	28.82/0.7980	27.07/0.8275	32.21/0.9350
DRRN [8]	298K	4356.99G	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.71/0.9379
MemNet [25]	678K	1706.67G	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376	32.51/0.9369
CARN [1]	1592K	76.15G	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	33.50/0.9440
SRResNet [15]	1554K	121.43G	<b>34.41/0.9274</b>	<b>30.36/0.8427</b>	<b>29.11/0.8055</b>	<b>28.20/0.8535</b>	33.54/0.9448
IMDN [8]	703K	45.81G	34.36/0.9270	30.32/0.8417	<b>29.09/0.8046</b>	28.17/0.8519	33.61/0.9445
RFDN [20]	541K	32.51G	<b>34.41/0.9273</b>	<b>30.34/0.8420</b>	<b>29.09/0.8050</b>	<b>28.21/0.8525</b>	<b>33.67/0.9449</b>
AIDN(ours)	330K	19.65G	<b>34.43/0.9274</b>	<b>30.35/0.8420</b>	<b>29.11/0.8051</b>	<b>28.25/0.8530</b>	<b>33.69/0.9451</b>
RDN [31]	22308K	-	34.71/0.9296	30.57/0.8468	29.26/0.8093	28.80/0.8653	34.13/0.9484
RCAN [30]	15629K	-	34.74/0.9299	30.65/0.8482	29.32/0.8111	29.09/0.8702	34.44/0.9499
SAN [2]	15859K	-	34.75/0.9300	30.59/0.8476	29.33/0.8112	28.93/0.8671	34.30/0.9494
Scale			$\times 4$				
SRCNN [3]	57K	60.57G	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.66/0.8505
FSRCNN [4]	12K	5.3G	30.71/0.8657	27.59/0.7535	26.98/0.7105	24.62/0.7280	27.90/0.8517
VDSR [12]	665K	704.14G	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8809
DRCN [13]	1774K	20660.11G	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.98/0.8816
LapSRN [14]	813K	171.72G	31.54/0.8850	29.19/0.7720	27.32/0.7280	25.21/0.7560	29.09/0.8845
DRRN [8]	297K	7812.53G	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.46/0.8960
MemNet [25]	677K	3060.23G	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630	29.42/0.8942
CARN [1]	1592K	104.48G	32.13/0.8937	<b>28.60/0.7806</b>	<b>27.58/0.7349</b>	26.07/0.7837	30.47/0.9084
SRResNet [15]	1518K	166.36G	32.17/0.8951	<b>28.61/0.7823</b>	<b>27.59/0.7365</b>	<b>26.12/0.7871</b>	30.48/0.9087
IMDN [8]	715K	58.53G	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
RFDN [20]	550K	33.13G	<b>32.24/0.8952</b>	<b>28.61/0.7819</b>	<b>27.57/0.7360</b>	26.11/0.7858	<b>30.58/0.9089</b>
AIDN(ours)	339K	20.27G	<b>32.26/0.8953</b>	<b>28.60/0.7818</b>	<b>27.58/0.7360</b>	<b>26.16/0.7864</b>	<b>30.59/0.9089</b>
RDN [31]	22271K	-	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028	31.00/0.9151
RCAN [30]	15592K	-	32.63/0.9002	28.87/0.7889	27.77/0.7436	26.82/0.8087	31.22/0.9173
SAN [2]	15822K	-	32.64/0.9003	28.92/0.7888	27.78/0.7436	26.79/0.8068	31.18/0.9169

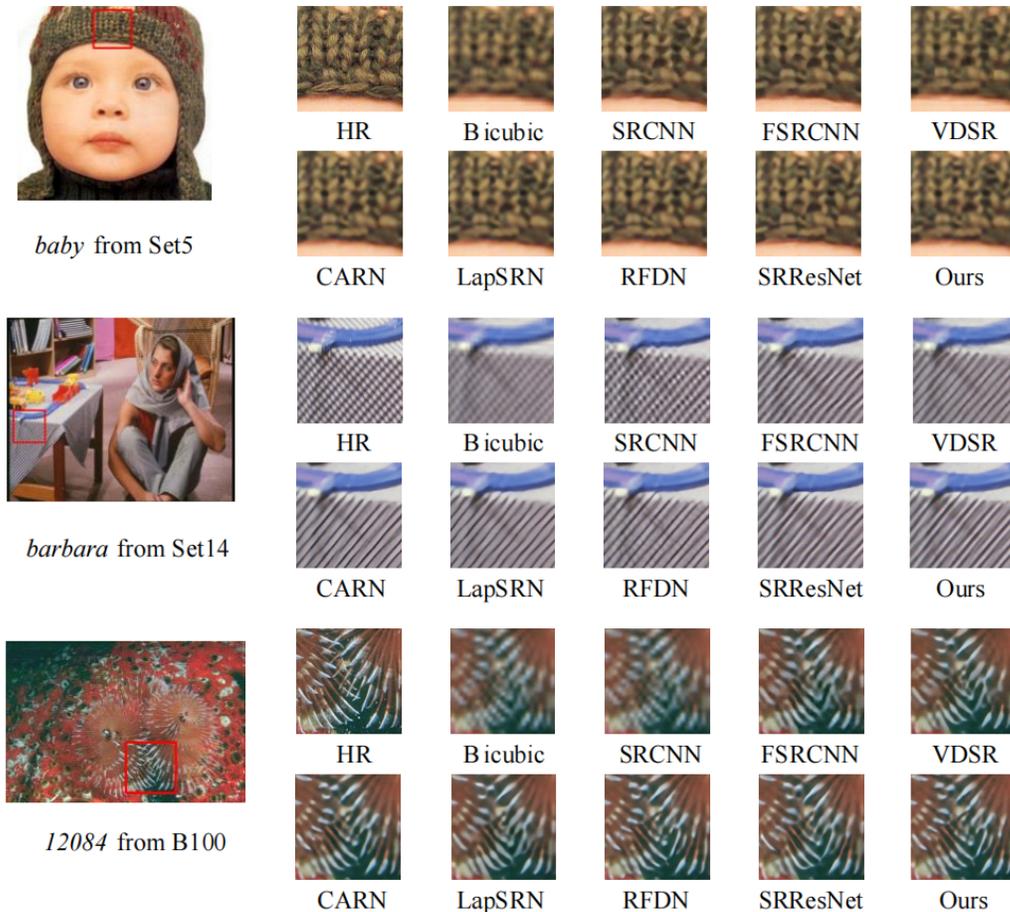


Figure 5. Visual comparison for upscale  $\times 4$

is obtained from the random crop on HR image. The batch size of training is set to 32. The training process is divided into two stages. In the first stage, AIDN is trained with L1 loss and Adam optimizer. The learning rate ranges from  $2 \times 10^{-4}$  to  $2 \times 10^{-6}$ . The first stage of training includes 1000 epochs. Then AIDN model is fine-tuned by L2 loss and Adam optimizer in the second stage training. The learning rate ranges from  $2 \times 10^{-5}$  to  $2 \times 10^{-6}$ . The second stage of training includes 500 epochs. Our proposed method is implemented using the pytorch framework and trained in NVIDIA GTX 1060Ti GPU.

### 4.3. Comparison with SRResNet and RFDN

The original intention of our proposed method is for efficient super resolution challenge. According to the requirements of the challenge, our goal is to maintain at least 29dB as SRResNet [15] in the validation dataset, and reduce the amount of computation as much as possible. In this section, we focus on our comparison with SRResNet [15] on five benchmark datasets with super resolution factors of  $\times 2$ ,  $\times 3$  and  $\times 4$  respectively. In addition, we also compared with

the champion model of AIM2020 efficient super resolution challenge to prove the effectiveness of our model. From Table 1, It can be clearly found that AIDN proposed by us is superior to CARN [1], SRResNet [15] and RFDN [20] in different upsampling multiples of five benchmark datasets. Importantly, our parameter quantities are 1/5 of CARN [1], 1/4 of SRResNet and 1/2 of RFDN [20].

### 4.4. Ablation study

In order to verify the effectiveness of our proposed AIEB and distilled information multiplexing, we designed the network structure of Figure 2 and carried out detailed ablation experiments. Figure 4 describes the four blocks respectively and shows the evaluation results in Table 2. It can be observed from the comparison results of the first two lines that under the same parameter scale, AIEB can increase by 0.08 dB compared with base model on PSNR. This shows that AIEB plays a good role in effective information extraction. From the last two rows of the Table 2, we can also find that PSNR can be increased by 0.04 dB by adding distillation information reuse.

## 4.5. Comparison with SOTA method

When the upsampling factors are 2, 3 and 4 respectively, we compare our proposed AIDN with advanced lightweight models, including SRCNN [3], FSRCNN [4], VDSR [12], DRCN [13], LapSRN [14], DRRN [8], MemNet [25], CARN [1], SRResNet [15], IMDN [8] and RFDN [20]. In addition, we also compared some SOTA models: RDN [31], RCAN [30] and SAN [2]. Table 3 shows the evaluation metrics (PSNR SSIM) of different methods on five benchmark datasets, and shows the parameters of each method. It can be observed from Table 3. Our proposed AIDN performs better than the current state of the art method with fewer parameters. In particular, CARN [1] and our model achieve the same super resolution performance, but CARN [1] has more than six times as many parameters as ours. Compared with the benchmark IMDN [8], the parameters of AIDN we proposed is less than 1 / 3 of IMDN [8], and also achieves the super resolution performance basically equivalent to IMDN [8]. Compared with RFDN [20] that won the first place in AIM2020 efficient super resolution challenge. The parameters of AIDN is 1 / 2 of RFDN [20], which also achieved same super resolution performance. In Figure 5, We show the visual comparison results of different methods. For 'barbara' images, our method can produce more accurate lines than other methods. For 'baby' and '12084' images, our method can achieve the same visual effect with fewer parameters than other methods.

## 5. Conclusion

In this paper, we propose a relatively lightweight image super resolution network (AIDN), and get the second place in the efficient super resolution challenge model complexity track of NTIRE2022. In particular, we propose a new feature distillation block (AIDB), which greatly reduces the amount of parameters while ensuring the accuracy of distilled information. In addition, the asymmetric feature enhancement block (AIEB) focuses on the features in the image from different directions and performs more effective enhancement. Moreover the nonlinear layer of AIEB further improves the nonlinear feature extraction ability. The multiplexing of distilled information makes up for the lost information caused by the reduction of channel dimension. Therefore, compared with SRResNet [15], our model maintains the same level in metrics (PSNR and SSIM) and greatly reduces the number of parameters. A large number of experiments show that our model AIDN can achieve comparable performance to state-of-the-art lightweight networks.

## References

[1] 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. 3, 5, 6, 7, 8

- [2] 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. 3, 6, 8
- [3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014. 1, 2, 6, 8
- [4] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016. 2, 5, 6, 8
- [5] Alireza Esmaeilzahi, M Omair Ahmad, and MNS Swamy. Srrmfrb: A deep light-weight super resolution network using multi-receptive field feature generation residual blocks. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2020. 1
- [6] Yong Guo, Jian Chen, Jingdong Wang, Qi Chen, Jiezhong Cao, Zeshuai Deng, Yanwu Xu, and Mingkui Tan. Closed-loop matters: Dual regression networks for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5407–5416, 2020. 3
- [7] 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
- [8] 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, 3, 4, 5, 6, 8
- [9] 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, 3
- [10] Lingxiu Jiang and Yue Zhou. Efficient channel attention feature fusion for lightweight single image super resolution. In *Journal of Physics: Conference Series*, volume 1828, page 012020. IOP Publishing, 2021. 1

- [11] Xinrui Jiang, Nannan Wang, Jingwei Xin, Xiaobo Xia, Xi Yang, and Xinbo Gao. Learning lightweight super-resolution networks with weight pruning. *Neural Networks*, 144:21–32, 2021. [1](#)
- [12] 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](#), [2](#), [6](#), [8](#)
- [13] 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](#), [6](#), [8](#)
- [14] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017. [2](#), [6](#), [8](#)
- [15] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. [1](#), [2](#), [5](#), [6](#), [7](#), [8](#)
- [16] Biao Li, Bo Wang, Jiabin Liu, Zhiquan Qi, and Yong Shi. s-lwsr: Super lightweight super-resolution network. *IEEE Transactions on Image Processing*, 29:8368–8380, 2020. [1](#)
- [17] Yawei Li, Shuhang Gu, Luc Van Gool, and Radu Timofte. Learning filter basis for convolutional neural network compression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5623–5632, 2019.
- [18] Yawei Li, Kai Zhang, Luc Van Gool, Radu Timofte, et al. Ntire 2022 challenge on efficient super-resolution: Methods and results. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2022. [1](#), [2](#), [3](#)
- [19] 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. [2](#)
- [20] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In *European Conference on Computer Vision*, pages 41–55. Springer, 2020. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#), [7](#), [8](#)
- [21] 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](#)
- [22] 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(20):21811–21838, 2017. [5](#)
- [23] Armin Mehri, Parichehr B Ardakani, and Angel D Sappa. Mprnet: Multi-path residual network for lightweight image super resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2704–2713, 2021. [1](#)
- [24] 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. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016. [2](#)
- [25] 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. [6](#), [8](#)
- [26] Zhihao Wang, Jian Chen, and Steven CH Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3365–3387, 2020. [3](#)
- [27] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010. [5](#)
- [28] Kai Zhang, Martin Danelljan, Yawei Li, Radu Timofte, Jie Liu, Jie Tang, Gangshan Wu, Yu Zhu, Xiangyu He, Wenjie Xu, et al. Aim 2020 challenge on efficient super-resolution: Methods and results. In *European Conference on Computer Vision*, pages 5–40. Springer, 2020. [1](#)

- [29] Xiangyu Zhang, Jianhua Zou, Kaiming He, and Jian Sun. Accelerating very deep convolutional networks for classification and detection. *IEEE transactions on pattern analysis and machine intelligence*, 38(10):1943–1955, 2015. [2](#)
- [30] 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. [3](#), [6](#), [8](#)
- [31] 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](#), [2](#), [6](#), [8](#)
- [32] Hengyuan Zhao, Xiangtao Kong, Jingwen He, Yu Qiao, and Chao Dong. Efficient image super-resolution using pixel attention. In *European Conference on Computer Vision*, pages 56–72. Springer, 2020. [1](#)