

Lightweight Image Super-Resolution with ConvNeXt Residual Network

Yong Zhang

Lanzhou University of Technology

Haomou Bai (✉ baihaomou@gmail.com)

Lanzhou University of Technology

Yaxing Bing

Lanzhou Jiaotong University

Xiao Liang


Lanzhou University of Technology

Research Article

Keywords: Artificial Intelligence, Machine Learning, Image reconstruction, Single-image Super-resolution

Posted Date: August 17th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1947449/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Neural Processing Letters on March 8th, 2023. See the published version at <https://doi.org/10.1007/s11063-023-11213-4>.

Abstract

Single image super-resolution (SISR) based on convolutional neural networks has been very successful in recent years. However, as the computational cost is too high, making it difficult to apply to resource-constrained devices, a big challenge for existing approaches is to find a balance between the complexity of the CNN model and the quality of the resulting SR. To solve this problem, various lightweight SR networks have been proposed. In this paper, we propose lightweight and efficient residual networks (IRN), which differ from previous lightweight SR networks that aggregate more powerful features by improving feature utilization through complex layer-connection strategies. The main idea is to simplify feature aggregation by using simple and efficient residual modules for feature learning, thus achieving a good trade-off between the computational cost of the model and the quality of the resulting SR. In addition, we revisit the impact of the activation function in the model and observe that different activation functions have an impact on the performance of the model. The experiment results show that IRN outperforms previous state-of-the-art methods in benchmark tests while maintaining a relatively low computational cost. The code will be available at <https://github.com/kptx666/IRN>.

1 Introduction

This paper focuses on the problem of single image super-resolution. Image super-resolution is a classic low-level vision task in computer vision that has a wide range of applications in security, surveillance, satellite, and medical imaging, and it can be used as a built-in module for other image recovery or recognition tasks. Single-image super-resolution refers to the reconstruction of visually appealing high-resolution images from corresponding low-resolution images. In the past few years, deep learning has greatly advanced the development of SR, and many deep neural network-based image SR methods have been proposed with great success. For example, Dong et al. [10] first proposed a super-resolution convolutional neural network with only three layers, SRCNN, and achieved superiority over previous non-deep learning methods. Subsequently, because deep convolutional neural networks [3] achieved good results in ImageNet classification, people were inspired to propose deeper and more complex architectures to improve the performance of SR methods. Kim et al. [5, 12] pushed the depth of SR networks to 20 and achieved better performance than SRCNN. EDSR networks [13] reached a depth of more than 160 layers. It was further demonstrated that deeper models are more beneficial to improving the performance of SR models. Although these SR networks greatly improved the quality of reconstructed images, their memory consumption and computational cost were huge, which made them difficult to deploy to resource-constrained devices, such as mobile devices. Therefore, improving the efficiency of SR models and designing lightweight models becomes critical.

To address these problems, a number of lightweight super-resolution models have been proposed. Ahn et al. [14] proposed a lightweight efficient cascaded residual network CARN-M with multiple residual connections, but its PSNR was too low. Hui et al. [4] proposed an information distillation network IDN, which achieved better performance with a smaller number of parameters. Subsequently, the Information Multiple Distillation Network (IMDN) [15] introduced an information multiple distillation block with a contrast-aware attention layer, which further improved the IDN. Wenbo et al. [17] proposed a linear combinatorial pixel adaptive regression network (LAPAR) that transformed direct LR to HR mapping learning into a linear coefficient regression task based on a dictionary of multiple predefined filter bases. FALSr [18] employed Network Architecture Search (NAS) techniques to implement lightweight super-resolution models. However, these SR models were not lightweight enough and the SR performance can be further improved.

For this purpose, this paper proposes a lightweight residual network (IRN) to better balance model performance and computational cost. It is computationally less expensive than IMDN, LAPAR-A and FALSr [15, 17, 18] and has better performance compared to them. Unlike most previous small parameter models that use recursive structures and information distillation, we design a residual block inspired by the ConvNeXt Block [19], which is shown to increase the depth of the network at a smaller computational cost, thus improving the performance of the network. Secondly, we introduce an effective Enhanced Spatial Attention (ESA) [16] module, which is used to improve the SR model's ability to collect a variety of fine-grained information. Specifically, we make use of more useful features (e.g. edges, corners, textures, etc.) for image recovery.

The contributions of this paper can be summarized as follows:

1. We introduce the ConvNeXt Block to construct the residual block and demonstrate its effectiveness against SR.
2. We deploy an effective attention module(ESA), to strengthen the model with an additional finite amount of computation.
3. Our proposed IRN integrates ConvNeXt Block and an effective attention module, which successfully enhances the compactness of the model and reduces the computational cost without sacrificing SR recovery accuracy.
4. Related Work

2 Related Work

2.1 Single image super-resolution

In recent years, with the rapid development of deep learning, more and more deep learning methods have been applied to SR tasks, which have greatly improved the performance of SR tasks. Dong et al. [10] first proposed the deep learning-based method SRCNN, a model that achieves better performance than traditional methods despite having only three layers. Although SRCNN achieves good results, its pre-input amplification of SRCNN achieves good results but its bi-trivial interpolation of LR images for amplification before input makes a large number of redundant computations. The authors subsequently improved SRCNN in FSRCNN [11] by removing this pre-processing and amplifying the image directly at the end of the network using transposed convolution to reduce the computational cost. To progressively reconstruct higher resolution images, Lai et al. proposed the Laplace operator pyramidal super-resolution network (LapSRN) for progressive upsampling networks [20]. There are other works such as MS-LapSRN [21] and Progressive upsampling SR (ProSR) [22] which also adopt this progressive upsampling SR framework and achieve relatively high performance. Shi et al. proposed an efficient sub-pixel convolution layer in ESPCN [23], where LR images were mapped through a series of features and then at the end of the network through a sub-pixel convolution module are amplified into an HR output. Due to the effectiveness of subpixel convolutional layers, later proposed networks have used subpixel convolutional layers as reconstruction modules and obtained better performance. Kim et al. [5, 12] obtained deeper layers of VDSR and DRCN by stacking convolutions using residual connections, resulting in a total of twenty layers and improved SR performance. Lim et al. [13] obtained a deeper layer of VDSR and DRCN by removing the batch normalized (BN) layer of residual blocks [1] were stacked to construct deeper and wider residual networks EDSR and MDSR and achieved significant performance improvement. Zhang et al. [27] proposed RDN based on EDSR by introducing dense connections [2, 7] to make full use of the information in all feature layers. They introduced channel attention module [25] into the residual block and proposed the very deep residual attention network (RCAN) [8]. They then introduced the non-local module into the residual block to construct the residual non-local attention network (RNaN) [26] for various image recovery tasks. Guo et al. [28] proposed a dual regression method to improve the performance of the SR model by introducing additional constraints. Liang et al. [30] proposed a Transformer architecture for image recovery based on the Swin Transformer [31], while Chen et al. [29] later proposed the SR model (HAT), which achieved significant improvements in the performance of the SR model.

2.2 Efficient SR Models

Although deep learning-based SR methods have achieved great success in terms of performance, their computational cost is too large and not suitable for application to resource-constrained devices, such as mobile devices. Therefore, many methods have been proposed to reduce the computational cost of SR models. For example, FSRCNN [11] reduced the redundant computation caused by direct bi-cubic interpolation of input images by SRCNN [10] by introducing inverse transpose convolution at the end of the network. DRCN [12] applies recurrent networks to SR models to reduce the number of parameters by reusing feature information multiple times. Ahn et al. [14] proposed a lightweight and efficient cascaded residual network CARN-M with multiple residual connections by using grouped convolution to reduce the computational effort of standard convolution. Hui et al. [4] proposed the Information Distillation Network (IDN), which splits the previously extracted features and then processes them separately to reduce the computational effort. The Information Multi Distillation Network (IMDN) [15] improved on the IDN by introducing a contrast-aware attention layer, thus improving the performance of the SR model. The Residual Feature Distillation Network (RFDN) [16] revisits the network architecture of the IMDN, further making the network lighter and improving the performance of the SR model by using feature distillation connections (FDC) and shallow residual blocks (SRB).

2.3 Attention model

The attention mechanism has become an important component in improving the performance of deep neural networks and was widely used in various computer vision tasks, such as image classification. The attention mechanism can be interpreted as focusing on the more useful information in the features. Hu et al. [25] proposed the (SE) block to exploit inter-channel attention at a lower computational cost, improving ResNet and achieving significant performance gains in image classification tasks. Wang et al. [33] improved the efficiency of the module and improved the performance by improving the fully connected layer in the SE block. CBAM [32] modified the SE block to enable the use of both spatial and channel attention.

In recent years, several attention-based SR models have also been proposed and have significantly improved the performance of SR. Zheng et al. [15] proposed IMDN with a contrast-aware channel attention mechanism (CCA) to enhance the ability of SR models to collect various fine-grained information. Zhang et al. [8] introduced the channel attention mechanism into SR models and proposed RCAN. Dai et al. [34] proposed a second-order attention network SAN to explore more powerful feature representations using second-order feature statistics. Wang et al. [24] proposed a non-local module to generate an attention graph using the computation of the correlation matrix between each spatial point in the feature graph, which is used to guide dense aggregation of contextual information.

3 Method

3.1 Network Architecture

In this section, we describe in detail our proposed lightweight residual network (IRN), the overall network architecture of which is shown in Figure 1.

Our IRN consists of three main components: the first shallow feature extraction block, multiple stacked residual blocks (IRBs) and the reconstruction module. We denote I^{LR} and I^{SR} as the input and output images of the IRN respectively. In the first stage we used a single 3×3 convolutional layer for shallow feature extraction, which can be represented as

$$F^0 = H_{ext} \left(I^{LR} \right),$$

where $H_{ext}(\bullet)$ denotes the convolution operation for shallow feature extraction and F^0 denotes the extracted feature map. We then use multiple IRBs in a cascade fashion for deep feature extraction, a process that can be represented as

$$F^n = H_{IRB}^n \left(H_{IRB}^{n-1} \left(\dots H_{IRB}^0 \left(F^0 \right) \dots \right) \right),$$

where $H_{IRB}^n(\bullet)$ denotes the nth IRB function and F^n is the nth output feature map.

In addition, we use a 3×3 convolutional layer to refine the deep features and then use the reconstruction module to generate I^{SR} , which can be expressed as

$$I^{SR} = H_{rm} \left(\left(H_r \left(F^n \right) + F^0 \right) \right),$$

H_{rm} represents the reconstruction module, which consists of a 3×3 convolution with a $3 \times S^2$ output channel and a sub-pixel convolution. In addition, H_r represents a 3×3 convolution operation. The model is optimized using the L_1 loss function, which can be expressed as

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \| H_{IRN} \left(I_i^{LR} \right) - I_i^{HR} \|_1$$

where $H_{IRN}(\bullet)$ is our IRN, θ is the model-learnable parameter, and $\| \bullet \|_1$ is the L_1 norm.

3.2 Residual blocks

In this subsection, we introduce the residual block (IRB). As shown in Fig. 2, the residual block consists of one 3×3 convolution and n ConvNeXt Blocks, where n = 1 or 3, only in the second IRB of the model n = 3, and n = 1 in the rest of the IRBs. We use the 3×3 convolution and ConvNeXt Blocks to extract features. In particular, our IRN uses fewer activation functions, with only the $4 \times dim$ layer in the ConvNeXt Block using the GELU [36] activation function, and none of the other layers using any activation function. The ConvNeXt Block [19] is an inverse bottleneck layer architecture that decomposes the standard convolution into depth-by-depth convolution and point-by-point 1×1 convolution, as shown in Fig. 3, and has a wider feature before activation. [35] A model with wider features before activation can significantly improve the performance of single image super resolution (SISR) for the same parameters and computational budget. We therefore introduce it into the model and experimental results show that it can significantly reduce the computational cost while maintaining SR model performance. Given an input feature of F_{in} , the structure is described as

$$F_{cdc1} = EF_1 \left(F_{in} \right)$$

$$F_{cdc2} = EF_{CB} \left(F_{cdc1} \right)$$

where EF_1 denotes a 3×3 convolutional block and EF_{CB} denotes a residual block made up of 1 or 3 ConvNext Blocks. F_{cdcj} is the feature extracted by the jth module. After two feature extraction modules, we add the final feature F_{cdc2} and the skipped feature F_{in} directly. This can be expressed as

$$F_{cdc} = F_{in} + F_{cdc2}$$

where F_{cdc} is the final refined output feature.

Next, we feed F_{cdc} into the ESA module [16] to obtain the final output of the IRB.

3.3 ESA Attention Module

As the effectiveness of the ESA module has been proven [16, 37], we introduce this module into our IRN. To keep the ESA module sufficiently lightweight, it applies a 1×1 convolutional layer at the beginning to perform the reduction of the channel dimension of the input features. Then a stepwise convolution and a maximum pooling layer are used to reduce the size of the feature map. After a set of convolutions to extract features, interpolation-based up-sampling is performed to recover the original feature map size. Finally, an attention mask is generated through the sigmoid layer. The specific architecture of the ESA module is shown in. Figure 4.

4 Experiments

4.1 Datasets and metrics

Following previous work, in our experiments, we used the DIV2K [38] dataset, which is widely used for image recovery tasks and contains 800 high-quality RGB training images to train the model. For testing, we used five widely used benchmark datasets: Set5 [39], Set14 [40], BSD100 [41], Urban100 [42] and Manga109 [43]. We use two metrics, peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [44], to evaluate the quality of super-resolution images. Based on the existing work, we calculate PSNR and SSIM on the Y channel of YCbCr converted from RGB.

4.2 Implementation details

Our model is trained on the RGB channel, and the LR images are generated by downsampling ($\times 2$, $\times 3$ and $\times 4$) the HR images in MATLAB using bicubic interpolation. In this paper, we use a randomly cropped HR image patch of size 192×192 from the HR image as input to our model, with the mini-batch size set to 64. We augment the training data with random horizontal flips and 90 rotations. Our model was trained using the ADAM optimizer [45] with momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$. The initial learning rate was set to 5×10^{-4} and was reduced by half after every 2×10^5 iterations. When training the final model, the $\times 2$ model was trained from scratch. After the model converges, we use it as a pre-trained model for other scales. In the IRN, we set the number of IRBs to 4. We implemented our network on the Pytorch framework and trained it on an NVIDIA RTX A5000 GPU.

4.3 Model analysis

In this subsection, we investigate the model parameters, the validity of the ESA, the effect of the activation function on the SR model and the validity of the IRN.

Model parameters. In order to construct a lightweight SR model, the parameters of the network are crucial. From Table 3, we can observe that our IRN achieves comparative or better performance compared to other state-of-the-art SR methods such as LAPAR-A (NeurIPS'21), SRFBN-S (CVPR'19), etc. We also visualize the trade-off analysis between performance and Multi-Adds/Parameters in Fig. 5 We can see that our IRN achieves a better trade-off between performance and computational cost.

Effectiveness of ESA. An ablation study was conducted and used to validate the effectiveness of the ESA module. As shown in Table 1, the IRN without ESA showed a significant performance degradation for a parameter drop of approximately 10%, and the complete IRN showed significant performance improvements on the Set5, Set14, BSD100, Urban100 and Manga109 datasets. The results show that the ESA module can effectively improve the performance of SR.

Table 1 Ablation studies of ESA

Method	Params[K]	Multi-Adds[G]	Set5		Set14		B100		Urban100		Manga109	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
IRN-woESA	470	26	32.05	0.8932	28.45	0.7998	27.44	0.7350	25.86	0.7807	30.20	0.9050
IRN	524	28	32.15	0.8942	28.53	0.7810	27.53	0.7361	25.98	0.7841	30.35	0.9069

A study of different activation functions. When introducing the ConvNeXt Block, we retain GELU as its activation function. However, most previous SR networks have used ReLU [46] or LeakyReLU [47] as the activation function. Therefore, we investigate the effects of these three

activation functions on the SR model. The results in Table 2 show that among these activation functions, GELU obtains a significant performance improvement. Therefore, we chose to retain GELU, as the activation function in our model.

Table 2 Quantitative comparison of different activation functions

Method	Set5		Set14		B100		Urban100		Manga109	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ReLU	32.06	0.8933	28.49	0.7804	27.50	0.7352	25.91	0.7819	30.27	0.9055
LeakyReLU	32.12	0.8938	28.48	0.7806	27.52	0.7355	25.94	0.7833	30.32	0.9061
GELU	32.15	0.8942	28.53	0.7810	27.53	0.7361	25.98	0.7841	30.35	0.9069

Comparison with state-of-the-art methods. We compare the proposed IRN with state-of-the-art lightweight SR methods, and Table 3 shows the quantitative comparison results for different scale factors. We also provide the number of parameters and Multi-Adds computed on an output of 1280×720 . We can observe that our IRN compares well with other state-of-the-art SR methods, including SRCNN [10], FSRCNN [11], VDSR [5], DRCN [12], MemNet [6], SRDenseNet [7], DRRN [48], LapSRN [20], SelNet [49], CARN-M [14], CARN [14], SRMDNF [50], SRFBN-S [51] and LAPAR-A [17], for $\times 3$ and $\times 4$ models, outperforming other comparative methods on most data sets, especially for the $\times 3$ model, where IRN uses fewer parameters and Multi-Adds, greatly outperformed other methods on all benchmark datasets.

Figure 6 shows a comparison of the visualization on the Set14 and Urban100 datasets at $\times 4$. For the Urban100 “img_62” image, we can see that the grid structure is better recovered. It also demonstrates the validity of our IRN

Table 3

Comparisons on multiple benchmark datasets for lightweight networks. The Multi-Adds is calculated corresponding to a 1280 × 720 HR image. Bold/red/blue: our/best/second best results

Scale	Method	Params	MultiAdds	Set5	Set14	BSD100	Urban100	Manga109
	SRCNN [10]	57K	53G	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946	35.74/0.9661
	FSRCNN [11]	12K	6G	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9694
	VDSR [5]	665K	613G	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9729
	DRCN [12]	1,774K	17,974G	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.63/0.9723
	MemNet [6]	677K	2,662G	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195	-
	DRRN [48]	297K	6,797G	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.92/0.9760
	LapSRN [20]	813K	30G	37.52/0.9590	33.08/0.9130	31.80/0.8950	30.41/0.9100	37.27/0.9740
×2	SelNet [49]	974K	226G	37.89/0.9598	33.61/0.9160	32.08/0.8984	-	-
	CARN-M [14]	412K	91G	37.53/0.9583	33.26/0.9141	31.92/0.8960	31.23/0.9193	-
	CARN [14]	1,592K	223G	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	-
	SRMDNF [50]	1,513K	348G	37.79/0.9600	33.32/0.9150	32.05/0.8980	31.33/0.9200	-
	SRFBN-S [51]	282K	680G	37.78/0.9597	33.35/0.9156	32.00/0.8970	31.41/0.9207	38.06/0.9757
	LAPAR-A [17]	548K	171G	38.01/0.9605	33.62/0.9183	32.19/0.8999	32.10/0.9283	38.67/0.9772
	IRN(Ours)	503K	106G	38.08/0.9607	33.64/0.9181	32.20/0.8999	32.11/0.9282	38.83/0.9773
	SRCNN [10]	57K	53G	32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989	30.59/0.9107
	FSRCNN [11]	12K	5G	33.16/0.9140	29.43/0.8242	28.53/0.7910	26.43/0.8080	30.98/0.9212
	VDSR [5]	665K	613G	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9310
	DRCN [12]	1,774K	17,974G	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.31/0.9328
	MemNet [6]	677K	2,662G	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376	-
	DRRN [48]	297K	6,797G	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.74/0.9390
×3	SelNet [49]	1,159K	120G	34.27/0.9257	30.30/0.8399	28.97/0.8025	-	-
	CARN-M [14]	412K	46G	33.99/0.9236	30.08/0.8367	28.91/0.8000	27.55/0.8385	-
	CARN [14]	1,592K	119G	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	-
	SRMDNF [50]	1,530K	156G	34.12/0.9250	30.04/0.8370	28.97/0.8030	27.57/0.8400	-
	SRFBN-S [51]	376K	832G	34.20/0.9255	30.10/0.8372	28.96/0.8010	27.66/0.8415	33.02/0.9404
	LAPAR-A [17]	594K	114G	34.36/0.9267	30.34/0.8421	29.11/0.8054	28.15/0.8523	33.51/0.9441
	IRN(Ours)	512K	48G	34.46/0.9276	30.37/0.8430	29.11/0.8056	28.18/0.8529	33.70/0.9452
	SRCNN [10]	57K	53G	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.66/0.8505
	FSRCNN [11]	12K	5G	30.71/0.8657	27.59/0.7535	26.98/0.7150	24.62/0.7280	27.90/0.8517
	VDSR [5]	665K	613G	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8809
	DRCN [12]	1,774K	17,974G	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.98/0.8816
	MemNet [6]	677K	2,662G	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630	-
	DRRN [48]	297K	6,797G	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.46/0.8960
	LapSRN [20]	813K	149G	31.54/0.8850	28.19/0.7720	27.32/0.7280	25.21/0.7560	29.09/0.8845
×4	SelNet [49]	1,417K	83G	32.00/0.8931	28.49/0.7783	27.44/0.7325	-	-
	SRDenseNet [7]	2,015K	390G	32.02/0.8934	28.50/0.7782	27.53/0.7337	26.05/0.7819	-

Scale	Method	Params	MultiAdds	Set5	Set14	BSD100	Urban100	Manga109
	CARN-M [14]	412K	33G	31.92/0.8903	28.42/0.7762	27.44/0.7304	25.62/0.7694	-
	CARN [14]	1,592K	91G	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	-
	SRMDNF [50]	1,555K	89G	31.96/0.8930	28.35/0.7770	27.49/0.7340	25.68/0.7730	-
	SRFBN-S [51]	483K	1,037G	31.98/0.8923	28.45/0.7779	27.44/0.7313	25.71/0.7719	29.91/0.9008
	LAPAR-A [17]	659K	94G	32.15/0.8944	28.61/0.7818	27.61/0.7366	26.14/0.7871	30.42/0.9074
	IRN(Ours)	524K	28G	32.21/0.8952	28.61/0.7822	27.59/0.7370	26.04/0.7852	30.49/0.9091

4.4 Running Time

As shown in Table 4, our method has the lowest number of parameters and running time compared to LAPAR-A (NeurIPS'21) and IMDN (ACM'19). For the average running time, as it is related to the optimization of the code and the computation of specific testbeds for different operators (more 1×1 convolutions are used in our IRN than in IMDN and LAPAR-A). Therefore, our method does not differ much from the runtime of IMDN (ACM'19).

Table 4 Comparison of our IRN with IMDN, LAPAR-A's 3x SR. Run times are the average of 10 runs on the Urban100 test set

Method	Params[K]	Runtime[ms]	Set5		Set14		BSD100		Urban100		Manga109	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
IMDN	703	92.6	34.36	0.9270	30.32	0.8417	29.09	0.8046	28.17	0.8519	33.61	0.9445
LAPAR-A	594	103.2	34.36	0.9267	30.34	0.8421	29.11	0.8054	28.15	0.8523	33.51	0.9441
IRN	512	89.4	34.46	0.9276	30.37	0.8430	29.11	0.8056	28.18	0.8529	33.70	0.9452

5 Conclusion

In this paper, we propose a lightweight and efficient single image super-resolution residual network (IRN). By using simple and efficient residual blocks, which are used to reduce the number of network layers and simplify the connections between layers, our network is made lighter and faster. In addition, we use effective ESA blocks to enhance the ability of model to collect fine grained information. We then investigate the effect of the activation function on the SR model to explore the best choice for our approach. Extensive experiments show that our proposed IRN strikes a good balance between model size, performance and computational cost compared to other lightweight SR models, so that it can be easily ported for use on mobile devices.

Declarations

Funding Open access funding provided by Lanzhou University of Technology.

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Availability of data and materials The ["2K resolution high quality images"] data that support the findings of this study are available in ["DIV2K dataset: DIVERse 2K resolution high quality images as used for the challenges @ NTIRE (CVPR 2017 and CVPR 2018) and @ PIRM (ECCV 2018)"], [<https://data.vision.ee.ethz.ch/cvl/DIV2K/>].

References

1. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 770–778
2. Gao Huang Z, Liu, van der Maaten, Kilian QW (2017) Densely connected convolutional networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 4700–4708
3. Karen Simonyan and Andrew Zisserman (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. In International Conference for Learning Representations (ICLR)
4. Zheng Hui X, Wang, Gao X(2018) Fast and Accurate Single Image Super-Resolution via Information Distillation Network. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 723–731
5. Kim J, Lee JK(2016) and Kyoung Mu Lee. Accurate image superresolution using very deep convolutional networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 1646–1654
6. Ying Tai J, Yang X, Liu, Xu C(2017) MemNet: A Persistent Memory Network for Image Restoration. In IEEE International Conference on Computer Vision (ICCV). 4539–4547
7. Tong T, Li G, Liu X, Gao Q(2017) Image Super-Resolution Using Dense Skip Connections. In IEEE International Conference on Computer Vision (ICCV). 4799–4807
8. Zhang Y, Li kunpeng, Li K, Wang L, Zhong B, Fu Y(2018) Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In European Conference on Computer Vision (ECCV). 286–301
9. Zhang Y, Tian Y, Kong Yu, Zhong B, Fu Y(2018) Residual Dense Network for Image Super-Resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2472–2481
10. Chao Dong CC, Loy K, He, Tang X(2014) Learning a deep convolutional network for image super-resolution. In European Conference on Computer Vision (ECCV). 184–199
11. Chao Dong CC, Loy, Tang X(2016) Accelerating the superresolution convolutional neural network. In European Conference on Computer Vision (ECCV). 391–407
12. Kim J, Lee JK(2016) and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 1637–1645
13. Lim B, Son S, Kim H, Nah S, Lee KM (2017) Enhanced deep residual networks for single image super-resolution. CVPR Workshops. IEEE Computer Society, pp 1132–1140
14. Ahn N, Kang B, Sohn K (2018) Fast, accurate, and lightweight super-resolution with cascading residual network. ECCV (10). Lecture Notes in Computer Science, vol 11214. Springer, pp 256–272
15. Hui Z, Gao X, Yang Y, Wang X(2019) : Lightweight image super-resolution with information multi-distillation network. In: ACM Multimedia. pp. 2024–2032. ACM
16. Liu J, Tang J, Wu G(2020), August Residual feature distillation network for lightweight image super-resolution. In European Conference on Computer Vision (pp. 41–55). Springer, Cham
17. Li W, Zhou K, Qi L, Jiang N, Lu J, Jia J (2020) Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. Adv Neural Inf Process Syst 33:20343–20355
18. Chu X, Zhang B, Ma H, Xu R, Li Q(2021), January Fast, accurate and lightweight super-resolution with neural architecture search. In 2020 25th International conference on pattern recognition (ICPR) (pp. 59–64). IEEE
19. Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S(2022) A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 11976–11986)
20. Lai W-S, Huang J-B, Ahuja N, Ming-Hsuan Y(2017) Deep laplacian pyramid networks for fast and accurate super-resolution. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 624–632
21. Lai W-S, Huang J-B, Ahuja N, Ming-Hsuan Y(2018) Fast and Accurate Image Super-Resolution with Deep Laplacian Pyramid Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence (2018)
22. Yifan Wang F, Perazzi B, McWilliams ASorkine-Hornung(2018) Olga Sorkin-Hornung, and Christopher Schroers. A Fully Progressive Approach to Single-Image Super-Resolution. In IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW). 977–986
23. Shi W, Caballero J, Huszar F, Totz J, Aitken AP, Bishop R, Rueckert D, Wang Z(2016) : Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR. pp. 1874–1883. IEEE Computer Society

24. Xiaolong Wang R, Girshick A, Gupta, He K(2018) Non-local Neural Networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 7794–7803
25. Jie Hu(2018) Li Shen, and Gang Sun. Squeeze-and-Excitation Networks. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 7132–7141
26. Zhang Y, Li K, Li K, Zhong B, Fu Y(2019) Residual Non-local Attention Networks for Image Restoration. In International Conference on Learning Representations (ICLR)
27. Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y (2020) Residual dense network for image restoration. IEEE Trans Pattern Anal Mach Intell 43(7):2480–2495
28. Guo, Y., Chen, J., Wang, J., Chen, Q., Cao, J., Deng, Z., ... Tan, M. (2020). Closed-loop matters: Dual regression networks for single image super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 5407–5416)
29. Chen X, Wang X, Zhou J, Dong C(2022) Activating More Pixels in Image Super-Resolution Transformer. arXiv preprint arXiv:2205.04437
30. Liang J, Cao J, Sun G, Zhang K, Van Gool L, Timofte R(2021) Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 1833–1844, 1, 2, 3
31. Ze Liu Y, Lin Y, Cao H, Wei HY, Zhang Z, Lin S(2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 10012–10022, 2, 3
32. Woo S, Park J, Lee J-Y(2018) and In So Kweon. CBAM: Convolutional Block Attention Module. In The European Conference on Computer Vision (ECCV). 3–19
33. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q(2020), June Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA (pp. 13–19)
34. Dai T, Cai J, Zhang Y, Xia ST, Zhang L(2019) Second-order attention network for single image super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11065–11074)
35. Yu J, Fan Y, Yang J, Xu N, Wang Z, Wang X, Huang T(2018) Wide activation for efficient and accurate image super-resolution.arXiv preprint arXiv:1808.08718
36. Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415, 2016. 3, 7
37. Liu J, Zhang W, Tang Y, Tang J, Wu G(2020) Residual feature aggregation network for image superresolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 2359–2368, 2, 3, 4, 5
38. 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
39. Marco Bevilacqua A, Roumy(2012) Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 5
40. Roman Zeyde M, Elad, Protter M(2010) On single image scale-up using sparse-representations. In International conference on curves and surfaces, pages 711–730. Springer, 5
41. Martin D, Fowlkes C, Tal D, Malik J(2001) volume 2, pages 416–423. IEEE, 2001. 5
42. Huang J-B, Singh A, Ahuja N(2015) Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 5197–5206, 5
43. Yusuke Matsui K, Ito Y, Aramaki A, Fujimoto T, Ogawa T, Yamasaki, Aizawa K (2017) Sketch-based manga retrieval using manga109 dataset. Multimedia Tools and Applications 76(20):21811–21838 5
44. Zhou Wang AC, Bovik HR, Sheikh, Eero P, Simoncelli. Image quality assessment: from error visibility to structural similarity.IEEE transactions on image processing, 13(4):600–612, 2004. 5
45. Diederik P(2014) Kingma and Jimmy Ba. Adam: A method for stochastic optimization.arXiv preprint arXiv:1412.6980,5
46. Vinod Nair and Geoffrey E Hinton (2010) Rectified linear units improve restricted boltzmann machines. In Icm1 4:7
47. Andrew L, Maas AY, Hannun, Andrew Y, Ng et al(2013) Rectifier nonlinearities improve neural network acoustic models. In Proc. icml, volume 30, page 3. Citeseer, 4, 7
48. Ying Tai J, Yang, Liu X(2017) Image superresolution via deep recursive residual network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3147–3155, 1, 2, 6, 8
49. Choi J-S, Kim M(2017) A deep convolutional neural network with selection units for super-resolution. In CVPRW, pages154–160,
50. Zhang K, Zuo W, Zhang L(2018) Learning a single convolutional super-resolution network for multiple degradations. In CVPR, pages3262–3271,

Figures

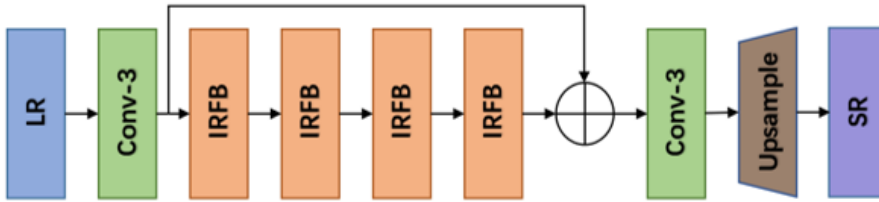


Figure 1

IRN

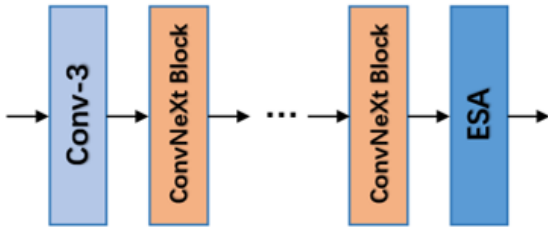


Figure 2

The architecture of IRB

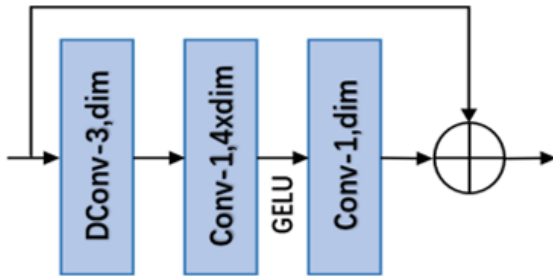


Figure 3

The architecture of ConvNeXt Block

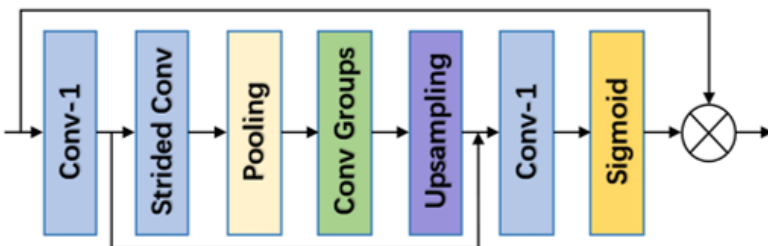


Figure 4

The architecture of ESA

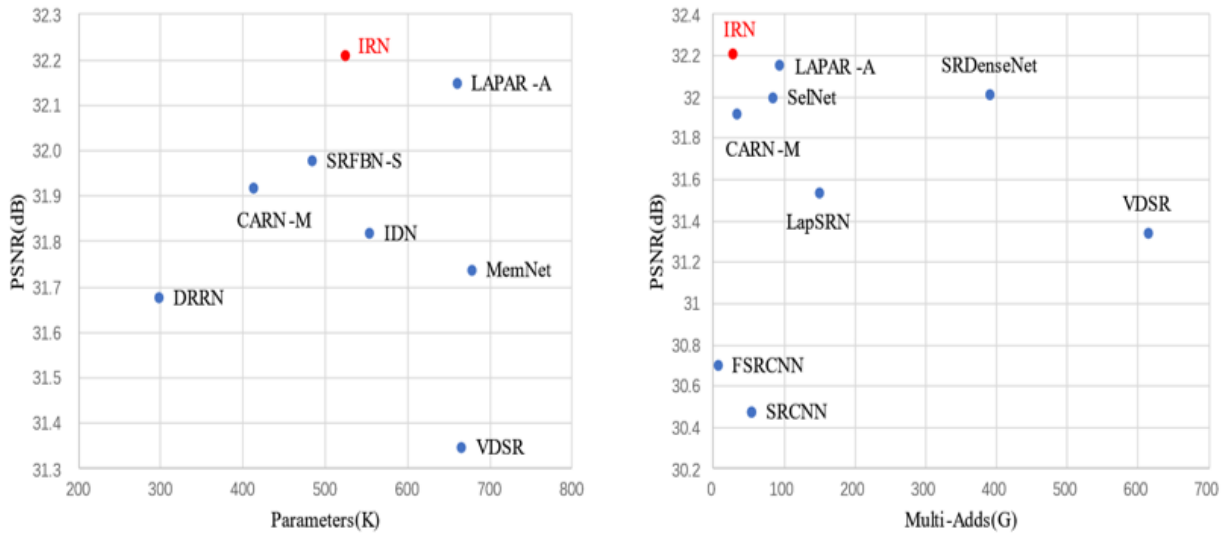


Figure 5

Illustration of PSNR, Multi-Adds and parameter numbers of different SISR models on the Set5 dataset for 4x SR

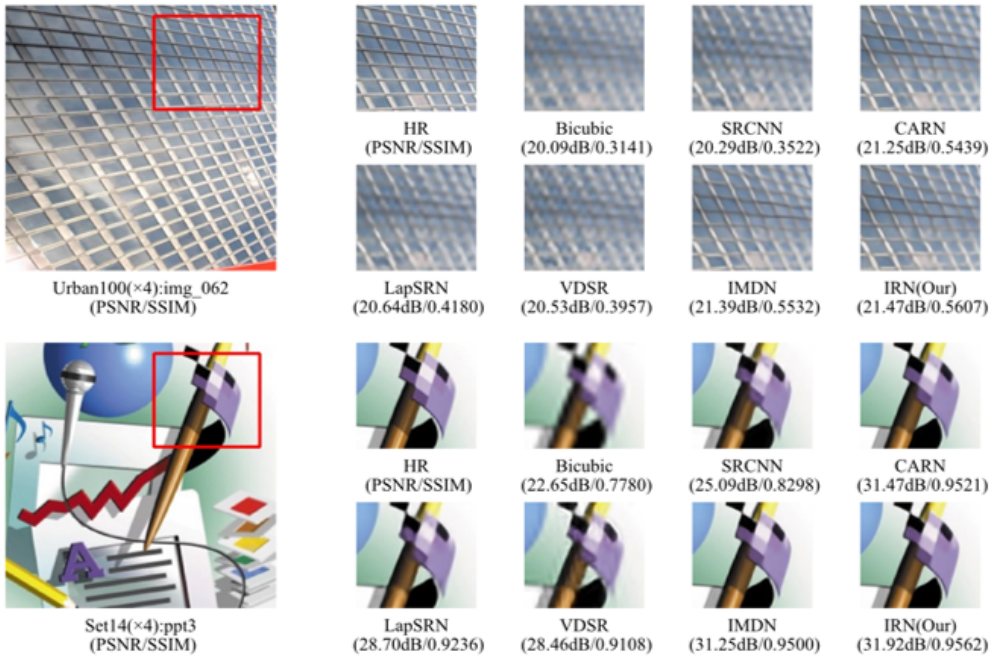


Figure 6

Visual comparisons of IRN with other SR methods on Set14 and Urban100 datasets