Conditional Convolution Residual Network for Efficient Super-Resolution

Yunsheng Guo[1],Jinyang Huang[1],Xiang Zhang[2],Xiao Sun[1],Yu Gu[3(✉)]

[1] School of Computer Science and Information Engineering, Hefei University of Technology, AnHui, China
guoyunsheng@mail.hfut.edu.cn,hjy@hfut.edu.cn,sunx@hfut.edu.cn

[2] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Anhui, China zhangxiang@ieee.org

[3] School of Cybers Science and Technology, University of Science and Technology of China, Anhui, China yugu.bruce@ieee.org

Abstract. With the continuous development of deep learning, single-image super-resolution (SISR) based on convolutional neural networks (CNNs) has made significant progress. Although CNN-based methods have achieved great success, these methods are difficult to apply to edge devices due to the need for large amounts of computing resources. To address this problem, the latest advancements in efficient SISR techniques focus on reducing the number of parameters and multiply-add operations (MAdds). In this paper, we propose a novel Conditional Convolution Residual Network (CCRN) to tackle this challenge. The main idea is to use conditional convolution instead of ordinary convolutional layers for residual feature learning and to combine Contrast-aware Channel Attention (CCA) and Enhanced Spatial Attention (ESA) mechanisms to improve the model's performance. The model's performance is ensured while reducing the computational complexity. Experimental results demonstrate that CCRN has fewer MAdds than existing SISR methods while achieving state-of-the-art performance.

Keywords: Efficient super-resolution · Conditional convolution · Attention mechanism

1 Introduction

Single Image Super-Resolution (SR) is a fundamental task in the field of computer vision, which aims to reconstruct high-resolution (HR) images from low-resolution (LR) images for better visual effects. With the development of deep learning, convolutional neural network-based methods have been widely introduced to the SR field to achieve high-quality super-resolution images.To improve the restoration quality of SR networks, existing SR networks typically employ large-scale models, which result in high computational complexity and make it challenging to apply them in real-world scenarios that require efficiency or real-time implementation, especially on edge devices.

To design lightweight neural networks, researchers have approached the problem from the perspectives of parameters and computational complexity and adopted different optimization strategies. For example, FSRCNN[1] reduces computation and parameter numbers by using upscaling modules and adhering to reducing the size of convolutions and features. Recursive learning has also been widely applied in many works, such as DRCN[2] and DRRN[3], to further reduce the number of parameters. However, due to their limited representation capabilities, these recursive methods also lead to performance degradation while consuming more computational resources. For instance, DRCN uses 17.9 trillion multiply-add operations (MAdds), while DRRN uses 6.8 trillion, which is difficult to afford for mobile devices. Therefore, to improve efficiency, some researchers have adopted different approaches, such as parameter sharing strategies[4], cascaded networks with grouped convolutions[5], information or feature distillation mechanisms[6], and attention mechanisms[7]. Although these methods employ compact architectures and improve mapping efficiency, there is still redundancy in convolution operations. Hence, researchers have shifted the focus from efficient SR to designing effective modules and dedicated networks to enhance the performance of SR networks further.

In this paper, we propose a novel lightweight super-resolution (SR) network, called Conditional Convolution Residual Network (CCRN). The network significantly reduces the MAdds of the network by optimizing convolution operations and introducing effective attention modules while achieving state-of-the-art performance.

Firstly, CCRN constructs the basic modules using Conditional Convolution[8]. This approach addresses the challenge of increasing model capacity by adding parameters, depth, and channels, which would otherwise result in greater computational demands and deployment difficulty. By inputting the convolution kernel parameters, Conditional Convolution breaks the static convolution characteristics, thereby improving the model's performance more efficiently. Secondly, to ensure the quality of reconstructed images, we introduce attention mechanisms in the process of feature extraction to select important pixel points at a fine-grained level, and better utilize pixel-level information in the image. Specifically, we add Enhanced Spatial Attention (ESA)[9] and Contrast-Aware Channel Attention (CCA)[6] modules at the end of each residual block to achieve this goal. Our proposed CCRN method significantly reduces the model's MAdds while maintaining SR performance in efficiency-oriented SR networks.

Overall, our main contributions can be summarized as follows:

- We introduce conditional convolution to construct basic modules and demonstrate their effectiveness in SR.
- We learn the importance of channels and space with two effective attention modules, ESA and CCA, respectively, and enhance the model's ability.
- The proposed CCRN, which integrates conditional convolution and effective attention modules, significantly reduces the network's computational complexity while maintaining SR performance.

2 Related Work

2.1 Deep Networks for SR

In recent years with the rapid development of deep learning techniques, convolutional neural networks (CNNs) have greatly advanced the development of low-level computer vision tasks[10]. Super-resolution (SR) tasks have also made increasingly significant progress. Since the pioneering work of Dong *et al.*[11], who proposed the SRCNN with a three-layer convolutional neural network that significantly outperformed traditional methods, a series of methods have been proposed to improve SR models. For example, Kim *et al.* proved [12] that deeper networks can achieve better performance by increasing the network depth to 20. Zhang et al.[13] introduced dense connections into the network, further enhancing the model's representation ability. Liang *et al.*[14] proposed a Transformer architecture for image restoration based on Swin Transformer[15], which achieved significant improvement and surpassed the state-of-the-art performance. [16] introduced a channel attention mechanism to utilize global statistics information for better performance.

Most of the above methods improve the quality by using more convolutional layers and attention mechanisms, ignoring resource-limited applications, which limits the practical application of these methods.

2.2 Efficient SR Models

Although the aforementioned methods have made significant progress in performance, most of them come with high computational costs, which has prompted researchers to develop more efficient methods for SR tasks. There have been many works aimed at designing more effective models for SR. Ahn *et al.*[5] proposed CARN-M, which is a residual network with a cascading mechanism that can reduce parameters and computations at the expense of lowered quality. Hui *et al.* proposed an information distillation network[17] that explicitly splits intermediate features for distilling and compressing local long-short path features. Based on IDN[17], IMDN[18]was introduced with a more reasonable feature distillation mechanism and effectively adaptive pruning strategy. By re-examining these distillation mechanisms, Liu *et al.*[19] proposed a novel channel-splitting strategy that utilizes convolutional layers for dimensionality change. Additionally, they designed shallow residual blocks to improve inference performance while maintaining parameter size.

3 Method

3.1 Network Architecture

The overall architecture of our proposed CCRN method is shown in Figure. 1. It inherits the architectures of IMDN and consists of four stages: shallow feature extraction, deep feature extraction, feature fusion, and upsampling module. The

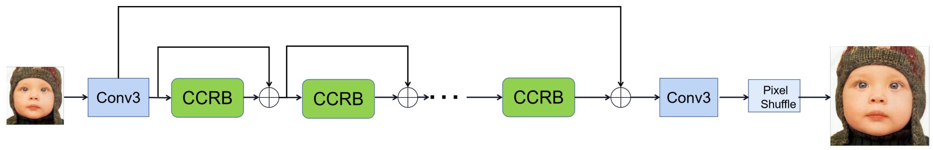

Fig. 1. CCRN Network Architecture

shallow feature extraction stage involves extracting coarse features from the LR image using a 3x3 convolution operation, which can also be used to supplement the residual information lost in the feature extraction process with subsequent blocks. Given an input image I_{LR}, this feature extraction process can be formulated as:

$$F_0 = L(I_{LR}) \tag{1}$$

where L represents the feature extraction function of the 3x3 convolution, and F_0 is the extracted feature map. Next, we use a cascaded approach with multiple CCRBs for deep feature extraction, which can be formulated as:

$$F_n = H^n_{CCRB}(H_{fuse}(F_{n-1}, F_{n-2})) \tag{2}$$

where $H^n_{CCRB}(\cdot)$ represents the n-th CCRB block, and F_n is the n-th output feature map. $H_{fuse}(\cdot)$ represents the fusion module, and to utilize residuals for learning, the input and output of the n-1 CCRB block are aggregated. In the feature fusion stage, the multi-distilled deep features F_n and the shallow features F_0 are fused together through residual connections.

$$F_{fuse} = H_{fuse}(F_0, F_n) \tag{3}$$

The reconstruction stage can be formulated as follows:

$$F_{rec} = H_{rec}(F_{fuse}) \tag{4}$$

where $H_{rec}(\cdot)$ consists of a 3x3 convolution layer and a pixelshuffle operation. The model is optimized using the L1 and L2 functions, and the specific optimization process is described in the experiments.

3.2 Conditional Convolution Residual Block

Inspired by the IMDB in IMDN, we designed a more efficient conditionally-convolutional residual block (CCRB) with a structure similar to IMDB. The overall architecture of CCRB is shown in Figure. 2(a).

A CCRB generally consists of three stages: feature distillation, feature condensation, and feature enhancement. In the first stage, for the input feature F_{in}, feature distillation can be formulated as:

$$\begin{aligned}
&F_{distilled1}, F_{coarse1} = DL_1(F_{in}), RL_1(F_{in}),\\
&F_{distilled2}, F_{coarse2} = DL_2(F_{coarse1}), RL_2(F_{coarse1}),\\
&F_{distilled3}, F_{coarse3} = DL_3(F_{coarse2}), RL_3(F_{coarse2}),\\
&F_{distilled4} = DL_4(F_{coarse3})
\end{aligned} \tag{5}$$

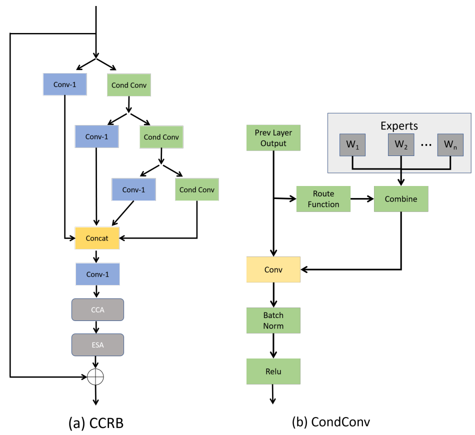

Fig. 2. The architecture of CCRB and CondConv

where DL represents the distillation layer that generates distilled features, and RL represents the refinement layer that further refines coarse features. In the feature condensation stage, the distilled features $F_{distilled1}$,$F_{distilled2}$,$F_{distilled3}$ and $F_{distilled4}$ are concatenated together and then compressed into a feature map with reduced dimensions using a 1x1 convolution.

$$F_{condensed} = H_{linear}(Concat((F_{distilled1}, \cdots, F_{distilled4})) \tag{6}$$

where $F_{condensed}$ is the compressed feature map, and $H_{linear}(\cdot)$ represents a 1x1 convolution layer. For the final stage, in order to enhance the model's representation ability while maintaining efficiency, we introduce a lightweight enhanced spatial attention (ESA) block[9] and a contrast-aware channel attention (CCA) block[6] as part of the CCRB.

$$F_{enhanced} = H_{ESA}(H_{CCA}(F_{condensed})) \tag{7}$$

where $F_{enhanced}$ is the enhanced feature map, $H_{ESA}(\cdot)$ and $H_{CCA}(\cdot)$ respectively represent the ESA and CCA modules, which have been shown to effectively enhance model ability from both spatial and channel perspectives.

3.3 Conditional Convolution

As shown in Figure. 2(b), in CondConv[8] each convolution kernel has the same dimension as the standard convolution kernel parameters. The ability improvement of conventional convolutional layers relies on increasing the kernel size and the number of channels, which further increases the overall computation of the network. However, the CondConv kernel is customized for each input sample, then the obtained kernel is used to perform convolution on that sample to obtain the corresponding output. Specifically, the convolution kernel in CondConv is parameterized by:

$$Output(x) = \sigma((\alpha_1 \cdot W_1 + \cdots + \alpha_n \cdot W_n) * x) \tag{8}$$

Each α is an example-dependent scalar weight computed using a routing function with learning parameters, n is the number of experts, and σ is the activation function. When we adjust the convolution layer to use CondConv, W_i is the same kernel in convolution as in normal convolution. The following routing function is used to compute α. This function is computationally efficient, distinguishes the input examples in a meaningful way, and is easy to interpret. We compute example-dependent routing weights α from layer inputs in three steps: global average pooling, fully connected layers, and sigmoid activation.

$$\alpha = Sigmoid(GlobalAveragePool(x) * R) \tag{9}$$

where R is the learning routing weight matrix that maps the pooled input to n expert weights. Normal convolutional operations operate only on local sensory fields, and the routing function described above allows to use of information from the global context in local operations.

In the previous efficient SR models, the capacity of the regular convolutional layers is generally increased by increasing the kernel height/width of the kernel or the number of input/output channels, but each additional parameter in the convolution requires additional multiplication proportional to the number of pixels in the input feature map, which can be large. This also increases the overall computational effort.

In CCRB we introduce the CondConv layer, where we compute a convolution kernel for each example as a linear combination of n experts before applying the convolution. It is crucial that each convolution kernel is computed only once, but is applied to many different positions in the input image. This means that by increasing n, we can increase the capacity of the network with only a small increase in inference cost; each additional parameter requires only 1 additional multiplication and addition. This greatly reduces the computational cost of increasing the capacity of the network and plays a big role in reducing the MAdds in Efficient SR.

3.4 ESA and CCA

We introduce both the enhanced spatial attention(ESA) module and the contrast-aware channel attention(CCA) module to pay more attention to the features related to the fine details of the image. The ESA mechanism operates at the end of each residual block to enforce features to focus more on the regions of interest. By aggregating these salient features together, we can obtain more representative features. The ESA mechanism as shown in Figure. 3(a) starts with a 1x1 convolutional layer to reduce the channel dimension, making the entire block very lightweight. To further expand the receptive field, we use a stride convolution (with a stride of 2), followed by a max pooling layer. An upsampling layer is employed to restore the spatial dimension, and a 1x1 convolutional layer is used

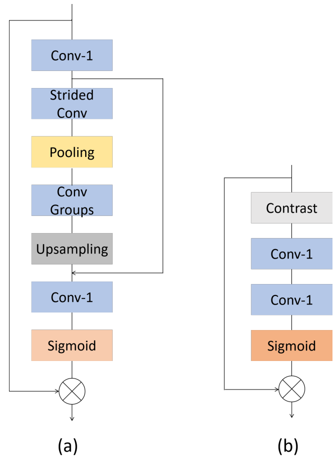

Fig. 3. The architecture of ESA and CCA

to recover the channel dimension. Finally, an attention mask is generated by a sigmoid layer. To utilize residual information, a skip connection is also employed to directly forward the high-resolution features before spatial downsampling to the end of each block.

In deep neural networks, different channels in different feature maps often represent different objects. Channel attention serves as an object selection process that can adaptively re-calibrate the weights of each channel to determine what to focus on. CCA as shown in Figure. 3(b) utilizes contrastive information, including the sum of mean and standard deviation, to calculate the weights for channel attention.

4 Experiment

4.1 Datasets and Metrics.

The training images consisted of 5000 images from LSDIR [20] and 800 images from DIV2K [21]. We employed four standard benchmark datasets, namely, Set5[22], Set14[23], B100[24], Urban100[25] to evaluate the performance of different methods. The average peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) on the Y channel (i.e., luminance) were used as evaluation metrics.

4.2 Implementation details of CCRN

The proposed CCRN consists of four CCRB blocks with 48 channels and employs two experts in conditional convolutions. All kernels in the deep convolutions have a size of 3. The batch size is set to 16, and each LR input patch has a size of

48×48. The model is optimized using the Adam optimizer with $\beta1 = 0.9$ and $\beta2 = 0.999$. The initial learning rate is set to 1×10^{-4} with drops by half every 200 epochs. We set the total epochs to 2000, first the L1 loss is employed for model optimization with 1000 epochs, and a total of 1×10^{6} iterations are performed. Then L2 loss is used to fine-tune the network with 1000 epochs. We implement our model using Pytorch on a GeForce RTX 3090 GPU.

Table 1. Results on method complexity (number of parameters, Multi-Adds). The multi-adds operation is calculated with 320×180 input size.

Method	Params[K]	Multi-Adds[G]
SRCNN	8	52.7
LapSRN	251	29.9
DRRN	298	6796.9
MemNet	678	2662.4
VDSR	665	612.6
IDN	553	31.1
CARN	1592	90.9
IMDN	715	41.0
CCRN(ours)	**752**	**14.15**

4.3 Study of the Basic Module

In our work, each feature extraction part of CCRB consists of three CondConv, and each CondConv can be set with different numbers of experts. We investigated the impact of changing the number of experts in CondConv. As shown in Table 2, with an increase in the number of experts, the SR performance improves, and the model's parameters increase, but the model's MAdds remain stable. This suggests that increasing the number of experts in CondConv can improve the model's performance without affecting the model's computational cost.

Figure. 4(b) demonstrate the outputs of feature maps with different layers. All feature maps are from the final CCRB module, each row displays 10 feature maps from the input of the CCA layer, the output of the CCA layer, and the output of the ESA layer, respectively. Through the attention mechanism, we can observe that the details of the image have been further explored, which is beneficial to enhancing the image quality for SR.

Table 2. Ablation results of various experts in CondConv(PSNR/SSIM)

Expert Number	Multi-Adds[G]	Params[K]	Set5[22]
1	14.15	503	31.55/0.8834
2	14.15	752	31.75/0.8875
3	14.15	1002	31.82/0.8887

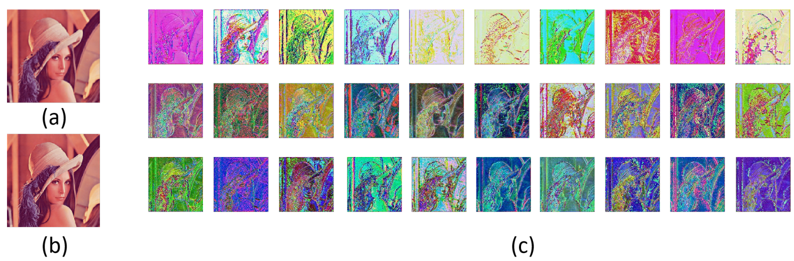

Fig. 4. Illustrations of (a)Low-resolution image (b)Super-resolution image (c)Feature Maps

Table 3. Average PSNR/SSIM for scale ×4 on datasets Set5, Set14, B100, Urban100 with bicubic degradation.(Compared to CARN, which achieves the best PSNR performance, we have obtained comparable results using only one-sixth of the MAdds (multiply-additions) utilized.)

Method	B100[24]	Set5[22]	Set14[23]	Urban100[25]
	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	25.96/0.6675	28.42/0.8104	26.00/0.7027	23.14/0.6577
SRCNN	26.90/0.7101	30.48/0.8626	27.50/0.7513	24.52/0.7221
LapSRN	27.32/0.7275	31.54/0.8852	28.09/0.7700	25.21/0.7562
DRRN	27.38/0.7284	31.68/0.8888	28.21/0.7720	25.44/0.7638
MemNet	27.40/0.7281	31.74/0.8893	28.26/0.7723	25.50/0.7630
VDSR	27.29/0.7251	31.82/0.8903	28.01/0.7674	25.18/0.7524
IDN	27.41/0.7297	32.13/0.8937	28.25/0.7730	25.41/0.7632
CARN	**27.58**/0.7349	**32.21**/0.8948	**28.60**/0.7806	**26.07**/0.7837
IMDN	27.56/0.7353	32.13/0.8948	28.58/0.7811	26.04/0.7838
CCRN(ours)	27.43/0.7305	31.75/0.8875	28.42/0.7756	25.69/0.7710

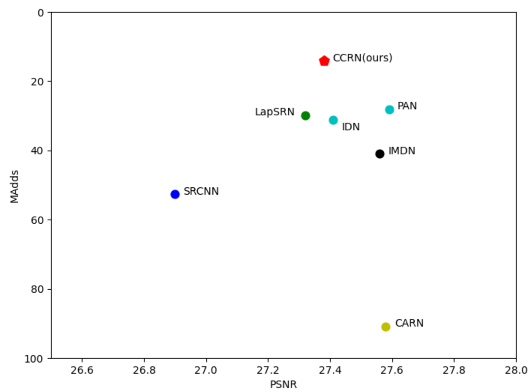

Fig. 5. MAdds and PSNR(based on B100[24] dataset)

4.4 Result and Discussion

The model complexity comparison and performance evaluation of CCRN and other Efficient SR models, namely,SRCNN[26],LapSRN[27],DRRN[3],MemNet[28], VDSR[12],IDN[17],CARN[5],IMDN[6] on the testing dataset are presented in Table 1 and Table 3 for $\times 4$ scales. The complexity of each model can be found in the second and third columns of Table 1, where the second column shows the number of parameters included in the model, and MAdds represents the number of Multi-Adds in the model, where one Multi-Add represents one multiplication and addition operation. A lower number of Multi-Adds indicates that the model requires less computation and thus has faster computational speed. It can be observed that CCRN has significantly fewer Multi-Adds than all the other models, which implies a great advantage in computational speed. Table 3 also shows that CCRN achieves good performance. Compared to IMDN, which suffers from a 0.13 PSNR loss on B100[24] at x4 scale, CCRN requires only one-third of the computation. As shown in Figure. 5. CCRN has the lowest number of Multi-Adds among all SR models while maintaining a high PSNR.

5 Conclusion

In this paper, we propose a lightweight network named Conditional Convolution Residual Network (CCRN) for single image super-resolution. Inspired by the Information Multi-distillation Network (IMDN) and Conditional Convolution (CondConv), the design of CCRN adopts a similar architecture to IMDN but introduces more efficient Conditional Convolution Residual Blocks (CCRB). Furthermore, effective ESA blocks and CCA blocks are used to enhance the representative ability of the model. Extensive experiments demonstrate that our method can achieve the same SR performance as advanced and efficient SR methods with much fewer MAdds, significantly reducing the computational cost of the model required for single image SR.

References

1. Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*, pages 391–407. Springer, 2016. 2
2. 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
3. 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, 10
4. 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
5. 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. 2, 3, 10
6. 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. 2, 5, 10
7. Haoyu Chen, Jinjin Gu, and Zhi Zhang. Attention in attention network for image super-resolution. *arXiv preprint arXiv:2104.09497*, 2021. 2
8. Brandon Yang, Gabriel Bender, Quoc V Le, and Jiquan Ngiam. Condconv: Conditionally parameterized convolutions for efficient inference. *Advances in Neural Information Processing Systems*, 32, 2019. 2, 5
9. 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. 2, 5
10. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 3
11. Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 3
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. 3, 10
13. 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. 3
14. 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. 3

15. 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. 3
16. 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
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. 3, 10
18. 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. 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. 3
20. Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. 7
21. 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. 7
22. Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012. 7, 9
23. 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. 7, 9
24. 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. 7, 9, 10
25. 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. 7, 9
26. 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. 10
27. 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. 10
28. 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. 10