Multimedia Tools and Applications
https://doi.org/10.1007/s11042-024-18677-z

Multi-scale cross-fusion for arbitrary scale image super resolution

Guangping Li[1] · **Huanling Xiao**[2] · **Dingkai Liang**[1] · **Bingo Wing-Kuen Ling**[1]

Received: 24 May 2023 / Revised: 17 January 2024 / Accepted: 19 February 2024

Abstract
Deep convolutional neural networks (CNNs) have great improvements for single image super resolution (SISR). However, most of the existing SISR pre-training models can only reconstruct low-resolution (LR) images in a single image, and their upsamling factors cannot be non-integers, which limits their application in practical scenarios. In this letter, we propose a multi-scale cross-fusion network (MCNet) to accomplish the super-resolution task of images at arbitrary scale. On the one hand, the designed scale-wise module (SWM) combine the scale information and pixel features to fullly improve the representation ability of arbitrary-scale images. On the other hand, we construct a multi-scale cross-fusion module (MSCF) to enrich spatial information and remove redundant noise, which uses deep feature maps of different sizes for interactive learning. A large number of experiments on four benchmark datasets show that the proposed method can obtain better super-resolution results than existing arbitrary scale methods in both quantitative evaluation and visual comparison.

Keywords Image super-resolution · Deep convolution neural networks · Multi-scale information

Huanling Xiao and Dingkai Liang contributed equally to this work

✉ Guangping Li
gpli@gdut.edu.cn

Huanling Xiao
2112003011@mail2.gdut.edu.cn

Dingkai Liang
2112003123@mail2.gdut.edu.cn

Bingo Wing-Kuen Ling
yongquanling@gdut.edu.cn

[1] School of Information Engineering, Guangdong University of Technology, Xiaoguwei, Guangzhou 510006, Guangdong, China

[2] College of Electronics and Information Engineering, Guangdong Ocean University, Haida Road, Zhanjiang 524003, Guangdong, China

Published online: 01 March 2024

1 Introduction

Image super resolution is a basic image processing technology, which aims to generate high resolution (HR) images on the basis of degraded low resolution (LR) images. In recent years, the single image super resolution (SISR) method based on deep convolutional neural networks (CNNs) has been significantly developed compared with the conventional SISR models [1–3, 3–14], and has been widely applied in various fields such as medical images [15, 16] and satellite imaging [17]. However, most existing SISR pre-training models can only perform single image restoration for LR images, which consumes additional computer resources. In addition, the fact that upsampling factors can only be integers limits its application in real-world scenarios.

In order to overcome the above problems, the up-sampling network is redesigned. Lim et al. [18] developed a multi-scale deep super resolution architecture (MDSR), which uses three different upsampling branches ($\times 2$, $\times 3$, $\times 4$) to generate HR images of different sizes from degraded images in the same model. In order to extend the scale factor to non-integer domains, Hu et al. [19] proposed a new advanced method for image reconstruction at arbitrary scale, called magnification-arbitrary network (Meta-SR), which uses several fully connected layers to predict the corresponding pixel values in HR images. This new network is a pioneering work in super resolution of arbitrary scale images. By using local implicit image function (LIIF) to learn the continuous representation of HR images, Chen et al. [20] achieved attractive SR results, which not only eliminated checkerboard artifacts in MetaSR, but also generated images with higher ($\times 6$, $\times 8$) resolution, while maintaining considerable visual perception. Lee et al. [21] use two-dimensional (2D) Fourier space to form a local texture estimator (LTE). In terms of backbone network, Wang et al. [22] developed dynamic scale-wise plug-in module (ArbSR) based on the existing SISR network to complete the task of image super resolution at arbitrary scale. Compared with the upsampling strategy LIIF, this specific neural implicit function can capture more image details. Li et al. [23] propose an enhanced dual branches network (EDBNet), which mix up pixel embedding and scale information to generate arbitrary-scale SR images in the upsampling network.

Compared to the traditional single scale upsampling module, the above arbitrary scale upsampling network method has better adaptability and flexibility. There is no denying that ArbSR does improve the backbone's representation ability to encode arbitrary-scale images with plug-and-play modules. However, it has a large number of parameters and image processing is slow. In addition, other methods have also been adjusted in the up-sampling module, but we think it can be further improved.

In this letter, we design a novel multi-scale cross-fusion network (MCNet), which has an excellent performance in arbitrary scale reconstruction. Firstly, the scale-wise module (SWM) combines the scale information and pixel features to effectively improve the representation capability of the backbone network for arbitrary scale images. Moreover, we design a powerful multi-scale cross-fusion module (MSCF) after the backbone network to enrich the spatial information and remove the redundant noise from the deep features. In our MSCF, deep feature maps in different sizes are used to conduct interactive learning from each other. The experiments with four benchmark datasets show the highly advantageous performance of our MCNet method.

The main contributions of this letter focus on the following aspects: 1) We propose a novel multi-scale cross-fusion network (MCNet), which not only removes the blurring artifacts for efficient and accurate image reconstruction but also delivers the most advanced results compared with other SR methods. 2) To see further improvement in feature representation ability,

we use the scale-wise module (SWM) to combine scale information with pixel features, effectively fusing two independent variables together. 3)We design a multi-scale cross-fusion module (MSCF) after the backbone network, which consists of two basic components: a) multi-downsampling convolution layer (MDConv) uses convolutional layers of different kernel sizes to generate smaller feature maps, and b) dual-spatial mask (DSM) is a dual spatial mask module that learns interactive information from the features with different scales.

2 Proposed method

2.1 Outline

As shown in Fig.1, our MCNet framework mainly consists of three parts: 1) feature extraction network, 2) multi-scale cross-fusion module (MSCF) and 3) arbitrary-scale upsampling network.

First, the extracted features F_d is obtained by performing a 3×3 convolutional layer and an existing SISR backbone network on the input LR image; i.e.,

$$F_d = E_\phi(Conv_{3\times3}(X)) \tag{1}$$

where E_ϕ denotes the backbone network with multiple stacked residual blocks [18] and novel SWM modules. We will discuss the new module in more depth in the next section. The second part of the MCNet framwork is our proposed MSCF, which makes a significant contribution to generating clean and abundant features F_{rid} given by

$$F_{rid} = Q(F_d) \tag{2}$$

where $Q(\cdot)$ will be described in more detail in a later section.

In the upsampling network, we incorporate scale information for image reconstruction by adding a new SGU module to another branch, which can accomplish a tailored image restoration task for our SR model. After the enrichment of features, F_{rid} and its mapping coordinate C in HR image space are used to facilitate the next stage on image upsampling network. Similar to the LTE [21], an HR image Y is generated through a continuous image upsampling module with local texture estimator G_{lte}; i.e.,

$$Y = \sum_{i=1}^{3} W_i \odot G_{lte}(F_i, C_i) \tag{3}$$

where i is the index of an offset latent code around F_{rid} and W_i is the corresponding weight of each coordinate.

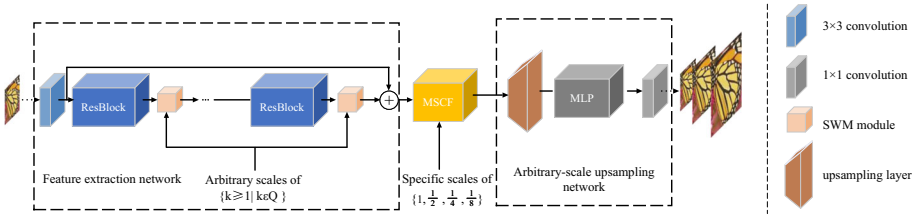

Fig. 1 The network structure of our proposed MCNet, which contains three main parts for: 1) feature extraction network, 2) cross-fusion module and 3) image reconstruction network

Consider a set of $(I_i^{LR}, I_i^{HR})_i^N$ that contains N $LR - HR$ pairs, where I_i^{LR} is an input LR image and I_i^{HR} stands for the corresponding ground-truth(GT) image. We choose the L_1 loss function to optimize our network during training.

$$\Theta^* = \arg\min_{\Theta} \frac{1}{N} \sum_{i=1}^{N} \|\Omega(I_i^{LR}) - I_i^{HR}\|_1 \tag{4}$$

where Ω denotes the set of learning parameters in our proposed model.

2.2 Scale-Wise Module (SWM)

Inspired by the idea of multi-modal [24], a great number of researches have taken full advantage of multimodal fusion algorithm to combine two independent variables. For example, the study of reading image information is to input two independent variables, such as image and text, into the backbone network to form two interleaving branches so that the corresponding information can communicate with each other. Among them, the researchers choose to design a fusion module between the two branches, which can perform multi-modal learning on two completely different variables to establish a close relationship (Fig. 2).

On the basis of the prior information above, we design a plug and play module, called scale-wise module, after each residual block of the EDSR [18] backbone network. Compared with ArbSR, this module requires less computation and fewer parameters, which can effectively combine the scale information with the image pixel features and fully improve the ability of backbone network to represent multi-scale images.

As shown in Fig. 3, we assume that F represents the pixel feature of the image and S is arbitrary-scale information, so that the working principle of the intelligent scale module can be expressed as

$$W_{FS} = \delta_s \left[f_1(F) \otimes f_2(S) \right] \tag{5}$$

where δ_s is the sigmoid activation function, $f_k(k = 1, 2, 3)$ represents different 1×1 convolution layers, and $\otimes$ denotes the matrix multiplication algorithm. W_{FS} is the pixel-scale weight matrix, which represents the result of the activation function mapping from 0 to 1 after the communication of image pixel features and scale information. This process is called pixel-scale attention manipulation. Then, the attention matrix W_{FS} is dotted with F through the convolution layer, passing more useful spatial information to the next residual block of

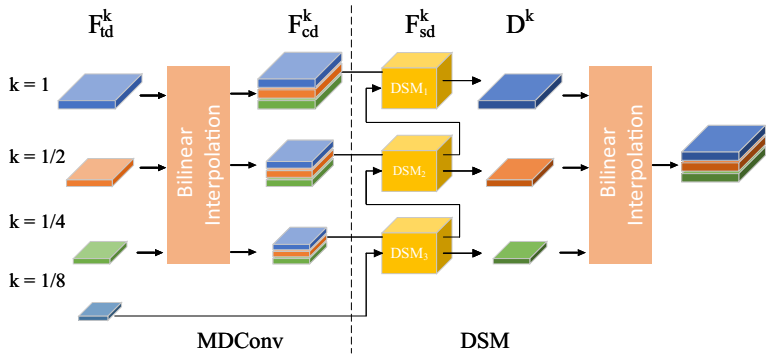

Fig. 2 Architecture of the multi-scale cross-fusion module (MSCF)

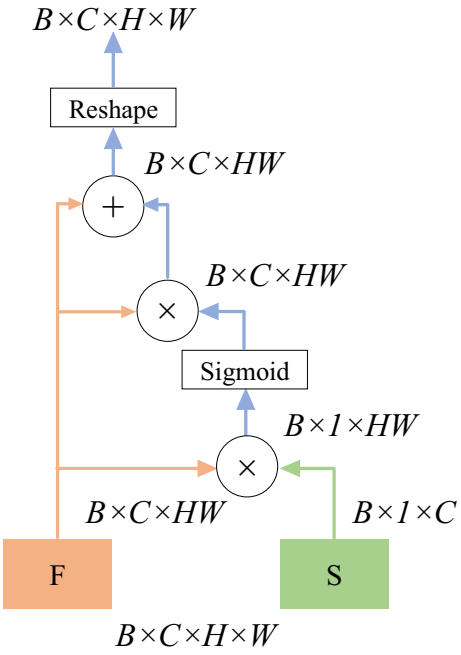

Fig. 3 Architecture of scale-wise module(SWM)

the EDSR [18].

$$F_S = W_{FS} \cdot f_3(F) + f_3(F) \tag{6}$$

It should be noted that the attention matrix $W_{FS} \in R^{B\times 1\times HW}$ and the pixel feature matrix $f_3(F) \in R^{B\times C\times HW}$, so the pixel-scale attention operations are spatial attention mechanisms that perform the same operations on all channel dimensions. The weight matrix W_{FS}, which is obtained by multiplying pixels and scale feature matrix, can be adapted to scale information to further discriminate the whole image space. Finally, we apply the reshape function to transform the tensor F_S with a shape of $B \times C \times HW$ into $B \times C \times H \times W$. This tensor will serve as the input for the next residual block in the encoder.

2.3 Multi-Scale Cross-Fusion module (MSCF)

To further improve the quality of the reconstruction images in backbone network, we design a powerful module consisting of multi-downsampling convolutional architecture (MDConv) and dual spatial mask (DSM). Referring in Fig. 2, in MDConv module, a set of convolutional layers are conducted to downsample the deep features F_d delivered by the SR backbone network; that is,

$$F_{td}^k = Conv \downarrow_k (F_d^k) \tag{7}$$

where $k(k = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1)$ represents the downsampling factor. F_{td} is the downsampled feature with a specific scale, which contains more plentiful global features of images. By performing the interpolation in space and concatenation in channel, the generated feature F_{td}^k is used to redefine the new feature map F_{cd}^k. Note that we use bilinear interpolation to make feature maps of different scales the same size. MDConv provides many feature maps with different receptive fields and structural information for the next step. Then, the multi-scale features

F_{cd} are fed into our MSCF sub-module dual spatial mask (DSM) in succession through performing the communication as follows:

$$D^{\frac{1}{4}}, C^{\frac{1}{4}} = DSM_1(F_{cd}^{\frac{1}{4}}, F_d^{\frac{1}{8}}) \tag{8}$$

$$D^{\frac{1}{2}}, C^{\frac{1}{2}} = DSM_2(F_{cd}^{\frac{1}{2}}, C^{\frac{1}{4}})$$

$$D^{1} = DSM_3(F_{cd}^{1}, C^{\frac{1}{2}})$$

$$F_{od} = Coi(D^{\frac{1}{4}}, D^{\frac{1}{2}}, D^{1}) \tag{9}$$

where D^k and C^k are the corresponding outputs by the DSM module, and Coi represents the corresponding operation of interpolation and concatenation. The operator $DSM_i(\cdot)$ denotes our dual spatial mask (DSM), which learns attention weights from two feature maps with different scales and its detailed structure is shown as follows:

$$F = F \cdot SM(C \uparrow^2) + F$$

$$C = C \cdot SM(F) + C \tag{10}$$

where F and C denote two different inputs of the mask module, $\uparrow^2$ is the operation for $\times 2$ upsampling. $SM(\cdot)$ is the spatial gate mechanism. Note that, two inputs of different sizes are adjusted to the same shape through the processing of our DSM. D^k is served as a part of the final output from MSCF, while C^k is used for the interactive learning in the next sub-module. They are utilized to learn additional textures and structures from each other.

3 Experiment results

3.1 Implementation details

As same as the setting in EDSR [18], we train our MCNet with DIV2K datasets. For testing, our MCNet are evaluated by using four standard benchmark datasets: Set5 [7], Set14 [25], B100 [26] and Urban100 [27]. During training, 16 degraded patches of size 48*48 are used as a batch input. For upsampling part, we sample random scale factors in uniform distribution *U(1, 4)*. Each example in a batch has different upsampling target. Adam [28] optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$ is utility in the MCNet for 1000 epochs. The learning rate is initialized to 1×10^{-4} and decreased by factor 0.5 at [200, 400, 600, 800].

3.2 Performance evaluation

Six SOTA SR networks are used to compare with our proposed MCNet method, including EDSR [18], Meta-SR [19], ArbSR [22], LIIF [20], LTE [21] and EDBNet [23]. Table 1 displays the Peak Signal-to-Noise Ratio (PSNR) values for four benchmark datasets at upscaling factors ranging from $\times 2$ to $\times 8$. It is important to note that EDSR [23] belongs to the category of single-scale image super-resolution models, and we only conducted training and testing on standard scales of $\times 2$, $\times 3$ and $\times 4$. We can find that our proposed MCNet significantly outperforms EDBNet [23] on the urban100 dataset. Specifically, compared with the EDBNet [23] model and our MCNet method, the PSNR results show improvements at medium scales of our model. Furthermore, we also show a visual comparison in Fig. 4. For the challenging details in "img044" and "img054", most previous work lost some crucial details when restoring the images. On the contrary, our MCNet achieves better results by

Table 1 Quantitative Results of Sate-of-The-Art Arbitrary-Scale SR Methods

Method	Scales	EDSR [18]	Meta-SR [19]	ArbSR [22]	LIIF [20]	LTE [21]	EDBNet [23]	MCNet
Set5 [7]	×2	38.05	37.97	38.04	37.97	38.04	38.07	**38.16**
	×3	34.39	34.37	34.43	34.41	34.46	34.53	**34.61**
	×4	32.13	32.04	32.07	32.24	32.25	32.29	**32.40**
	×6	–	28.68	28.70	28.94	28.96	28.98	**29.05**
	×8	–	26.70	26.72	26.97	27.06	27.07	**27.16**
Set14 [25]	×2	33.60	33.60	33.68	33.61	33.65	33.70	**33.79**
	×3	30.32	30.31	30.35	30.33	30.37	30.43	**30.51**
	×4	28.58	28.53	28.58	28.61	28.67	28.67	**28.73**
	×6	–	26.31	26.34	26.48	26.50	26.52	**26.60**
	×8	–	24.79	24.79	24.97	24.97	25.00	**25.07**
BSD100 [26]	×2	32.19	32.17	32.22	32.17	32.19	32.20	**32.32**
	×3	29.09	29.09	29.13	29.09	29.12	29.15	**29.24**
	×4	27.57	27.56	27.59	27.59	27.60	27.62	**27.71**
	×6	–	25.76	25.79	25.84	25.85	25.87	**25.93**
	×8	–	24.72	24.73	24.81	24.82	24.83	**24.89**
Urban100 [27]	×2	32.14	32.11	32.24	32.15	32.24	32.34	**32.46**
	×3	28.17	28.12	28.24	28.22	28.30	28.36	**28.43**
	×4	26.06	25.95	26.05	26.14	26.19	26.27	**26.35**
	×6	–	23.59	23.69	23.77	23.83	23.87	**23.96**
	×8	–	22.29	22.37	22.45	22.53	22.55	**22.64**

Boldface Indicates the Best (PSNR (dB))

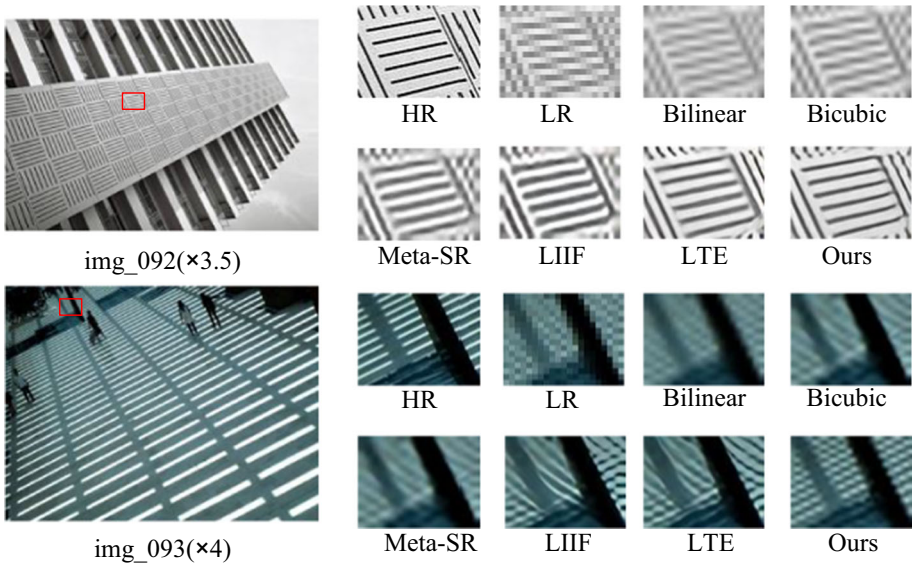

Fig. 4 Qualitative comparison of different methods on Urban100 datasets

Table 2 Memory usage and time consumption compared with other arbitray-scale SR models for $\times 2$ upsampling

Query	Method	Params.	Mem.(MiB)	Time(ms)
262144	MetaSR	1.7M	2369	81.8
(512 × 512)	LIIF	1.6M	2257	177.5
	LTE	1.7M	2191	238.4
	EDBNet	1.8M	2219	245.3
	MCNet	2.0M	2293	263.7

recovering more detailed components. In addition, as shown the cost consumption of four arbitrary-scale image super-resolution models in Table 2, we can find that the MCNet model only increases a little additional computation resources. In a word, compared with other arbitrary scale super-resolution methods, our model has the most advanced image reconstruction performance although it adds extra computational cost.

3.3 Ablation study

To confirm the effectiveness of the scale-wise module (SWM), we compared ArbSR's Scale-Aware Feature Adaption (SAFA) to the plug-and-play module in this letter. Table 3 shows that SWM has a very small number of parameters and consumes less computer resources. Moreover, SWM module has better performance for SR image restoration than the SAFA module of ArbSR. The PSNR results is tested on Urban100 with $\times 4$ upsampling. All in all, it is very rare that our SWM module is more superior than SAFA module with very little resource consumption.

We all know that if the image is downsampled continuously, the small scale image will have more comprehensive global information and less noise. In order to generate cleaner and richer high-resolution images, we design a Multi-Scale Cross-Fusion Module (MSCF) after the feature extraction network of the model, which includes multi-downsampling convolutional architecture (MDConv) and dual spatial mask (DSM). After several down-sampling processes, MDConv can obtain feature maps of various sizes . The multi-scale texture and structure information in different feature maps provides an important information basis for later learning of DSM. Diverse information uses DSM interactive learning to absorb more semantic information from each other, so that the output high-dimensional feature map contains clean and rich texture and structure information.

Table 4 shows the ablation experiments of the DSM and MDConv. In our two variants, we validate the effectiveness by performing on the dataset Urban100 with six scale factors from 2 to 8. All networks are pre-trained on the EDSR [18] backbone for 1000 epochs. The definitions of -D, -M indicate that MCNet removes the corresponding components of DSM, and MDConv. Compared MCNet with MCNet(-D), We observe that using DSM achieves further improvement particularly with the upsampling scales that in training distribution, which is consistent with our motivation. To confirm the validation of MDConv, we also compare MCNet to MCNet(-M), which enhances the quality of both in-scale and out-of-scale factors.

Table 3 Computer resource consumption compared with ArbSR method for $\times 2$ upsampling

Method	Params.	Mem.(MiB)	Time(ms)	PSNR
ArbSR(-SAFA)	4.2M	2297	296.4	26.28
MCNet(-SWM)	1.7M	2281	223.8	26.31

Table 4 An Ablation Investigation of Two Variants Performed on the Dataset Urban100

Variant	In-scale			Out-of-scale	
	2×	3×	4×	6×	8 ×
MCNet	32.46	28.43	26.35	23.96	22.64
MCNet(-D)	32.39	28.41	26.32	23.91	22.59
MCNet(-M)	32.38	28.39	26.30	23.88	22.57

4 Conclusion

In this letter, we propose a novel scale-guidance fusion network (MCNet) for the existing SISR network with arbitrary scaling factors. The designed scale module (SWM) integrates the scale information and pixel features to effectively improve the representation ability of arbitrary scale images. In addition, the multi-scale cross-fusion module (MSCF) cleans redundant noises of deep feature maps and provides abundant space embedding for subsequent image restoration. The comprehensive evaluation has demonstrated that our MCNet achieves superior performance compared to state-of-the-art arbitrary-scale works.

Author Contributions Guangping Li: Conceptualization, Methodology, Visualization Writing - Review and Editing. Huanling Xiao: Data Curation, Writing - Original Draft, Software. Dingkai Liang: Writing - Review and Editing, Supervision. Bingo Wing-Kuen Ling: Writing - Review and Editing, Supervision.

Funding This work is supported by Science and Technology Planning Project of Daya Bay (Grant No. 2020010203) and National Natural Science Foundation of China (Grant No. 61601130)

Availability of data and materials The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethical Approval The authors' research does not address ethical issues.

Competing interests The authors declared no conflict of interest.

References

1. Dong C, Loy CC, He K, Tang X (2015) Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell 38(2):295–307
2. Shi W, Caballero J, Huszár 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: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1874–1883
3. Kim J, Lee JK, Lee KM (2016) Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1637–1645
4. Dong C, Loy CC, Tang X (2016) Accelerating the super-resolution convolutional neural network. In: European conference on computer vision, Springer, pp 391–407
5. Tong T, Li G, Liu X, Gao Q (2017) Image super-resolution using dense skip connections. In: Proceedings of the IEEE international conference on computer vision, pp 4799–4807
6. Zhang Y, Li K, Li K, Wang L, Zhong B, Fu Y (2018) Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European conference on computer vision (ECCV), pp 286–301
7. Bevilacqua M, Roumy A, Guillemot C, Alberi-Morel ML (2012) Low-complexity single-image super-resolution based on nonnegative neighbor embedding

8. Kim J, Lee JK, Lee KM (2016) Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1646–1654
9. Lai W-S, Huang J-B, Ahuja N, Yang M-H (2017) Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 624–632
10. Kim J-H, Lee J-S (2018) Deep residual network with enhanced upscaling module for super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 800–808
11. Hui Z, Wang X, Gao X (2018) Fast and accurate single image super-resolution via information distillation network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 723–731
12. Shuai Y, Wang Y, Peng Y, Xia Y (2018) Accurate image super-resolution using cascaded multi-column convolutional neural networks. In: 2018 IEEE International conference on multimedia and expo (ICME), IEEE, pp 1–6
13. Chang K, Li M, Ding PLK, Li B (2020) Accurate single image super-resolution using multi-path wide-activated residual network. Signal Process 172:107567
14. Li F, Cong R, Bai H, He Y (2020) Deep interleaved network for image super-resolution with asymmetric co-attention. arXiv preprint arXiv:2004.11814
15. Huang Y, Shao L, Frangi AF (2017) Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 6070–6079
16. Mahapatra D, Bozorgtabar B, Garnavi R (2019) Image super-resolution using progressive generative adversarial networks for medical image analysis. Comput Med Imaging Graph 71:30–39
17. Lu T, Wang J, Zhang Y, Wang Z, Jiang J (2019) Satellite image super-resolution via multi-scale residual deep neural network. Remote Sensing 11(13):1588
18. Lim B, Son S, Kim H, Nah S, Mu Lee K (2017) Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 136–144
19. Hu X, Mu H, Zhang X, Wang Z, Tan T, Sun J (2019) Meta-sr: A magnification-arbitrary network for super-resolution. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1575–1584
20. Chen Y, Liu S, Wang X (2021) Learning continuous image representation with local implicit image function. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 8628–8638
21. Lee J, Jin KH (2022) Local texture estimator for implicit representation function. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 1929–1938
22. Wang L, Wang Y, Lin Z, Yang J, An W, Guo Y (2021) Learning a single network for scale-arbitrary super-resolution. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 4801–4810
23. Li G, Xiao H, Liang D (2022) Enhanced dual branches network for arbitrary-scale image super-resolution. Electron Lett
24. Li L, Han L, Ding M, Cao H (2023) Multimodal image fusion framework for end-to-end remote sensing image registration. IEEE Trans Geosci Remote Sens 61:1–14
25. Zeyde R, Elad M, Protter M (2010) On single image scale-up using sparse-representations. In: international conference on curves and surfaces, Springer, pp 711–730
26. Martin D, Fowlkes C, Tal D, Malik J (2001) 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, IEEE, 2:416–423
27. 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, pp 5197–5206
28. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980