Image Super-Resolution Postprint Based on Deep Residual Back-Projection Attention Network

Authors: Hu Gaopeng, Chen Ziliu, Wang Xiaoming, Zhang Kaifang, Huang Zengxi, Du Yajun

Date: 2020-09-28T00:00:00+00:00

Abstract

To address the issues of insufficient feature information exploitation, difficulty in determining the interdependent relationships among channels of feature maps, and reconstruction errors when reconstructing high-resolution (HR) images, which exist in most single image super-resolution (SISR) reconstruction methods, we propose an image super-resolution (SR) algorithm based on a deep residual back-projection attention network. Specifically, it utilizes the idea of residual learning to alleviate training difficulty and fully exploit the feature information of images, employs a back-projection learning mechanism to learn the interdependent relationships between low- and high-resolution images, and additionally introduces an attention mechanism to dynamically allocate different attention resources to each feature map, thereby excavating more high-frequency information and learning the dependencies among channels of feature maps. Experimental results demonstrate that compared with most single image super-resolution methods, the proposed method not only achieves significant improvements in objective metrics but also produces reconstructed prediction images with richer texture information.

Full Text

Preamble

Vol. 38 No. 3 *Application Research of Computers* (ChinaXiv Partner Journal) Accepted Paper

Image Super-Resolution Based on Deep Residual Back-Projection Attention Network

Hu Gaopeng[1], *Chen Ziliu*[1], *Wang Xiaoming*[1,2]*†*, *Zhang Kaifang*[1], *Huang Zengxi*[1], *Du Yajun*[1]
([1]School of Computer & Software Engineering; [2]Robotics Research Center, Xihua University, Chengdu 610039, China)

Abstract: Most single image super-resolution (SISR) reconstruction methods suffer from insufficient feature information exploitation, difficulty in determining interdependencies among feature map channels, and reconstruction errors when generating high-resolution (HR) images. To address these issues, this paper proposes an image super-resolution (SR) algorithm based on a deep residual back-projection attention network. The method leverages residual learning to ease training difficulty and fully exploit image feature information, employs back-projection learning mechanisms to capture mutual dependencies between low- and high-resolution images, and introduces an attention mechanism to dynamically allocate different attention resources to feature maps, thereby discovering more high-frequency information and learning dependencies among feature map channels. Experimental results demonstrate that compared with most single image super-resolution methods, the proposed approach achieves significant improvements in objective metrics while producing reconstructed images with richer texture information.

Keywords: attention mechanism; super-resolution; back projection; residual learning; convolutional neural network

0 Introduction

Single image super-resolution aims to reconstruct HR images from a single low-resolution (LR) input. This technique has found widespread applications in real-world scenarios, such as enhancing face recognition accuracy in surveillance videos, producing higher-quality video in HDTV, and obtaining high-resolution medical images. Through continuous exploration and development, numerous machine learning and deep learning-based SR methods have emerged. Given the clear advantages of deep learning approaches, this paper focuses on deep learning-based SISR tasks.

Current deep learning SR algorithms can be categorized by their upsampling strategies into front-end upsampling methods [1–5], back-end upsampling methods [6–9], progressive upsampling methods [10], and iterative upsampling methods [11]. Dong et al. [1] first applied convolutional neural networks to SR tasks, proposing the Super-Resolution Convolutional Neural Network (SRCNN) method, which uses bicubic interpolation to upsample LR images to a target scale before learning the mapping to HR images through a three-layer CNN. Subsequently, Kim et al. [2] proposed Very Deep Super-Resolution (VDSR), which introduced residual learning to mitigate gradient vanishing/explosion, deepened the network to 20 layers, and employed a larger receptive field to extract features from LR images, thereby accelerating convergence and improving reconstruction quality.

Tai et al. [6] proposed Deep Recursive Residual Network (DRRN), which utilized local and global residual strategies with recursive mechanisms to deepen the network to 52 layers while maintaining parameter efficiency, achieving further performance improvements. Kim [7] et al. introduced Deep Recursive Convolutional Network (DRCN), which used recursive sharing of network parameters to reduce training difficulty. Dong et al. [8] proposed Fast SRCNN (FSRCNN), employing deconvolution at the back-end of the CNN to enlarge image dimensions. Shi et al. [9] proposed Efficient Sub-Pixel Convolutional Neural Network (ESPCN), using sub-pixel convolution at the network' s back-end to map learned LR features to the target resolution. Both FSRCNN and ESPCN demonstrated that back-end upsampling strategies effectively reduce computational complexity while improving HR spatial resolution.

Lai et al. [10] proposed Laplacian Pyramid Networks (LapSRN), which integrates Laplacian pyramid concepts to progressively learn high-frequency components of images while only applying bicubic interpolation to low-frequency components. This strategy accelerates learning progress while enhancing reconstruction quality.

Haris et al. [11] subsequently proposed Deep Back-Projection Networks for Super-Resolution (DBPN), which employs continuous iterative upsampling and downsampling to learn mappings between HR and LR images, using an error feedback mechanism to correct reconstruction errors. Although this method achieved good results through interconnected up-down sampling strategies and error feedback, its deep architecture produced overly smooth HR images and overlooked the differential contributions of feature maps generated at different stages to the final HR prediction.

To address these limitations, this paper proposes a Residual Back-Projection Attention Network that integrates residual concepts with attention mechanisms to mitigate high-frequency information loss and learn interdependencies among feature maps. The main contributions are twofold: (a) We combine iterative back-projection with residual learning to propose a residual back-projection method that effectively reduces training difficulty, minimizes information loss during training, and preserves high-frequency features. (b) We introduce and develop an attention mechanism, proposing a Global Attention unit that automatically allocates attention resources to feature maps and their channels from different stages of residual back-projection blocks, enabling the discovery of more high-frequency information during HR image reconstruction.

1.1 Residual Learning

When training very deep networks, gradient vanishing can occur during backpropagation due to initialization parameters being close to zero, often degrading rather than improving performance. He et al. [12] proposed ResNet to address this through residual learning. The core idea assumes that in a deep network, redundant layers should ideally perform identity mapping (i.e., maintain identical input and output). However, learning identity mapping is difficult. ResNet avoids this by using the architecture shown in Figure 1, where when the residual term $F(x) = 0$, the construction $H(x) = x$ becomes trivial. This transforms learning $H(x) = x$ into learning $F(x) = H(x) - x$ as a residual term.

Figure 1 [Figure 1: see original paper] Residual learning

1.2 Deep Back-Projection Networks

Haris et al. [11] proposed Deep Back-Projection Network (DBPN), which uses iterative back-projection to learn mappings between LR and HR images and employs error feedback to correct reconstruction errors. The DBPN architecture consists of an initial feature extraction unit, multiple interconnected back-projection units, and a reconstruction layer. They further developed Dense DBPN (D-DBPN) by introducing dense connections, as shown in Figure 2. For an input LR image, shallow features F_0 are first extracted, then multiple iterative up- and down-projection units learn reconstruction errors between HR and LR features, and finally, HR feature maps from all previous stages are concatenated for image reconstruction. Each back-projection unit contains upsampling and downsampling operations implemented through deconvolution and convolution layers, respectively.

Figure 2 Schematic diagram of deep back-projection network

---

2 Our Method

Although D-DBPN [11] has achieved satisfactory results, several issues remain. The deep architecture causes feature information loss and gradient vanishing. While dense connections were used to improve DBPN and deepen the network, they concatenate feature maps at the dimensional level without fully exploiting feature information in HR and LR spaces. Additionally, D-DBPN treats hierarchical features equally, ignoring dependencies among different layers and redundant information across stages. To address these limitations, we propose the Residual Back-Projection Attention Network.

2.1 Network Architecture

The proposed Residual Back-Projection Attention Network is illustrated in Figure 3, comprising four modules: a shallow feature extraction unit, Residual

Back-Projection (RBP) units, a Global Attention (GA) unit, and a reconstruction unit. The shallow feature extraction unit uses two convolutional layers to extract shallow features F_0 from the input LR image. Given input LR image I^{LR} and predicted HR image I^{HR}, the shallow feature extraction is formulated as:

$$F_0 = \phi(I^{LR}) \tag{1}$$

where ϕ denotes convolution on the input LR image. The extracted shallow features F_0 are then fed into residual back-projection units to discover deeper high-frequency features:

$$F_{RBP} = H_{RBP}(F_0) \tag{2}$$

where H_{RBP} represents the residual back-projection unit. Our residual back-projection unit contains N back-projection units connected via residual links, with each unit consisting of an up-projection block and a down-projection block. These interconnected blocks learn reconstruction errors between various upsampling and downsampling operators, using error feedback to correct mappings between HR and LR features. Each projection unit concatenates HR features from all previous units, and residual learning is introduced via skip connections to enhance feature learning (detailed in Section 3.2).

Subsequently, the global attention mechanism learns to allocate attention resources among concatenated HR feature maps and their channels from all projection units:

$$F_{GA} = H_{GA}(F_{RBP}) \tag{3}$$

where H_{GA} denotes the Global Attention unit. The GA unit automatically allocates different attention resources to feature maps and channels based on their contributions to the final reconstruction and their interdependencies. Finally, the features from the GA unit are reconstructed using a single convolutional layer:

$$I^{HR} = H_{REC}(F_{GA}) = H_{REC}(H_{GA}(H_{RBP}(\phi(I^{LR})))) \tag{4}$$

where the reconstruction unit comprises a single convolutional layer, and H_{RBPAN} and H_{REC} represent the residual back-projection network and reconstruction unit, respectively.

Section 3.1 analyzes the appropriate loss function for our method. Due to the sparsity of L1 norm and its proven faster convergence [20], we select L1 norm as our optimization objective. Given a training set $\{I_i^{LR}, I_i^{HR}\}_{i=1}^{N}$ containing N LR-HR image pairs, our optimization goal is to minimize the L1 loss function:

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N} \|H_{RBPAN}(I_i^{LR}) - I_i^{HR}\|_1 \tag{5}$$

where θ represents all network parameters. We use stochastic gradient descent to minimize the L1 loss. Since the shallow feature extraction and reconstruction units are straightforward, we focus on detailing the residual back-projection unit and global attention unit.

2.2 Residual Back-Projection Unit

This section details the residual back-projection unit, which contains T back-projection units and one up-projection block. Each back-projection unit includes an up-projection block and a down-projection block, with dense connections between units. Since dense concatenation occurs at the dimensional level, linear mapping is employed within each projection block for feature fusion. To fully exploit deep feature information, residual learning is introduced:

$$F_{DF,t} = \psi(F_{DF,t-1} + \varphi(F_{DF,1}, F_{DF,2}, ..., F_{DF,t-1})) \tag{6}$$

where $H_{RBP,t}$ and ψ denote the t-th back-projection unit and element-wise sum function, respectively. $F_{DF,t}$, $F_{DF,t-1}$, and $F_{DF,t-2}$ represent deep features from the t-th, $(t-1)$-th, and $(t-2)$-th back-projection units, while $F_{IN,t-1}$ is the input to the $(t-1)$-th unit. Skip connections add the input features of the $(t-1)$-th unit to the element-wise sum of previous units' outputs, forming the input to the t-th unit.

The input to the global attention mechanism is the concatenated HR feature maps from all projection units. Let $X \in \mathbb{R}^{C\times m\times n}$ denote the feature maps, where the t-th back-projection unit outputs HR features $H_t \in \mathbb{R}^{m\times n}$ with dimension m. For input features of size $H \times W \times C$, we apply global pooling for channel-wise statistics:

$$z_c = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_c(i,j) \tag{7}$$

where $x_c(i,j)$ represents the feature at position (i,j) in the c-th feature map, and z_c is the statistic for the c-th channel. The aggregated features then pass through convolutional layers, ReLU activation, another convolution, and finally a gating function to compute channel attention weights. These weights are element-wise multiplied with input feature maps. The complete process is formulated as:

$$X' = \delta(W_U \cdot f(W_D \cdot Z)) \otimes X$$

where f and δ represent ReLU and Sigmoid functions, W_U and W_D are weight matrices, X and X' are input and attention-weighted features, and Z denotes globally pooled features.

Each back-projection unit comprises an up-projection block and down-projection block, whose structures are shown in Figures 4 and 5, respectively.

a) Up-projection Block: The input concatenates outputs from previous down-projection blocks. For the t-th up-projection block, the input is $[L_{t-1}, L_{t-2}, ..., L_0]$. A concatenation layer combines these inputs, followed by a 1×1 convolution to reduce dimensions and obtain feature H_t. Upsampling and downsampling produce H_t^U and L_t^D, respectively. The error $e_t^L = L_{t-1} - L_t^D$ is computed and used to correct the mapping between HR and LR features.

b) Down-projection Block: Similarly, inputs concatenate residual learning results from previous up-projection blocks. After concatenation and linear mapping to obtain feature map L_t, downsampling and upsampling operations compute reconstruction error $e_t^H = H_t - H_t^U$, which guides LR feature reconstruction.

Figure 4 Up-projection block structure
Figure 5 [Figure 5: see original paper] Down-projection block structure

2.3 Global Attention Unit

During HR image reconstruction, outputs from up-projection blocks at each stage are used. However, features learned at different stages exhibit individual differences, and channels within the same stage contain distinct information, contributing differently to final reconstruction. To address this, we introduce a global attention mechanism that allocates varying attention to each channel of HR features from projection units, enabling discovery of more detailed information.

The Global Attention unit structure is shown in Figure 6, where $\otimes$ denotes element-wise multiplication. The module input is the concatenated HR features from all back-projection units.

Figure 6 [Figure 6: see original paper] Schematic diagram of global attention unit

---

3 Experiments

We trained our network on 800 images from the DIV2K dataset [13] and evaluated it on four standard benchmark datasets: Set5 [14], Set14 [15], BSD100 [16], and Urban100 [17]. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) [18] served as objective evaluation metrics. For different upscaling factors, we used different deconvolution kernel sizes: 6×6 for scale factor 2, 8×8 for scale factor 4, and 12×12 for scale factor 8. The initial learning rate was set to 1×10^{-4}, reduced by a factor of 10 every 6×10^5 iterations. We used L1 loss with Adam optimizer [19] (momentum 0.9). All experiments were conducted on an Nvidia TITAN X (Pascal) GPU and Intel(R) Xeon(R) W-2125 CPU.

3.1 Comparison of Norm Loss and MSE Loss Functions

The loss function defines the network' s learning objective and significantly impacts results. In SISR tasks, Mean Squared Error (MSE) and L1 norm loss functions are commonly used. While MSE tends to yield higher PSNR, recent research [20] shows L1 norm accelerates convergence. To select the optimal loss function for our method, we constructed Model I with MSE loss and Model II with L1 loss, keeping network depth, kernel sizes, and other settings identical (scale factor 4, 6 back-projection units). Testing on Set5 and Set14 datasets produced the results shown in Figures 7 and 8.

On Set5, L1 loss achieved average PSNR and SSIM improvements of 0.10 dB and 0.0007 over MSE loss. On Set14, the improvements were 0.73 dB and 0.0227. These results confirm L1 loss as the more suitable optimization objective for our method.

Figure 7 Comparison of PSNR between L1-norm loss and MSE loss on different datasets
Figure 8 [Figure 8: see original paper] Comparison of SSIM between L1-norm loss and MSE loss on different datasets

3.2 Model Analysis

Our Deep Residual Back-Projection Attention Network integrates residual learning and attention mechanisms to improve prediction quality. To validate each component's effectiveness, we constructed three models by sequentially removing residual connections and attention mechanisms: Model III (original D-DBPN [11]), Model IV (with only global attention), and Model V (with both residual learning and global attention). All models used 6 back-projection blocks with kernel sizes 6×6, 8×8, and 12×12 for scale factors 2, 4, and 8, respectively. Results are shown in Table 1.

Table 1 demonstrates that Models IV and V significantly outperform Model III in PSNR and SSIM across all datasets. At scale factor 4 on Set5, Models IV and V improve PSNR/SSIM by 1.46 dB/0.039 and 1.38 dB/0.038, respectively. On Set14, improvements are 0.82 dB/0.087 and 1.55 dB/0.108. At scale factor 8, Models IV and V show gains of 0.79 dB/0.076 and 1.27 dB/0.081 on Set5, and 1.09 dB/0.145 and 1.13 dB/0.146 on Set14. Except for one case (Set5, scale factor 4), Model V consistently outperforms Model IV, confirming the overall superiority of the full method.

Table 1 Performance of Models III, IV, and V on Set5 and Set14 datasets

3.3 Parameter Scale Comparison

To evaluate parameter efficiency, we compared our method' s parameter count with state-of-the-art algorithms including SRCNN [1], VDSR [2], FSRCNN [8], LapSRN [10], D-DBPN [11], EDSR [20], MemNet [21], and RCAN [22]. Figures 9 and 10 show parameter scale comparisons for scale factors 4 and 8 on Set5.

At scale factor 4, our method achieves 0.81 dB higher PSNR than the second-best algorithm while maintaining a smaller parameter count. At scale factor 8, our parameter count increases but still yields 0.88 dB PSNR improvement over the runner-up. Overall, our method demonstrates superior PSNR performance with reasonable parameter efficiency.

Figure 9 Parameter scale comparison of x4 models of mainstream algorithms on Set5 dataset
Figure 10 [Figure 10: see original paper] Parameter scale comparison of x8 models of mainstream algorithms on Set5 dataset

3.4 Comparison with State-of-the-Art Methods

We validated our method against leading algorithms on four benchmark datasets: Set5 and Set14 from the University of Liège and Bell Labs, BSDS100 from UC Berkeley, and Urban100 from Huang et al. Compared methods include SRCNN [1], VDSR [2], FSRCNN [11], LapSRN [10], D-DBPN [11], EDSR [20], MemNet [21], RCAN [22], SCN [23], SRMDNF [24], and RDN [25]. We evaluated PSNR, SSIM, and visual quality at scale factors 2, 4, and 8, using publicly available code retrained on the same dataset for fair comparison.

Tables 2–4 present quantitative results. At scale factor 2, our method achieves the best PSNR on all datasets except Set5, with the most significant improvement on BSDS100 (2.93 dB over RCAN) and optimal SSIM across all datasets. At scale factors 4 and 8, our method consistently achieves the best PSNR and SSIM performance, again with substantial gains on BSDS100. The advantages become more pronounced at larger scaling factors.

Table 2 Average performance of x2 models of various SISR algorithms on different datasets
Table 3 Average performance of x4 models of various SISR algorithms on different datasets
Table 4 Average performance of x8 models of various SISR algorithms on different datasets

For visual comparison, we selected representative images "women" from Set5, "comic" from Set14, "119082" from BSDS100, and "img_{037}" from Urban100, showing reconstruction results at scale factor 4 in Figures 11–14. Bicubic interpolation produces blurry results lacking detail. VDSR and LapSRN show modest improvements but still miss details. EDSR, D-DBPN, and RCAN achieve better quality but lack sharp edges compared to our method. Both objective metrics and subjective visual quality confirm our method' s superior performance.

Figure 11 [Figure 11: see original paper] Comparison of reconstruction results for "women" from Set5 dataset
Figure 12 [Figure 12: see original paper] Comparison of reconstruction results for "comic" from Set14 dataset
Figure 13 [Figure 13: see original paper] Comparison of reconstruction results for "119082" from BSDS100 dataset
Figure 14 [Figure 14: see original paper] Comparison of reconstruction results for "img_{037}" from Urban100 dataset

---

4 Conclusion

This paper proposes a Deep Residual Back-Projection Global Attention Network that integrates iterative back-projection with residual learning and global attention mechanisms. This approach alleviates insufficient feature utilization and high-frequency information loss while exploiting differences among feature maps to discover more useful high-frequency information for reconstruction. Experimental results demonstrate our algorithm' s superiority in PSNR and SSIM metrics, producing predictions with richer details and better visual quality.

References

[1] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks [J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38(2): 295-307.
[2] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks [C]// Proceedings of the 2016 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016. 1646-1654.
[3] Wang Xiaoming, Huang Feng, Liu Shaopeng, Xu Tao. Improved single-image self-learning super-resolution reconstruction method [J]. Computer Application Research, 2019, 36(08): 2534-2538.
[4] Shao Wenze, Wei Zhihui. Variational super-resolution reconstruction of multi-frame images based on anisotropic MRF modeling [J]. Chinese Journal of Electronics, 2009, 37(6): 1256-1263.
[5] Li Zhan, Zhang QingFeng, Meng Xiaohua, et al. Super-resolution reconstruction of multi-resolution image sequences [J]. Acta Automatica Sinica, 2012, 38(11): 1804-1814.
[6] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network [C]// Proc of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017.
[7] Kim J, Kwon L J, MU L K. Deeply-recursive convolutional network for image super-resolution [C]// Proceedings of the 2016 IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016. 1637–1645.

[8] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network [C]// Proceedings of the 2016 European Conference on Computer Vision. Amsterdam, Netherlands: Springer Verlag, 2016: 391-407.
[9] Shi W, Caballero J, HUSZÁR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1874-1883.
[10] Lai W S, Huang J B, Ahuja N, et al. Deep laplacian pyramid networks for fast and accurate super-resolution [C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 624-632.
[11] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution [C]// Proceeding of the 2018 IEEE International Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 1664-1673.
[12] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition [C]// Proceeding of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, 2016: 770-778.
[13] Timofte R, Agustsson E, Van G L, et al. NTIRE 2017 challenge on single image super-resolution: Methods and results [C]// Proceeding of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2017: 1110-1121.
[14] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [C]// Proceeding of the 2012 British Machine Vision Conference. Durham: BMVA Press, 2012: 135.1–135.10.
[15] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations [C]// Proceeding of the 7th International Conference on Curves and Surfaces. Berlin: Springer, 2010: 711-730.
[16] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics [C]// proceeding of the 2001 Eighth IEEE International Conference on Computer Vision. Canada: IEEE 2001.
[17] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars [C]// Proceedings of the 2015 Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015. 5197-5206.
[18] YE Y X, Shan J, Bruzzone L, et a1. Robust registration of multimodal remote sensing images based on structural similarity [J]. IEEE Transactions on Geoseience and Remote Sensing, 2017, 55(5): 2941–
[19] Kingma D P, Ba J. Adam: a method for stochastic optimization [EB/OL]. [2014-12-22]. https://arxiv.org/abs/1412.6980.
[20] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution [C]// CVPRW 2017: Proceeding of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Washington, DC: IEEE Computer Society, 2017: 1132–1140.
[21] Tai Y., Yang J, Liu X, et al. Memnet: a persistent memory network for

image restoration [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Venice, Italy: IEEE 2017.
[22] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks [C]// Proceedings of the 2018 European Conference on Computer Vision. Munich Germany: Springer Verlag 2018: 294-310.
[23] Wang Z, Liu D, Yang J, et al. Deep networks for image super-resolution with sparse prior [C]// Proceeding of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE 2015: 370-378.
[24] Zhang K, ZUO W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations [C]// Proceeding of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE 2018: 3262-3271.
[25] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution [C]// proceeding of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Athens, Greece: IEEE 2018: 2472-2481.