Image Super Resolution Reconstruction Based on Improved Invertible Rescaling Net

Taiping Mo
Guilin University of Electronic Technology

Dehong Chen (cdehong@163.com)
Guilin University of Electronic Technology https://orcid.org/0000-0003-3038-4968

Research Article

Keywords: Invertible network, Wavelet transform, Super resolution, Attention mechanism

Posted Date: July 4th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-545193/v1

Image super resolution reconstruction based on improved invertible rescaling net

Taiping Mo[1]· Dehong Chen[1,*]

1 School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541000, Guangxi, China

Abstract – The Invertible Rescaling Net (IRN) is modeling image downscaling and upscaling as a unified task to alleviate the ill-posed problem in the super-resolution task. However, the ability of potential variables of the model embedded high-frequency information is general, which affects the performance of the reconstructed image. In order to improve the ability of embedding high-frequency information and further reduce the complexity of the model, the potential variables and feature extraction of key components of IRN are improved. Attention mechanism and dilated convolution are used to improve the feature extraction block, reduce the parameters of feature extraction block, and allocate more attention to the image details. The high frequency sub-band interpolation method of wavelet domain is used to improve the potential variables, process and save the image edge, and enhance the ability of embedding high frequency information. Experimental results show that compared with IRN model, improved model has less complexity and excellent performance.

Keywords - Invertible network· Wavelet transform· Super resolution· Attention mechanism

1. Introduction

Single image super-resolution reconstruction (SISR) (Timofte et al. 2018) is to obtain enlarged or original size high-resolution (HR) image from a reduced single low resolution (LR) image. Because of the loss of many image details of LR image, ill posedness occurs to the reconstruction of SR image. Therefore, improving the ill posed problem is the focus and challenge to SISR.

In 2014, Dong et al. (2014) first used convolutional neural network to propose SRCNN algorithm on single image super-resolution. The algorithm reconstructs the image through three-layer convolution layer network: image block extraction, feature representation and feature nonlinear mapping. The algorithm lays the foundation of convolution neural network applied to image super-resolution task.

Since SRCNN was proposed, many scholars focus on the super-resolution task based on deep learning, and constantly propose new optimization algorithms to improve the performance of reconstructed image. Kim et al. (2016a; b) introduced residual network and recurrent neural network, and proposed VDSR and DRCN algorithm respectively. The algorithms solve the problem of gradient explosion or disappearance, expands the receptive field and deepens the network structure. Lim et al. (2017) removed the batch normalization operation, extracted more feature information, and proposed an up-sampling structure MDSR to obtain different multiple output results. Inspired by the generated countermeasure network (Creswell et al. 2018), Ledig et al. (2017) proposed SRGAN algorithm by introducing generation adversarial network, proposed new loss functions content loss and countermeasure loss, and obtained more realistic texture reconstruction image. Inspired by invertible network, Xiao et al. (2020) proposed IRN algorithm, which unifies the modeling image downscaling and upscaling tasks, captures and embeds the lost high-frequency information, and is compatible with image downscaling operation to obtain images with good reconstruction visual quality.

However, it is found that the potential variables of IRN model cannot save the high-frequency information on LR image. As a result, the effect of embedded high-frequency information is general, which affects the performance of SR image reconstruction. After studying the components of IRN model, we improve IRN from two parts. Firstly, the potential variables of the model are improved by using high frequency subbands interpolation in wavelet domain (Anbarjafari and Demirel 2010), which can save and process the edge values of LR image and enhance the ability of embedding high frequency information. Secondly, we use attention mechanism and dilated convolution to design feature extraction block, which can reduce the model parameters and improve the attention to image details during training.

2. Related theory

This section briefly introduces the main theoretical basis of implementation super-resolution algorithm. Parts A and B introduce the network structure and attention mechanism of the algorithm.

2.1 Invertible neural network

Invertible neural network is a kind of neural network of bijective structure, effective reversibility. The network has easy to handle Jacobian determinant. Neural network $f: \mathbb{R}^D \to \mathbb{R}^L$, mapping data point $x \in \mathbb{R}^D$ to a potential variable $z \in \mathbb{R}^L$, each $z \in \mathbb{R}^L$ has a unique $x \in \mathbb{R}^D$ such that $f(x) = z$.

The neural network is composed of a series of building blocks. In this study, the affine coupling layer of reference (Dinh et al. 2016) is used to build the neural network, as shown in Fig. 1. The input is divided into two parts $[J_1, J_2]$, which are transformed by Jacobian affine transformation of upper / lower triangle:

$$K_1 = J_1 + \phi(J_2) \tag{1}$$

* ∗ Corresponding author. e-mail: cdehong@163.com (Dehong Chen)

$$K_2 = J_2 \odot \exp(\varphi(K_1)) + \eta(K_1) \tag{2}$$

Output [K_1, K_2] connected to the next building block, internal function ϕ, φ, η represents an arbitrary neural network, when the coupling layer is reversed:

$$J_1 = K_1 - \phi(J_2) \tag{3}$$

$$J_2 = (K_2 - \eta(K_1)) \oslash \exp(\varphi(K_1)) \tag{4}$$

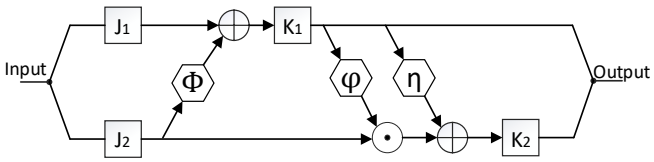


Fig. 1 Invertible neural network building block.

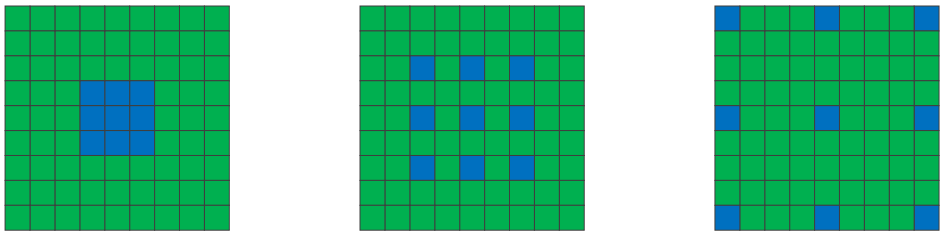
2.2 Attention mechanism

Attention mechanism is a kind of signal processing mechanism, which obtains the focus of attention by scanning global, allocates more attention to the focus, pays attention to the details of the target, and suppress other irrelevant information. Attention mechanism is mainly divided into soft attention and hard attention. Soft attention selectively ignores part of the information to reweight and aggregate the rest of the information. Hard attention only focuses on the information of a certain position in the sequence and ignores the rest of the information. At present, neural networks usually use soft attention mechanism, which is widely used in deep learning tasks such as natural language processing (Al-Makhadmeh and Tolba 2020), image processing (Jumbo et al. 2021) and speech recognition (Li et al. 2021).

Attention mechanism can be used in super-resolution tasks to focus on the high-frequency details of the image, suppress the image noise and improve the performance of the reconstructed image. (Lu et al. 2018) used channel attention to calibrate input elements, and aggregated multi-level features and shared recursive unit parameters to improve reconstruction quality. (Liu et al. 2018) used attention method to distinguish texture region and smooth region to restore high-frequency details effectively. (Zhang et al. 2018) used the attention mechanism to give different weights to the feature graph to increase the difference and improve the feature representation ability of the network. (Dai et al. 2019) uses second-order channels to enhance feature expression and feature learning to capture long-distance context space information. (Chen et al. 2021) Propose an image super-resolution reconstruction method using attention mechanism with feature map to improve the low-frequency and high-frequency components of feature map. In this study, we propose a feature extraction network based on attention mechanism to improve the performance of high-frequency feature extraction.

3. Algorithm implementation

This section mainly introduces the core part of IRN algorithm improvement: feature extraction based on attention mechanism, potential variables based on high frequency subbands interpolation in wavelet domain, model framework of improved algorithm.



(a) 1-dilated conv (b) 2-dilated conv (c) 4-dilated conv

Fig. 2 Dilated convolution with 3 × 3 convolution cores, with different dilation rates of 1,2,4

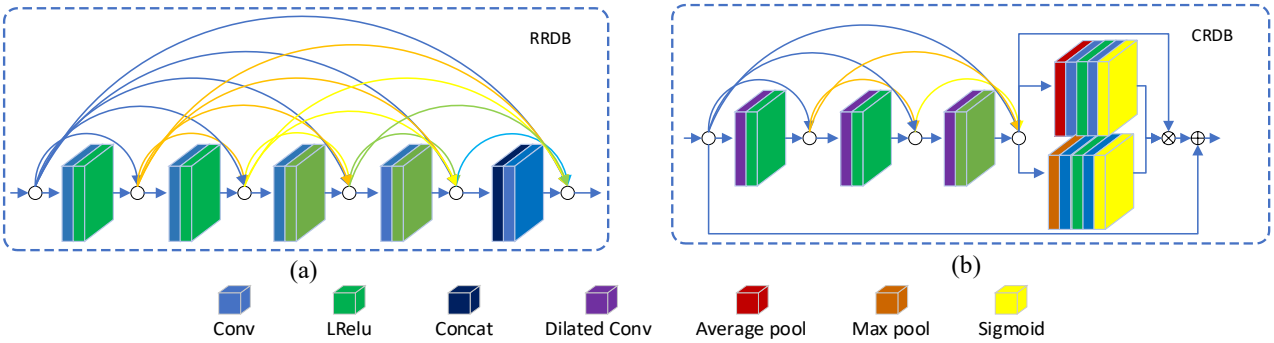


Fig. 3 Comparison of residual dense block structure. (a) RRDB block used by IRN model. (b) We propose dense residual block

3.1 Dense residual block

IRN model uses RRDB (Xintao Wang et al. 2018) module to extract image feature information. In order to improve the ability of neural network to extract image features, the feature extraction structure of RRDB is improved on two directions: First of all, we use dilated convolution (Yanjie Wang et al. 2020) instead of standard convolution, choose the convolution core

size of 3 × 3, and set the dilated rate of 1, 2 and 4 respectively to avoid the gridding effect, as shown in Fig. 2. The purpose is to expand the receptive field and extract the deep feature information of the image without increasing the model parameters. Secondly, the attention mechanism is added after the dilated convolution dense connection. We use the channel domain of CBAM (Woo et al. 2018) module to allocate more attention to the details of the image. The purpose is to pay attention to the details of the image and capture more high-frequency information on the image when extracting image features.

Fig. 3 shows the comparison between the feature extraction block of IRN module and our proposed feature extraction block. In Fig. 3 (b), two convolution layers with dense connection are reduced, and the standard convolution layer is replaced by the dilated convolution layer. After dense connection, CBAM channel domain attention mechanism is added. Compared with RRDB, the proposed CRDB improves the attention on attention to image details, effectively reduces the parameters of dense residual blocks, and improves the performance of image feature extraction.

3.2 High frequency subbands interpolation

In reverse process of IRN model, the potential variable $\mathcal{Z}$ generates a set of data randomly using Gaussian distribution to simulate high-frequency information, and reconstructs SR image with LR image. Because of the randomness of potential variable $\mathcal{Z}$, it cannot replace the high-frequency information on LR image. On the contrary, it affects the quality of reconstructed SR image. In reference (Anbarjafari and Demirel 2010), a super-resolution reconstruction method based on wavelet domain high-frequency sub-band interpolation is proposed. Inspired by this method, we use wavelet domain high-frequency sub-band interpolation to improve the potential variable $\mathcal{Z}$, as shown in Fig. 4.

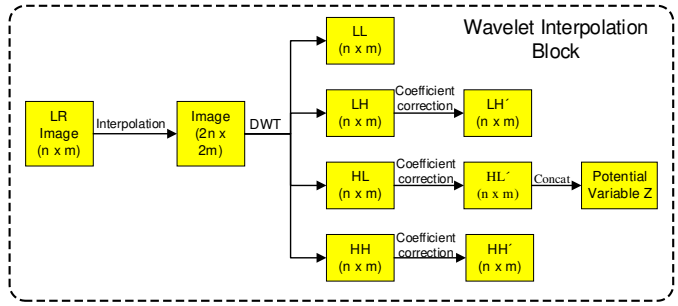


Fig. 4 Wavelet high-frequency subbands interpolation block

LR image is amplified by interpolation, and then the high frequency information on LR image is obtained by wavelet transform. According to the comparison between the LR image and the low-frequency information LL after wavelet transformed, the coefficients of LH, HL and HH are modified, and finally the high-frequency information in three directions is spliced as the potential variable $\mathcal{Z}$. Using wavelet domain high frequency subbands interpolation to improve potential variable $\mathcal{Z}$ can effectively use the high frequency information of LR image to obtain better SR image quality.

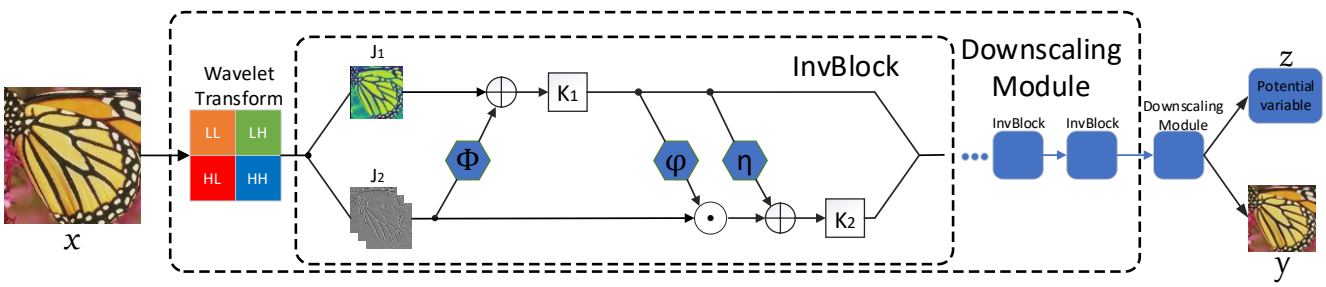


Fig. 5 The model frame diagram of this paper.

3.3 Invertible structure

Fig. 5 shows the model framework of this paper, which is composed of a wavelet transform and several reversible building blocks stacked into a downscaling module, and two downscaling modules constitute the whole invertible neural network model. The downscaling of each downscaling module is 2 times. In order to facilitate the model calculation, the three learning functions ϕ, φ and η of the reversible building block are the same, and CRDB block (Fig. 3) is selected as the learning function.

In the forward process of the model, $\mathcal{X}$ is downscaled to $\mathcal{Y}$, and each downscaling module is decomposed into $\mathcal{X}$ by two-dimensional wavelet transform as the first layer. The low frequency components LL is used as the J1 of invertible building block, and the high frequency components LH, HL and HH are used as the J2 of invertible building block. The output is calculated by alternating coupling on multiple invertible building blocks. $\mathcal{Y}$ embedded potential variables $\mathcal{Z}$ reconstruct $\mathcal{X}$ through model reverse learning.

4. Experiment analysis

4.1 Training details

The forward loss function and reverse loss function of the training model use L2 and L1 respectively. The total number of iterations of the model is 5×10^5, and the initial learning rate is 2×10^{-4}. When the number of iterations reaches [10K, 20K, 30K, 40K], the learning rate decays, and the decay rate $\gamma = 0.5$. The optimizer Adam is used to update the network parameters until the model converges.

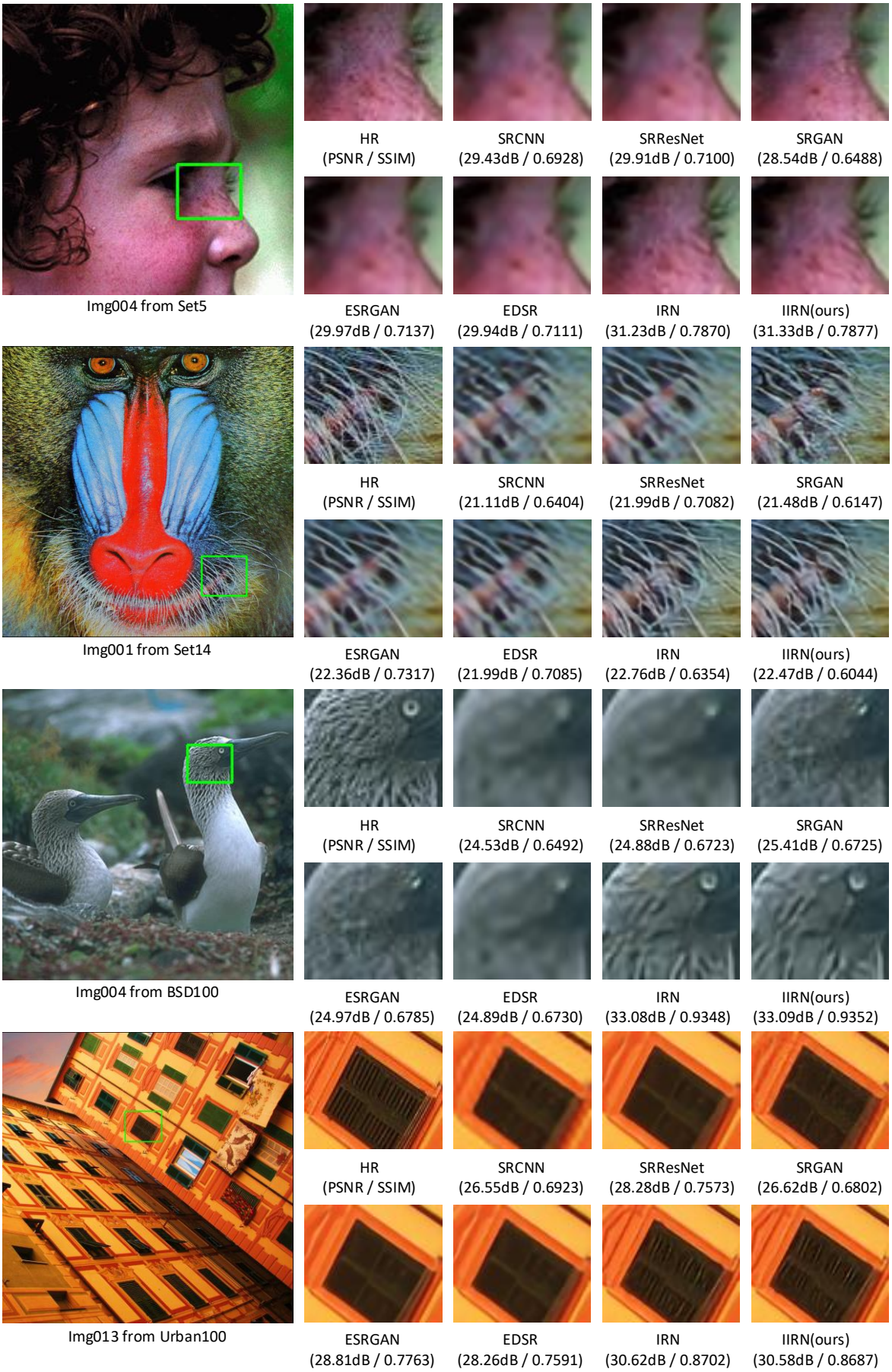


Fig. 6 qualitative comparison results

In this paper, DIV2K (Agustsson and Timofte 2017) data set is used as the training set. There are 800 high-resolution images in DIV2K data set, which provide low resolution images with corresponding scale of ×4. The low-resolution images are generated by bicubic interpolation method. Because the number of pictures in DIV2K training set is too less, and it is necessary to maintain the consistency of all experimental training data, our carries out random translation, flipping, clipping et al. operations on the images of DIV2K training set, and expands the training set to 32593, and unifies them into 480×480 pixels, which meets the demand of training data. We use widely used benchmark data sets Set5 (Bevilacqua et al. 2012), Set14 (Zeyde et al. 2010), BSD100 (Martin et al. 2001), Urban100 (Huang et al. 2015) as test sets, and use peak signal-to-noise ratio (PSNR) (Tanchenko 2014) and structural similarity ratio (SSIM) (Sara et al. 2019) on RGB channel to qualitatively evaluate the model.

4.2 Analysis

This algorithm is compared with several advanced SISR methods. Fig. 6 shows the qualitative comparison results of our model and other super-resolution models under ×4 super-resolution reconstruction. It can be seen from the figure that our SR image and IRN model have the same pleasant visual effect. Compared with other super-resolution models, it has more detailed texture (such as human skin, animal fur and building structure), and the PSNR and SSIM values of the image are also higher than other models.

In the scale factor ×4 data set test, the values of PSNR and SSIM of the four benchmark data sets of our model are basically the same as those of IRN model, and higher than other super-resolution models. This shows that the model has the same super-resolution reconstruction performance as IRN model, and is better than other SR models. The comparison results of qualitative evaluation are shown in Table 1.

Table 1 PSRN/SSIM comparison results

Scale	Method	Set5 (PSNR/SSIM)	Set14 (PSNR/SSIM)	BSD100 (PSNR/SSIM)	Urban100 (PSNR/SSIM)
×4	SRCNN	28.43/0.81	25.51/0.71	25.43/0.67	22.76/0.68
	SRResNet	30.23/0.85	26.75/0.75	26.22/0.70	24.57/0.76
	SRGAN	27.94/0.78	24.75/0.67	24.21/0.61	22.76/0.69
	ESRGAN	30.64/0.86	27.06/0.76	26.41/0.71	25.28/0.78
	EDSR	30.22/0.85	26.76/0.75	26.23/0.70	24.51/0.76
	IRN	**33.11/0.91**	30.04/0.86	29.76/**0.86**	28.60/0.88
	IIRN(our)	32.99/**0.91**	**30.12/0.87**	**29.77/0.86**	**28.91/0.89**

Under the same conditions, compared with the IRN model, the parameters of our improved algorithm model are reduced by 1.65k, the number of floating-point operations is reduced by 393.19m, and the average running time of the model is reduced by 3ms. The comparison results are shown in Table 2. The experimental results show that our improved algorithm can greatly reduce the complexity of the algorithm model, reduce the computational cost of the model, speed up the reconstruction speed of the image, and the performance of the reconstructed SR image is the same as that of the IRN model.

Table 2 Complexity comparison results of various SISR models

Method	Time	Model size	FLOPs
SRCNN	130ms	20.1K	0.52G
SRResNet	147ms	1.52M	4.07G
SRGAN	160ms	36.4M	4.07G
ESRGAN	262ms	16.7M	28.73G
EDSR	139ms	1.52M	3.18G
IRN	88ms	4.35K	936.28M
IIRN(our)	85ms	2.71K	543.09M

5. Conclusions

This paper proposes an improved super-resolution reconstruction algorithm. On the basis of residual-in-residual dense block, attention mechanism is added, and dilated convolution replaces standard convolution. This paper proposes a dense residual connection block based on attention mechanism, which reduces the size of the model and allocates more attention to image details. This paper proposes a high-frequency sub-band interpolation method based on wavelet domain to improve the potential variable and save the high-frequency information of the image, which solves the problem that the high-frequency information cannot be saved in the inverse process of the model. The experimental results show that the improved model reduces the parameters and complexity by about 40% compared with the IRN model, while maintaining the best performance of the reconstructed image.

Acknowledgments This work was supported by the National Natural Science Foundation of China under Grant (61863007) and Guangxi Natural Science Foundation under Grant (2020GXNSFDA238029) and Innovation Project of GUET Graduate Education (2021YCXS125) and Guangxi Key Laboratory of Automatic Detecting Technology and Instruments (YQ21107).

Compliance with ethical standards

Conflict of interest We all declare that we have no conflict of interest in this paper.

References

Agustsson E,Timofte R (2017) NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2017, pp 1122-1131

Al-Makhadmeh Z,Tolba A (2020) Automatic hate speech detection using killer natural language processing optimizing ensemble deep learning approach. Computing 102(2):501-522

Anbarjafari G,Demirel H (2010) Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image. ETRI J 32(3):390-394

Bevilacqua M, Roumy A, Guillemot C,Morel A (2012) Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding. In: Proceedings of the 23rd British Machine Vision Conference (BMVC). BMVA Press

Chen Y, Liu L, Phonevilay V, Gu K, Xia R, Xie J, et al. (2021) Image super-resolution reconstruction based on feature map attention mechanism. Applied Intelligence (8):1-14

Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B,Bharath A A (2018) Generative Adversarial Networks: An Overview. IEEE Signal Process Mag. 35(1):53-65

Dai T, Cai J, Zhang Y, Xia S-T,Zhang L (2019) Second-Order Attention Network for Single Image Super-Resolution. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp 11057-11066

Dinh L, Sohl-Dickstein J,Bengio S (2016) Density estimation using real nvp. arXiv preprint

Dong C, Loy C C, He K,Tang X (2014) Learning a deep convolutional network for image super-resolution. In: European Conference on Computer Vision. Springer, 2014, pp 184-199

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. IEEE, 2015, pp 5197-5206

Jumbo O E, Asfour S, Sayed A M,Abdel-Mottaleb M (2021) Correcting Higher Order Aberrations Using Image Processing. IEEE Trans. Image Process. 30(99):2276-2287

Kim J, Lee J K,Lee K M (2016a) Accurate image super-resolution using very deep convolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2016, pp 1646-1654

Kim J, Lee J K,Lee K M (2016b) Deeply-recursive convolutional network for image super-resolution. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016, pp 1637-1645

Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, et al. (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2017, pp 4681-4690.

Li G, Liang S, Nie S, Liu W,Yang Z (2021) Deep neural network-based generalized sidelobe canceller for dual-channel far-field speech recognition. Neural Networks 141:225-237

Lim B, Son S, Kim H, Nah S,Mu Lee K (2017) Enhanced Deep Residual Networks for Single Image Super-Resolution. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2017, pp 1132-1140

Liu Y, Wang Y, Li N, Cheng X, Zhang Y, Huang Y, et al. (2018) An attention-based approach for single image super resolution. In: 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018, pp 2777-2784

Lu Y, Zhou Y, Jiang Z, Guo X,Yang Z (2018) Channel Attention and Multi-level Features Fusion for Single Image Super-Resolution. In: 2018 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2018, pp 1-4

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. IEEE, 2001, pp 416-423

Sara U, Akter M,Uddin M S (2019) Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. Journal of Computer and Communications 7(3):8-18

Tanchenko A (2014) Visual-PSNR measure of image quality. J. Visual Commun. Image Represent. 25(5):874-878

Timofte R, Gu S, Wu J,Van Gool L (2018) Ntire 2018 challenge on single image super-resolution: Methods and results. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2018, pp 965-96511

Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, et al. (2018) ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In: Proceedings of the European Conference on Computer Vision (ECCV) Workshops.

Wang Y, Hu S, Wang G, Chen C,Pan Z (2020) Multi-scale dilated convolution of convolutional neural network for crowd counting. Multimedia Tools Applications 79(1):1057-1073

Woo S, Park J, Lee J Y,Kweon I S (2018) CBAM: Convolutional Block Attention Module. In: Proceedings of the European Conference on Computer Vision.

Xiao M, zheng S,Liu C (2020) Invertible Rescaling Image In: European Conference on Computer Vision. Springer, 2020, pp 126-144

Zeyde R, Elad M,Protter M (2010) On single image scale-up using sparse-representations. In: International conference on curves and surfaces. Springer, 2010, pp 711-730

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.