



Image super-resolution via channel attention and spatial attention

Enmin Lu¹ · Xiaoxiao Hu¹

Accepted: 20 April 2021 / Published online: 7 June 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Deep convolutional networks have been widely applied in super-resolution (SR) tasks and have achieved excellent performance. However, even though the self-attention mechanism is a hot topic, has not been applied in SR tasks. In this paper, we propose a new attention-based network for more flexible and efficient performance than other generative adversarial network(GAN)-based methods. Specifically, we employ a convolutional block attention module(CBAM) and embed it into a dense block to efficiently exchange information throughout feature maps. Furthermore, we construct our own spatial module with respect to the self-attention mechanism, which not only captures long-distance spatial connections, but also provides more stability for feature extraction. Experimental results demonstrate that our attention-based network improves the performance of visual quality and quantitative evaluations.

Keywords Attention mechanism · Spatial attention

1 Introduction

Image super-resolution is one of most attractive topics in computer vision and aims to provide more image information by recovering details from low-resolution images. In reality, this task has many industrial applications, such as medical imaging [29], traffic surveillance [18], and film restoration [35]. The ultimate goal of super-resolution (SR) is to build suitable and reliable connections between low-resolution (LR) images and super-resolution (SR) images with appealing visual effects. However, there is a lack of high-frequency information in LR images, which means that some details we may see in high resolution(HR) images are ignored and do not lead to a unique solution to the SR problem. To address this ill-posed problem, numerous methods have been proposed, and recently, the attention mechanism has been one of the concerns in this field.

In recent decades, interpolated-based [23] and example-based [7] methods have been suitable for this task. It is efficient and easy to implement but does not handle complex patterns well. With the development of computing power,

deep network architectures work as a new powerful tool in many parts of computer vision fields, and SRCNN [4] was the first deep architecture used in SR tasks. Currently, ResNet is widely used in such tasks [12, 21, 37]. Christian Ledig et al. [11] employed perception and content losses to improve the realistic effect of reconstructed images. On the basis of super-resolution residual network(SRResnet), Xintao et al. [27] introduced the DenseNet Structure and achieved the state-of-the-art performance in terms of the peak signal to noise ratio(PSNR).

In the process of exploration, many researchers have attempted to build deeper or more complicated models to handle SR tasks. This kind of approach is efficient, but it may weaken the connection between the global image content and the local extracted information [26]. Moreover, during information transmission, different feature maps in the same layers cannot exchange information until they are transferred to the next layer, which may incur time spending and harmness of minutiae extraction. Inspired by the attention mechanism [25], we propose an attention-based network for efficient information extraction and SR image construction. Our method is mainly based on cross-channel feature correlation and non-local feature extraction. More specifically, we employ a convolution block attention module to auto adjust every feature map according to their own presentation. Additionally, considering that only the mean and the maximum values of different channels are utilized in each pixel, the original spatial attention block in the convolutional block attention module may not be very

✉ Xiaoxiao Hu
huxx@lzu.edu.cn

Enmin Lu
luem18@lzu.edu.cn

¹ Lanzhou University, Lanzhou, China

elastic to weigh each pixel position, so we try to introduce an attention mechanism into the SR task to achieve a better presentation.

We summarize our contributions as follows: 1) we find a new way to combine attention module with super-resolution network to improve the process of SR image construction; 2) we introduce a self-attention mechanism to our model and it achieves better performance metrics and more stable visual effects.

2 Related work

Attention model The attention mechanism was first applied in the computer vision field. Its original idea is to shift limited visual emphasis on important information to boost efficiency and save resources. Normally, the attention mechanism uses a mask as an implementation method, which calculates the aligned weight layer to mark areas of interest in the feature map. The formation of the area mask is dynamically auto-adjusted through the training process. Different network structures, such as non-local networks [26], self-attention [20], and transformers [25], aim at making better weight masks. First, the attention model uses only hard-attention and soft-attention models. Then self-attention is considered as a better way to learn weight layers automatically. Recently, transformers with multi-head self-attention structures have been proposed, and many researchers are exploring new ways to apply this structure in the computer vision field.

Since Volodymyr et al. [17] applied attention processing in deep learning frameworks in 2014, the idea has led to an increase in attention-based research [14, 22, 31, 34]. The attention mechanism is not only easy to understand and implement, but also convenient to embed into convolution-based network. Its properties have attracted the interest of many researchers to explore its potential applications. For example, Xu et al. [30] added attention model into an encoder-decoder structure and first proposed definitions of soft attention and hard attention. Bello et al. [2] augment convolutional networks with self-attention mechanism to lead to consistent improvements in image classification and object detection tasks. Wang et al. [26] applied the idea of non-local similarity in CNNs to denoise images. In addition, some researchers have applied this idea in natural language processing. Liu et al. [13] utilize attention mechanism to give different focus to the information extracted by the hidden layers of bidirectional long short-term memory network (LSTM). All these works imply the strength of the attention model.

There have also been some works regarding attention for image super-resolution. Zhang et al. [8] proposed very deep residual channel attention networks. Liu [15] proposed using

a spatial attention block to learn the cross-correlation across features at different layers. However, to our knowledge, we have not found combinations of multi-angle attention mechanism and generator adversarial networks in this domain, which is the main motivation of this paper.

Super-resolution neural network In this part, we will focus on single image super-resolution.

The first deep neural network applied in SR tasks is the SRCNN [4]. This network is an end-to-end method that achieves better image quality and processing speed than the traditional image scaling method. The SRCNN first rescales the LR image to the target image size and then uses three convolutional layers to fit the non-linear mapping process and finally outputs the HR image. Then, the FSRCNN [5] has been proposed to improve processing speed by skipping the rescaling step and replacing the old large convolution kernel size with a relatively small kernel size. On the basis of SRCNNs, VDSR [8] has been developed. The authors of this method considered that there was a basic similarity between LR input and HR input, and the main difference between them is information in the high-frequency domain. Therefore, based on this idea, it is natural to introduce the ResNet structure into SR tasks. Since then, Resnet has been used as an important part of SR tasks and has long influenced on subsequent research [9, 21]. DRCNs [9] use the basis of RNNs and skip-connections and deeper network increases the receptive field of the network and improves efficiency. RED [21] chooses another method that employs the encoder-decoder structure. Every convolutional layer of this network has its own corresponding deconvolutional layer and skip connections exist between the encoder-decoder layer pairs. DenseNet [6] takes the matter from the feature map and achieves feature reuse by stacking different feature map in the channel direction. It decreases the number of parameters and strengthens the propagation of features. SRDenseNet [24] applies DenseNet in SR tasks and uses a number of dense blocks to learn high-level features.

However, most of the methods discussed above are mainly designed to address low scale ratio SR tasks. When dealing with scale ratios greater than $\times 4$, the result of the model seems too smooth and unrealistic for some details. To mitigate this problem, generative adversarial network is applied in SR tasks and an SRGAN has been proposed [11]. The SRGAN borrows the idea of a GAN and separately forms discriminator network and a generator network. The discriminator network is designed to judge whether the received image is real, and the generator network is trained to generate fake images that look real enough to fool the discriminator network. The SRGAN utilizes content and adversarial losses to provide a more realistic view of image. To further improve the visual quality

of details, enhanced super-resolution generative adversarial networks(ESRGAN) [27] was proposed, which combines the idea of DenseNet with an SRGAN. Additionally, it employs a loss function of a relativistic GAN to have the discriminator predict relative realism instead of absolute realism. This change aims at increasing the stability of the whole network and improving the quality of the generated image. Moreover, ESRGAN uses the features before activation to measure the content loss. The authors believed that in this way the model can provide stronger supervision for brightness consistency and texture recovery.

3 Model architecture

In this section, we mainly introduce our network architecture in detail. The whole GAN includes a generator network and a discriminator network. We use the basic structure of super-resolution residual network(SRResNet) as our backbone network and use the residual in residual structure as basic feature extraction block. In every basic block, we integrate DenseNet with a modified convolutional block attention module(CBAM) [28] structure to improve performance. In addition, we use a pretrained VGG-128 network as our basic discriminator network Fig. 1.

3.1 Basic block

The basic block can be divided into two parts. The first part is used for feature extraction. Due to the excellent performance of DenseNet applied in computer vision, we use DenseBlock as the first part in every basic block, which

is constructed by convolution layers and concatenate layers. All convolution layers in DenseNet use 3x3 kernels, and every feature map has the same shape so that different layers can be concatenated in the channel dimension.

In addition, we integrate the modified CBAM structure in the first part. The modified CBAM structure plays a role as a regulator, which automatically retunes the attention allocation to achieve a better visual effect. Given an intermediate feature map $F \in R^{C \times H \times W}$ after DenseNet, the modified CBAM forms a 1D channel attention mask $M_c \in R^{C \times 1 \times 1}$ and 2D spatial attention mask $M_s \in R^{1 \times H \times W}$. The whole process can be described as below

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \oplus F' \end{aligned} \quad (1)$$

where \otimes represents Hadamard product. As can be seen, the channel mask and spatial mask are all inferred from feature maps. Hence, different feature map values result in different masks. Even though there exists a large diversity between training images, the model can adjust mask values to fit the situation and ensure the stability of feature map formation process. By this way, it can increase the flexibility of the whole model Fig. 2.

3.2 Attention module

Channel attention module Each channel of the feature maps can be considered a feature detector [32]. The channel attention mask is mainly used for learning inter-channel connections of the feature maps, which means that this attention intends to determine which layers of

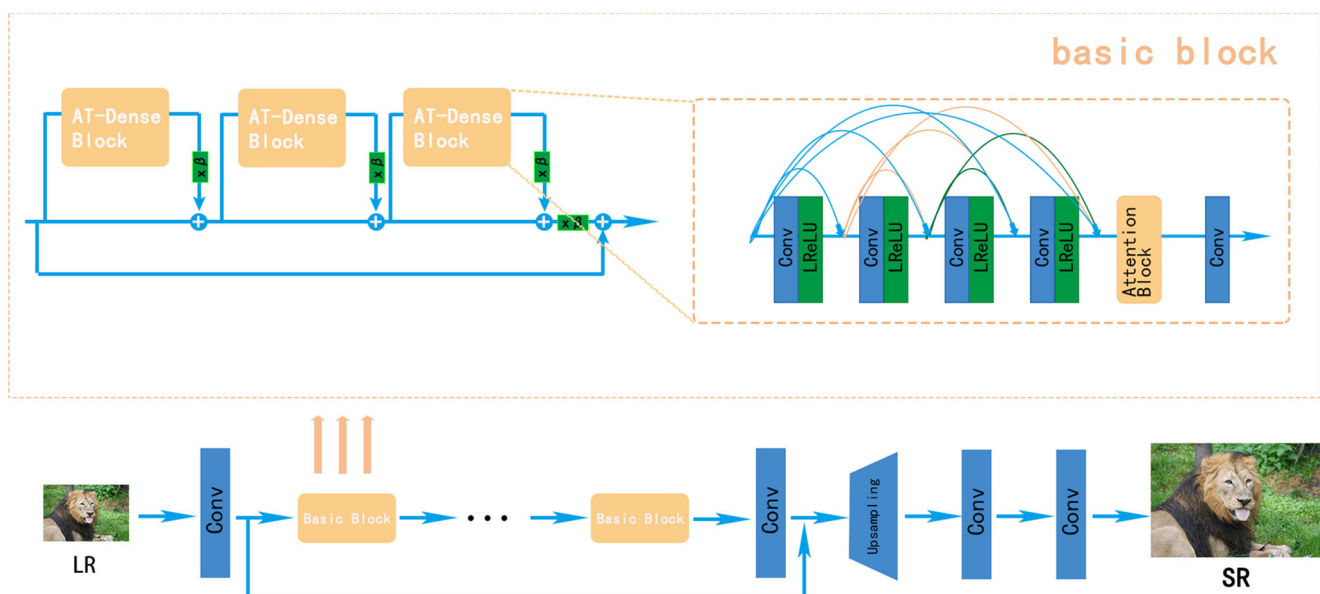


Fig. 1 The generator network employs the basic frame of SRResNet [11] and designs unique basic block structure to improve generator network performance

Modified Convolutional Block Attention Module

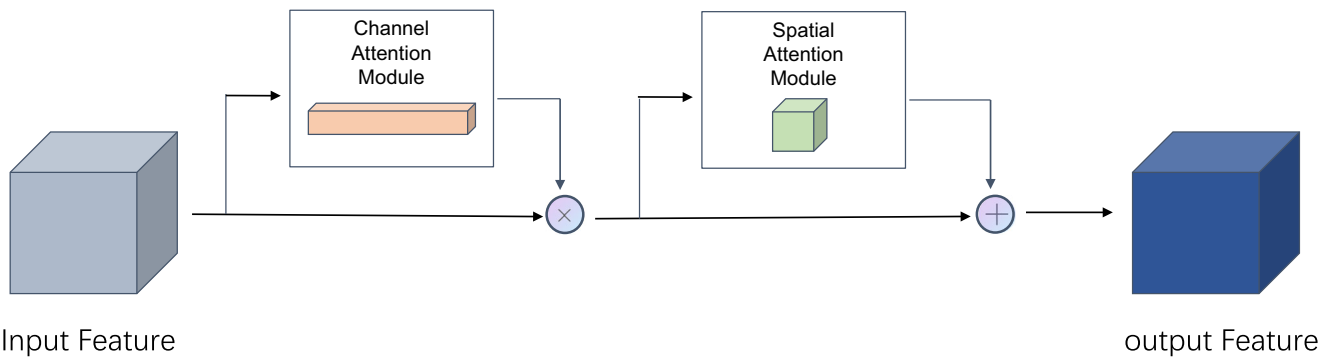


Fig. 2 Structure of the attention block combined with the channel attention block and spatial attention block

the feature maps are more meaningful. To compute the channel attention mask, we follow the initial settings of the CBAM, which uses max-pooling and average-pooling of every feature map to aggregate information throughout this map. Then we construct a shared network so that both avg-pooling and max-pooling features are transferred to it. The shared network is a kind of multilayer perceptron (MLP) with two hidden layers. The number of neurons in the first hidden layer is set to $R^{C/r \times 1 \times 1}$ controlled by a hyperparameter r , which is the reduction ratio to prevent excessive parameters. The number of neurons in the second hidden layer is the same as the shape of the input feature map. In this way, we can make the mask match the corresponding input. The network is shared both with max-pooling features and avg-pooling features. Finally, the outputs of the shared network are added together and sent to the sigmoid output neuron. In short, the whole operation process can be described as follows Fig. 3:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (2)$$

Spatial attention module Different from channel attention, spatial attention focuses on 'where' the input feature maps carry more specific details. The initial spatial settings of

the spatial attention module in the CBAM are the same as those of the channel attention module. Since direct usage of max-pooling and avg-pooling may create too much information loss, we suggest a more flexible structure based on self-attention mechanism Fig. 4.

The spatial attention value at a position can be formulated by the weighted sum of the values at other positions, and we can reduce it to the following expression:

$$Y_i = \frac{1}{C(X_i)} \sum_j s(X_i, X_j) h(X_j) \quad (3)$$

where i represents the target position and j enumerates all the positions in the same feature map. The function h calculates the representation value at any position j and the function s measures the similarity between the signal value at position i and position j . $C(X_i)$ is a kind of normalized factor at a position i . Expression in the network structure is similar to that in a non-local network. Additionally, residual learning is embedded in the attention module. We combine the attention values with raw input values and use a learned hyperparameter α to balance attention concern against raw image intensity.

$$Z_i = X_i + \alpha Y_i \quad (4)$$

Channel Attention Module

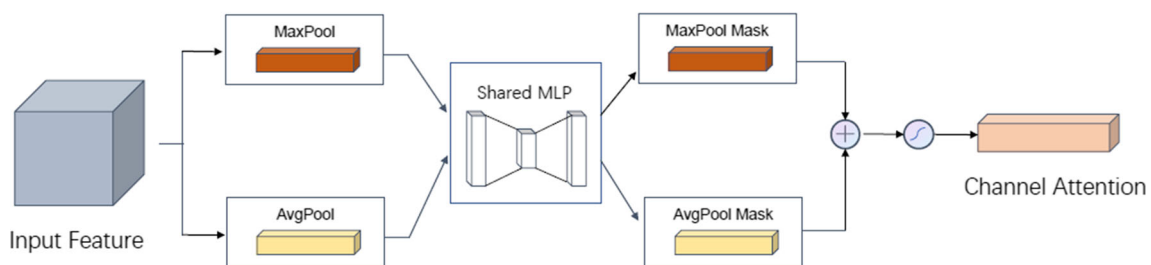


Fig. 3 the Specific structure of channel attention module

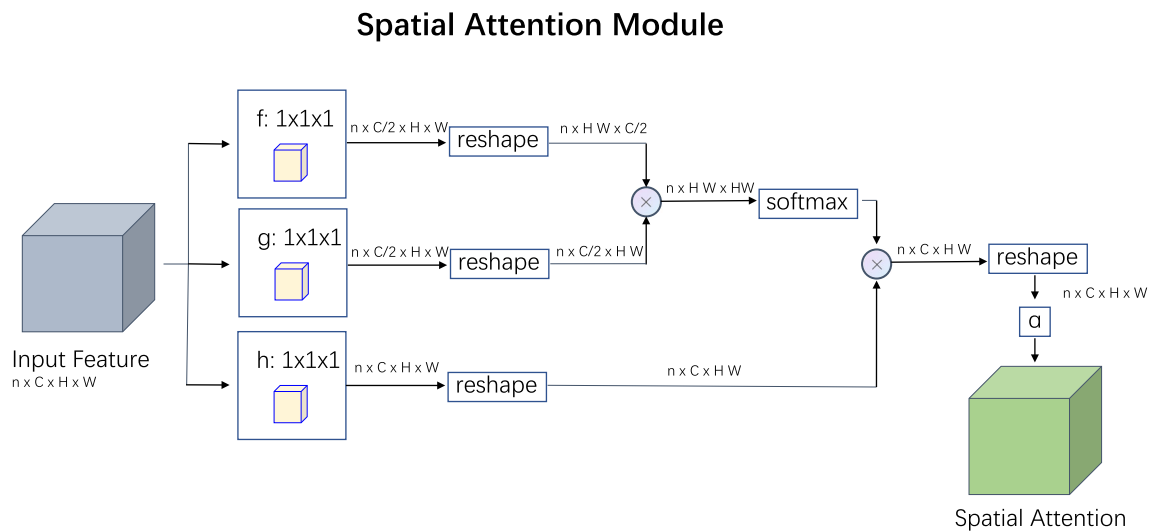


Fig. 4 Specific structure of Spatial Attention Module embedded with self-attention mechanism

In other words, the spatial attention module is a kind of comprehensive information module that utilizes global spatial information. In this way, we can provide extra information to enhance the visual effects in the attention area and not create information loss.

4 Experiments

4.1 Datasets

The DIV2K dataset [1], which is a widely used high-quality (2K resolution) dataset for image restoration tasks, is our training dataset. It is divided into three parts: 800 images for training, 100 images for validation, 100 images for testing. In training process, we use numbered 1-800 for training model. In testing process, DIV2K validation dataset and four other datasets (Set5 [3], Set14 [33], BSD100

[16] and Urban100 [7]) are used to help our research generalizability to diverse contextual conditions.

4.2 Implementation details

Following an ESRGAN [27], we experiment on all the datasets with a scaling factor of $\times 4$ between the LR and HR images. The LR images are obtained by downsampling the HR images with the bicubic interpolated method. To make an appropriate comparison with ESRGAN, we decide to keep our preprocessing process consistent with the ESRGAN settings. The mini-batch size is 16 and the spatial size of cropped the HR patch is set to 128×128 . We also use the same data augmentation processes (random horizontal flips and 90 degree rotations).

We try several different ways to implement the pretraining procedures and make comparisons. The first method is to choose models with a CBAM, and the second method is

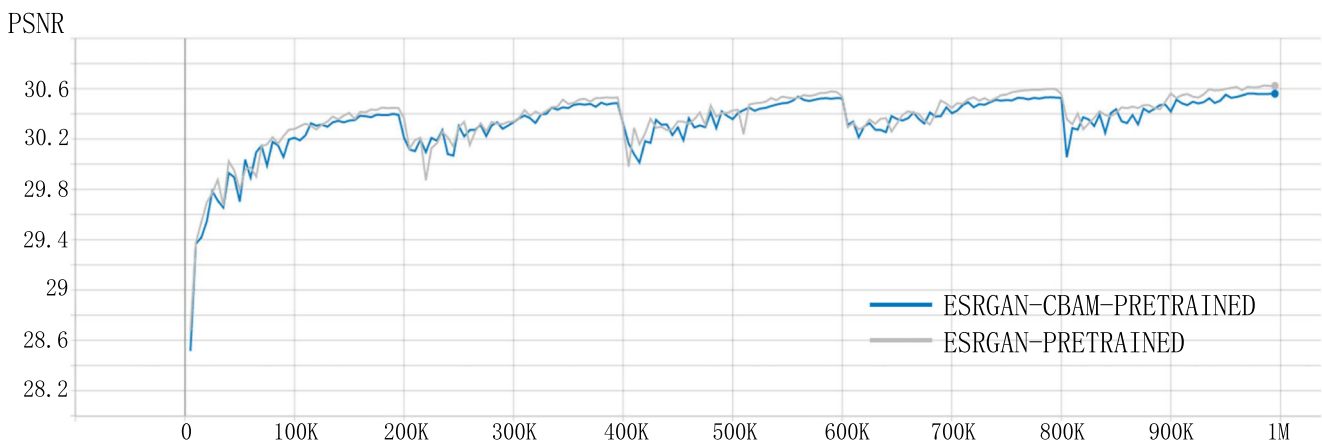


Fig. 5 Training PSNR curve in Set5 dataset of ESRGAN model with/without CBAM

to choose models without a CBAM. The hyperparameters of the different methods are kept the same. As Fig. 5 shows, we find that after the same number of iterations, the training curves for the PSNR are close and the PSNR performance of the models without a CBAM is slightly higher than that of the models with a CBAM. Additionally, the model with a CBAM takes longer to train than the model without a CBAM. Therefore, considering the efficiency, we choose a training model without a CBAM as our pretraining strategy.

Considering that the main focus of this work on the attention block in the basic block, and the structures of the other layers in the basic block are similar to the RRDB used in ESRGAN, we choose to divide our training process into two steps: 1) Employ the ESRGAN structure and train a PSNR-oriented model to form pretrained weights. The number of iterations is 1×10^6 and the learning rate is initialized as 2×10^{-4} and decayed by a factor of 2 every 2×10^5 mini-batch updates. 2) Load the pretrained model weights in our model and retrain the whole model. The number of iterations is 1.5×10^5 . The learning rate is set to 1×10^{-4} and halved at [25k, 50k, 75k, 100k] iterations. The entire network is trained using the Adam optimizer [10] where $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 1 \times 10^{-6}$. We implement our model with the PyTorch framework. In the test process, constrained by GPU capability, we divide some high resolutions images into several smaller patches as raw inputs and stitch the corresponding SR outputs together to make final images.

4.3 Quantitative results

We compare our model with several CNN-based SR methods including SRGAN [11], ESRGAN [27], EDSR [12], SRFeat [19], RCAN [36], that care more about

visual presentation. Table 1 summarizes the quantitative comparison results of the different SR methods.

We can see from Table 1 that the AT-ESRGAN** model achieves the best PSNR. To achieve a better visual effect, we rebuild our AT-ESRGAN* model by adjusting the AT-ESRGAN* loss to be the same as the ESRGAN loss, which is a mixture of the relative GAN loss, L1 loss and perception loss. This behavior is aimed at creating a balance between visual performance in reality and the quantification index. The quantitative results show that the AT-ESRGAN* model still outperforms the SRGAN and ESRGAN models; all of these methods are based on GANs, even though we do not actually expand the deeper model structure.

4.4 Qualitative results

Figure 6 shows that for regular structures such as wings textures or floor bricks, our model with an attention block provides a better performance closer to reality from a visual standpoint. As the attention block can gather information from different channels and spatial locations, it acts more similar to a stabilizer that makes every generated pixel value fit around the other generated pixels and retrains the original spatial feature. Additionally, some generated images show that it provides fewer undesirable artifacts than the other GAN-based methods.

5 Conclusion

We have presented an attention-based model that provides more stable and reliable results than other GAN-based methods. We construct a novel block architecture that embeds attention mechanisms into the traditional DenseNet

Table 1 Quantitative evaluation results (PSNR/structural similarity index measure (SSIM)) of the x4 scaling SR method in different datasets (the best results are marked in bold; * represents our final model, and ** represents our method oriented by the PSNR loss)

Datasets name	Evaluation Index	SRGAN	ESRGAN	AT-ESR GAN*	AT-ESR GAN**	RCAN	SRFeat	EDSR
Set5	PSNR	28.156110	29.023770	29.193397	30.561635	30.145991	29.644453	30.063253
	SSIM	0.811047	0.825541	0.837497	0.868528	0.932121	0.927945	0.931813
Set14	PSNR	25.130720	25.224044	25.126460	26.938502	26.435175	26.167116	26.409623
	SSIM	0.676109	0.681895	0.683483	0.749156	0.869440	0.864037	0.869386
BSD100	PSNR	24.511538	24.617141	24.306996	26.338953	26.236326	26.020048	26.211117
	SSIM	0.635368	0.649095	0.639716	0.716579	0.894988	0.891693	0.894740
Urban100	PSNR	22.862387	23.310677	23.458035	25.141755	24.597606	23.928324	24.452474
	SSIM	0.697274	0.723122	0.731640	0.783386	0.884757	0.871665	0.883135
DIV2K (validation)	PSNR	26.981501	27.262692	27.271958	29.261175	28.967265	28.531927	28.892560
	SSIM	0.754077	0.765433	0.772612	0.825877	0.918936	0.913192	0.918552

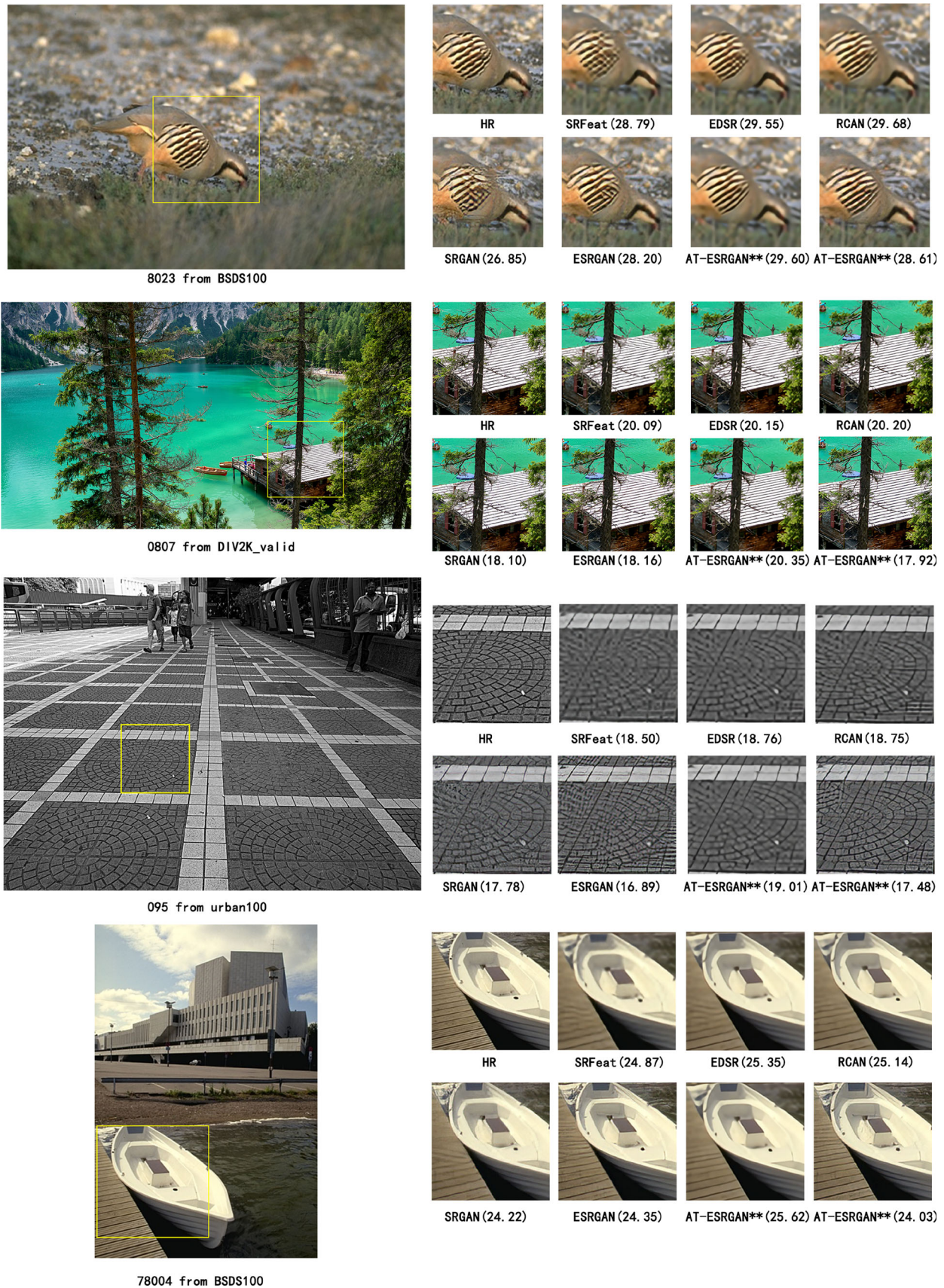


Fig. 6 Comparison of x4 super resolution images reconstructed by SRGAN, ESRGAN, AT-ESRGAN**, AT-ESRGAN*, RCAN, SRFeat, EDSR

structure. By utilizing CBAM, we employ an attention mechanism in two ways: channel attention and spatial attention. Furthermore, we introduce a self-attention mechanism into the process of constructing spatial attention to offer a greater degree of model stability. To verify its efficacy, we conduct experiments with several perception-oriented models and confirm that introducing an attention mechanism into the residual model improves performance. We hope our work can provide a new idea of the combination of attention mechanisms and super-resolution tasks.

References

- Agustsson E, Timofte R (2017) Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 126–135
- Bello I, Zoph B, Vaswani A, Shlens J, Le QV (2019) Attention augmented convolutional networks. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 3286–3295
- Bevilacqua M, Roumy A, Guillemot C, Alberi-morel ML (2012) Low-complexity single-image super-resolution based on nonnegative neighbor embedding
- Dong C, Loy CC, He K, Tang X (2014) Learning a deep convolutional network for image super-resolution. In: European conference on computer vision. Springer, pp 184–199
- Dong C, Loy CC, Tang X (2016) Accelerating the super-resolution convolutional neural network. In: European conference on computer vision. Springer, pp 391–407
- Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4700–4708
- Huang JB, 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
- Kim J, Kwon Lee J, Mu Lee K (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
- Kim J, Kwon Lee J, Mu Lee K (2016) Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1637–1645
- Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv:1412.6980
- Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z et al (2017) Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4681–4690
- 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
- Liu G, Guo J (2019) Bidirectional lstm with attention mechanism and convolutional layer for text classification. *Neurocomputing* 337:325–338
- Liu T, Yu S, Xu B, Yin H (2018) Recurrent networks with attention and convolutional networks for sentence representation and classification. *Appl Intell* 48(10):3797–3806
- Liu ZS, Wang LW, Li CT, Siu WC, Chan YL (2019) Image super-resolution via attention based back projection networks. In: 2019 IEEE/CVF international conference on computer vision workshop (ICCVW). IEEE, pp 3517–3525
- 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, vol 2. IEEE, pp 416–423
- Mnih V, Heess N, Graves A et al (2014) Recurrent models of visual attention. In: Advances in neural information processing systems, pp 2204–2212
- Molina-Cabello MA, Luque-Baena RM, Lopez-Rubio E, Thurnhofer-Hemsi K (2018) Vehicle type detection by ensembles of convolutional neural networks operating on super resolved images. *Integrated Comput-Aided Eng* 25(4):321–333
- Park SJ, Son H, Cho S, Hong KS, Lee S (2018) Srfeat: Single image super-resolution with feature discrimination. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 439–455
- Shaw P, Uszkoreit J, Vaswani A (2018) Self-attention with relative position representations. arXiv:1803.02155
- Tai Y, Yang J, Liu X (2017) Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3147–3155
- Tian C, Zhu X, Hu Z, Ma J (2020) Deep spatial-temporal networks for crowd flows prediction by dilated convolutions and region-shifting attention mechanism. *Appl Intell* 50(10):3057–3070
- Timofte R, De Smet V, Van Gool L (2014) A+: Adjusted anchored neighborhood regression for fast super-resolution. In: Asian conference on computer vision. Springer, pp 111–126
- 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
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008
- Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7794–7803
- Wang X, Yu K, Wu S, Gu J, Liu Y, Dong C, Qiao Y, Change Loy C (2018) ESRGAN: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV), pp 0–0
- Woo S, Park J, Lee JY, So Kweon I (2018) Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV), pp 3–19
- Wu Y, Ma Y, Liu J, Du J, Xing L (2019) Self-attention convolutional neural network for improved mr image reconstruction. *Inf Sci* 490:317–328
- Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015) Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning, pp 2048–2057
- Yu H, Wang J, Huang Z, Yang Y, Xu W (2016) Video paragraph captioning using hierarchical recurrent neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4584–4593
- Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: European conference on computer vision. Springer, pp 818–833
- 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

34. Zhang J, Bargal SA, Lin Z, Brandt J, Shen X, Sclaroff S (2018) Top-down neural attention by excitation backprop. *Int J Comput Vis* 126(10):1084–1102
35. Zhang Q, Ding Y, Yu B, Xu M, Li C (2019) Old film image enhancements based on sub-pixel convolutional network algorithm. In: Tenth international conference on graphics and image processing (ICGIP 2018), vol 11069, p 110693k. International society for optics and photonics
36. 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
37. Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y (2018) Residual dense network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2472–2481

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.