WLSAN: Wavelet-Layer-Spatial Attention Network for Single Image Super Resolution

Ruixin Peng †, Yuyang Su †, Bingxuan Yu †, Chunyan Yang *

School of Mathematics and Statistics, Yunnan University, Kunming, Yunnan, 650500, China

* Corresponding author e-mail: yangchy@ynu.edu.cn

†These authors contributed equally

Abstract. Single Image Super Resolution has been a hot topic in recent years, which has wide application prospects. Some recent works use attention in SR to capture the inter-channel and inter-spatial relationships of feature maps. How- ever, these works cannot focus on the image features at different frequency do- mains. At the same time, the loss function of some SR models will make the model more inclined to reconstruct the smooth part and cannot consider the de- tails in multiple directions of the image. To address these problems, we propose a Wavelet-Layer-Spatial Attention Network (WLSAN), which makes use of both essential global context and local information to improve accuracy and visual quality. Specifically, the proposed WLSAN consists of wavelet attention to model inter-dependencies by decomposing the image signals and non-local attention, including channel-spatial-layers. Meanwhile, we combine L1 with improved Sobel Loss, which enhances the stability during training and constrains different aspects of the HR image generation. Extensive experiments demonstrate that the proposed WLSAN exceeds many classical methods and states of the arts.

Keywords: Single Image Resolution; Attention Mechanism; Wavelet Transform; Mixed Loss.

1. Introduction

The task of Single Image Super Resolution (SISR) is to generate high-resolution (HR) images with necessary edge structure and texture details from a single low-resolution (LR) input. This is a highly ill-posed problem because there are always multiple possible high-resolution solutions for low-resolution images. SISR is widely used in computer vision applications such as security and surveillance imaging, satellite imaging, and medical imaging.

In 2014, Dong [1] et al. first applied the deep learning method to the super resolution. With the continuous development of deep learning models. SRResNet[3], Mem Net[2] RDN[4] with advanced network design further improves the quality of image super-resolution. In recent years, the Attention Mechanism has been introduced to exploit the correlation such as channels (CA) or spatial regions (SA), which can improve the representation ability of the network.

While some attention-based works [18][19] can focus on more useful information, we found two defects in them: First, the same attention degree corresponds to different frequency-domain features of channels and spatial regions, respectively. This may cause key high-frequency features in certain regions of channels and layers to be ignored or low- frequency features to be amplified, which is not conducive to reconstructing images with rich details of high-frequency features. Second, these models only focus on pixel loss. This can easily lead to the resulting image being too smooth to focus on the details of the feature edges. Therefore, these methods still have difficulties preserving image texture and recovering texture details.

To address these issues, we propose a novel Wavelet Spatial Layer Attention Mechanism (WLSAN) inspired by DWSR [5] and HAN [6]. WLSAN captures image texture details by introducing wavelet transform and combines spatial attention and layer attention mechanisms to explore inter-channel correlations. Meanwhile, we follow Zheng [32], combining pixel error and Advanced Sobel Loss as loss functions without introducing parameters to reconstruct image texture details better. The main contributions of this paper are as follows:

1) We introduce wavelet transform to extract images' multi-scale and multi-directional texture information. At the same time, by introducing channel attention and layer attention, the weight of hierarchical features and the channel spatial correlation of each layer feature are learned.

2) Since our high-resolution image generation involves multiple pipeline processes, we change the single pixel-level loss to a weighted mixture loss to constrain different aspects of the generation process.

3) Our proposed WLSAN further improves the generation of SR image texture through cooperative wavelet transform and attention mechanisms. Extensive experiments show that our algorithm performs better than existing SISR methods.

2. Related work

2.1 Image Super-Resolution

Single Image Super Resolution reconstruction is a classic application of computer vision. With the development of convolutional neural networks in recent years, many deep learning based super resolution methods have been proposed. Dong et al. [1] pro- posed a super-resolution method based on a convolutional neural network (SRCNN), the pioneering deep learning work in image super-resolution reconstruction. Simonyan et al. [7] found that the deepening of the network can lead to improved performance. Kim et al. proposed a super resolution model VDSR with an intense convolutional net- work [8], and later they proposed DRCN with a recursive structure [9]. Lai et al. pro- posed LapSRN [10] by introducing the Laplacian pyramid, allowing the network to perform multi-scale super resolution simultaneously in a feed forward. Gao et al. pro- posed a lightweight information multi-distillation network (IMDN) [11] by constructing the cascaded information multi-distillation blocks (IMDB). Liu et al. proposed a novel residual feature aggregation (RFA) [12] framework that groups several residual modules together and directly forwards the features on each local residual branch by adding skip connections. Liu et al. proposed a new holistic attention network (HAN) [6], which consists of a layer attention module (LAM) and a channel-spatial attention module (CSAM), to model the holistic inter-dependencies among layers.

The above methods are mainly based on pixel loss minimization, but it was later found that pixel loss is inaccurate for reconstruction quality. Therefore, researchers de- signed various loss functions (such as content loss [13] and adversarial loss [3]) to pro- duce more realistic and higher quality results, researchers often combined the weighted average [10], [3], [14], [15], [16] multiple loss functions to constrain different aspects of the generation process. Inspired by these works, we introduce a new loss function composed of multiple loss functions to stabilize the training process.

2.2 Visual Attention in Super Resolution

The attention mechanism helps the network ignore irrelevant information and focus on important information [17]. Considering the inter-dependence and interaction of feature representations between different channels, Hu et al. proposed a "squeeze excitation" module to improve learning by explicitly modeling channel inter-dependencies [18]. Zhang et al. [19] proposed the stochastic resonance mechanism by combining the channel attention mechanism with the stochastic resonance mechanism, significantly improving representation ability. Dai et al. [20] further proposed a second-order channel adaptation (SOCA) module to adaptive refine features using second-order feature statistics.

It is well-known that the wavelet transform (WT) [21], [22] efficiently represents images by decomposing the image signal into high-frequency sub-bands representing texture details and low-frequency sub-bands containing global topological information. Models using this strategy tend to significantly reduce the model size and computational cost while maintaining competitive performance [23], [24]. Inspired by this, based on HAN, we combine the attention mechanism with wavelet transformation to better capture the spatial information of images.

3. Proposed Method

3.1 Overall Network Structure

Features extraction. Our model firstly performs feature extraction on the original LR image without interpolation. LR images are directly fed into the network to extract shallow features:

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

Then, we use the backbone of RCAN [19] to extract deep features. The model con- sists of multiple concatenated residual groups. The structure of the residual group can be expressed as follows:

$$F_{RB} = F_0 + conv(\sum_{i=1}^{N} R_i(F_o)) \tag{2}$$

Where R_i represents the $i\text{-}th$ channel attention block.

WLSAN attention. After extracting the overall features through the residual group, we extract the four sub-bands of the image through wavelet transform and summarize the features of different frequency domains through the channel-spatial attention mechanism, supplemented by local residual connections and a small number of convolution blocks. The following formula can represent it:

$$F = M_l(concatenate\ (F_1, F_2, \cdots, F_N) + M_c(F_N) \tag{3}$$

Where M_l stands for layer attention, which weights feature according to the correlation matrix. M_c stands for channel-spatial attention module (CSAM), and its output is only the feature graph of one channel.

Image reconstruction. To improve computational efficiency and make full use of deep learning techniques inspired by [25][26], we use sub-pixel layers as a post-sampling method to upsample low-dimensional features and reconstruct the HR image by aggregating the feature maps.

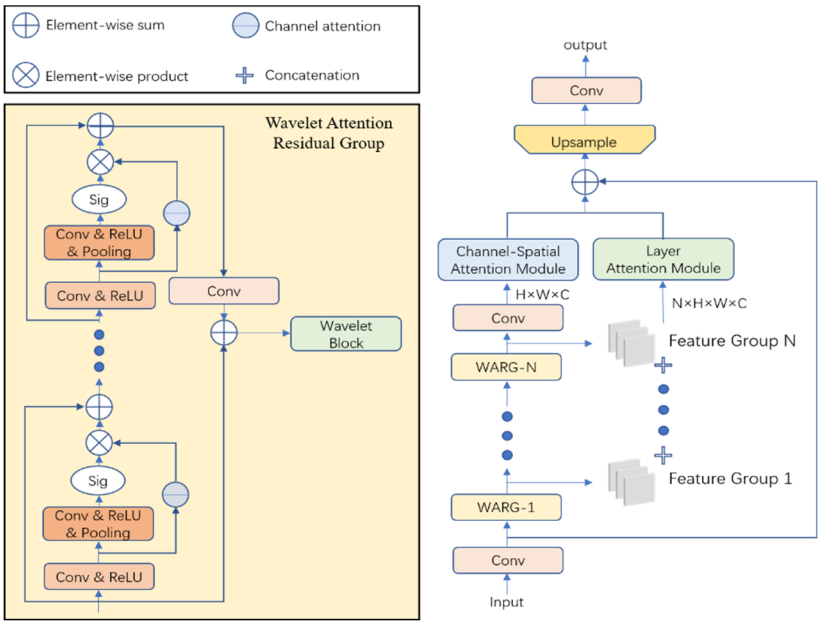

Fig 1. Overall Network Structure

3.2 Attention Mechanism

We were inspired by HAN [6] and introduced channel-spatial attention and layer attention to non-local attention.

Channel-Spatial Attention. The specific structure of the channel-spatial attention mechanism is shown in Figure 2. The input is $F_N \in R^{H\times W\times C}$ then three-dimensional convolution operation is performed on it. And the operation is expressed as the following formula:

$$W_{csa} = F_N * W_{3D} \tag{4}$$

Where * represents the convolution operation, $W_{3D} \in R^{3\times3\times3}$ is the convolution kernel, and the stride is 1. Then the output W_{csa} is feature weighted as the following formula:

$$F_{CS} = \beta\sigma(W_{csa}) \odot F_N + F_N \tag{5}$$

Where $\sigma(\cdot)$ is the sigmoid function.$\odot$ means element-wise multiplication, and the scaling factor is initialized to 0, and it is gradually improved in the following iterations.

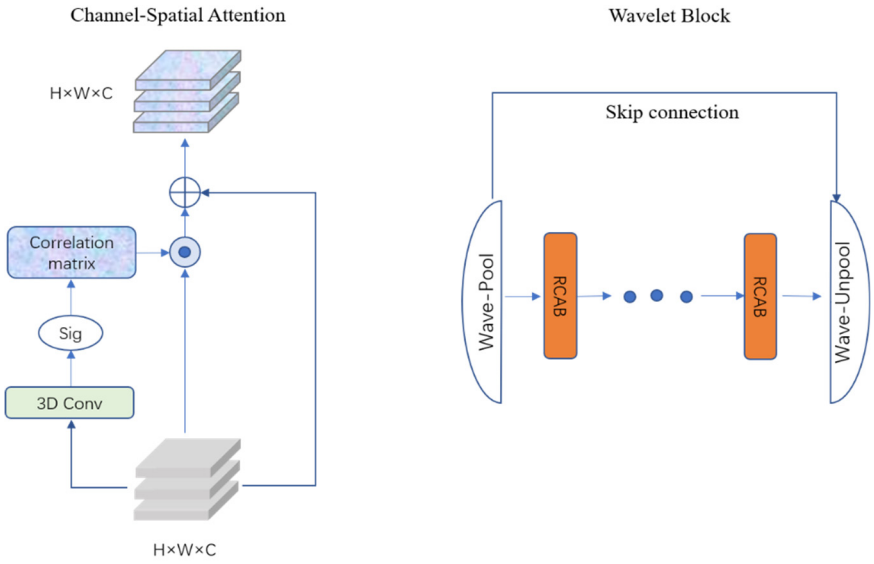

Fig 2. Architecture of the Channel-Spatial Attention model and Wavelet Block

Layer Attention. We use the LA module to better use information between layers, which avoids the network 'forgetting' shallow features by calculating dependencies between layers. Compared with the recurrent network, feature reuse can be enhanced better. To a certain extent, we can make the network more powerful by emphasizing important features and avoiding feature degradation caused by the deep network.

The network structure of LA is shown in Figure 3. Generally speaking, its composition can be divided into three parts. Firstly, we take the intermediate features of each layer as input, which size is $N \times H \times W \times C$, all the intermediate features are combined into a two-dimensional matrix, which size is $N \times HWC$ and each row represents the output of the corresponding layer. Then, we calculate attention between features:

$$s_{ij} = \text{softmax}(x_i x_j) \tag{6}$$

Where s_{ij} indicates the extent to which the layer attention attends to the $i\text{-}th$ layer when synthesizing the $j\text{-}th$ layer from N residual groups. Then we recalculate the element values of the feature map through the weights calculated above:

$$F_i = \alpha \sum_{i=1}^{N} s_{ij} x_i \tag{7}$$

Where α is a reconstruction weight that is dynamically adjusted in the network. Finally, we reshape the output into the same 3d shape as the input and add elements to the initial input to get the final layer output:

$$FL_i = F_i + x_i \tag{8}$$

As a result, the special conquest of the summation of the scholarship concentrates on the main part of the internet.

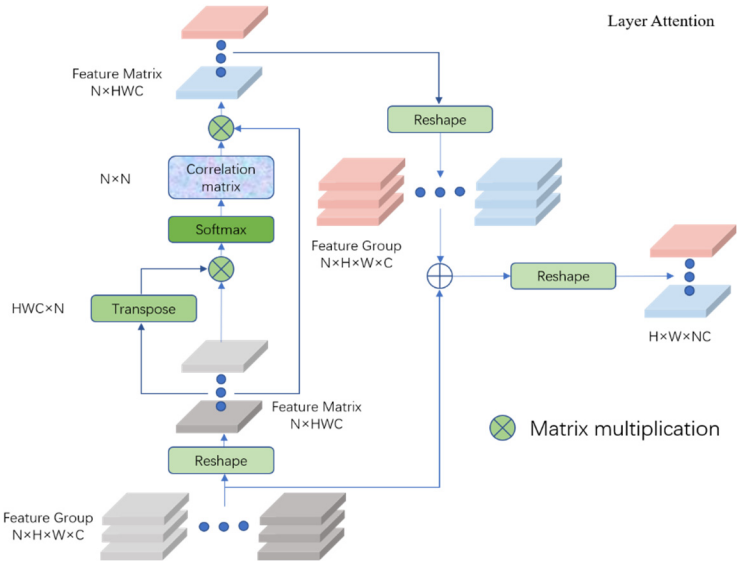

Fig 3. Architecture of the Layer Attention model

3.3 Wavelet Attention

To enable the deep super resolution network to adaptively distinguish high-frequency and low-frequency information, we propose a wavelet-based attention module and use a long skip connection to further improve the utilization of high-frequency details.

As shown in Figure 2, by considering inter-dependencies between channels, the features of the channel directions are adaptively redivided to improve the reconstruction results.

Image data can be regarded as a two-dimensional signal whose index is $[m, n]$ where $x[m, n]$ represents the pixel value of the $m\text{-}th$ row and $n\text{-}th$ column of the image. We use a Haar[27] wavelet to perform a two-dimensional discrete wavelet transform on the image. The principle is shown in Figure 4.

The input image is passed through a high-pass filter along with the columns:

$$G_H[n] = \begin{cases} 1, n = 0 \\ -1, n = 1 \\ 0, otherwise \end{cases} \tag{9}$$

And a low pass filter along the row:

$$G_L[n] = \begin{cases} 1, n = 0,1 \\ 0, otherwise \end{cases} \tag{10}$$

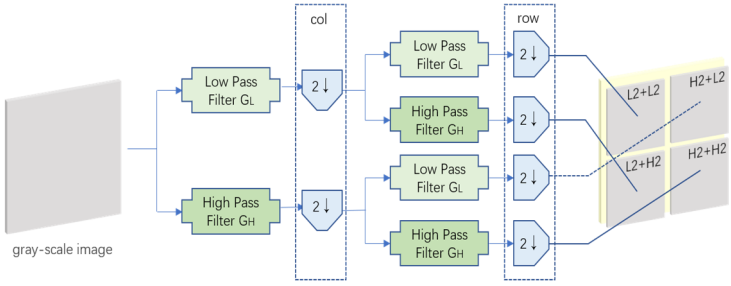

Fig 4. The principle of Haar wavelet transform

As shown in Figure 4, after filtering, the image size is reduced to half of the original size, and this process is repeated to get the four components of the input signal.

Inspired by the above principles, we combine the wavelet transform with the channel attention mechanism to improve the relevance and expressiveness of features. First, we pass the LR images through the 2d-DWT for the Haar wavelet to decompose the signal and produce four LR wavelet Sub-Bands, which can be denoted as:

$$I_{sub-bands} = WT(I_{LR}) \tag{11}$$

Where $I_{sub-bands}$ is a collection that covers the average, vertical, horizontal, and diagonal details of the LR image. Then we feed them into RCAN blocks which contain channel attention mechanism to get high-level representations of different features. To make features more sparse and more accessible for the network to learn a sparse map instead of a dense map, we use a long skip connection at the model's tail. Finally, we use the 2d-IDWT [27], which can trace back the 2d-DWT procedure by inverting the steps in Figure 4, which can be denoted as:

$$\mathrm{I_f} = IWT\left(I_{sub\_bands}\right) \tag{12}$$

Where I_f is the final feature map.

3.4 Loss Function

Compared to L1 loss, the L2 loss is more sensitive to larger errors but more tolerant to small errors, so the results tend to be smooth. In practice, L1 loss shows better performance and convergence than L2 loss [27], [28], [29]. Therefore, we choose L1 loss as the basis function. However, since the L1 Loss does not take into account image quality (e.g., perceptual quality [13], texture [30]), the results tend to lack high-frequency details, and for over-smoothed textures [3], [13], [31], are not visually satisfying. And [6] only uses distance as the loss function:

$$L(\Theta) = \frac{1}{m}\sum_{i=1}^{m}\left\|H_{HAN}\left(I_{SR}^{i}\right) - I_{HR}^{i}\right\|_1 = \frac{1}{m}\sum_{i=1}^{m}\left\|I_{SR}^{i} - I_{HR}^{i}\right\|_1 \tag{13}$$

Inspired by [32], we introduce a function composed of an improved filter to capture more image structural information and semantic information to obtain images with higher perceptual quality and more stable training results. The two-direction template of the traditional operator can only detect horizontal, vertical, and edge directions close to them and is invalid for 45-degree and 135-degree oblique edges and edges close to the two directions. In practical applications, the contours of most images are relatively complex and are especially dominated by oblique edges, with a small proportion of horizontal and vertical edges.

Compared with the classic Sobel Filter, the Improved Sobel Filter (ISL) adds two templates and two directions to readjust the weights of the original two templates and increase the weights of the new template in the direction of the oblique edge. It can be expressed as the following formula:

$$ISL\left(\hat{Z},Z\right) = \frac{1}{N}\sum\left|Sobel^{*}(Z) - Sobel^{*}\left(\hat{Z}\right)\right| \tag{14}$$

Where T represent the real image, $\tilde{T}$ re represents of the output of our attention model, and $Sobel$ represents the improved Sobel edge detection filter. We use a weighted average of hyperparameters for ISL and L1, which can be expressed as:

$$Loss\left(\hat{Z},Z\right) = \mathcal{L}_1\left(\hat{Z},Z\right) + \lambda.ISL\left(\hat{Z},Z\right) \tag{15}$$

4. Experiment

4.1 Datasets and Training Details

We used the Flickr2K [33] for training, and the model was evaluated with standard and popular benchmarks, including Set5 [34] and Set14 [35]. We also included BSD100 [36] and a dataset of urban landscape Urban100[37].

The model implementation of this paper mainly builds PyTorch deep learning framework based on an RTX3090 GPU and environments on a Linux terminal virtual machine. We keep the learning rate as $1e^{(-5)}$, and trained for 250 epochs. We used $\lambda_1 = 0.1$ and $\lambda_2 = 0.9$ for weights of pixel loss and Improved Sobel Loss. In each epoch, we used an ADAM optimizer. We did not use any data augmentation, and fixed the number of residual groups as 10.

4.2 Comparison with the State of the Arts

Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) are used as the image-quality metrics to evaluate the model of the datasets. We used the pretrained HAN[6] model to initialize the model parameters and get the test images using the Bicubic (BI) Degradation Model to compare the metrics of 13 state-of-the-art methods with ours. Table 1 shows quantitative results and indicates that our model can improve the PSNR and SSIM of the existing attention-based model on all four datasets.

It can be seen from Table 1 that our proposed model achieves great PSNR and SSIM scores on all test sets, which exceeds many of the existing classical deep learning-based methods. Specifically, our WLSAN yields the best quantitative results, better than the attention-based methods of RCAN and SA, and has a leap forward on a large scale compared to HAN.

4.3 Visual Results

We also compare the visual effects of different methods with the BI model from the BSD100 dataset on 4x scale. Our WLSAN model visually outperforms other classical methods. As can be seen from Figure 5, the classical network that used residual block as the backbone (VDSR, LapSRN) has poor reconstruction and cannot restore the needle-shaped mushroom leaf in the first row of Figure 5 and the grid of the buildings in the second row. They also suffer from blurring artifacts, as well as unpleasantly ant aliased edges. Although attention-based networks (RCAN, HAN) can alleviate it to a certain extent, but still misses the detailed texture. In contrast, our WLSAN effectively exploits the details and the internal image statistics to super-resolve the high-frequency contents (for example, our method can recover the delicate texture of mushroom caps and grid structure of the building that is difficult to observe with the human eyes). From the reconstruction metrics of a single image, our method outperforms HAN, which further demonstrates the effectiveness of our improved method.

Table 1. Quantitative results with BI degradation model. The best and second best results are highlighted in **bold** and *underlined*

Methods	Scale	Set5		Set14		BSD100		Urban100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	×2	33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403
SRCNN	×2	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946
FSRCNN	×2	37.05	0.956	32.66	0.9090	31.53	0.8920	29.88	0.9020
VDSR	×2	37.53	0.9591	33.05	0.9130	31.9	0.8960	30.77	0.9140
LapSRN	×2	37.52	0.9597	33.08	0.9130	31.08	0.8950	30.41	0.9101
MemNet	×2	37.78	0.9602	33.28	0.9142	32.08	0.8978	31.31	0.9195
EDSR	×2	38.11	0.9601	33.92	0.9195	32.32	0.9013	32.93	0.9351
SRMDNF	×2	37.79	0.9600	33.32	0.9159	32.05	0.8985	31.33	0.9204
D-DBPN	×2	38.09	0.9600	33.85	0.9190	32.27	0.9000	32.55	0.9324
RDN	×2	38.24	0.9614	34.01	0.9212	32.34	0.9017	32.89	0.9353
RCAN	×2	38.27	0.9614	34.12	0.9216	32.41	0.9027	33.34	0.9384

SRFBN	×2	38.11	0.9609	33.82	0.9196	32.29	0.9010	32.62	0.9328
SAN	×2	*38.31*	*0.9620*	34.07	0.9213	*32.42*	*0.9028*	33.10	0.9370
HAN	×2	38.27	0.9614	*34.16*	*0.9217*	32.41	0.9027	*33.35*	*0.9385*
WLSAN(ours)	**×2**	**38.34**	**0.9651**	**34.19**	**0.9274**	**32.42**	**0.9078**	**33.36**	**0.9422**
Bicubic	×4	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577
SRCNN	×4	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221
FSRCNN	×4	30.72	0.8660	27.61	0.7550	26.98	0.7150	24.62	0.7280
VDSR	×4	31.35	0.8830	28.02	0.7680	27.29	0.0726	25.18	0.7540
LapSRN	×4	31.54	0.8850	28.19	0.7720	27.32	0.7270	25.21	0.7560
MemNet	×4	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630
EDSR	×4	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033
SRMDNF	×4	31.96	0.8925	28.35	0.7787	27.49	0.7337	25.68	0.7731
D-DBPN	×4	32.47	0.8980	28.82	0.7860	27.72	0.7400	26.38	0.7946
RDN	×4	32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028
RCAN	×4	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087
SRFBN	×4	32.47	0.8983	28.81	0.7868	27.72	0.7409	26.60	0.8015
SAN	×4	*32.64*	*0.9003*	*28.92*	0.7888	27.78	0.7436	26.79	0.8068
HAN	×4	*32.64*	0.9002	28.90	*0.7890*	*27.80*	*0.7442*	*26.85*	*0.8094*
WLSAN(ours)	**×4**	**32.68**	**0.9063**	**28.95**	**0.8045**	**27.8**	**0.7604**	**26.91**	**0.8201**
Bicubic	×8	24.40	0.6580	23.10	0.5660	23.67	0.5480	20.74	0.5160
SRCNN	×8	25.33	0.6900	23.76	0.5910	24.13	0.5660	21.29	0.5440
FSRCNN	×8	20.13	0.5520	19.75	0.4820	24.21	0.5680	21.32	0.5380
SCN	×8	25.59	0.7071	24.02	0.6028	24.30	0.5698	21.52	0.5571
VDSR	×8	25.93	0.7240	24.26	0.6140	24.49	0.5830	21.70	0.5710
LapSRN	×8	26.15	0.7380	24.35	0.6200	24.54	0.5860	21.81	0.5810
MemNet	×8	26.16	0.7414	24.38	0.6199	24.58	0.5842	21.89	0.5825
MSLapSRN	×8	26.34	0.7558	24.57	0.6273	24.65	0.5895	22.06	0.5963
EDSR	×8	26.96	0.7762	24.91	0.6420	24.81	0.5985	22.51	0.6221
D-DBPN	×8	27.21	0.7840	25.13	0.6480	24.88	0.6010	22.73	0.6312
RCAN	×8	27.31	0.7878	25.23	0.6511	24.98	0.6058	23.00	0.6452
SAN	×8	27.22	0.7829	25.14	0.6476	24.88	0.6011	22.70	0.6314
HAN	×8	*27.33*	*0.7884*	*25.24*	*0.6510*	*24.98*	*0.6059*	*22.98*	*0.6437*
WLSAN(ours)	**×8**	**27.40**	**0.7920**	**25.42**	**0.6729**	**25.05**	**0.6273**	**23.19**	**0.6658**

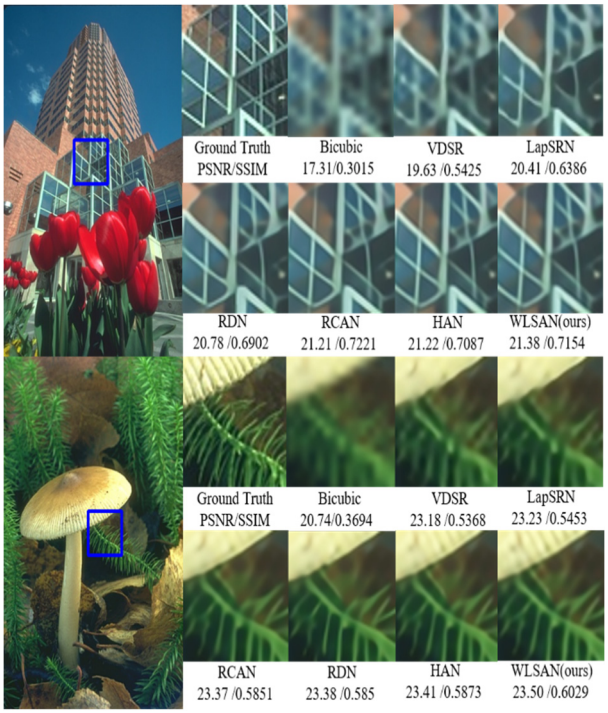

Fig 5. Visual result for 4x SR with BI model on the BSD100dataset

5. Ablation Study

In this section, we test Set5 with different scales to study the effect of the improvement of WLSAN on the model's reconstruction quality. As can be seen from Table 2, at all scales, our proposed Improved Sobel loss, and wavelet attention improve the accuracy. Among them, the use of wavelet attention increased by 0.06db compared with the average without use; Using only the Improved Sobel Loss increases by 0.02db compared to the unused average. If both are used, the average increase is 0.09db.

6. Conclusion

This work improved the attention-based model in single-image super resolution. We captured the interaction of the local informative feature representations using wavelet transformation and further enhanced the ability through a mixed weighted summation of spatial and layer attention combined with advanced ISL loss. This work also helps improve training stability and get results with good accuracy and visual quality compared with other methods. Our future work will address generating high-resolution images for real-world scenarios such as unknown degradation and missing paired LR-HR images.

Table 2. Reconstruction performance of different structures

ASL	WV	Set5 x2	Set5 x4	Set5 x8
×	×	38.327	32.561	27.275
×	√	38.342	32.652	27.367
√	×	38.329	32.596	27.391
√	√	38.347	32.681	27.407

References

[1] Dong, C., Loy, C. C., He, K., & Tang, X. (2014, September). Learning a deep convolutional network for image super-resolution. In European conference on computer vision (pp. 184-199). Springer, Cham.

[2] Tai, Y., Yang, J., Liu, X., & Xu, C. (2017). Memnet: A persistent memory network for image restoration. In Proceedings of the IEEE international conference on computer vision (pp. 4539-4547).

[3] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (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).

[4] 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).

[5] Guo, T., Seyed Mousavi, H., Huu Vu, T., & Monga, V. (2017). Deep wavelet prediction for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 104-113).

[6] Niu, B., Wen, W., Ren, W., Zhang, X., Yang, L., Wang, S., ... & Shen, H. (2020, August). Single image super-resolution via a holistic attention network. In European conference on computer vision (pp. 191-207). Springer, Cham.

[7] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[8] Kim, J., Lee, J. K., & Lee, K. M. (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] Kim, J., Lee, J. K., & Lee, K. M. (2016). Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1637-1645).

[10] 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).

[11] Hui, Z., Gao, X., Yang, Y., & Wang, X. (2019, October). Lightweight image super-resolution with information multi-distillation network. In Proceedings of the 27th acm international conference on multimedia (pp. 2024-2032).

[12] Liu, J., Zhang, W., Tang, Y., Tang, J., & Wu, G. (2020). Residual feature aggregation network for image super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2359-2368).

[13] Johnson, J., Alahi, A., & Fei-Fei, L. (2016, October). Perceptual losses for real-time style transfer and super-resolution. In European conference on computer vision (pp. 694-711). Springer, Cham.

[14] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. Computer Vision and Pattern Recognition. IEEE, Piscataway.

[15] Wang, X., Yu, K., Dong, C., & Loy, C. C. (2018). Recovering realistic texture in image super-resolution by deep spatial feature transform. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 606-615).

[16] Guo, Y., Chen, Q., Chen, J., Huang, J., Xu, Y., Cao, J., ... & Tan, M. (2018). Dual reconstruction nets for image super-resolution with gradient sensitive loss. arXiv preprint arXiv:1809.07099.

[17] Bai, Y., Zhang, Y., Ding, M., & Ghanem, B. (2018). Sod-mtgan: small object detection via multi-task generative adversarial network. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 206-221).

[18] 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).

[19] Dai, T., Cai, J., Zhang, Y., Xia, S. T., & Zhang, L. (2019). Second-order attention network for single image super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11065-11074).

[20] Daubechies, I. (1992). Ten lectures on wavelets. Society for industrial and applied mathematics.

[21] Mallat, S. (1999). A wavelet tour of signal processing. Elsevier.

[22] Bae, W., Yoo, J., & Chul Ye, J. (2017). Beyond deep residual learning for image restoration: Persistent homology-guided manifold simplification. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 145-153).

[23] Liu, P., Zhang, H., Zhang, K., Lin, L., & Zuo, W. (2018). Multi-level wavelet-CNN for image restoration. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 773-782).

[24] Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3365-3387.

[25] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & 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).

[26] Abraham, E. (2018, November). Moiré pattern detection using wavelet decomposition and convolutional neural network. In 2018 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1275-1279). IEEE.

[27] Ahn, N., Kang, B., & Sohn, K. A. (2018). Fast, accurate, and lightweight super-resolution with cascading residual network. In Proceedings of the European conference on computer vision (ECCV) (pp. 252-268).

[28] 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).

[29] Zhao, H., Gallo, O., Frosio, I., & Kautz, J. (2016). Loss functions for image restoration with neural networks. IEEE Transactions on computational imaging, 3(1), 47-57.

[30] Sajjadi, M. S., Scholkopf, B., & Hirsch, M. (2017). Enhancenet: Single image super-resolution through automated texture synthesis. In Proceedings of the IEEE international conference on computer vision (pp. 4491-4500).

[31] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4), 600-612.

[32] Zheng, B., Yuan, S., Slabaugh, G., & Leonardis, A. (2020). Image demoireing with learnable bandpass filters. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3636-3645).

[33] 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).

[34] Bevilacqua, M., Roumy, A., Guillemot, C., & Alberi-Morel, M. L. (2012). Low-complexity single-image super-resolution based on nonnegative neighbor embedding.

[35] Zeyde, R., Elad, M., & Protter, M. (2010, June). On single image scale-up using sparse-representations. In International conference on curves and surfaces (pp. 711-730). Springer, Berlin, Heidelberg.

[36] Martin, D., Fowlkes, C., Tal, D., & Malik, J. (2001, July). 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, pp. 416-423). IEEE.

[37] 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).