Single Image Super-resolution Based On Non-subsampled Shearlet Transform

Ming Tan 825443411@qq.com
Liang Chen * cl_0827@126.com
Xuan Wu xwu08198@gmail.com
Yi Wu wuyi@fjnu.edu.cn
Fujian Provincial Key Laboratory of Photonics Technology, Fujian Normal University, Fuzhou, China.

Editors: Berrin Yanıkoğlu and Wray Buntine

Abstract

With the development of deep learning, breakthroughs in single image super-resolution have been achieved. However, most existing methods are limited to using only spatial domain information or only frequency domain information, and the rich information of the image in the frequency domain space is not fully utilized, so it is still difficult to recover satisfactory texture details. In this paper, we propose a method to fuse the frequency domain and spatial domain information. Our method uses a two-branch network to extract the spatial domain information and the frequency domain information separately and uses a fusion module to fuse the different information in the two domains. We also use the Non-Subsampled Shearlet Transform (NSST) to preserve the texture directionality well, and design two NSST-based directional texture enhancement modules, which are embedded in different parts of the network, to enhance the recovery of texture details in the image reconstruction process. Quantitative and qualitative experimental results show that the method outperforms advanced single-image super-resolution methods in recovering images.

Keywords: Non-subsampled Shearlet Transform;image super-resolution;frequency domain;

1. Introduction

Image super-resolution (SR) refers to the process of recovering high-resolution (HR) images from low-resolution (LR) images, and is an important class of image processing technologies in computer vision and image processing. In general, this problem is very challenging and has inherent uncertainties as there are always multiple HR images corresponding to a single LR image. Image restoration is an anti-problem for the degradation process, because some important content information about the image is lost during the image degradation process. Therefore, in order to restore high-quality images, the rich information contained in degraded images should be fully utilized.

For the reconstruction of natural images, both spatial domain information and frequency domain information can be recovered. First of all, most of the current research only uses spatial domain information or only uses frequency domain information, and the method of integrating spatial domain information and frequency domain information needs to be

∗. Corresponding author.

Figure 1: Natural images show textures with many similarities. The directionality of oblique textures (red squares) and horizontal textures (cyan squares) can be used as a basis for finding similar textures.

further studied. Secondly, most of the studies are based on the reconstruction of spatial information, although with the deepening of the research, the reconstruction effect of the image has been improved to a certain extent, but it ignores the rich information of the image in the frequency domain space. The reconstruction methods based on frequency domain space mainly include methods based on Fourier transform, discrete cosine transform, and wavelet transform, which only convert the image into the frequency domain, and simply divide the frequency domain information into high-frequency and low-frequency information to participate in the image reconstruction process, without further refining and mining the frequency domain information.

In order to alleviate the above problems, the following options are proposed: Firstly,the two-branch network is used to reconstruct the frequency domain information and the spatial domain information respectively, and the fusion module is used to realize the fusion of the spatial domain information and the frequency domain information, so as to solve the problem of incomplete utilization of low-resolution image information in a single domain in image reconstruction. Secondly,this paper proposes an image super-resolution method based on Non-Subsampled Shearlet Transform(NSST) Easley et al. (2008), which refines the frequency domain features of the image, separates the texture directionality that is helpful for image reconstruction, extracts the direction information of high-frequency features in the frequency domain of the image through the NSST, and makes full use of the network model through the texture enhancement module to realize the enhancement of the texture in the process of image reconstruction. Our contributions are as follows:

1) A model that integrates frequency domain information and spatial domain information is proposed, which allows our method to make full use of the rich information contained in degraded images.

2) A texture enhancement structure based on NSST is designed, which introduces the directionality of texture into the image reconstruction process, which can alleviate the problem that the edge texture area is difficult to recover during the image reconstruction process.

3) Our system is an end-to-end trainable model that does not require any pre-training phase. The experimental results also show that the proposed method outperforms state-of-the-art methods with few parameters.

2. Related works

2.1. Single image super-resolution

Single image super-resolution based on deep neural networks started with SRCNN Dong et al. (2015) and has been rapidly developed in recent years. Since then, various network models have been proposed to achieve better image super-resolution. Kim et al. (2016) proposed VDSR , which uses a deeper network depth to increase the perceptual field of the network model and improves super-resolution accuracy by introducing residual learning to alleviate the problem of training difficulties caused by deeper network depth. Lim et al. (2017) trained a very deep super-resolution network using optimised residual blocks EDSR , which further improved the super-resolution performance of the residual network. By making full use of the information at each level in the convolutional layer, Zhang et al. (2018b) proposed RDN to achieve high quality image SR using the rich information flow from the real environment. To enhance the network's discriminative learning capability, Zhang et al. (2018a) proposed a deep residual channel attention network (RCAN), which adaptively rescales the features of each channel through a channel attention mechanism, allowing the network to focus on more useful channels and enhance the discriminative learning capability. Hui et al. (2019) proposed a lightweight information multi-distillation network (IMDN) , which gradually extracts hierarchical features through the distillation module, and uses the fusion module to aggregate them according to the importance of the candidate features, so that the network can obtain better reconstruction results at the same time. Consumes only a small amount of computing resources. Zhou et al. (2022) designed the Vast-receptive-field Pixel attention network (VapSR) by improving the attention mechanism, and achieved similar performance to other networks with fewer parameters. SwinIR Liang et al. (2021) promotes Swin Transformer for the SR task. HAT Chen et al. (2023) refreshes state-of-the-art performance through hybrid attention schemes and pre-training strategy.

2.2. Frequency domain based image super-resolution

Numerous studies have shown that the frequency information of HR images can be expressed by CNNs. Li et al. (2018) used the convolution theorem to convert image super-resolution networks to the frequency domain. They converted the convolution in the spatial domain to the product in the frequency domain and the non-linearity in the spatial domain to the convolution in the frequency domain, which improved the computational efficiency. Guo

et al. (2019) integrated DCT as a convolutional DCT (CDCT) layer into the network and proposed the DCT deep SR network (ORDSR), which achieved the best performance at that time. Xu et al. (2022) innovatively designed the DCT space cube to achieve multi-level feature decomposition of LR images, and then modified the adaptive non-local dual attention mechanism to achieve adaptive high-frequency feature extraction, which can effectively recover the high-frequency information of images. Esmaeilzehi et al. (2021b) proposed a lightweight multi-domain SR network (SRNSSI), which designed a multi-domain residual block including a spatial domain feature processing module for learning spatial information and a frequency domain feature processing module for learning spectral information, improving the performance of lightweight SR networks. Zou et al. (2022) proposed the joint wavelet sub-bands guided network (JWSGN). Different frequency information of the image is separated by DWT and then recovered by a multi-branch network. The edge extraction module was used to estimate the edge feature map using the similarity of high frequency sub-bands to recover finer edge details. By introducing the Fourier transform, Wang et al. (2023) designed a spatial-frequency interaction network (SFMNet) to capture global face structure information using the Fourier transform to achieve an image size acceptance domain. Complementary spatial and frequency information is merged with each other to enhance the model's capabilities. Most of these frequency domain-based methods only mechanically introduce frequency domain information into the image reconstruction network, or simply divide the frequency domain information into high and low frequencies for separate processing, without further mining the frequency domain information and ignoring the directionality of the image texture corresponding to the high frequency information in the frequency domain, which apparently plays a key role in the reconstruction of the image edge texture.

3. Motivation

It is well known that in image restoration, smooth areas of an image are relatively easier to recover, while edge textures are more difficult to recover, corresponding to high frequency parts in the frequency domain. Edge texture regions that are more difficult to extract in the spatial domain are easier to separate in the frequency domain. The SR method of fusing frequency domain and spatial domain information has not been fully investigated, so we try to provide the network model with the feature information of edge texture in the frequency domain by fusing the reconstruction of frequency domain information with the reconstruction of spatial domain information to further enhance the reconstruction of texture details. There is a wealth of line information in the real world, and the vast majority of man-made buildings follow a pattern of symmetry and repetition , as this corresponds to the natural human aesthetic, and the same pattern exists in nature. As a result, real-world photographs often contain some local details with similar texture structure, such as the textures in the red or blue boxes in Figure 1. For image reconstruction tasks, exploiting the local texture similarity of an image can be beneficial in solving the problem that some areas are severely degraded and difficult to reconstruct. Some recent studies on non-local Wang et al. (2018) and Self-attention in transformer Liang et al. (2021); Conde et al. (2022); Chen et al. (2023) have demonstrated that the internal similarity of images can help image reconstruction, and for texture details with similarity, their internal similarity can be used

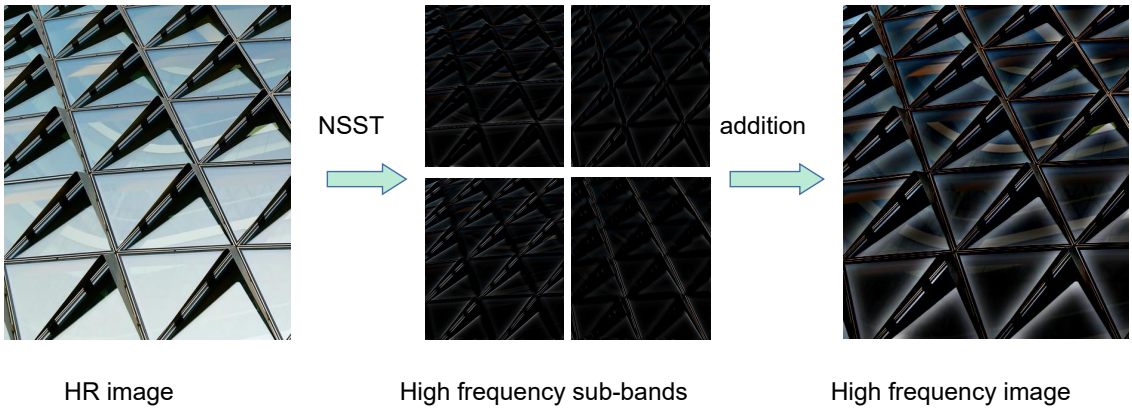

Figure 2: Visualization of NSST extracted directional textures. Here 1 level of NSST was applied to each HR image to generate 4 high frequency subbands.

to complement each other in the reconstruction process, and the better preserved texture regions in low-resolution images can complement the regions where similar texture details are severely lost. How to accurately and efficiently filter out the areas with similar textures in the image and the target area? The similar textures in the red boxes have the same directionality, as shown in Figure 1(with some similar textures framed in red and cyan boxes). Therefore, we try to extract similar textures by their directionality. Based on the use of NSST sub-band in Liu et al. (2021) to represent the human body image a priori, which can have a good preservation of the directionality of the image texture information and enhance the details of the reconstructed human body image, we embed the NSST method in the model to achieve the preservation of the directionality of the texture of the image itself during feature extraction.The different high frequency sub-bands in the NSST Easley et al. (2008) represent edges along different directions and therefore allow better differentiation between different textures. For an image, the non-subsampled shearlet transform is defined as:

$$\left\{\psi_{j,l,k} = |\det A|^{j/2}\psi\left(S^l A^j x - k\right) : l, j \in Z, k \in Z^2\right\} \tag{1}$$

where ψ is a collection of basis function and satisfiifies $\psi \in L^2(R^2)$, j, l, and k are scale, direction, and shift parameter, respectively. The A and S are both 2×2 invertible matrices, and $|\det A| = 1$. The A is an anisotropic matrix, dominating the scaling of shearlet, while the S is a shearing matrix, which controls the orientation of shearlet. Following standard settings of Liu et al. (2019), $A = A_0 = [4, 0; 0, 2]$ or $A = A_1 = [2, 0; 0, 4]$, and $S = S_0 = [1, 1; 0, 1]$ or $S = S_1 = [1, 0; 1, 1]$. Therefore, the shearlet transform functions can be written as:

$$\hat{\psi}^{(0)}_{j,l,m}(x) = 2^{j\frac{3}{2}}\psi^0\left(S_0^l A_0^j x - k\right), \tag{2}$$

$$\hat{\psi}^{(1)}_{j,l,m}(x) = 2^{j\frac{3}{2}}\psi^1\left(S_1^l A_1^j x - k\right) \tag{3}$$

where j $\leq 0, -2^j \leq 1 \leq 2^j - 1$. NSST is the shift-invariant version of the shearlet transform. The NSST generally consists of two parts: a multiscale decomposition and a multidirectional decomposition. The former uses a non-downsampling Laplace pyramid transform to achieve the multiscale decomposition, while the latter uses a combination of several different shear filters to achieve it. As shown in the Figure 2, after NSST decomposition of the HR image, we obtain several high-frequency subbands, each representing texture details along different directions. The enhancement of the image texture reconstruction is achieved by feeding the high-frequency subbands into the texture enhancement module to emphasise the directional information contained in the texture.

4. Method

In view of the difficulty of recovering edge texture information in super-resolution tasks, we believe that the reason is that the edge texture information in the current solution is not fully explored and utilized. Therefore, we try to design a deep learning network model that emphasizes edge texture information. On the one hand, we use the two-branch structure to reconstruct the spatial domain and frequency domain features respectively, and use the fusion structure to fuse the frequency domain features with the spatial domain features, so as to expand more diversified texture information for the network model. On the other hand, considering that the edge texture information that is difficult to separate in the spatial domain can be separated in the frequency domain by simple and intuitive methods, we use the high-frequency information in the frequency domain to emphasize the edge texture information in the network model. Since the NSST can retain the local feature information of the image in different directions and scales, it can extract the high-frequency information with directionality, which will help the recovery of edge texture information in the image super-resolution. We embed the NSST into the network model, and further design the Directional Texture Enhancement Module(DTEM) to achieve the emphasis on edge texture information by transforming the features in the training process to obtain the direction information of the texture.

4.1. Network architecture

The general architecture of the network is shown in the Figure 3. The network takes degraded low quality images as input, processes the images within the network and outputs recovered high quality images. Specifically, this network contains three components. 1) The feature extraction layer is implemented as multiple Residual Channel Attention Blocks (RCAB) Zhang et al. (2018a) cascaded with Spatial Frequency Domain Blocks (SFB) Zhang et al. (2022) blocks to transform the input image into a feature mapping. 2) The texture emphasis component is implemented through a two-branch network with the frequency branch as a secondary task, which is based on the IMDB Hui et al. (2019) with the addition of VapB Zhou et al. (2022) to enhance the network distillation capability and Non-local is added to expand the global information; The spatial domain branch is used as the main task, and the directional texture extraction module NSSB is added to the basic IMDB block to preserve the directionality of the texture in the features. The two-branch network achieves the fusion of spatial and frequency domain features by means of a fusion module. Skip connections are applied to the texture emphasis module. 3) The image reconstruction

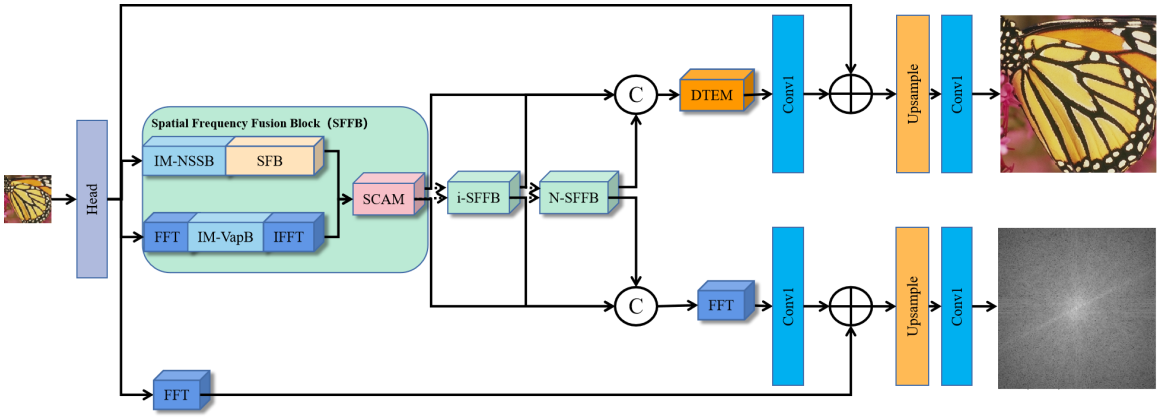

Figure 3: The architecture of the proposed NSSTNet in which FFT and IFFT are Fourier transform and inverse Fourier transform. NSSTNet consists of a spatial branch (top branch) and a frequency branch (bottom branch). The former aims to extract local features of the image and achieve enhancement of directional textures using the NSST approach. The latter, on the other hand, focuses on capturing global information through the Fourier transform and achieving an image size acceptance domain.

part, using the rich features computed by the previous operations, is estimated on the recovered image. In particular, for the spatial domain branch, we inserted a Directional Texture Enhancement Module (DTEM) to achieve an emphasis on the directional texture of the image.

4.2. Overview

In this subsection, we describe the pipeline in detail. For the LR image I^{LR}, we extract features by feeding them from two branches into two convolutional layers, generating F_f and F_s corresponding to the frequency and spatial branches. Then, the extracted features are fed into N Spatia Frequency Fusion Blocks (SFFB) to extract multi-scale features that,

$$F_s^{\mathrm{i}}, F_f^i = H_{SFFB}^i \left(F_s^{i-1}, F_f^{i-1} \right) \tag{4}$$

where H_{SFFB} is a function of the i-th SFFB. After N SFFBs, F_s^N is fed into the directional texture emphasis module(DTEM), which further enhances the image texture by:

$$F_s = H_{\mathrm{DTEM}} \ \left(H_C \left(F_s^1, \ldots, F_s^i, \ldots, F^N \right) \right), \tag{5}$$

$$F_f = H_C \left(F_f^1, \ldots, F_f^i, \ldots, F_f^N \right)^s \tag{6}$$

H_C denotes concatenate; F_s, F_f are then fed into the reconstruction layer, (consisting of convolutional layers) which is divided into two branches to recover the image I_F^{SR} and I_S^{SR}

, as shown in Figure 3. In order for the model to perform SR, the model is supervised by pixel-level and frequency-level loss functions that,

$$L = \lambda \left\| I_F^{SR} - H_{FFT}\left(I^{HR}\right) \right\|_1 + \left\| I_S^{SR} - I^{HR} \right\|_1 \tag{7}$$

where λ is the trade-off parameters and H_{FFT} indicates the conversion of the image to the frequency domain.

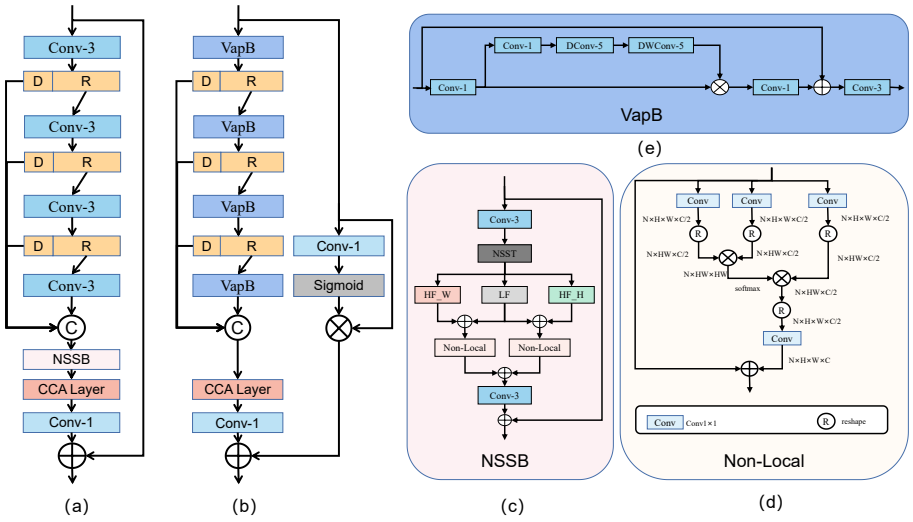

Figure 4: The structure of the spatial branch basic block IM-NSSB (a) and the frequency domain branch basic block IM-VapB (b), where D,R denote distill and remain respectively. The detailed structure of the texture enhancement module NSSB is shown in (c), and (d) and (e) show the detailed structure of the Non-local and VapB used in the network structure, respectively.

4.3. Spatia Frequency Fusion Block

The i-th spatia frequency fusion block (SFFB) in NSSTNet is described in detail here. In detail, at the i-th SFFB, the F_s^{i-1} and F_f^{i-1} generated by the i-1-th SFFB are fed into the spatial branch and the frequency branch, respectively,

$$F_f^i, F_s^i = H_{SCAM}\left(H_{\text{VapB}}\left(F_f^{i-1}\right), H_{SFB}\left(f_{\text{NSSTB}}\left(F_s^{i-1}\right)\right)\right) \tag{8}$$

where H_{VapB} and H_{NSSTB} correspond to the frequency block (IM-VapB) and the spatial block (IM-NSSB) in the frequency and spatial branches, respectively, and F_f^i and F_s^i are the extracted features.H_{SCAM} denotes the fusion module of frequency domain features with spatial domain features. IM-NSSB is used for the extraction of spatial domain branching features. Inspired by recent advances in self-attention and visual transformer, we use an elaborate NSSB module added to IMDB in IM-NSSB to enhance the extraction of image

texture features by spatial domain branching. Among them, NSSB uses the NSST method to decompose the image texture into horizontal and vertical, and exploits the self-correlation of textures in the same direction through the Non-local structure respectively, and fuses the texture information in different directions through a convolution layer. The structure is shown in Fig. 4(a). IM-VapB improves on the feature extraction part of IMDB Hui et al. (2019). We modified the original IMDB to use only a simple layer of convolution for feature distillation and replaced it with an enhanced spatial attention structure VapB, to improve the network's ability to extract feature information and make the distilled features more effective. We also added a skip-connected part to link the shallow and deep features. The exact structure is shown in Fig. 4(b).

With the Fourier transform, the frequency branch can capture global dependencies through the receptive domain of the image size, while the spatial branch can extract local dependencies. Given that global and local dependencies are complementary and can facilitate SR, we used SCAM Chu et al. (2022) to fuse the features of the frequency branch with those of the spatial branch, which is placed after each IM-NSSB with the IM-VapB to exploit their complementary nature.

4.4. Based on the design of the NSST texture emphasis module

We design two ways to take full advantage of NSST to implement the network model's emphasis on directional textures. Among them, NSSB is embedded in the spatial branch to enable the network to accurately extract the directionality of image features, and we refine the texture details into textures with directionality and separate them by direction.The structure is shown in Fig. 4(c). The low-frequency information reflects the structural features of the image, and the high-frequency information reflects the texture detail characteristics of the image, and we enable the network to fully explore and utilize different types of features by merging the structural features with the texture detail features. Through the non-local algorithm structure, the receptive field is increased, so that the texture in all directions can make full use of the common features of textures in the same direction, and jointly promote the recovery of texture features. Considering that the number of multi-directional features obtained by NSST increases exponentially with the direction factor, and the sparsity of the obtained features of each channel also increases simultaneously, which greatly aggravates the difficulty of training the network model, we fuse the texture features in all directions into representative horizontal and vertical directional features as textures. We define the NSSB as follows:

$$F_{NSSB} = F_{\text{in}} + H_{\text{conv}}\ \left(H_{NL}\left(f^{H}_{NSST}\left(H_{\text{conv}}\ \left(F_{\text{in}}\ \right)\right)\right) + H_{NL}\left(f^{W}_{NSST}\left(H_{\text{conv}}\ \left(F_{\text{in}}\ \right)\right)\right)\right) \quad (9)$$

where F_{in} denotes the input features, H_{NL} denotes Non-local blocks (Figure 4(d)), and f^{H}_{NSST} and f^{W}_{NSST} denote the extraction of horizontal and vertical textures, respectively.

In addition, we have designed the directional texture enhancement module DTEM, as shown in Figure 5, the input features are decomposed by NSST into texture features with orientation, and the different orientations are reinforced by different branches for the features, with the reinforced features fused by additions. Skip connections are used to keep the network learning direction consistent. Here is a detailed description of DTEM in NSSTNet. In detail,after the second stage of spatial frequency domain information

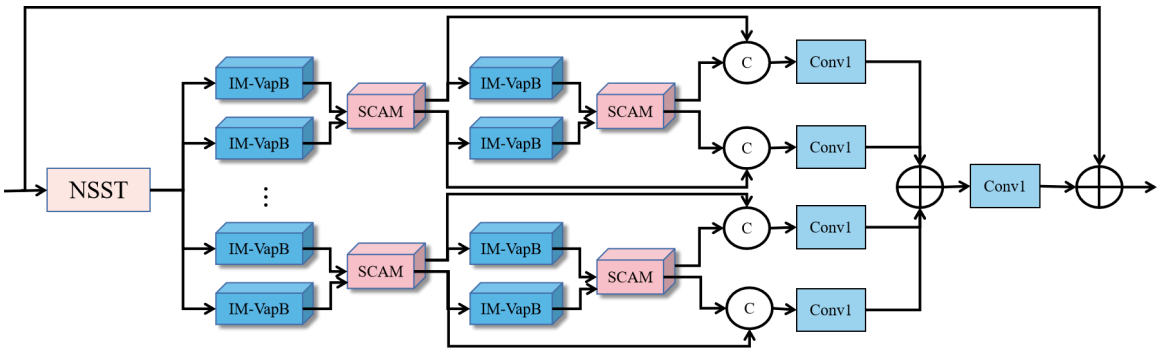

Figure 5: Structure of the directional texture enhancement module(DTEM). Where C denotes concatenate.

fusion, F_s is fed into the NSSB to achieve further enhancement of directional textures.F_s is decomposed by NSST into M directional texture components F_s^j, $j = 2n+1, n \in N$. F_s^j is fed into different branches and fused:

$$F_{s,k}^j, F_{s,k}^{j+1} = H_{SCAM}\left(H_{\text{VapB}}\left(F_{s,k-1}^j, F_{s,k-1}^{j+1}\right)\right) \tag{10}$$

Finally the reinforced directional textures are summed, adjusted by a convolutional layer, and skip connections are added to maintain consistency in model learning:

$$F_{es} = H_{\text{conv}}\left(\sum_{j=1}^{M} H_{conv}\left(H_C\left(F_{s,1}^j, \ldots, F_{s,k}^j, \ldots, F_{s,K}^j\right)\right)\right) \tag{11}$$

5. Experiments

The training images is DIV2K Agustsson and Timofte (2017) dataset containing 800 images. We evaluated our model on widely used benchmark datasets: Set5 Bevilacqua et al. (2012), Set14 Zeyde et al. (2012), BSD100 Martin et al. (2001), Urban100 Huang et al. (2015) and Manage109 Matsui et al. (2017). Common data enhancement methods were used on the training dataset.Specifically, we used a random combination of 0°, 90°, 180°, 270° random rotations and horizontal flips for data enhancement. The mean peak signal-to-noise ratio (PSNR) Jinjin et al. (2020) and structural similarity Wang et al. (2004) (SSIM) on the luminance (Y) channel were used as evaluation metrics.

Implementation details. We provide the model NSSTNet contains 6 blocks of SFFBs with directional branches of 4 in DTEM. The number of cascade combination blocks of RCAB and SFB used in the Head section is 6.

Training details. Our model was trained by the ADAM optimizer with a momentum parameter = 0.9 and an initial learning rate set to 2e-4 and halved every 200 iterations. We applied the PyTorch framework to implement the proposed network on a desktop computer.

Comparison with SOTA Methods. We compared the proposed NSSTNet with published SR methods including SRCNN Dong et al. (2015), LapSRN Lai et al. (2017), IMDN Hui

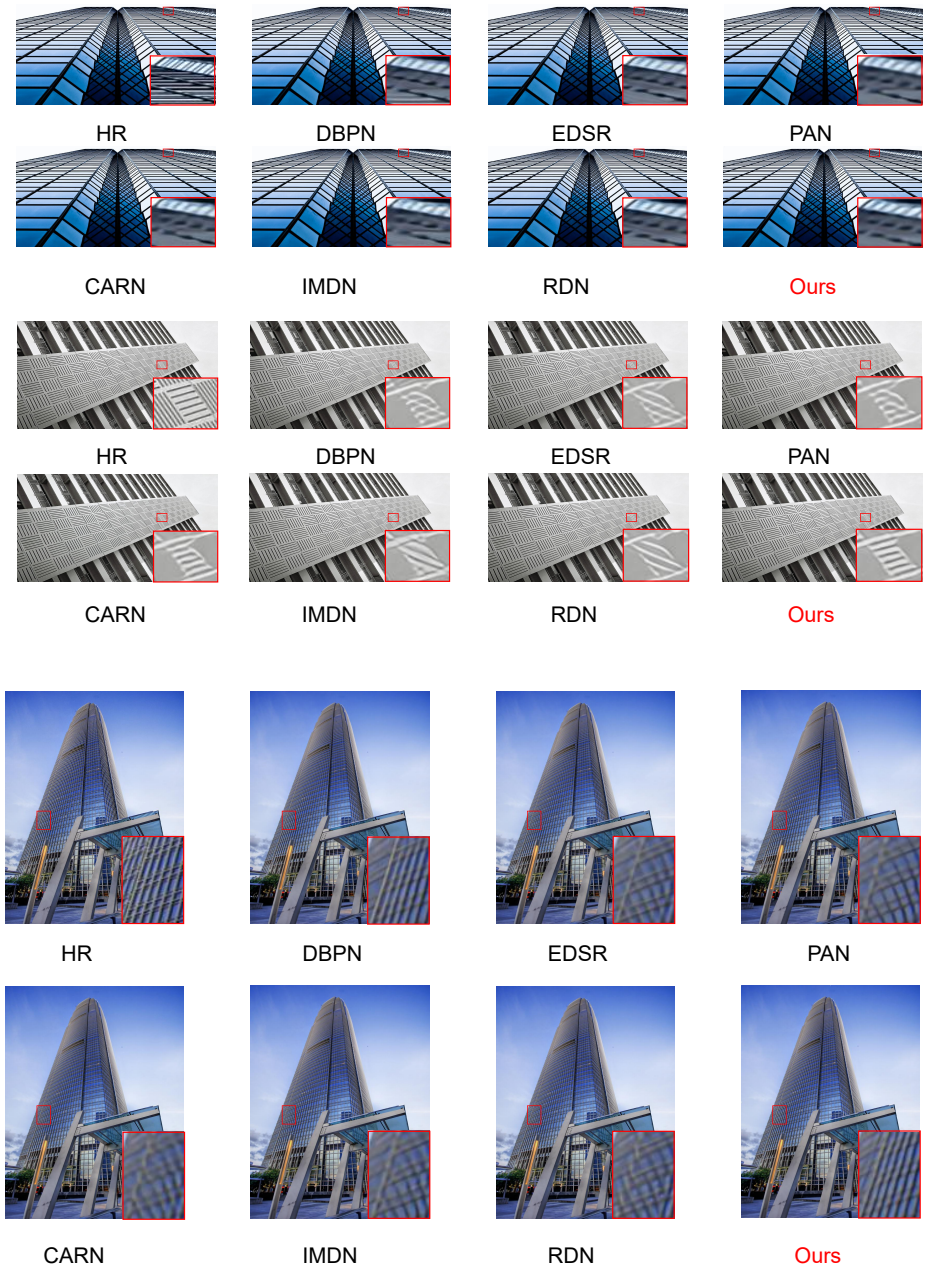

Figure 6: Qualitative comparison with the leading algorithms: DBPN Haris et al. (2018), EDSR Lim et al. (2017), PAN Zhao et al. (2020), CARN Ahn et al. (2018), IMDN Hui et al. (2019) and RDN Zhang et al. (2018b) on $\times 4$ task. From the figure, we can see that our method can generate finer details of the image and achieve outstanding performance.

et al. (2019),PAN Zhao et al. (2020), DCT-FANet Xu et al. (2022), SRNSSI Esmaeilzehi et al. (2021a), WDRN Xin et al. (2020), LapSRN Lai et al. (2017), DBPN Haris et al. (2018), RDN Zhang et al. (2018b),EDSR Lim et al. (2017),DRN Guo et al. (2020), RCAN Zhang et al. (2018a), on ×4. Table 1 shows the quantitative comparison results.The visualisation results are shown in Figure 6. As can be seen from the quantitative results and the visualisation, our method has a better performance in terms of reconstructing details.

Table 1: Quantitative comparison with state-of-the-art methods on benchmark datasets on ×4. The best and second-best performance are in red and blue colors respectively.

Method	Set5 PSNR/SSIM	Set14 PSNR/SSIM	BSD100 PSNR/SSIM	Urban100 PSNR/SSIM	Manga109 PSNR/SSIM
SRCNN	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221	27.58/0.8555
LapSRN	31.54/0.8850	28.19/0.7720	27.32/0.7270	25.21/0.7560	29.09/0.8900
IMDN	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
PAN	32.13/0.8948	28.61/0.7822	27.59/0.7363	26.11/0.7854	30.51/0.9095
DCT-FANet	32.09/0.7154	28.57/0.5658	27.51/0.5251	26.08/0.5667	30.18/0.6149
ESRNSSI	32.34/0.8969	28.83/0.7883	27.68/0.7410	26.13/0.7884	-
DBPN	32.47/0.8983	28.75/0.7859	27.67/0.7396	26.38/0.7947	30.90/0.9134
RDN	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028	31.00/0.9151
EDSR	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033	31.03/0.9149
DRN	32.61/0.8974	28.89/0.7876	27.74/0.7389	26.71/0.8038	31.33/0.9150
RCAN	32.63/0.9002	28.87/0.7889	27.77/0.7436	26.82/0.8087	31.22/0.9173
ours	32.65/0.8991	28.90/0.7869	27.73/0.7416	26.91/0.8088	31.34/0.9167

5.1. Ablation experiments

We designed ablation experiments to verify the effectiveness of the proposed modules. The experimental results for the two modules designed based on NSST are shown in Table 2. Using Set5 as the test set, PSNR as the measurement metric and IMDN as the base, the addition of the NSSB module was able to boost 0.17 dB, the addition of the two-branch strategy and the directional texture enhancement module boosted 0.04 dB and 0.12dB. DTEM* in the table indicates that the M of DTEM is set to 4. As can be seen from the results, with the addition of our two proposed NSST-based texture enhancement blocks, the directionality of the textures is exploited, which in turn effectively enhances the reconstruction of the model.

Table 2: Ablation study of the proposed NSSTNet.

NSSB	✘	✔	✔	✔	✔
Double branch	✘	✘	✔	✔	✔
DTEM	✘	✘	✘	✔	✔
DTEM*	✘	✘	✘	✘	✔
PSNR	32.04	32.21	32.25	32.37	32.41

In addition, we also verified the effect of the frequency domain branching basic block IM-VapB, as shown in Table 3, the addition of the skip-connected SA and VapB parts improved 0.07dB and 0.1dB respectively, demonstrating that the used IM-VapB can better extract texture features.

Table 3: Ablation study of the proposed IM-VapB.

SA	✘	✔	✔
VapB	✘	✘	✔
PSNR	32.04	32.11	32.21

6. Conclusion

In this paper, we develop a single image super-resolution network based on the NSST. The proposed NSSTNet is a two-branch network consisting of a spatial branch and a frequency branch, with the spatial branch extracting spatial domain features and the frequency branch reconstructing the image from the frequency domain perspective. We use a fusion structure to fuse the two domain features to make full use of the spatial and frequency domain information of the degraded image. To enhance the reconstruction of the texture part of the image, we elaborate two image texture enhancement modules using the property that NSST can preserve the directionality of the image texture, and embed them into the spatial domain branch of the network model to improve the image super-resolution performance. Experimental results show that the method can achieve better performance than previous SR methods.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China under Grant No. 61901117, A21EKYN00884B03, in part by the Natural Science Foundation of Fujian Province under Grant No. 2023J01083.

References

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017.

Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European conference on computer vision (ECCV)*, pages 252–268, 2018.

Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012.

Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22367–22377, 2023.

Xiaojie Chu, Liangyu Chen, and Wenqing Yu. Nafssr: Stereo image super-resolution using nafnet. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1239–1248, June 2022.

Marcos V Conde, Ui-Jin Choi, Maxime Burchi, and Radu Timofte. Swin2SR: Swinv2 transformer for compressed image super-resolution and restoration. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2022.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.

Glenn Easley, Demetrio Labate, and Wang-Q Lim. Sparse directional image representations using the discrete shearlet transform. *Applied and Computational Harmonic Analysis*, 25 (1):25–46, 2008.

Alireza Esmaeilzehi, M. Omair Ahmad, and M.N.S. Swamy. Srnssi: A deep light-weight network for single image super resolution using spatial and spectral information. *IEEE Transactions on Computational Imaging*, 7:409–421, 2021a. doi: 10.1109/TCI.2021.3070522.

Alireza Esmaeilzehi, M Omair Ahmad, and MNS Swamy. Srnssi: a deep light-weight network for single image super resolution using spatial and spectral information. *IEEE Transactions on Computational Imaging*, 7:409–421, 2021b.

Tiantong Guo, Hojjat Seyed Mousavi, and Vishal Monga. Adaptive transform domain image super-resolution via orthogonally regularized deep networks. *IEEE transactions on image processing*, 28(9):4685–4700, 2019.

Yong Guo, Jian Chen, Jingdong Wang, Qi Chen, Jiezhang Cao, Zeshuai Deng, Yanwu Xu, and Mingkui Tan. Closed-loop matters: Dual regression networks for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5407–5416, 2020.

Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018.

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015.

Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In *Proceedings of the 27th acm international conference on multimedia*, pages 2024–2032, 2019.

Gu Jinjin, Cai Haoming, Chen Haoyu, Ye Xiaoxing, Jimmy S Ren, and Dong Chao. Pipal: a large-scale image quality assessment dataset for perceptual image restoration. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XI 16*, pages 633–651. Springer, 2020.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017.

Junxuan Li, Shaodi You, and Antonio Robles-Kelly. A frequency domain neural network for fast image super-resolution. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844, 2021.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

Yu-Nan Liu, Shan-Shan Zhang, Yu Sang, and Si-Miao Wang. Improving image retrieval by integrating shape and texture features. *Multimedia Tools and Applications*, 78:2525–2550, 2019.

Yunan Liu, Shanshan Zhang, Jie Xu, Jian Yang, and Yu-Wing Tai. An accurate and lightweight method for human body image super-resolution. *IEEE Transactions on Image Processing*, 30:2888–2897, 2021.

David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. 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*, volume 2, pages 416–423. IEEE, 2001.

Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76:21811–21838, 2017.

Chenyang Wang, Junjun Jiang, Zhiwei Zhong, and Xianming Liu. Spatial-frequency mutual learning for face super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22356–22366, 2023.

Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7794–7803, 2018.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Jingwei Xin, Jie Li, Xinrui Jiang, Nannan Wang, Heng Huang, and Xinbo Gao. Wavelet-based dual recursive network for image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2):707–720, 2020.

Ruyu Xu, Xuejing Kang, Chunxiao Li, Hong Chen, and Anlong Ming. Dct-fanet: Dct based frequency attention network for single image super-resolution. *Displays*, 74:102220, 2022.

Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*, pages 711–730. Springer, 2012.

Dafeng Zhang, Feiyu Huang, Shizhuo Liu, Xiaobing Wang, and Zhezhu Jin. Swinfir: Revisiting the swinir with fast fourier convolution and improved training for image super-resolution. *ArXiv*, abs/2208.11247, 2022.

Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018a.

Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018b.

Hengyuan Zhao, Xiangtao Kong, Jingwen He, Yu Qiao, and Chao Dong. Efficient image super-resolution using pixel attention. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pages 56–72. Springer, 2020.

Lin Zhou, Haoming Cai, Jinjin Gu, Zheyuan Li, Yingqi Liu, Xiangyu Chen, Yu Qiao, and Chao Dong. Efficient image super-resolution using vast-receptive-field attention. *arXiv preprint arXiv:2210.05960*, 2022.

Wenbin Zou, Liang Chen, Yi Wu, Yunchen Zhang, Yuxiang Xu, and Jun Shao. Joint wavelet sub-bands guided network for single image super-resolution. *IEEE Transactions on Multimedia*, 2022.