mathematics

MDPI

Article

Single Image Super-Resolution Reconstruction with Preservation of Structure and Texture Details

Yafei Zhang [1], Yuqing Huang [1], Kaizheng Wang [2,*], Guanqiu Qi [3], Jinting Zhu [4]

[1] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China
[2] Faculty of Electrical Engineering, Kunming University of Science and Technology, Kunming 650500, China
[3] Department of Computer Information Systems, State University of New York at Buffalo State, Buffalo, NY 14222, USA
[4] School of Natural and Computational Sciences, Massey University at Auckland, Auckland 0632, New Zealand
* Correspondence: kz.wang@foxmail.com or kz.wang@kust.edu.cn

Abstract: In recent years, deep-learning-based single image super-resolution reconstruction has achieved good performance. However, most existing methods pursue a high peak signal-to-noise ratio (PSNR), while ignoring the quality of the structure and texture details, resulting in unsatisfactory performance of the reconstruction results in terms of human subjective perception. To solve this issue, this paper proposes a structure- and texture-preserving image super-resolution reconstruction method. Specifically, two different network branches are used to extract features for image structure and texture details. A dual-coordinate direction perception attention (DCDPA) mechanism is designed to highlight structure and texture features. The attention mechanism fully considers the complementarity and directionality of multi-scale image features and effectively avoids information loss and possible distortion of image structure and texture details during image reconstruction. Additionally, a cross-fusion mechanism is designed to comprehensively utilize structure and texture information for super-resolution image reconstruction, which effectively integrates the structure and texture details extracted by the two branch networks. Extensive experiments verify the effectiveness of the proposed method and its superiority over existing methods.

Keywords: deep neural networks; super-resolution image reconstruction; structure and texture details; attention mechanism

MSC: 68U10

Citation: Zhang, Y.; Huang, Y.; Wang, K.; Qi, G.; Zhu, J. Single Image Super-Resolution Reconstruction with Preservation of Structure and Texture Details. *Mathematics* **2023**, *11*, 216. https://doi.org/10.3390/math11010216

Academic Editor: Konstantin Kozlov

Received: 29 November 2022
Revised: 23 December 2022
Accepted: 27 December 2022
Published: 1 January 2023

1. Introduction

Image super-resolution (SR) focuses on recovering the corresponding high-resolution images from degraded low-resolution images. Image SR techniques can be used in face recognition [1], medical imaging [2,3], satellite imagery [4] and so on. As an important image-processing method, image SR has received extensive attention from researchers. Existing SR methods can be roughly divided into four categories: interpolation-based methods [5–8], inverse reconstruction-based methods [9–12], traditional machine-learning-based methods [13–22], and deep-learning-based methods [23–31].

Interpolation-based methods use known adjacent pixel information to generate unknown pixels. This type of method is simple to implement, but its detail recovery ability is poor, and the reconstruction results are prone to blur and aliasing. Assuming that the low-resolution image is obtained by a series of degradations of the high-resolution image, inverse reconstruction-based methods use the reverse reconstruction algorithm to restore the high-resolution image. Based on the degradation model of the image, this kind of method is often developed in combination with prior knowledge of the image [9,11] and obtains a

high-resolution image by optimizing the model. Although this kind of method preserves more image details, the reconstruction performance is susceptible to image degradation and regularization performance. Additionally, when the resolution is improved by more than four times, the reconstruction performance is usually not good. Traditional learning-based methods use the training data to learn the non-linear mapping relationship between a low-resolution image and its high-resolution version. Then, the learned mapping relationship is used to guide high-resolution image reconstruction. Freeman et al. [32] first proposed the learning of this relationship in Markov neural networks, but this method requires a lot of time to construct a training set and image reconstruction is also time-consuming. Chang et al. [33] assumed that both high- and low-resolution image patches have manifolds with the same geometric structure and proposed a super-resolution reconstruction method based on neighborhood embedding. Since dictionary learning has shown excellent performance in computer vision [14,34] and medical imaging [35,36], Yang et al. [37] proposed a super-resolution reconstruction method based on sparse representation. This method assumes that the high- and low-resolution images have the same coding coefficients under the respective dictionary, and reconstructs the high-resolution image by applying the coding coefficients of the low-resolution image under the low-resolution dictionary to the high-resolution dictionary. In view of the effectiveness of sparse representation and its superiority for image processing [38–41] and recognition [42–44], image super-resolution reconstruction based on sparse representation became a very popular method at that time. Although this type of method has achieved breakthroughs in performance, the optimization process of the model is too complicated, time-consuming and memory-intensive.

In 2014, Dong et al. [17] first applied deep learning to image super-resolution reconstruction and proposed super-resolution convolutional neural networks (SRCNN). Since then, deep-learning-based methods have become the mainstream methods for image super-resolution reconstruction [23–31,45,46]. This kind of method usually uses a single-branch network under the constraint of loss function so that the network can extract the texture details required to restore the low-resolution image to the high-resolution image, thereby effectively improving the visual effect of the reconstructed image. However, texture features are easily lost in the process of image convolution in real cases. This is extremely disadvantageous for recovering texture details in high-resolution images. In order to improve the visual performance of reconstructed images, Ledig et al. [13] proposed super-resolution generative adversarial networks (SRGAN), which use the generative adversarial loss to constrain the generator and discriminator. Hence, the generator obtains texture details that are closer to the real world. Since then, Wang et al. [47] further improved the detail recovery ability of SRGAN by introducing dense residual blocks. However, the discriminator may introduce instability in the optimization process during the adversarial process, causing structure inconsistency in the reconstructed image, resulting in unnatural artifacts and geometric distortions in the recovered super-resolution image. Therefore, Ma et al. [24] proposed a structure-preserving super-resolution (SPSR) method to alleviate the above problem, which uses the gradient information of the image to guide the low-resolution image restoration, while maintaining the advantages of GAN-based methods to generate results that achieve satisfactory performance in terms of human subjective perception.

In order to reconstruct the structure and texture details of images with high quality, this paper proposes a structure- and texture-preserving image super-resolution method. In the proposed method, the structure and texture details of the image are extracted by different network branches. In this process, a dual-coordinate direction perception attention (DCDPA) mechanism is designed to highlight structure and texture features. This mechanism not only makes full use of the complementarity of multi-scale image features, but also considers the directionality of image features in the horizontal and vertical coordinates. Therefore, both information loss and possible distortion of the image structure and texture details are effectively avoided during image reconstruction. To comprehensively utilize the structure and texture information to reconstruct high-resolution images, this paper proposes a cross-fusion mechanism to integrate structure and texture features.

In the network structure, the proposed network framework is mainly composed of a structure feature extraction module (SFEM), a texture detail feature extraction module (TDFEM), and feature highlight and aggregation modules (FHAM). The SFEM and TDFEM are responsible for the learning of image structure features and texture detail features, respectively. The FHAM enhances the structure and texture detail features in different directions in the image through the designed DCDPA mechanism and promotes information interaction between them through the cross fusion of the structure and texture detail features to improve the feature representation ability. The network is optimized by reconstruction loss, texture loss, and gradient loss. This paper makes three main contributions as follows:

- An image super-resolution reconstruction network for structure and texture restoration is designed. This network uses the dual-stream structure to extract the structure and texture details of the image, respectively, and enhances these two features layer-by-layer through the FHAM to improve the network's ability to extract features.
- The FHAM is designed to highlight the structure and texture features in the image. In order to enable the network to capture the salient structure and texture features in the image, this paper proposes a DCDPA mechanism in this module. This mechanism makes full use of the directionality of image features to highlight the features of different directions in the image. Moreover, a cross-fusion block (CFB) is proposed to promote the interactive fusion of structure and texture features.
- A large number of experimental results show that the proposed method contains salient structure and texture details in its super-resolution reconstruction results, which outperform state-of-the-art methods in visual quality and performance evaluation.

2. Relate Work

2.1. Deep-Learning-Based Super-Resolution Reconstruction Methods

In recent years, deep learning has been widely used in single-image super-resolution reconstruction. Dong et al. [17] proposed the SRCNN. Compared with traditional methods, SRCNN achieves better super-resolution reconstruction results. Since simply stacking convolutional layers can lead to gradient explosion, Kim et al. [22] proposed an accurate image super-resolution using very deep convolutional networks (VDSR). The networks avoid the problem of gradient explosion and speed up the network training by adding jump connections. DRRN [48] and Memnet [49] proposed by Tai et al. introduced recursive blocks and memory blocks, respectively, to stack more convolutional layers. However, the above-mentioned networks all use predefined upsampling. Due to resolution changes using traditional methods, the network complexity increases. In addition, since the traditional upsampling methods do not involve network parameter training, the image cannot be reconstructed by convolution, activation function, etc. Therefore, the upsampling process magnifies the defects in low-resolution images. To solve this problem, Dong et al. [25] proposed a fast convolutional neural-network-based super-resolution reconstruction (FSRCNN) method, which performs feature extraction on low-resolution images, reduces the complexity of the network, and applies deconvolution to upsampling and image reconstruction. Shi et al. [26] proposed an efficient sub-pixel convolutional layer to achieve upsampling. In recent years, Zhu et al. [27] proposed a lightweight feature separation and fusion network, adopting a block-processing structural framework, which enables the network to fully utilize and enrich features at different levels. Liu et al. [28] proposed a multi-scale skip-connection network (MSN) to improve the visual quality of super-resolution images, utilizing convolution kernels of different sizes to capture multi-scale features of low-resolution images. Huang et al. [29] proposed an attention network with detail fidelity, which achieves super-resolution image reconstruction while obtaining detail fidelity. Mehri et al. [30] proposed the multi-path residual network (MPRNet), which adaptively learns the most valuable features to learn more high-frequency information about the image. Tan et al. [31] proposed an image super-resolution method via a self-calibrated feature fusion network (SCFFN) to achieve a better balance between network performance

and applicability. However, the above methods all consider the texture feature information and structure feature information of the image as a whole. In reality, the texture features are easily lost during the image convolution process. This is extremely disadvantageous for reconstructing high-resolution images rich in texture feature information.

2.2. Visual-Perception-Based Super-Resolution Reconstruction Methods

To improve the perceptual quality of images, in recent years, a large number of super-resolution reconstruction models that benefit the perceptual performance of human eyes have been proposed [13,47,50]. These methods use a variety of information compensation methods to compensate the texture features of the super-resolution reconstruction results. While reconstructing the structure information, they prevent the loss of texture details during the super-resolution reconstruction process and improve the perceptual performance of images. A photo-realistic single image super-resolution using a generative adversarial network (SRGAN) proposed by Ledig et al. [13] uses the discriminator and the generator for adversarial learning and enables the generator to generate texture details that are close to the real world. In order to reduce the artifacts caused by generative confrontation in SRGAN, Wang et al. [47] proposed an enhanced super-resolution generative adversarial network (ESRGAN), which introduced dense residual blocks into the basic network framework and abandoned the batch normalization operation to generate more realistic and natural images. Wu et al. [51] proposed a perceptual generative adversarial network for single-image super-resolution (SRPGAN), in which a robust perceptual loss based on the model's discriminator was proposed to recover texture details. Wang et al. [50] proposed an image super-resolution reconstruction method for texture recovery by deep spatial feature transform, which integrates a semantic prior into low-resolution images and uses a spatial feature transform (SFT) layer to further improve texture recovery. A new dense block proposed by Chen et al. [52] uses complex connections between each layer to build a more powerful generator and applies new feature maps to compute the perceptual loss to make the output image more realistic and natural. Ma et al. [24] proposed a structure-preserving image super-resolution method, which uses the gradient information of the image to guide image recovery. Fu et al. [53] proposed a method to guide the image super-resolution reconstruction by weak texture information. This network consists of a main network and two auxiliary prediction networks. The main network extracts and combines the main distinct depth features to assist the prediction network in extracting weak texture information. Cai et al. [54] proposed a texture and detail preservation network (TDPN), which not only focuses on local region feature recovery, but also texture and detail preservation. Meng et al. [55] proposed a gradient information distillation network that not only maintains the advantage of fast and lightweight, but also improves the model performance through information distillation. All of the above-mentioned methods have made great contributions to improving the visual perception of the reconstructed images. However, unnatural artefacts still appear in the recovered super-resolution images. Therefore, this paper proposes a super-resolution reconstruction network that fuses structure and texture information. By enhancing the structure and texture features in the image and cross-fusion, the extraction ability of the network for structure and texture features is improved, thereby reducing artefacts in the reconstructed image and improving the perceptual quality of the reconstructed image.

3. The Proposed Network Framework

The framework of the proposed image super-resolution reconstruction network for structure and texture detail recovery is shown in Figure 1. The model mainly comprises three modules: SFEM, TDFEM, and FHAM. The network structure of both SFEM and TDFEM is the same, but the parameters are not shared. They both consist of a 5×5 convolutional layer and five sets of residual blocks (RBs). Convolutional layers and RBs extract shallow and deep features of the input image, respectively. The input of the SFEM is the original low-resolution image. Additionally, the texture map of the original low-

resolution image is obtained by the local binary pattern (LBP) method [56]; the texture map and the original image are concatenated as the input of the TDFEM. The FHAM consists of five gated fusion blocks (GFBs). The GFB enhances the structure and texture information of the image, respectively, by introducing a DCDPA mechanism and fusing the enhanced structure and texture information through cross-fusion. Finally, the reconstructed super-resolution image is obtained by deconvolution.

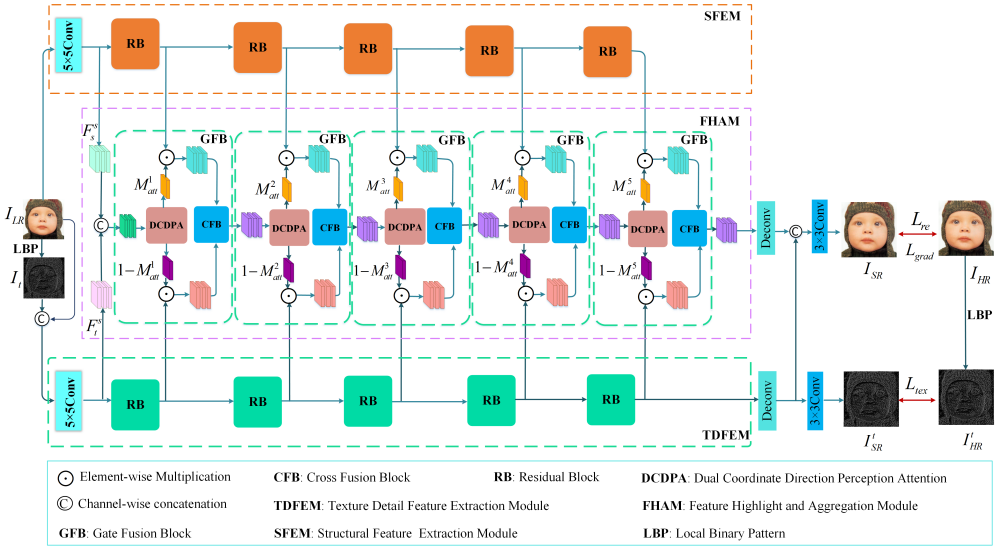

Figure 1. The framework of the proposed network.

3.1. Structural Feature Extraction Module and Texture Detail Feature Extraction Module

The network structure of both SFEM and TDFEM is the same, including shallow-feature extraction and deep-feature extraction. The network adopts a 5×5 convolution layer and five RBs to extract the shallow and deep features of the image, respectively. Assuming that $\boldsymbol{I}$ represents the input of the 5×5 convolutional layer, the shallow feature $\boldsymbol{F}^s$ obtained by convolution is:

$$\boldsymbol{F}^s = sconv(\boldsymbol{I}, k = 5) \tag{1}$$

where $sconv$ represents equal-sized convolution operations, and k represents the size of the convolution kernel.

Let $\boldsymbol{I}_{LR}$ denote the input original low-resolution image. For the SFEM $\boldsymbol{I} = \boldsymbol{I}_{LR}$, the obtained shallow structure feature is $\boldsymbol{F}_s^s$. In order to prevent the subjective image quality degradation caused by loss of texture detail information during the super-resolution reconstruction process, the LBP method is used to obtain the texture map $\boldsymbol{I}_t$ of the original image $\boldsymbol{I}_{LR}$ and to concatenate them with the input of the TDFEM. Therefore, for the TDFEM, $\boldsymbol{I} = concat(\boldsymbol{I}_{LR}, \boldsymbol{I}_t)$, where $concat$ denotes the concatenation operation. The obtained shallow texture detail feature is $\boldsymbol{F}_t^s$. The RB is used to extract the deep features of the image. The relationship between the input and output of the i-th ($i = 1, 2, \ldots, 5$) RB can be expressed as:

$$\boldsymbol{F}_r^i = sconv_g(\boldsymbol{F}_r^{i-1}) + \boldsymbol{F}_r^{i-1} \tag{2}$$

where $sconv_g$ represents a convolution group, which consists of six 3×3 convolutional layers. $\boldsymbol{F}_r^{i-1}$ and $\boldsymbol{F}_r^i$ represent the input and output of the i-th RB, respectively.

3.2. Feature Highlight and Aggregation Module and Feature Reconstruction

To ensure that the network can realize the fusion of image structure and texture features and prevent the loss of structure and texture details, this paper proposes the FHAM. The module consists of five GFBs, as shown in Figure 1. The GFB is composed of DCDPA and CFB.

3.2.1. Dual-Coordinate Direction Perception Attention

To enable the network to capture structure and texture details in different directions from the image, a DCDPA mechanism is used, which includes direction perception coordinate attention (DPCA) and global perception coordinate attention (GPCA), as shown in Figure 2.

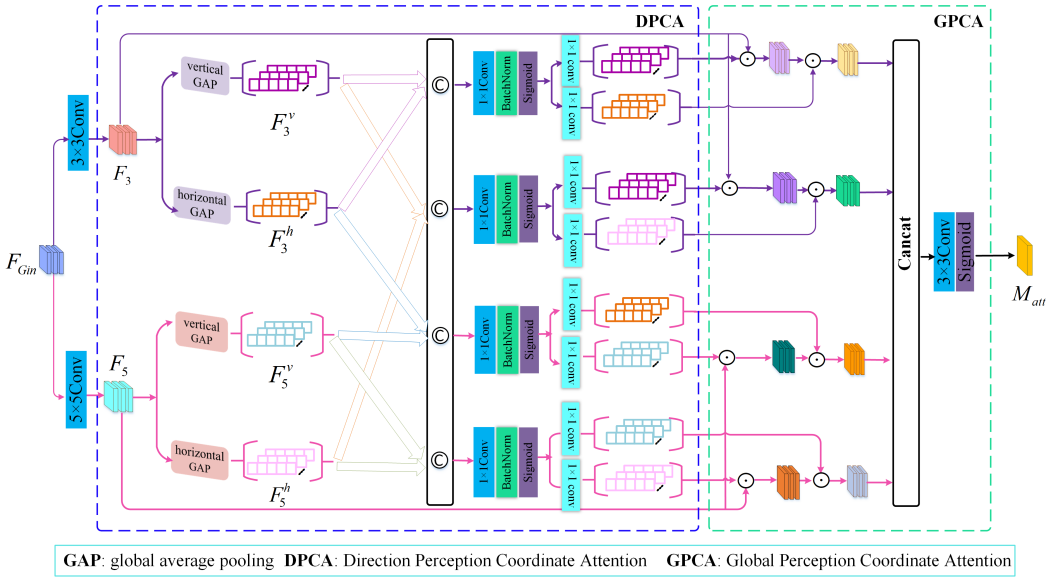

Figure 2. Dual-coordinate direction perception attention.

The input F_{Gin} of the DCDPA module contains the texture detail and structure features of the image. Specifically, for the first GFB, its input is obtained by concatenating the shallow structure feature F_s^s and the shallow texture detail feature F_t^s. For other GFBs, the corresponding input is the output of the previous GFB. F_{Gin} passes through 3×3 and 5×5 convolutions, respectively, to obtain the features F_l ($l = 3, 5$) of different scales. F_l obtains the horizontal feature F_l^h and the vertical feature F_l^v through global average pooling (GAP) in the horizontal and vertical directions, respectively:

$$\begin{aligned} F_l^h &= P_h(F_l) \\ F_l^v &= P_v(F_l) \end{aligned} \tag{3}$$

where $l = 3, 5$, and P_h and P_v represent the GAP in the horizontal and vertical directions, respectively. Set $F_l \in \mathbb{R}^{3\times H\times W}$, then $F_l^v \in \mathbb{R}^{3\times 1\times W}$, $F_l^h \in \mathbb{R}^{3\times H\times 1}$. To concatenate features in different directions, it is necessary to reshape F_l^h into $3 \times 1 \times H$ dimensions. The features in different directions are concatenated along the horizontal direction to obtain four sets of joint features as follows:

$$\begin{aligned} F_{3,3}^{hv} &= concat(F_3^h, F_3^v) \\ F_{3,5}^{hv} &= concat(F_3^h, F_5^v) \\ F_{5,3}^{hv} &= concat(F_5^h, F_3^v) \\ F_{5,5}^{hv} &= concat(F_5^h, F_5^v) \end{aligned} \tag{4}$$

where *concat* represents the concatenation operation. To obtain the perceptual coordinate attention weights in different directions, 1×1 convolution and *sigmoid* are applied to process the joint features to obtain attention maps in different directions. Taking $F_{3,3}^{hv}$ as an

example, the attention maps in the horizontal and vertical directions obtained after $\boldsymbol{F}_{3,3}^{hv}$ is processed are $\boldsymbol{M}_{3,3}^{h} \in \mathbb{R}^{3\times1\times W}$ and $\boldsymbol{M}_{3,3}^{v} \in \mathbb{R}^{3\times1\times W}$:

$$\left[\boldsymbol{M}_{3,3}^{h}, \boldsymbol{M}_{3,3}^{v}\right] = split(sigmoid(BN(sconv(\boldsymbol{F}_{3,3}^{hv}, k=1)))) \tag{5}$$

where *split* represents the segmentation operation, *sigmoid* represents the sigmoid activation function, and *BN* represents the batch normalization layer. After applying 1×1 convolution and dimension expansion in the horizontal and vertical directions to $\boldsymbol{M}_{3,3}^{h}$ and $\boldsymbol{M}_{3,3}^{v}$, the perceptual coordinate attention maps $\tilde{\boldsymbol{M}}_{3,3}^{h} \in \mathbb{R}^{3\times W\times H}$ and $\tilde{\boldsymbol{M}}_{3,3}^{v} \in \mathbb{R}^{3\times W\times H}$ in different directions are obtained, respectively. Similarly, other attention maps $\left[\tilde{\boldsymbol{M}}_{3,5}^{h}, \tilde{\boldsymbol{M}}_{3,5}^{v}\right]$, $\left[\tilde{\boldsymbol{M}}_{5,3}^{h}, \tilde{\boldsymbol{M}}_{5,3}^{v}\right]$, and $\left[\tilde{\boldsymbol{M}}_{5,5}^{h}, \tilde{\boldsymbol{M}}_{5,5}^{v}\right]$, can be obtained.

The features in different directions in an image are enhanced using DPCA in both the vertical and horizontal directions. $\boldsymbol{F}_3$ is enhanced with attention maps $\left[\tilde{\boldsymbol{M}}_{3,3}^{h}, \tilde{\boldsymbol{M}}_{3,3}^{v}\right]$ and $\left[\tilde{\boldsymbol{M}}_{3,5}^{h}, \tilde{\boldsymbol{M}}_{3,5}^{v}\right]$, respectively, and the enhanced features are $\tilde{\boldsymbol{F}}_{3,3}$ and $\tilde{\boldsymbol{F}}_{3,5}$:

$$\begin{aligned} \tilde{\boldsymbol{F}}_{3,3} &= \boldsymbol{F}_3 \odot \tilde{\boldsymbol{M}}_{3,3}^{h} \odot \tilde{\boldsymbol{M}}_{3,3}^{v} \\ \tilde{\boldsymbol{F}}_{3,5} &= \boldsymbol{F}_3 \odot \tilde{\boldsymbol{M}}_{3,5}^{h} \odot \tilde{\boldsymbol{M}}_{3,5}^{v} \end{aligned} \tag{6}$$

where $\odot$ denotes point-wise multiplication.

Similarly, $\boldsymbol{F}_5$ is enhanced with attention maps $\left[\tilde{\boldsymbol{M}}_{5,3}^{h}, \tilde{\boldsymbol{M}}_{5,3}^{v}\right]$ and $\left[\tilde{\boldsymbol{M}}_{5,5}^{h}, \tilde{\boldsymbol{M}}_{5,5}^{v}\right]$, respectively, and the enhanced features are $\tilde{\boldsymbol{F}}_{5,3}$ and $\tilde{\boldsymbol{F}}_{5,5}$, respectively. The GPCA map $\boldsymbol{M}_{att}$ is generated through the concatenation, convolution and sigmoid activation function of the features $\tilde{\boldsymbol{F}}_{3,3}$, $\tilde{\boldsymbol{F}}_{3,5}$, $\tilde{\boldsymbol{F}}_{5,3}$ and $\tilde{\boldsymbol{F}}_{5,5}$:

$$\boldsymbol{M}_{att} = sigmoid(sconv(concat(\tilde{\boldsymbol{F}}_{3,3}, \tilde{\boldsymbol{F}}_{3,5}, \tilde{\boldsymbol{F}}_{5,3}, \tilde{\boldsymbol{F}}_{5,5}))) \tag{7}$$

The output $\boldsymbol{F}_s^i$ of the i-th RB of the structure feature extraction module is enhanced by the GPCA map $\boldsymbol{M}_{att}^i$ ($i = 1, 2, \ldots, 5$) in the i-th GFB, and the output $\boldsymbol{F}_t^i$ of the i-th RB of the TDFEM is enhanced by the complementary attention map $1 - \boldsymbol{M}_{att}^i$. The enhanced structure feature and texture feature are $\boldsymbol{F}_{se}^i$ and $\boldsymbol{F}_{te}^i$, respectively, as follows:

$$\begin{aligned} \boldsymbol{F}_{se}^i &= \boldsymbol{F}_s^i \odot \boldsymbol{M}_{att}^i \\ \boldsymbol{F}_{te}^i &= \boldsymbol{F}_t^i \odot (1 - \boldsymbol{M}_{att}^i) \end{aligned} \tag{8}$$

3.2.2. Cross-Fusion Block

To make full use of the structure and texture detail information enhanced by the DCDPA mechanism, a cross-fusion block (CFB) is proposed, as shown in Figure 3. Specifically, the structural feature $\boldsymbol{F}_{se}$ and texture feature $\boldsymbol{F}_{te}$ generate the spatial attention map $\boldsymbol{M}_s$ through concatenation, convolution and the sigmoid:

$$\boldsymbol{M}_s = sigmoid(sconv(concat(\boldsymbol{F}_{se}, \boldsymbol{F}_{te}), k=3)) \tag{9}$$

Using $\boldsymbol{M}_s$ to further enhance the structure features, $\boldsymbol{F}_{st}$ is obtained by the cross-fusion of the enhanced structure features and texture features. Similarly, $\boldsymbol{F}_{ts}$ is obtained by the cross-fusion of the structure features and the enhanced texture results by $\boldsymbol{M}_s$:

$$\begin{aligned} \boldsymbol{F}_{st} &= \boldsymbol{F}_{se} \odot \boldsymbol{M}_s + \boldsymbol{F}_{te} \\ \boldsymbol{F}_{ts} &= \boldsymbol{F}_{te} \odot \boldsymbol{M}_s + \boldsymbol{F}_{se} \end{aligned} \tag{10}$$

The final fusion result is obtained by the concatenation of $\boldsymbol{F}_{st}$ and $\boldsymbol{F}_{ts}$:

$$\boldsymbol{F}_{Gout} = concat(\boldsymbol{F}_{st}, \boldsymbol{F}_{ts}) \tag{11}$$

The fusion result F^i_{Gout} of the i-th GFB is used as the input of the $i+1$-th GFB, i.e., $F^{i+1}_{Gin} = F^i_{Gout}$.

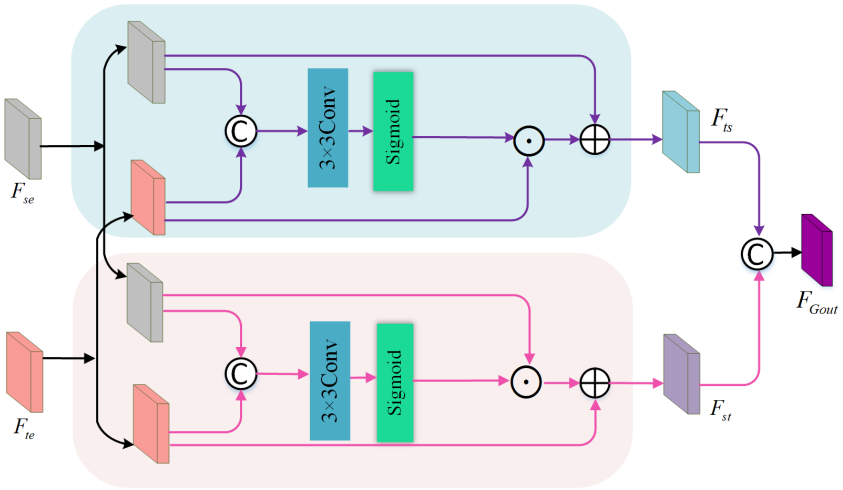

Figure 3. Cross-fusion block.

3.2.3. Feature Reconstruction

To obtain the reconstructed image, the output F^d_{Gout} of the last layer in the FHAM and the output F^d_t of the last layer in the TDFEM are upsampled by deconvolution to obtain F^h and F^h_t, respectively. The texture detail feature map F^h_t is concatenated with F^h to obtain features with further enhanced details; the high-resolution image I_{SR} is finally reconstructed by 3×3 convolution:

$$\begin{aligned} F^h &= Deconv(F^d_{Gout}) \\ F^h_t &= Deconv(F^d_t) \\ I_{SR} &= sconv(concat(F^h, F^h_t), k = 3) \end{aligned} \tag{12}$$

where *Deconv* stands for deconvolution.

3.3. Loss Function

The model is trained by a joint loss including reconstruction loss, gradient loss, and texture loss to achieve good visual performance.

Reconstruction loss: In this paper, the labeled high-resolution image, and the high-resolution image after super-resolution reconstruction, are constrained by the reconstruction loss as follows:

$$L_{re} = \|I_{HR} - I_{SR}\|_1 \tag{13}$$

where I_{HR} and I_{SR} are the labeled high-resolution image and reconstructed super-resolution image, respectively.

Gradient loss: To ensure that the reconstruction result has the same gradient information as the labeled image, the following gradient loss function to further optimize the reconstructed image is used.

$$L_{grad} = \frac{1}{H \times W} \|\nabla I_{HR} - \nabla I_{SR}\|_1 \tag{14}$$

where ∇ is the gradient operator and H and W represent the height and width of the image, respectively.

Texture loss: To make the texture features in the reconstructed image and the labeled image consistent, a texture loss is introduced as follows:

$$L_{tex} = \|I^t_{HR} - I^t_{SR}\|_1 \tag{15}$$

where I_{HR}^{t} and I_{SR}^{t} are the texture image extracted from the labeled high-resolution image by the LBP operator and the texture image reconstructed by the texture detail feature extraction module, respectively. So, the joint loss is summarized as follows:

$$L_{joint} = L_{re} + \lambda_{grad} L_{grad} + \lambda_{tex} L_{tex} \tag{16}$$

where λ_{grad} and λ_{tex} represent the weight parameters.

4. Experiment and Result Analysis

4.1. Experimental Setup

The Pytorch 1.7 framework and NVIDIA GeForce RTX 2080ti 12 GB GPU were used in the experiments. The training images were from the DIV2K dataset [57], which includes 800 training images, 100 validation images, and 100 test images. Four benchmark datasets, Set5 [58], Set14 [59], B100 [60], Urban100 [61], were used to evaluate the performance of the different methods. The Set5, Set14, and B100 datasets contained 5, 10, and 100 natural images, respectively. There were 100 urban building images in the Urban100 dataset. The high-resolution testing images in the testing set were downsampled by 2×, 3× and 4× bicubic interpolation to obtain the corresponding low-resolution images for super-resolution reconstruction testing. When training the network model with a 2× upsampling factor, to reduce memory usage and save running time, the size of the high-resolution image was 256 × 256 and the size of the low-resolution image was 128 × 128. When training the network model with a 3× upsampling factor, the size of the high-resolution image was 192 × 192, and the size of the low-resolution image was 64 × 64. When training the network model with a 4× upsampling factor, the size of the high-resolution image was 180 × 180 and the size of the low-resolution image was 45 × 45. In the training process, the Adam optimizer [62] was selected to train all modules. Following [54], the learning rate of the model training was set to 10^{-4}. The convergence curves of the training loss during the training for scale 2 magnification on the DIV2K dataset are shown in Figure 4. It can be seen from Figure 4 that the training loss gradually decreased as the training epochs increased. When it reached 1800 epochs, the loss curve converged. Allowing for some margin, the epoch for the model training was set to 2000. The parameters of the entire network were trained through back propagation.

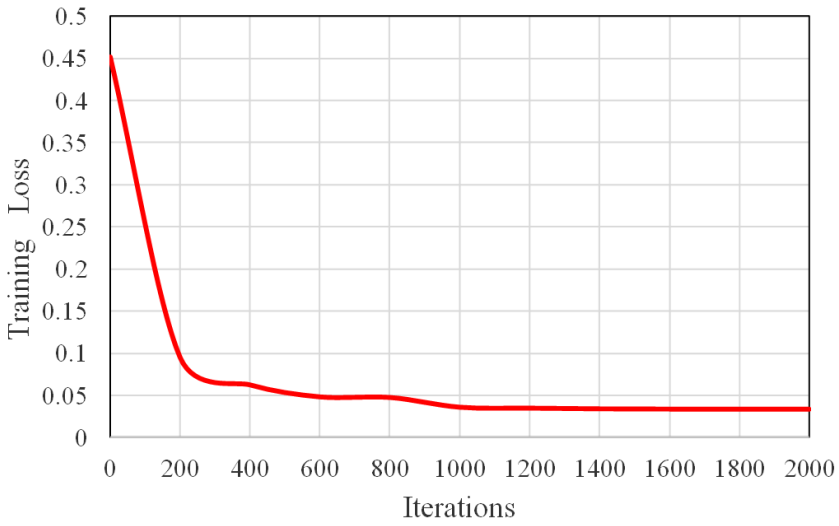

Figure 4. Training loss curve.

Two objective evaluation indicators, the peak signal-to-noise ratio (PSNR) and a structural similarity index measure (SSIM) were used to evaluate the quality of the super-resolution reconstruction results. The PSNR was used to measure the mean-square error between the reconstructed image and the original image; its unit is dB. The larger the value

of PSNR, the less the image distortion, that is, the higher the image reconstruction quality. The SSIM was used to measure the structural similarity between the reconstructed image and the original image. A higher value of SSIM indicates that the reconstructed image is closer to the original image, that is, it shows better image reconstruction performance.

Given two images, one being the ground truth (GT) $\boldsymbol{X}$, the other the super-resolution reconstruction result $\boldsymbol{Y}$, the PSNR [63] can be defined as:

$$PSNR = 10\log_{10}(\frac{(2^n-1)^2}{MSE}) \tag{17}$$

where MSE is the mean square error of the GT image $\boldsymbol{X}$ and the reconstructed image $\boldsymbol{Y}$. $MSE = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}(\boldsymbol{X}(i,j)-\boldsymbol{Y}(i,j))^2$. H and W are the height and width of the image, respectively. n is the bits of a pixel which is generally set to 8 (i.e., a gray level of 256). The unit of the PSNR is the decibel (dB).

The SSIM [64] is defined as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{18}$$

where μ_x and μ_y are the mean values of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. σ_x and σ_y are the standard deviations of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. σ_{xy} is the covariance of $\boldsymbol{X}$ and $\boldsymbol{Y}$. c_1 and c_2 are set to constants to avoid system error caused by a zero denominator. The value of the SSIM is between 0 and 1.

4.2. Quantitative Comparison

To verify the effectiveness of the proposed method, it was compared with state-of-the-art super-resolution methods, such as SRCNN [65], FSRCNN [25], VDSR [22], LapSRN [66], CDC [67], and SREFBN [68]. SRCNN and FSRCNN were trained on the 91-image dataset [17]. VDSR, LapSRN, CDC and SREFBN were trained on the DIV2K dataset. The above methods were all validated on the Set5, Set14, B100, and Urban100 datasets. The quantitative evaluation results are shown in Tables 1–3 with the best results marked in bold. According to Tables 1–3, the proposed method achieved most of the best evaluation results.

Table 1. Quantitative comparison of 2× super-resolution results obtained by different methods.

Method	Scale	Set5		Set14		B100		Urban100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SRCNN	2	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946
FSRCNN	2	37.05	0.9560	32.66	0.9090	31.53	0.8920	29.88	0.9020
VDSR	2	37.53	0.9590	33.05	0.9130	31.90	0.8960	30.77	0.9140
LapSRN	2	37.52	0.9591	33.08	0.9130	31.08	0.8950	30.41	0.9101
CDC	2	32.35	0.8766	29.20	0.7932	28.58	0.7856	26.06	0.7766
SREFBN	2	37.92	0.9593	33.59	0.9175	32.17	0.8991	32.13	**0.9279**
Proposed	2	**38.07**	**0.9713**	**33.60**	**0.9442**	**32.20**	**0.9036**	**32.19**	0.8549

Table 2. Quantitative comparison of 3× super-resolution results obtained by different methods.

Method	Scale	Set5		Set14		B100		Urban100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SRCNN	3	32.75	0.9090	29.30	0.8215	28.41	0.7863	26.24	0.7989
FSRCNN	3	33.18	0.9140	29.37	0.8240	28.53	0.7910	26.43	0.8080
VDSR	3	33.67	0.9210	29.78	0.8320	28.83	0.7990	27.14	0.8290
LapSRN	3	33.82	0.9227	29.87	0.8320	28.82	0.7980	27.07	0.8280
CDC	3	25.29	0.2966	23.97	0.6224	24.74	0.6165	21.96	0.6087
SREFBN	3	34.33	0.9257	30.28	0.8407	29.06	**0.8038**	28.03	**0.8494**
Proposed	3	**35.12**	**0.9432**	**30.37**	**0.8557**	**29.13**	0.7941	**28.27**	0.7249

Table 3. Quantitative comparison of 4× super-resolution results obtained by different methods.

Method	Scale	Set5		Set14		B100		Urban100	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
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	0.7026	25.18	0.7540
LapSRN	4	31.54	0.8850	28.19	0.7720	27.32	0.7270	25.21	0.7560
CDC	4	25.41	0.7177	24.01	0.6165	24.19	0.6085	21.50	0.5610
SREFBN	4	32.01	0.8917	28.52	**0.7790**	27.50	**0.7332**	25.87	0.7787
Proposed	4	**32.27**	**0.9039**	**28.59**	0.7624	**27.60**	0.7075	**26.13**	**0.7893**

4.3. Qualitative Comparison

To compare the visual quality of the super-resolution reconstruction results obtained by different methods, images from the Set5, Set14, BSD100, and Urban100 datasets were selected for the visual display of the 2×, 3×, and 4× super-resolution reconstruction results and ground truth (GT), as shown in Figures 5–7. In terms of visual performance, the detailed textures of the reconstructed images obtained by the SRCNN, FSRCNN, VDSR, and CDC methods were relatively blurred, while the reconstruction results obtained by the LapSRN, CDC, and SREFBN methods contained detailed features that did not match the actual texture details. The proposed method could fit real natural images well, enhancing the brightness of reconstructed high-resolution images, while recovering textures, so that the reconstructed images had rich texture details. In summary, the proposed method has certain advantages in terms of visual performance and quantitative evaluation.

Figure 5. Visual comparisons with different methods for 2× super-resolution on the Set5, Set14, B100, and Urban100 datasets.

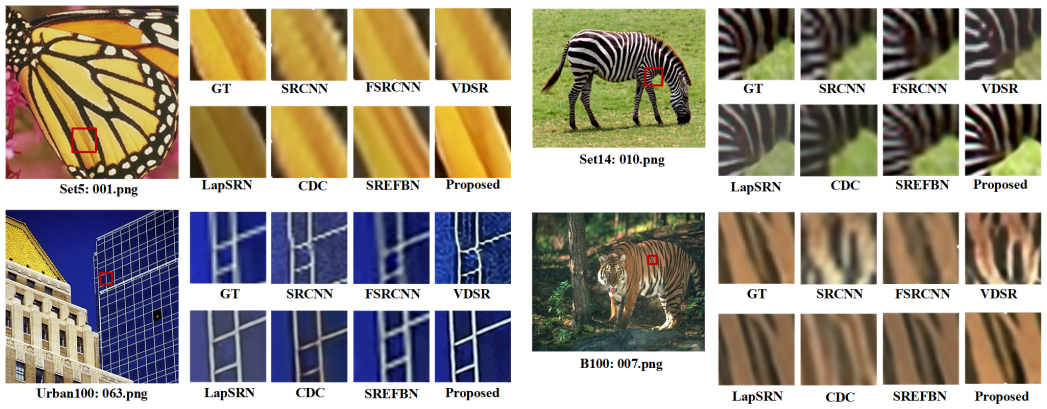

Figure 6. Visual comparisons with different methods for 3× super-resolution on the Set5, Set14, B100, and Urban100 datasets.

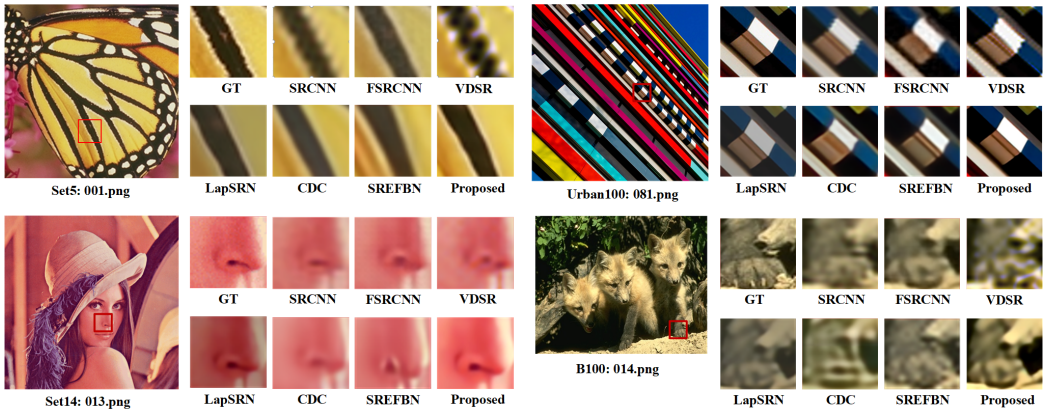

Figure 7. Visual comparisons with different methods for 4× super-resolution on the Set5, Set14, B100, and Urban100 datasets.

4.4. Ablation Study

To verify the effectiveness of the proposed TDFEM, DCDPA, and CFB methods on the quality of the reconstructed super-resolution images, ablation experiments were performed. Taking the network without DCDPA, CFB and TDFEM as the benchmark (Base), the Base network plus TDFEM was marked as Base + TDFEM, the Base + TDFEM network plus DCDPA was marked as Base + TDFEM + DCDPA, and the Base+TDFEM+DCDPA network plus CFB was marked as Base + TDFEM + DCDPA + CFB. Ablation experiments were performed on the images from the Urban100 dataset at 2× magnification. The quantitative evaluation results of the ablation experiments are shown in Table 4 and the corresponding visual performance is shown in Figure 8.

Table 4. Ablation experiment performance.

Method	Base	TDFEM	DCDPA	CFB	PSNR/SSIM
Base	√				31.38/0.8333
Base + TDFEM	√	√			31.51/0.8456
Base + TDFEM + DCDPA	√	√	√		31.85/0.8509
Base + TDFEM + DCDPA + CFB	√	√	√	√	32.19/0.8549

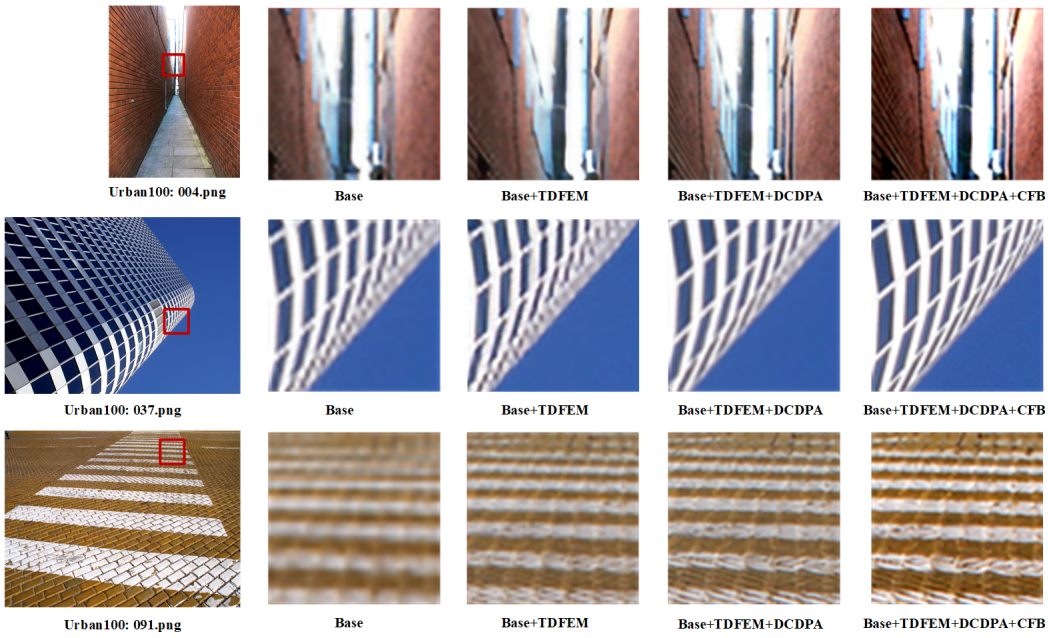

Figure 8. Comparison of ablation experiment results.

4.4.1. The Effectiveness of TDFEM

TDFEM was used to extract texture information separately to prevent the loss of texture detail information in the process of super-resolution reconstruction, which led to the deterioration in subjective image quality. So, TDFEM was introduced into the proposed model. In the TDFEM, the LBP operator is used to extract the texture details of the original image and the obtained texture detail map is concatenated with the original image, which enhances the network's ability to extract texture detail information. This helps to improve the perceptual quality of reconstructed super-resolution images. To demonstrate its effectiveness, the reconstruction results of the 'Base' model were compared with the reconstruction results of the 'Base + TDFEM' model. As shown in Figure 8, the 'Base + TDFEM' model achieved good performance in texture detail extraction during the reconstruction process. Additionally, as shown in Table 4, the utilization of 'TDFEM' also led to a significant improvement in the objective evaluation results, which confirmed the effectiveness of 'TDFEM' in the proposed model.

4.4.2. The Effectiveness of DCDPA

In the proposed model, DCDPA is used to capture the structure and texture feature regions of interest and to prevent the loss of feature information. This attention mechanism makes full use of the complementarity and directionality of multi-scale image features and effectively avoids information loss and the possible distortion of image structure and texture details during image reconstruction. This is conducive to improving the visual quality of the reconstructed image. To verify the effectiveness of 'DCDPA', the reconstruction results of the 'Base + TDFEM + DCDPA' model were compared with those of the 'Base + TDFEM' model. In Figure 8, the structure and texture details of the reconstructed images obtained by the 'Base + TDFEM + DCDPA' are highlighted. Moreover, according to the values of the objective evaluation metrics shown in Table 4, compared with 'Base + TDFEM', the performance of 'Base + TDFEM + DCDPA' was improved. So, the effectiveness of DCDPA was demonstrated.

4.4.3. The Effectiveness of CFB

To prevent the loss of texture details and ensure that the texture detail information enhanced by the DCDPA was fully utilized by the later layers, a CFB was proposed. This module facilitates the information interaction between structure and texture detail features by cross-fusion of them to enhance the network's ability to mine structure and texture features. This contributes to improving the quality of the reconstructed image. To verify the effectiveness of 'CFB', the reconstruction results of 'Base + TDFEM + DCDPA + CFB' were compared with those of 'Base + TDFEM + DCDPA'. As shown in Figure 8, the 'Base + TDFEM + DCDPA + CFB' model more fully integrated structure and texture features in the reconstruction process and the corresponding evaluation results presented in Table 4 were also enhanced.

4.5. Parameter Analysis

4.5.1. Influence of the Number of GFBs

This section discusses the influence of the number of GFBs on the performance of the proposed model. The proposed model with different numbers of GFBs was tested on the DIV2K dataset. Table 5 shows the performance of the model with different numbers of GFBs (denoted by n). 'Params' indicates the number of parameters. As shown in Table 5, PSNR and SSIM achieved the best performance at $n = 5$. When n was 7 or 9, the number of model parameters increased and the model's performance decreased. Therefore, n was set to 5.

Table 5. Performance of the model with different numbers of GFBs.

GFB Modules	$n = 3$	$n = 5$	$n = 7$	$n = 9$
PSNR	37.10	38.07	37.40	37.24
SSIM	0.9674	0.9713	0.9688	0.9676
Params (M)	18.4	21.1	26.1	30.2

4.5.2. Hyper-Parameter Selection

In Equation (16), two hyper-parameters λ_{grad} and λ_{tex} need to be set. The impact of one hyper-parameter is analyzed by fixing another hyper-parameter. For the DIV2K dataset, the impact of λ_{grad} and λ_{tex} on the super-resolution reconstruction results was analyzed by PSNR and SSIM, respectively; the corresponding experimental results are shown in Figure 9. As shown in Figure 9, when $\lambda_{grad} = 0.01$ and $\lambda_{tex} = 0.001$, both PSNR and SSIM achieved the optimal values. Therefore, $\lambda_{grad} = 0.01$ and $\lambda_{tex} = 0.001$ were set for model training.

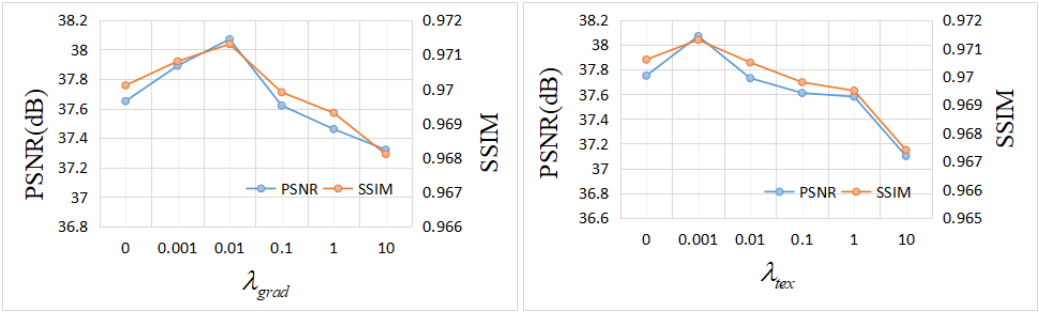

Figure 9. Hyper-parameter analysis on the DIV2K dataset.

5. Conclusions

A super-resolution reconstruction network that preserves structure and texture details was designed. The proposed network uses different network branches for feature extraction of the image structure and texture details, respectively, and enhances these two features layer-by-layer through the FHAM to improve the network's ability to extract features. To fully extract the information of different directions in the image, a dual-coordinate direction perception attention mechanism is proposed. The proposed attention mechanism fully considers the complementarity and directionality of the multi-scale image features, which can avoid information loss and possible distortion of the image structure and texture details. Additionally, a cross-fusion mechanism to effectively integrate the structure and texture features extracted by the network was designed, which can enhance the network's ability to mine structure and texture features. This is helpful to improve the quality of the reconstructed image. The experimental results show that the super-resolution reconstruction results of the proposed method produced good structure and texture details. The proposed method outperformed state-of-the-art methods in both visual quality and performance evaluation. However, the proposed method uses synthetic datasets to train networks in a supervised way. In the synthetic datasets, a low-resolution image is generated by bicubic downsampling its high-resolution counterpart. In practical scenarios, the high-resolution images are unavailable and image degradation models are unknown. Therefore, how to improve the quality of super-resolution reconstructed images in real scenes will be the direction of our future research.

Author Contributions: Conceptualization, K.W. and Y.H.; methodology, Y.Z. and Y.H.; software, Y.H.; validation, J.Z.; formal analysis, Y.H.; investigation, Y.Z. and Y.H.; data curation, Y.H.; writing—original draft preparation, Y.Z., Y.H. and G.Q.; writing—review and editing, Y.Z., Y.H. and G.Q.; visualization, Y.H. and J.Z.; supervision, K.W.; project administration, Y.Z. and K.W.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China No. 62161015.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Program and test data are available at https://github.com/YafeiZhang-KUST/PSTD, accessed on 28 November 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. ElSayed, A.; Mahmood, A.; Sobh, T. Effect of Super Resolution on High Dimensional Features for Unsupervised Face Recognition in the Wild. In Proceedings of the 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 10–12 October 2017.
2. Zhao, X.; Zhang, Y.; Zhang, T.; Zou, X. Channel splitting network for single MR image super-resolution. *IEEE Trans. Image Process.* **2019**, *28*, 5649–5662. [CrossRef] [PubMed]
3. Li, H.; Yu, Z.; Mao, C. Fractional differential and variational method for image fusion and super-resolution. *Neurocomputing* **2016**, *171*, 138–148. [CrossRef]
4. Meishvili, G.; Jenni, S.; Favaro, P. Learning to have an ear for face super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1364–1374.
5. Keys, R. Cubic convolution interpolation for digital image processing. *IEEE Trans. Acoust. Speech Signal Process.* **1981**, *29*, 1153–1160. [CrossRef]
6. Zhang, L.; Wu, X. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Trans. Image Process.* **2006**, *15*, 2226–2238. [CrossRef] [PubMed]
7. Sanchez-Beato, A.; Pajares, G. Noniterative interpolation-based super-resolution minimizing aliasing in the reconstructed image. *IEEE Trans. Image Process.* **2008**, *17*, 1817–1826. [CrossRef] [PubMed]
8. Zhou, F.; Yang, W.; Liao, Q. Interpolation-based image super-resolution using multisurface fitting. *IEEE Trans. Image Process.* **2012**, *21*, 3312–3318. [CrossRef] [PubMed]
9. Sun, J.; Xu, Z.; Shum, H.Y. Image super-resolution using gradient profile prior. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AL, USA, 23–28 June 2008; pp. 1–8.
10. Mairal, J.; Bach, F.; Ponce, J.; Sapiro, G.; Zisserman, A. Non-local sparse models for image restoration. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 2272–2279.
11. Tai, Y.W.; Liu, S.; Brown, M.S.; Lin, S. Super resolution using edge prior and single image detail synthesis. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 2400–2407.
12. Zhang, K.; Gao, X.; Tao, D.; Li, X. Single image super-resolution with non-local means and steering kernel regression. *IEEE Trans. Image Process.* **2012**, *21*, 4544–4556. [CrossRef]
13. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 105–114.
14. Timofte, R.; De Smet, V.; Van Gool, L. Anchored neighborhood regression for fast example-based super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 1920–1927.
15. Chen, C.L.P.; Liu, L.; Chen, L.; Tang, Y.Y.; Zhou, Y. Weighted couple sparse representation with classified regularization for impulse noise removal. *IEEE Trans. Image Process.* **2015**, *24*, 4014–4026. [CrossRef]
16. Liu, L.; Chen, L.; Chen, C.L.P.; Tang, Y.Y.; Pun, C. M. Weighted joint sparse representation for removing mixed noise in image. *IEEE Trans. Cybern.* **2017**, *47*, 600–611. [CrossRef]
17. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 184–199.
18. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.
19. Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.
20. Liu, J.; Zhang, W.; Tang, Y.; Tang, J.; Wu, G. Residual feature aggregation network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 2356–2365.
21. Niu, B.; Wen, W.; Ren, W.; Zhang, X.; Yang, L.; Wang, S.; Zhang, K.; Cao, X.; Shen, H. Single image super-resolution via a holistic attention network. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 191–207.
22. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.

23. Soh, J.W.; Park, G.Y.; Jo, J.; Cho, N.I. Natural and realistic single image super-resolution with explicit natural manifold discrimination. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 8114–8123.
24. Ma, C.; Rao, Y.; Cheng, Y.; Chen, C.; Lu, J.; Zhou, J. Structure-preserving super resolution with gradient guidance. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 7766–7775.
25. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 391–407.
26. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
27. Zhu, K.; Chen, Z.; Wu, Q.; Wang, N.; Zhao, J.; Zhang, G. FSFN: Feature separation and fusion network for single image super-resolution. *Multimed. Tools Appl.* **2021**, *80*, 31599–31618. [CrossRef]
28. Liu, J.; Ge, J.; Xue, Y.; He, W.; Sun, Q.; Li, S. Multi-scale skip-connection network for image super-resolution. *Multimed. Syst.* **2021**, *27*, 821–836. [CrossRef]
29. Huang, Y.; Li, J.; Gao, X.; Hu, Y.; Lu, W. Interpretable detail-fidelity attention network for single image super-resolution. *IEEE Trans. Image Process.* **2021**, *30*, 2325–2339. [CrossRef] [PubMed]
30. Mehri, A.; Ardakani, P.B.; Sappa, A.D. MPRNet: Multi-path residual network for lightweight image super resolution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 2703–2712.
31. Tan, C.; Cheng, S.; Wang, L. Efficient image super-resolution via self-calibrated feature fuse. *Sensors* **2022**, *22*, 329. [CrossRef]
32. Freeman, W.T.; Pasztor, E.C.; Carmichael, O.T. Learning low-level vision. *Int. J. Comput. Vis.* **2000**, *40*, 25–47. [CrossRef]
33. Chang, H.; Yeung, D.Y.; Xiong, Y. Super-resolution through neighbor embedding. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; Volume 1, p. I.
34. Timofte, R.; De Smet, V.; Van Gool, L. A+: Adjusted anchored neighborhood regression for fast super-resolution. In Proceedings of the Asian Conference on Computer Vision, Singapore, 1–5 November 2014; pp. 111–126.
35. Rahim, T.; Novamizanti, L.; Ramatryana, I.N.A.; Shin, S.Y.; Kim, D.S. Total variant based average sparsity model with reweighted analysis for compressive sensing of computed tomography. *IEEE Access* **2021**, *9*, 119158–119170. [CrossRef]
36. Rahim, T.; Novamizanti, L.; Ramatryana, I.N.A.; Shin, S.Y. Compressed medical imaging based on average sparsity model and reweighted analysis of multiple basis pursuit. *Comput. Med Imaging Graph.* **2021**, *90*, 101927. [CrossRef] [PubMed]
37. Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AL, USA, 24–26 June 2008; pp. 1–8.
38. Li, H.; He, X.; Tao, D.; Tang, Y.; Wang, R. Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning. *Pattern Recognit.* **2018**, *79*, 130–146. [CrossRef]
39. Xie, M.; Wang, J.; Zhang, Y. A unified framework for damaged image fusion and completion based on low-rank and sparse decomposition. *Signal Process. Image Commun.* **2021**, *29*, 116400. [CrossRef]
40. Zhang, Y.; Yang, M.; NanLi.; Yu, Z. Analysis-synthesis dictionary pair learning and patch saliency measure for image fusion. *Signal Process.* **2020**, *167*, 107327. [CrossRef]
41. Li, H.; Wang, Y.; Yang, Z.; Wang, R.; Li, X.; Tao, D. Discriminative Dictionary Learning-Based Multiple Component Decomposition for Detail-Preserving Noisy Image Fusion. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 1082–1102. [CrossRef]
42. Li, H.; Yan, S.; Yu, Z.; Tao, D. Attribute-Identity Embedding and Self-Supervised Learning for Scalable Person Re-Identification. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 3472–3485. [CrossRef]
43. Yan, S.; Zhang, Y.; Xie, M.; Zhang, D.; ZhengtaoYu. Cross-domain person re-identification with pose-invariant feature decomposition and hypergraph structure alignment. *Neurocomputing* **2022**, *467*, 229–241. [CrossRef]
44. Li, H.; Xu, J.; Yu, Z.; Luo, J. Jointly Learning Commonality and Specificity Dictionaries for Person Re-Identification. *IEEE Trans. Image Process.* **2020**, *29*, 7345–7358. [CrossRef]
45. Li, H.; Cen, Y.; Liu, Y.; Chen, X.; Yu, Z. Different Input Resolutions and Arbitrary Output Resolution: A Meta Learning-Based Deep Framework for Infrared and Visible Image Fusion. *IEEE Trans. Image Process.* **2021**, *30*, 4070–4083. [CrossRef] [PubMed]
46. Xiao, W.; Zhang, Y.; Wang, H.; Li, F.; Jin, H. Heterogeneous Knowledge Distillation for Simultaneous Infrared-Visible Image Fusion and Super-Resolution. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 5004015. [CrossRef]
47. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018; pp. 63–79.
48. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2790–2798.
49. Tai, Y.; Yang, J.; Liu, X.; Xu, C. Memnet: A persistent memory network for image restoration. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4549–4557.
50. Wang, X.; Yu, K.; Dong, C.; Loy, C.C. Recovering realistic texture in image super-resolution by deep spatial feature transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 606–615.

51. Wu, B.; Duan, H.; Liu, Z.; Sun, G. SRPGAN: Perceptual generative adversarial network for single image super resolution. *arXiv* **2017**, arXiv:1712.05927.
52. Chen, B.X.; Liu, T.J.; Liu, K.H.; Liu, H.H.; Pei, S.C. Image super-resolution using complex dense block on generative adversarial networks. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 2866–2870.
53. Fu, B.; Wang, L.; Wu, Y.; Wu, Y.; Fu, S.; Ren, Y. Weak texture information map guided image super-resolution with deep residual networks. *Multimed. Tools Appl.* **2022**, *81*, 34281–34294. [CrossRef]
54. Cai, Q.; Li, J.; Li, H.; Yang, Y.H.; Wu, F.; Zhang, D. TDPN: Texture and Detail-Preserving Network for Single Image Super-Resolution. *IEEE Trans. Image Process.* **2022**, *31*, 2375–2389. [CrossRef]
55. Meng, B.; Wang, L.; He, Z.; Jeon, G.; Dou, Q.; Yang, X. Gradient information distillation network for real-time single-image super-resolution. *J.-Real-Time Image Process.* **2021**, *18*, 333–344. [CrossRef]
56. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [CrossRef]
57. Agustsson, E.; Timofte, R. Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1122–1131.
58. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Morel, M.L.A. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In Proceedings of the Electronic Proceedings of the British Machine Vision Conference, Surrey, UK, 3–7 September 2012; pp. 1–10.
59. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In Proceedings of the International Conference on Curves and Surfaces, Avignon, France, 24–30 June 2010; pp. 711–730.
60. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In Proceedings of the IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; pp. 416–423.
61. Huang, J.B.; Singh, A.; Ahuja, N. Single image super resolution from transformed self-exemplars. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5197–5206.
62. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
63. Fang, Z.; Zhao, M.; Yu, Z.; Li, M.; Yang, Y. A guiding teaching and dual adversarial learning framework for a single image dehazing. *Vis. Comput.* **2022**, *38*, 3563–3575. [CrossRef]
64. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
65. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [CrossRef]
66. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5835–5843.
67. Wei, P.; Xie, Z.; Lu, H.; Zhan, Z.; Ye, Q.; Zuo, W.; Lin, L. Component divide-and-conquer for real-world image super-resolution. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 101–117.
68. Ketsoi, V.; Raza, M.; Chen, H.; Yang, X. SREFBN: Enhanced feature block network for single-image super-resolution. *IET Image Process.* **2022**, *16*, 3143–3154. [CrossRef]