remote sensing

MDPI

Technical Note

Image Reconstruction of Multibranch Feature Multiplexing Fusion Network with Mixed Multilayer Attention

Yuxi Cai, Guxue Gao, Zhenhong Jia and Huicheng Lai *

College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China; cai@stu.xju.edu.cn (Y.C.); gaoyangshuang123@stu.xju.edu.cn (G.G.); jzhh@xju.edu.cn (Z.J.)
* Correspondence: lai@xju.edu.cn

Abstract: Image super-resolution reconstruction achieves better results than traditional methods with the help of the powerful nonlinear representation ability of convolution neural network. However, some existing algorithms also have some problems, such as insufficient utilization of phased features, ignoring the importance of early phased feature fusion to improve network performance, and the inability of the network to pay more attention to high-frequency information in the reconstruction process. To solve these problems, we propose a multibranch feature multiplexing fusion network with mixed multilayer attention (MBMFN), which realizes the multiple utilization of features and the multistage fusion of different levels of features. To further improve the network's performance, we propose a lightweight enhanced residual channel attention (LERCA), which can not only effectively avoid the loss of channel information but also make the network pay more attention to the key channel information and benefit from it. Finally, the attention mechanism is introduced into the reconstruction process to strengthen the restoration of edge texture and other details. A large number of experiments on several benchmark sets show that, compared with other advanced reconstruction algorithms, our algorithm produces highly competitive objective indicators and restores more image detail texture information.

Keywords: super-resolution reconstruction; feature reuse; multistage fusion

Citation: Cai, Y.; Gao, G.; Jia, Z.; Lai, H. Image Reconstruction of Multibranch Feature Multiplexing Fusion Network with Mixed Multilayer Attention. *Remote Sens.* **2022**, *14*, 2029. https://doi.org/10.3390/rs14092029

Academic Editors: Junjun Jiang and Jaime Zabalza

Received: 3 March 2022
Accepted: 21 April 2022
Published: 23 April 2022

1. Introduction

For a long time, single-image super-resolution reconstruction (SISR) has been a classic research problem in the field of computer bottom vision. It aims to use certain technical means to restore the corresponding high-resolution image containing rich texture information from the degraded low-resolution image. SISR has been widely used in remote sensing, public security, medicine and other fields. There are many different high-resolution images in the real scene; after different degradation, the same low-resolution image is obtained. Therefore, SISR is a typical ill-posed problem. Benefiting from the vigorous development of machine learning, researchers have proposed many reconstruction algorithms based on deep learning, such as Cross-SRN [1], MRFN [2], MADNet [3] and so on, which have achieved better results than traditional algorithms.

Dong et al. [4] constructed a shallow end-to-end nonlinear mapping neural network for the first time and achieved better results than traditional algorithms such as interpolation. Subsequently, researchers have built a series of very deep neural networks, but with the deepening of the number of network layers and the further improvement of the reconstruction effect, they are also faced with the problem that the network is difficult to train. Kim et al. [5] constructed VDSR by introducing residual learning. With the help of jump connection, the network was further deepened, and the performance of model was improved again. Inspired by this, Lim et al. [6] removed the commonly used batch normalization layer in the network and constructed EDSR using small-scale residual blocks, which not only saved the storage cost, but also improved the reconstruction effect. Liu et al. [7] believe that the features after residual learning will form more complex fusion features, which

makes the network ignore the cleaner residual features generated in the middle. Therefore, Liu transmits the residual features generated in the middle to the end of the basic block for local fusion through jump connection, which greatly improves the reconstruction effect. Cross-SRN [1] uses a hierarchical feature network to detect and save structural information in the way of multiscale feature fusion. MRFN [2] extracts and fuses hierarchical features from local to global through multireceptive-field block and dense connections. MADNet [3] uses dual residual-path block to take advantage of hierarchical features from the original low-resolution image.

Most CNN-based algorithms have achieved excellent results, but they still face some problems. Usually, most of the extracted phased intermediate features will be processed by a series of stacked convolution layers, and the generated intermediate features are rarely used, or only used by the network once; they cannot be further processed by the network, resulting in a certain degree of feature waste. Then, the features transferred to the next layer are processed by convolution layer to form new features more complex than the original features. Ma et al. [8] have shown that the same feature will show different information in different feature extraction stages of network. At the same time, due to different receptive fields, the same feature will also be extracted with different information, which contributes to the reconstruction results to varying degrees. Most algorithms focus on the final generated features and ignore the reuse of intermediate features, which leads to the decline of network performance to a certain extent. In addition, another little attention is paid to the stage feature fusion early. Most reconstruction algorithms only fuse the features in different stages at the end of the basic block to form complex features with relatively rich high-frequency information. However, the feature fusion in different stages in the early stage can not only aggregate the hierarchical information under different sensory fields, but also further deepen the fusion features. It indirectly realizes the extraction and fusion of multiscale features and further expands the receptive field of the network. In addition, some algorithms use multiple branches to process features, but there is a lack of direct and effective information exchange between the branches to guide the network to focus on regions of interest in a coordinated manner. In addition, these algorithms ignore the importance of feature fusion at different stages in the early stages to enhance the network effectiveness. Finally, due to the use of deconvolution to realize up sampling, it will bring different degrees of artifacts to the reconstructed image and affect the reconstruction effect. Most reconstruction algorithms use subpixel convolution to realize up sampling of feature size. In the reconstruction process, these algorithms cannot make the network effectively focus on high-frequency information such as local edge texture, resulting in poor recovery effect.

To solve the above problems, we designed MBMFN, which realizes the multiple utilization of features and multistage local feature fusion, so that the network can learn a more discriminative feature representation. Meanwhile, with the help of the attention mechanism in the reconstruction process, we can better recover the edge texture and other details of the image. Our contribution mainly includes the following points:

1. A multibranch feature, reuse fusion attention block, is proposed, which realizes the reuse of features and the fusion of multistage local features through the interaction of information between multiple branches and enriches the hierarchical types of phased features.
2. A lightweight enhanced residual channel attention (LERCA) is proposed, which can pay more attention to the high-frequency information in low resolution space. We use 1×1 convolution to establish the interdependence between channels, which avoids the loss of some channel information caused by channel compression, and the module is more lightweight.
3. In the reconstruction process, the attention mechanism is introduced, and the U-LERCA is constructed combined with LERCA, which enhances the sensitivity of the network to key information. In particular, few people have studied the attention strategies in the reconstruction process.

4. We constructed a multibranch feature multiplexing fusion network with mixed multi-layer attention, which achieves a good recovery effect. At the same time, the network achieves a good balance between parameters and performance.

2. Related Work

2.1. Phased Feature Fusion

With the vigorous development of deep learning, many lightweight reconstruction algorithms have been proposed and achieved good results in objective indicators and subjective vision. SMSR [9] uses a sparse mask module to generate a space mask and a channel mask to locate redundant computing and then dynamically skips redundant computing with the help of sparse mask convolution to achieve efficient image reconstruction. In order to strengthen the fusion of different levels of features, Hui et al. introduced the idea of distillation in IDN [10] and divided the features processed by convolution layer into two parts through channel splitting operation. Part of the features continue to be further processed by convolution layer, and the remaining features are spliced with the original input features and transmitted to the end of the enhancement module by jumping for cross fusion of different local features in order to strengthen network learning of LR contour region. On this basis, Hui continues to deeply study the application of distillation idea and local feature fusion and put forward IMDN [11]. The network uses a channel splitting strategy to distill fine features layer by layer and aggregates different levels of features through splicing and a 1×1 convolution layer but uses channel splitting operation for feature distillation, which brings a certain degree of inflexibility to the network. To this end, Liu et al. proposed RFDN [12], which uses a 1×1 convolution layer instead of a channel splitting operation to carry out compressed feature distillation, while using a convolution layer with residual to replace the original convolution layer. By doing so, the network becomes lighter, and the performance is further improved, but the network only extracts features at a fixed scale and cannot effectively aggregate feature information of different scales. Wang et al. proposed MSFIN [13], which restores the input image to the target size by interpolation algorithm, uses a 3×3 convolution layer to down-sample different image sizes and then uses three branches to process the features of different image sizes in parallel. In order to make up for the lack of information exchange between branches, deconvolution is used to restore the feature size, and the up-sampled features are fused with the features of the corresponding stage of the previous branch. Through this design, the network effectively integrates features of different sizes and benefits from it, but the network runs in high-resolution space, resulting in a lot of memory and computational overhead. Cross-SRN [1] uses multiple branches to extract hierarchical features in the basic block, but each branch processes the same number of features, so it is unable to establish different levels of feature relationship, and there is a lack of information exchange between branches. The basic block of MRFN [2] can make the network feel the hierarchical information in different receiving domains through three parallel branches, but there is a lack of effective information communication between the branches. MADNet [3] obtains hierarchical features at different stages through four branches with different void rates, but the network ignores the importance of early feature fusion in the base block to improve network performance, and in addition, each branch cannot effectively guide network to focus on regions of interest.

2.2. Attention Mechanism

Attention mechanism can effectively guide neural network to focus on the most important information in the input features and strengthen the network's learning and expression of these information. It has been widely used in various computer vision tasks, including image segmentation, target tracking, image restoration and so on, and has shown great advantages in improving the performance of network. Hu et al. introduced the attention mechanism to the image classification task and proposed the SENet [14] network, which explicitly models the interdependence between channels, which can adaptively

correct channel features and retain valuable features, so the network performance is further improved. Roy et al. [15] use a 1 × 1 convolution layer instead of a full connection layer in the SENet to model the dependence between channels. Wang et al. [16] generate channel weights through lighter and faster one-dimensional convolution. Recently, Hui et al. [11] proposed the Contrast-aware channel attention, and the network not only achieves a good objective index but also restores more detailed information such as edge texture. Niu et al. [17] constructed a hybrid attention mechanism to adaptively capture key information. Hou et al. [18] constructed coordinated attention by embedding location information into channel attention, which captures not only cross-channel information but also direction-aware and location-aware information, which can help the network to locate and identify targets of interest more precisely.

3. Paper Method

In this part, we first introduce the multibranch feature multiplexing fusion network proposed in this paper and then introduce each component of the network in detail.

3.1. Overall Framework

As shown in Figure 1, the multibranch feature multiplexing fusion network with mixed multilayer attention (MBMFN) is composed of a shallow feature extraction block, a multibranch feature multiplexing fusion attention block (MBMFB) and reconstruction block. In this paper, $\mathbf{I_{LR}}$, $\mathbf{I_{SR}}$ are represented as the input and output images of the network, respectively. According to the research of [19,20], in this paper, only a 3 × 3 convolution layer is used for shallow feature extraction of $\mathbf{I_{LR}}$:

$$\mathbf{F}_0 = \mathbf{Conv}_{3\times3}(\mathbf{I_{LR}}) \tag{1}$$

where $\mathbf{Conv}_{3\times3}(\cdot)$ represents convolution operation with convolution kernel of 3 × 3, and $\mathbf{F}_0$ represents the shallow features extracted by the convolution layer. Then, the $\mathbf{F}_0$ is used as the input of the MBMFB to deepen the features in order to learn a more discriminating feature representation.

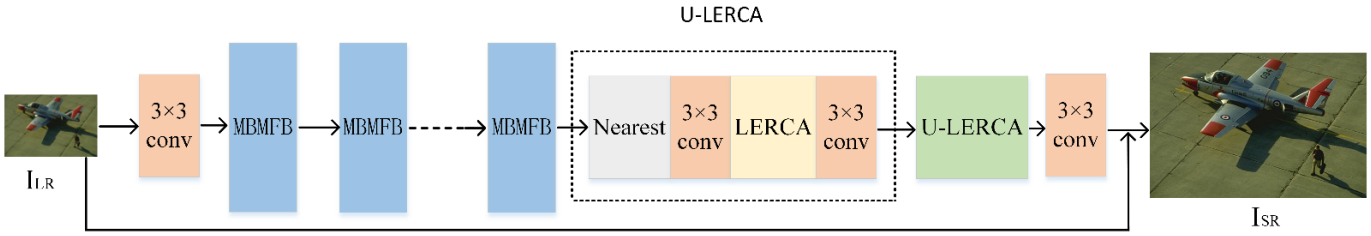

Figure 1. Overall block diagram of multibranch feature multiplexing fusion network with mixed multilayer attention.

Assuming that there are d MBMFBs, the output features of the d-th MBMFB are expressed as:

$$\mathbf{F_d} = \mathbf{H_{MBMFB,d}}(\mathbf{F_{d-1}}) = \mathbf{H_{MBMFB,d}}(\mathbf{H_{MBMFB,d-1}}(\ldots(\mathbf{H_{MBMFB,1}}(\mathbf{F}_0))\ldots)) \tag{2}$$

where $\mathbf{H_{MBMFB,d}}(\cdot)$ represents the d-th MBMFB with composite function. $\mathbf{F_d}$ represents the local fusion feature extracted after the d-th MBMFB processing. More details about MBMFB are described in detail in Section 3.2.

After multiple MBMFBs processing, we send the learned discriminating features into the reconstruction module with attention so as to restore to the corresponding target size. The operation process is expressed as follows:

$$\mathbf{F_n} = \mathbf{H^n_{U-LERCA}}\left(\mathbf{H^{n-1}_{U-LERCA}}\left(\cdots\mathbf{H^0_{U-LERCA}}(\mathbf{F_d})\cdots\right)\right) \tag{3}$$

where $\mathbf{H}^{\mathbf{n}}_{\mathbf{U-LERCA}}$ represents the n-th U-LERCA block, $\mathbf{F_n}$ represents the output features of the n-th U-LERCA, and more information about U-LERCA are introduced in Section 3.4.

In order to make up for the problem of losing part of the underlying information in the continuous deepening processing of features, we use the traditional interpolation algorithm to sample the $\mathbf{I_{LR}}$ to the corresponding size and supplement the information by jumping connection, so as to generate the final $\mathbf{I_{SR}}$:

$$\mathbf{I_{SR}} = \mathbf{Conv}_{3\times3}(\mathbf{F_n}) + \mathbf{H_{up}}(\mathbf{I_{LR}}) \tag{4}$$

where $\mathbf{H_{up}}(\cdot)$ represents the up-sampling operation of bilinear interpolation.

According to the previous research work, we use $\mathbf{L}_1$ loss function to optimize the network parameters. Given a training set $\left\{\mathbf{I}^{\mathbf{j}}_{\mathbf{LR}}, \mathbf{I}^{\mathbf{j}}_{\mathbf{HR}}\right\}_{\mathbf{j=1}}^{\mathbf{j=N}}$ that contains many image pairs, where N represent the number of training image pairs, the $\mathbf{L}_1$ loss function with parameters used in this paper is expressed as follows:

$$\mathrm{L}(\theta) = \frac{1}{N}\sum_{\mathrm{i}}^{N} \left\| \mathbf{H_{MBMFN}}\left(\mathbf{I}^{\mathbf{j}}_{\mathbf{LR}}\right) - \mathbf{I}^{\mathbf{j}}_{\mathbf{HR}} \right\|_1 \tag{5}$$

where θ represents network parameters that need to be optimized, and $\mathbf{H_{MBMFN}}(\cdot)$ represents the multibranch feature multiplexing fusion network with mixed multilayer attention.

3.2. Multibranch Feature Multiplexing Fusion Attention Block

In order to realize the reuse of local features and promote the fusion of multilevel features, a semiparallel multibranch feature reuse fusion attention block is designed in this paper, which not only expands the receiving domain of network but also avoids the deepening of network. The internal structure of the block is shown in Figure 2.

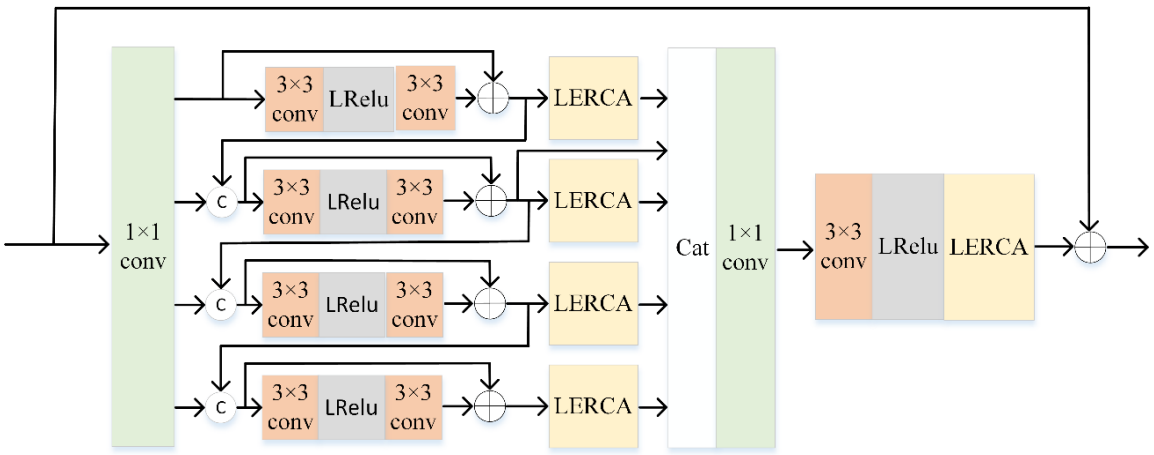

Figure 2. Internal structure diagram of multibranch feature multiplexing fusion attention block.

First of all, we use a 1 × 1 convolution layer for feature distillation to reduce the amount of computation and redundant feature information of network, and then the refined features are sent to the four branches of parallel processing for feature extraction and fusion.

Each branch contains a residual block and LERCA. With the help of the residual block composed of two 3 × 3 convolution layers, the features after distillation are further deepened. In order to enhance the feature extraction ability of network and focus the network on the key features, we designed LERCA in order to enhance the network's learning of key features. More details on the LERCA are introduced in Section 3.3. We added LERCA to each branch to strengthen the network to extract the key information of

hierarchical features in different channels so that the network can learn more discriminating feature representation.

The output features after the residual block in the first branch and the previous distillation features are spliced together as the input of the second branch. With the help of the convolution layer of the second branch, the network integrates the hierarchical features of different stages, realizes the reuse of features and expands the acceptance domain of the network. Therefore, the 3×3 convolution layer shoulders the important task of feature fusion and extraction. Because the input of the second branch is spliced by the features of two stages, with the help of jump connection, the network adds the features of different stages to the same stage features and further forms rich hierarchical features. By analogy, through feature reuse and feature processing of three or four branches, the network effectively integrates the features of multiple stages and forms more abundant hierarchical features.

Although channel attention can enhance the network's attention to information-rich channels, there is some key information in those suppressed channels. In order to make up for the loss of this information, we take the output feature after the residual block in the second branch as the basic bottom feature (as shown by the blue line in Figure 2) and splice it together with the output features of the other four branches in a local form. Then, the 1×1 convolution layer is used for feature aggregation. Finally, the attention mechanism is used to strengthen the learning of important features in order to improve the representation ability of the network. In addition, at the end of MBMFB, we use residual connections to jump forward the original information so as to benefit from residual learning and speed up the back propagation of gradient in network optimization.

3.3. Lightweight Enhanced Residual Channel Attention

As shown in Figure 3, although RCAB [21] proposed by Zhang et al. can adaptively adjust channel features response according to the interdependence between channels and achieve quite good performance, RCAB compresses the channel, resulting in varying degrees of channel information loss. Secondly, RCAB uses a 3×3 convolution layer with a large number of parameters for many times, which brings a lot of memory overhead to the network and is not suitable for lightweight network reconstruction. Therefore, LERCA is proposed in this paper.

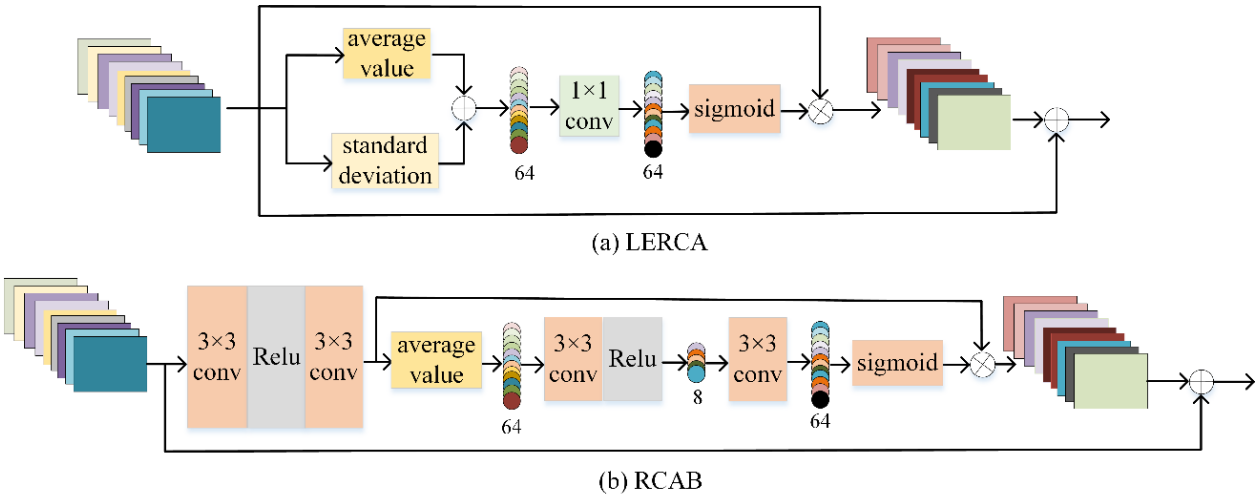

Figure 3. (**a**) Lightweight enhanced residual channel attention (**b**) residual channel attention block (8 and 64 in the figure represent different number of channels, respectively).

The LERCA proposed in this paper differs from RCAB in the following three ways: First, we removed the two convolution layers in front of RCAB. Behjati et al. [22] think that channel attention may discard some relevant detailed features, which will be difficult to

regain at a deeper network level. At the same time, the 3×3 convolution layer will bring many parameters to the network. In order to supplement the relevant detailed features and meet the needs of the lightweight network, we remove the previous two convolution layers. Due to the existence of convolution layer, it is difficult for RCAB to compensate for the missing details information through jump connection, but our LERCA can easily compensate the lost feature information. Second, we use the sum of global average and standard deviation instead of global average pooling. Compared with the global average pooling, the sum of the global average and standard deviation can more fully characterize the characteristics of the channel. Finally, the 1×1 convolution layer is used to replace the two 3×3 convolution layers in RCAB. RCAB uses two 3×3 convolution layers for channel compression and recovery respectively, which destroys the channel characteristics and causes the loss of channel information to a certain extent. We directly use a 1×1 convolution layer to model the interdependence between channels, meanwhile reducing the number of parameters and achieving the goal of lightening the module.

In short, we use the sum of the global average and standard deviation of features to characterize the channel characteristics, directly model the relationship between channels with the help of the 1×1 convolution layer and then generate an attention mask through the Sigmoid function. The module can not only reduce the loss of channel information but also help the network to recover more details such as image edge texture. With the help of jump connection, the network effectively makes up for the relevant information lost due to channel attention, and this information can no longer be regained in the deeper layers of the network.

3.4. Reconstruction Block (U-LERCA)

In some previous reconstruction algorithms, the reconstruction block of network is usually composed of a convolution layer and a subpixel convolution layer, or the traditional interpolation algorithm is used to realize the up-sampling operation, and the attention mechanism is rarely introduced into the reconstruction process. As a result, the key features are difficult to play their due roles. On the other hand, for large-scale reconstruction tasks, such as $\times 4$, if there is a lack of enough high-frequency information, it is difficult for the network to achieve satisfactory reconstruction results. In view of the above two points, we combined LERCA to build U-LERCA, whose internal structure is shown in Figure 1, which is composed of the nearest neighbor interpolation algorithm, LERCA and two convolution layers.

In U-LERCA, we abandon the subpixel convolution and use nearest neighbor interpolation algorithm to achieve the up-sampling operation, mainly because the subpixel convolution layer brings many parameters to network but also because it cannot achieve the corresponding recovery effect. We use the convolution layer to further establish the correlation between channels after interpolation. In order to distinguish the importance between channels and to play the role of some key features, we use LERCA to enhance the network's attention to key channels to achieve the purpose of improving network performance.

In order to cope with the fact that under the large-scale reconstruction task, the network still has enough high-frequency information available, while ensuring that each step of the up-sampling is optimal, we adopt a distributed strategy to carry out step-by-step up-sampling until the target size is reached. For example, the $\times 4$ reconstruction task is divided into two cascaded $\times 2$ reconstruction tasks in this paper. At the same time, thanks to the addition of the attention mechanism in the reconstruction process, the network achieves a good reconstruction performance.

4. Experiment and Analysis

In this paper, the DIV2K dataset is used for network training. In the testing phase, four standard benchmark datasets are adopted: Set5, Set14, B100, Urban100. In this paper, the image is converted from RGB color space to YCbCr color space, and only the Y channel is trained and tested. Compared with deep no-reference image quality analysis methods,

such as [23], we use the widely used Peak Signal-to-Noise Ratio (PSNR) and Structure Similarity Index Measure (SSIM) to quantitatively analyze the reconstruction results.

4.1. Experimental Environment and Parameter Setting

In the reconstruction tasks of $\times 2$, $\times 3$ and $\times 4$, we randomly extract 24 image patches of 192×192 as the input of the network. One epoch is formed by 1000 back propagation iterations, and the initial learning rate is 0.0002. After each 200 epochs, learning rate decays to half of the original. Six MBMFBs are used in the network, and the activation function Leaky-Relu is set to 0.05. The Adam algorithm is used to optimize the network gradient. Under the framework of Pytorch deep learning, we construct the MBMFN algorithm. The experimental hardware platform is NVIDIA Tesla V100-PCIE-16GB, and the software environment is Windows10 operating system.

4.2. Ablation Experiment

In order to verify effectiveness of the network design, we carried out ablation experiments on some of the important blocks. All ablation experiments were based on the training of 400 epochs under the $\times 4$ reconstruction task.

Verify the impact of the location of information input between branches in MBMFB on network performance: As shown in Table 1, where Basic Branch refers to the branch in the second branch of MBMFB that skips LERCA, as shown in the blue line in Figure 2, this branch is mainly used to supplement underlying information. Before Residual Way (BRW) refers to the input of features before residual to the next branch. After Residual Way (ARW) refers to the input of features after residual to the next branch. After Attention Way (AAW) refers to the input of features processed by LERCA to the next branch.

Table 1. Influence of information input location between branches in basic blocks on network performance.

Branch Input Type	BRW	ARW	AAW	PSNR(Set5)	SSIM(Set5)
Basic Branch					
No	√	×	×	32.209	0.8940
	×	√	×	32.231	0.8945
	×	×	√	32.216	0.8939
Yes	√	×	×	32.223	0.8943
	×	√	×	32.247	0.8945
	×	×	√	32.243	0.8942

As can be seen from Table 1, with or without Basic Branch, ARW achieved better results than other input patterns, which may benefit from residual learning. In the case of no Basic Branch, the effect of AAW is significantly lower than that of ARW, which may be due to the fact that some of the key information in the suppressed channel features cannot be effectively expressed after LERCA processing, resulting in performance degradation. Compared with the situation with Basic Branch, ARW achieved a good improvement effect, which is due to the fact that Basic Branch supplements the neglected feature information and makes up for the loss of underlying information caused by the attention mechanism.

Verify the validity of LERCA: In order to verify the effectiveness of LERCA, we replace LERCA in MBMFB with Squeeze-and-Excitation (SE), Channel attention (CA), Residual channel attention (RCA) and Contrast-aware channel attention (CCA), respectively, in which RCA only adds residual connections on the basis of CA. The experimental results are shown in Table 2.

It can be clearly seen from Table 2 that various attention mechanisms significantly improved the performance of the network. Compared with MBMFB-CA, PSNR and SSIM of MBMFB-LERCA on Set5 test set increased by 0.014 db and 0.0012, respectively, which is due to LERCA's uncompressed processing of the channel, reducing the loss of channel information and protecting channel characteristics. Compared with MBMFB-RCA, PSNR

and SSIM of MBMFB-LERCA increased by 0.041 db and 0.001, respectively, on Set5, which may be related to the fact that LERCA uses the sum of channel global average and standard deviation to characterize channel characteristics. Compared with MBMFB-CCA, PSNR and SSIM of MBMFB-LERCA increased by 0.039 db and 0.0013, respectively, indicating that LERCA is more effective than CCA.

Table 2. Performance comparison of various attention mechanisms in basic blocks.

Attention Type	MBMFB-No	MBMFB-SE	MBMFB-CA	MBMFB-RCA	MBMFB-CCA	MBMFB-LERCA
PSNR(Set5)	32.164	32.214	32.233	32.206	32.208	32.247
SSIM(Set5)	0.8937	0.8943	0.8942	0.8944	0.8941	0.8954

Verify the effectiveness of adding the attention mechanism in the reconstruction phase: In order to prove that the introduction of attention mechanism in the reconstruction process has a good improvement effect on the network, we carried out experimental verification. The experimental results are shown in Table 3, in which U-Nearest-LERCA-$\times$4 means that in the reconstruction process, the nearest neighbor interpolation is used to realize the up-sampling operation; the LERCA is added; and the up sampling is directly $\times$4 without step-by-step processing. U-Nearest-$\times$2-$\times 2^{\text{Weight sharing}}$ means that the nearest neighbor interpolation is used to realize the up-sampling operation in the reconstruction process; LERCA is not added; and the $\times$4 reconstruction task is divided into two cascaded $\times$2 reconstruction task, with the two reconstruction blocks using the weight-sharing strategy. U-subpixel refers to the up-sampling operation using subpixel convolution.

Table 3. Effectiveness of attention mechanism and distributed processing in the reconstruction phase.

Up-Sampling Pattern	U-Nearest-$\times$4	U-Nearest-LERCA-$\times$4	U-Nearest-$\times$2-$\times 2^{\text{Weight sharing}}$	U-Nearest-LERCA-$\times$2-$\times 2^{\text{Weight sharing}}$	U-Nearest-LERCA-$\times$2-$\times 2^{\text{No Weight sharing}}$	U-Subpixel
PSNR(Set5)	32.207	32.221	32.219	32.247	32.246	32.216
SSIM(Set5)	0.8941	0.8941	0.8942	0.8945	0.8945	0.8943
Parameter	1220 K	1224 K	1220 K	1224 K	1291 K	1250 K

Compared with the reconstruction process without attention mechanism, U-Nearest-LERCA-$\times$4 has higher 0.014 dB than U-Nearest-$\times$4 on PSNR. Similarly, U-Nearest-LERCA-$\times$2-$\times 2^{\text{Weight sharing}}$ has higher, 0.028, dB than U-Nearest-$\times$2-$\times 2^{\text{Weight sharing}}$, and it also has higher, 0.031, dB than U-Subpixel. On SSIM, U-Nearest-LERCA-$\times$2-$\times 2^{\text{Weight sharing}}$ is 0.0003 higher than U-Nearest-$\times$2-$\times 2^{\text{Weight sharing}}$. This directly and effectively proves that the introduction of LERCA attention mechanism in the reconstruction process can improve the reconstruction performance of the network.

Effectiveness of step-by-step processing in reconstruction phase: This shows the effectiveness of the distributed processing strategy in the reconstruction process, and the experimental results are shown in Table 3. Compared with the single-step processing, U-Nearest-$\times$2-$\times 2^{\text{Weight sharing}}$ has higher, 0.012, dB than U-Nearest-$\times$4 on PSNR; similarly, U-Nearest-LERCA-$\times$2-$\times 2^{\text{Weight sharing}}$ has higher, 0.026, dB than U-Nearest-LERCA-$\times$4. On SSIM, U-Nearest-LERCA-$\times$2-$\times 2^{\text{Weight sharing}}$ is 0.0004 higher than U-Nearest-LERCA-$\times$4. Obviously, the use of distributed processing strategy can ensure that each reconstruction task has enough high-frequency information available, thus improving the performance of the network.

In order to further reduce the network parameters, we use the weight-sharing strategy for the reconstruction blocks in the reconstruction process. As can be seen from Table 3, on PSNR, U-Nearest-LERCA-$\times$2-$\times 2^{\text{Weight sharing}}$ has higher, 0.001, dB than U-Nearest-LERCA-$\times$2-$\times 2^{\text{No Weight sharing}}$, and the number of parameters is 67 K less.

4.3. Comparison with Other Advanced Algorithms

In order to prove effectiveness of this network, we compare MBMFN with other advanced lightweight super-resolution reconstruction algorithms with up-sampling factors of ×2, ×3 and ×4. Include SRCNN [4], VDSR [5], DRCN [24], MemNet [25], CARN [26], IMDN [11], DNCL [27], MRFN [2], FilterNeL [28], MADNet-L_F [3], RFDN [12], CFSRCNN [29], SeaNet-baseline [30], MSFIN [13], SMSR [9], Cross-SRN [1]. Experimental results are shown in Table 4. Except for the algorithm in this paper, the results of other algorithms come from published papers.

Table 4. Average PSNR/SSIM of BI degradation models ×2, ×3 and ×4, and the optimal results are shown in bold.

Scale	Method	Year	Set5		Set14		B100		Urban100	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×2	SRCNN	2016	36.66	0.9542	32.42	0.9063	31.36	0.8879	29.50	0.8946
	VDSR	2016	37.53	0.9587	33.03	0.9124	31.90	0.8960	30.76	0.9140
	DRCN	2016	37.63	0.9588	33.04	0.9118	31.85	0.8942	30.75	0.9133
	MemNet	2017	37.78	0.9597	33.28	0.9142	32.08	0.8978	31.31	0.9195
	CARN	2018	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256
	IMDN	2019	38.00	0.9605	33.63	0.9177	32.19	0.8996	32.17	0.9283
	DNCL	2019	37.65	0.9599	33.18	0.9141	31.97	0.8971	30.89	0.9158
	MRFN	2019	37.98	**0.9611**	33.41	0.9159	32.14	0.8997	31.45	0.9221
	FilterNeL	2020	37.86	0.9610	33.34	0.9150	32.09	0.8990	31.24	0.9200
	MADNet-L_F	2020	37.85	0.9600	33.39	0.9161	32.05	0.8981	31.59	0.9234
	RFDN	2020	38.05	0.9606	33.68	0.9184	32.16	0.8994	32.12	0.9278
	CFSRCNN	2020	37.79	0.9591	33.51	0.9165	32.11	0.8988	32.07	0.9273
	SeaNet-baseline	2020	37.99	0.9607	33.60	0.9174	32.18	0.8995	32.08	0.9276
	SMSR	2021	38.00	0.9601	33.64	0.9179	32.17	0.8990	32.19	0.9284
	Cross-SRN	2021	38.03	0.9606	33.62	0.9180	32.19	**0.8997**	32.28	0.9290
	MBMFN		**38.05**	0.9599	**33.78**	**0.9193**	**32.21**	0.8996	**32.44**	**0.9303**
×3	SRCNN	2016	32.75	0.9090	29.28	0.8209	28.41	0.7863	26.24	0.7989
	VDSR	2016	33.66	0.9213	29.77	0.8314	28.82	0.7976	27.14	0.8279
	DRCN	2016	33.82	0.9226	29.76	0.8311	28.80	0.7963	27.15	0.8276
	MemNet	2017	34.09	0.9248	30.00	0.8350	28.96	0.8001	27.56	0.8376
	CARN	2018	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493
	IMDN	2019	34.36	0.9270	30.32	0.8417	29.09	0.8046	28.17	0.8519
	DNCL	2019	33.95	0.9232	29.93	0.8340	28.91	0.7995	27.27	0.8326
	MRFN	2019	34.21	0.9267	30.03	0.8363	28.99	0.8029	27.53	0.8389
	FilterNeL	2020	34.08	0.9250	30.03	0.8370	28.95	0.8030	27.55	0.8380
	MADNet-L_F	2020	34.14	0.9251	30.20	0.8395	28.98	0.8023	27.78	0.8439
	RFDN	2020	34.41	0.9273	30.34	0.8420	29.09	0.8050	28.21	0.8525
	CFSRCNN	2020	34.24	0.9256	30.27	0.8410	29.03	0.8035	28.04	0.8496
	SeaNet-baseline	2020	34.36	**0.9280**	30.34	**0.8428**	29.09	0.8053	28.17	0.8527
	SMSR	2021	34.40	0.9270	30.33	0.8412	29.10	0.8050	28.25	0.8536
	Cross-SRN	2021	34.43	0.9275	30.33	0.8417	29.09	0.8050	28.23	0.8535
	MBMFN		**34.52**	0.9273	**30.41**	0.8426	**29.12**	**0.8052**	**28.36**	**0.8553**

Table 4. *Cont.*

Scale	Method	Year	Set5		Set14		B100		Urban100	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
×4	SRCNN	2016	30.48	0.8628	27.49	0.7503	26.90	0.7101	24.52	0.7221
	VDSR	2016	31.35	0.8838	28.01	0.7674	27.29	0.7251	25.18	0.7524
	DRCN	2016	31.53	0.8854	28.02	0.7670	27.23	0.7233	25.14	0.7510
	MemNet	2017	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630
	CARN	2018	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837
	IMDN	2019	32.21	0.8948	28.58	0.7811	27.56	0.7353	26.04	0.7838
	DNCL	2019	31.66	0.8871	28.23	0.7717	27.39	0.7282	25.36	0.7606
	MRFN	2019	31.90	0.8916	28.31	0.7746	27.43	0.7309	25.46	0.7654
	FilterNeL	2020	31.74	0.8900	28.27	0.7730	27.39	0.7290	25.53	0.7680
	MADNet-L_F	2020	32.01	0.8925	28.45	0.7781	27.47	0.7327	25.77	0.7751
	RFDN	2020	32.24	0.8952	28.61	0.7819	27.57	0.7360	26.11	0.7858
	CFSRCNN	2020	32.06	0.8920	28.57	0.7800	27.53	0.7333	26.03	0.7824
	SeaNet-baseline	2020	32.18	0.8948	28.61	0.7822	27.57	0.7359	26.05	0.7896
	SMSR	2021	32.12	0.8932	28.55	0.7808	27.55	0.7351	26.11	0.7868
	MSFIN	2021	32.28	**0.8957**	28.57	0.7813	27.56	0.7358	26.13	0.7865
	Cross-SRN	2021	32.24	0.8954	28.59	0.7817	27.58	**0.7364**	26.16	0.7881
	MBMFN		**32.31**	0.8952	**28.68**	**0.7829**	**27.60**	0.7363	**26.26**	**0.7899**

From Table 4, we can see clearly that in the reconstruction tasks of ×2, ×3, ×4, compared with other advanced algorithms such as Cross-SRN and SMSR, our MBMFN has produced excellent results on PSNR and SSIM. In addition, under the ×4 reconstruction task, for PSNR, MBMFN has higher 0.07 dB, 0.09 dB and 0.10 dB than Cross-SRN on Set5, Set14 and Urban100 test sets, respectively. On Set14 and Urban100, MBMFN is 0.0012 and 0.0018 higher than Cross-SRN on SSIM, respectively. Compared with MADNet-L_F, the PSNR of our algorithm is 0.3 dB higher on Set5 and 0.0027 higher on SSIM. On Urban100, the PSNR of our algorithm is even 0.49 dB higher. This shows that the early-stage feature fusion and the information exchange between branches that we applied in MBMFN have a nonunderestimated contribution to improve the network performance. In particular, for the Urban100 test set containing a large number of detailed texture images, MBMFN is 0.034 higher than MSFIN and 0.031 higher than SMSR on SSIM, which shows that our MBMFN has great advantages in reconstructing high-frequency information such as edge texture.

In order to more intuitively verify that MBMFN has a high reconstruction ability for edge texture and other information, we visualize the reconstruction results under the ×4 reconstruction task, and the visual effect comparison is shown in Figure 4. On Set5 and Set14, our algorithm recovers more edge texture detail with relative clarity and hierarchy compared to other good algorithms. In the enlarged comparison of local images of Urban100_092 in Figure 4, the image reconstructed by CARN has serious line distortion and blur, and IMDN and RFDN also have different degrees of local line distortion. Although SMSR avoids line distortion, the reconstructed image has some blur. In contrast, our MBMFN restores more edge texture information of buildings. It is also closer to the original high-resolution image.

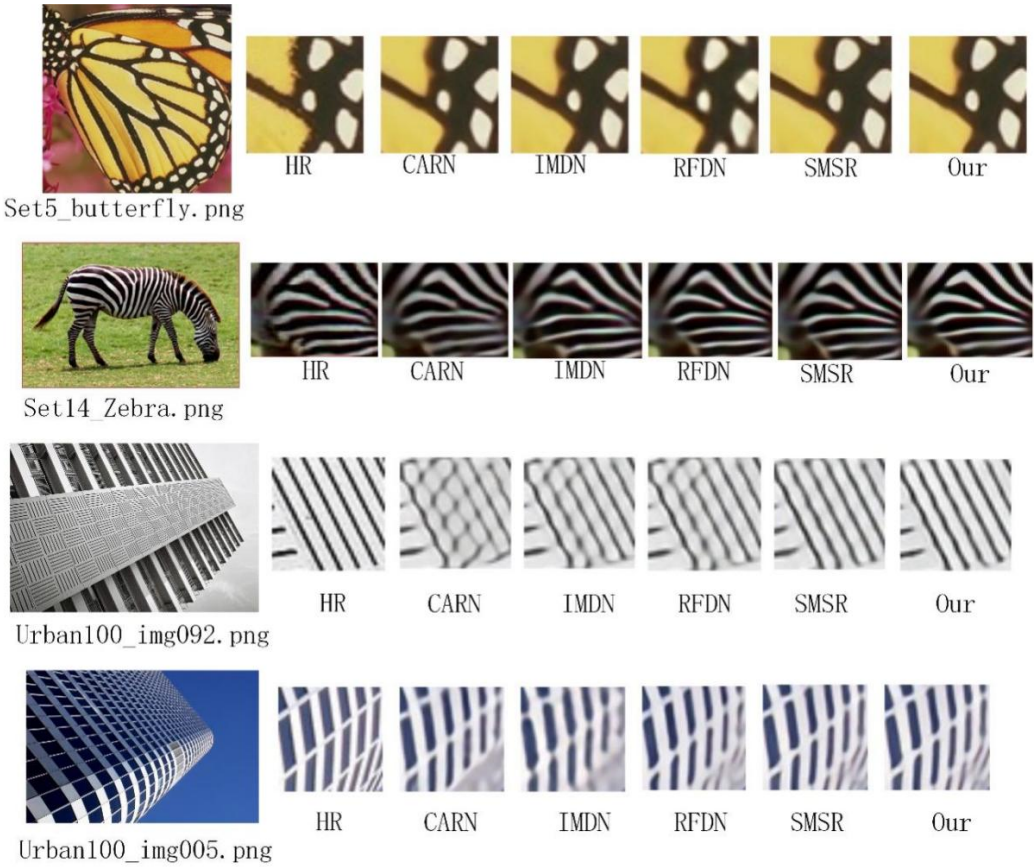

Figure 4. Visual comparison on test datasets (Set5_butterfly.png refers to the image named butterfly in the Set5 test set. Urban100_img092.png refers to the 092nd image in the Urban100 test set).

In addition, in order to more intuitively compare the relationship between the performance of each algorithm and network parameters, under ×4 reconstruction task, we visualized the corresponding relationship between PSNR and parameters of some algorithms on Set5 test set. As shown in Figure 5, compared with some other advanced SR algorithms, MBMFN achieves the best performance under the condition that the number of network parameters is kept within a reasonable range. In particular, MBMFN achieves a good balance between parameters and performance. This makes it possible to apply it to small devices with limited memory and computing.

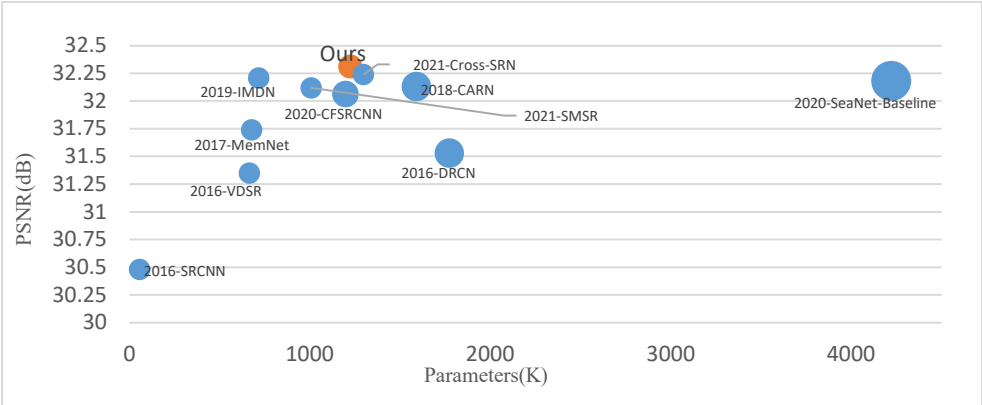

Figure 5. Comparison of the relationship between model parameters and performance (2016-SRCNN refers to the SRCNN algorithm that appeared in 2016).

5. Conclusions

In this paper, we proposed a multibranch feature multiplexing fusion network with mixed multilayer attention to realize SISR and design a semiparallel multibranch feature multiplexing fusion attention block, which processes the local features extracted in different stages through multiple branches in parallel and exchanges the feature information of different branches with the help of jump connection. It realizes the multiple utilization of local features and multistage feature fusion, expands the network receptive field and avoids the problem that the network is difficult to train due to the deepening of the network. In addition, we designed a lightweight enhanced residual channel attention block to improve the sensitivity of the network to information rich channels so as to learn more discriminative feature representation. We used the sum of global average and standard deviation of features to describe channel information more comprehensively and used a 1×1 convolution layer to directly model the relationship between channels, which avoids the loss of channel information caused by channel compression. This block can help the network recover more image details. In the reconstruction phase, we introduced an attention mechanism and adopted a distributed step-by-step up-sampling method to strengthen the network's attention and utilization of the high-frequency information in the features. Of note, few people study attention strategies during the reconstruction phase. We also used the weight-sharing strategy to reduce the number of network parameters. In addition, as far as we know, we should be the first to combine the weight-sharing strategy with the step-by-step up-sampling strategy in the reconstruction phase. A large number of experimental results show that our MBMFN achieved excellent performance in both objective indicators and subjective vision. This demonstrates that MBMFN has the potential for future work in the reuse of features and the utilization and fusion of multilevel features in the early stage.

In addition, the performance of our algorithm needs to be further improved and refined under further storage and computation constraints. In the future, we will further study lightweight and effective reconstruction algorithms so that they can be used on a large scale for small, mobile devices.

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

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. U1803261, Grant No. U1903213).

Data Availability Statement: Our training set DIV2k datasets can be obtained from available online: https://data.vision.ee.ethz.ch/cvl/DIV2K/ (accessed on 18 April 2022). Set5, Set14, B100, Urban 100 can be obtained from: https://arxiv.org/abs/1909.11856/ (accessed on 18 April 2022). In addition, in order to facilitate learning and reproduce the experimental results, our code can be obtained in the following link: https://github.com/Cai631/MBMFN (accessed on 18 April 2022).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Liu, Y.; Jia, Q.; Fan, X.; Wang, S.; Ma, S.; Gao, W. Cross-SRN: Structure-Preserving Super-Resolution Network with Cross Convolution. *IEEE Trans. Circuits Syst. Video Technol.* **2021**. [CrossRef]
2. He, Z.; Cao, Y.; Du, L.; Xu, B.; Yang, J.; Cao, Y.; Tang, S.; Zhuang, Y. MRFN: Multi-Receptive-Field Network for Fast and Accurate Single Image Super-Resolution. *IEEE Trans. Multimed.* **2019**, *22*, 1042–1054. [CrossRef]
3. Lan, R.; Sun, L.; Liu, Z.; Lu, H.; Pang, C.; Luo, X. MADNet: A Fast and Lightweight Network for Single-Image Super Resolution. *IEEE Trans. Cybern.* **2020**, *51*, 1443–1453. [CrossRef] [PubMed]

4. 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] [PubMed]
5. 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.
6. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
7. Liu, J.; Zhang, W.; 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, 13–19 June 2020; pp. 2359–2368.
8. Ma, X.; Guo, J.; Tang, S.; Qiao, Z.; Chen, Q.; Yang, Q.; Fu, S. DCANet: Learning connected attentions for convolutional neural networks. *arXiv* **2020**, arXiv:2007.05099.
9. Wang, L.; Dong, X.; Wang, Y.; Ying, X.; Lin, Z.; An, W.; Guo, Y. Exploring sparsity in image super-resolution for efficient inference. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4917–4926.
10. Hui, Z.; Wang, X.; Gao, X. Fast and accurate single image super-resolution via information distillation network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 723–731.
11. Hui, Z.; Gao, X.; Yang, Y.; Wang, X. Lightweight image super-resolution with information multi-distillation network. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2024–2032.
12. Liu, J.; Tang, J.; Wu, G. Residual feature distillation network for lightweight image super-resolution. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; Springer: Cham, Switzerland, 2020; pp. 41–55.
13. Wang, Z.; Gao, G.; Li, J.; Yu, Y.; Lu, H. Lightweight Image Super-Resolution with Multi-scale Feature Interaction Network. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; pp. 1–6.
14. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 201; pp. 7132–7141.
15. Roy, A.G.; Navab, N.; Wachinger, C. Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2018; pp. 421–429.
16. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
17. 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 *Computer Vision—ECCV 2020*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; pp. 191–207.
18. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, Tennessee, 19–25 June 2021; pp. 13713–13722.
19. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
20. Qin, J.; Huang, Y.; Wen, W. Multi-scale feature fusion residual network for single image super-resolution. *Neurocomputing* **2020**, *379*, 334–342. [CrossRef]
21. 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.
22. Behjati, P.; Rodriguez, P.; Mehri, A.; Hupont, I.; Tena, C.F.; Gonzalez, J. Hierarchical Residual Attention Network for Single Image Super-Resolution. *arXiv* **2020**, arXiv:2012.04578.
23. Mukherjee, S.; Valenzise, G.; Cheng, I. Potential of deep features for opinion-unaware, distortion-unaware, no-reference image quality assessment. In *International Conference on Smart Multimedia*; Springer: Cham, Switzerland, 2019; pp. 87–95.
24. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
25. Tai, Y.; Yang, J.; Liu, X.; Xu, C. MemNet: A Persistent Memory Network for Image Restoration. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 4539–4547.
26. Ahn, N.; Kang, B.; Sohn, K.A. Fast, accurate, and lightweight super-resolution with cascading residual network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 252–268.
27. Xie, C.; Zeng, W.; Lu, X. Fast single-image super-resolution via deep network with component learning. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 3473–3486. [CrossRef]
28. Li, F.; Bai, H.; Zhao, Y. FilterNet: Adaptive Information Filtering Network for Accurate and Fast Image Super-Resolution. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 1511–1523. [CrossRef]
29. Tian, C.; Xu, Y.; Zuo, W.; Zhang, B.; Fei, L.; Lin, C.-W. Coarse-to-fine CNN for image super-resolution. *IEEE Trans. Multimed.* **2020**, *23*, 1489–1502. [CrossRef]
30. Fang, F.; Li, J.; Zeng, T. Soft-edge assisted network for single image super-resolution. *IEEE Trans. Image Processing* **2020**, *29*, 4656–4668. [CrossRef] [PubMed]