

Article

Lightweight Dual Mutual-Feedback Network for Artificial Intelligence in Medical Image Super-Resolution

Beibei Wang ¹, Binyu Yan ^{1,*}, Gwanggil Jeon ^{2,*} , Xiaomin Yang ¹, Changjun Liu ¹ and Zhuoyue Zhang ¹ ¹ College of Electronics and Information Engineering, Sichuan University, Chengdu 610065, China² Department of Embedded Systems Engineering, Incheon National University, Incheon 22012, Republic of Korea

* Correspondence: yanby@scu.edu.cn (B.Y.); gjeon@inu.ac.kr (G.J.)

Abstract: As a result of hardware resource constraints, it is difficult to obtain medical images with a sufficient resolution to diagnose small lesions. Recently, super-resolution (SR) was introduced into the field of medicine to enhance and restore medical image details so as to help doctors make more accurate diagnoses of lesions. High-frequency information enhances the accuracy of the image reconstruction, which is demonstrated by deep SR networks. However, deep networks are not applicable to resource-constrained medical devices because they have too many parameters, which requires a lot of memory and higher processor computing power. For this reason, a lightweight SR network that demonstrates good performance is needed to improve the resolution of medical images. A feedback mechanism enables the previous layers to perceive high-frequency information of the latter layers, but no new parameters are introduced, which is rarely used in lightweight networks. Therefore, in this work, a lightweight dual mutual-feedback network (DMFN) is proposed for medical image super-resolution, which contains two back-projection units that operate in a dual mutual-feedback manner. The features generated by the up-projection unit are fed back into the down-projection unit and, simultaneously, the features generated by the down-projection unit are fed back into the up-projection unit. Moreover, a contrast-enhanced residual block (CRB) is proposed as each cell block used in projection units, which enhances the pixel contrast in the channel and spatial dimensions. Finally, we designed a unity feedback to down-sample the SR result as the inverse process of SR. Furthermore, we compared it with the input LR to narrow the solution space of the SR function. The final ablation studies and comparison results show that our DMFN performs well without utilizing a large amount of computing resources. Thus, it can be used in resource-constrained medical devices to obtain medical images with better resolutions.

Keywords: attention module; dual mutual feedback; lightweight; medical image super-resolution; unity feedback



Citation: Wang, B.; Yan, B.; Jeon, G.; Yang, X.; Liu, C.; Zhang, Z.

Lightweight Dual Mutual-Feedback Network for Artificial Intelligence in Medical Image Super-Resolution. *Appl. Sci.* **2022**, *12*, 12794. <https://doi.org/10.3390/app122412794>

Academic Editor: Cosimo Nardi

Received: 28 October 2022

Accepted: 9 December 2022

Published: 13 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The aim of SR is to learn a mapping function from input low-resolution (LR) images to output high-resolution (HR) images. High-resolution medical images are very important for doctors in terms of making accurate diagnoses of lesions; thus, SR for medical images has recently received a great deal of attention. However, image super-resolution remains a challenge, as LR images lose a certain amount of information as compared to HR images [1]. Many researchers have tried to find a solution to this critical issue [2–5].

On the basis of deep learning, Dong et al. proposed the SR convolutional neural network (SRCNN) [2], which utilizes the convolutional neural network (CNN) architecture and is vastly superior to other traditional methods. Thereafter, Dong et al. proposed fast SR convolutional neural networks (FSRCNNs) [5], which up-sample feature maps using deconvolution in the last layer of the network and provide more accurate estimates with less computation. The deconvolutional layer generates HR features by enlarging feature maps. Then, the subpixel convolutional layer was proposed by Shi et al. [6], which expands

the number of feature channels in order to store more pixels and rearrange them to generate HR features. The Laplacian pyramid super-resolution network (LapSRN) [7] up-samples LR feature maps progressively, which enables it to reconstruct multi-scale SR images in one training session.

To further improve SR performance, deep networks were introduced into SR. The very deep SR convolutional neural network (VDSR) [8] proposed by Kim et al. is the first deep multiple-scale model. It bypasses interpolated LR images to the end by residual learning. Then, on the basis of VDSR [8], the authors proposed a deeply recursive convolutional network (DRCN) [9], which trains the network using a recursive-supervised strategy and achieves a similar performance to VDSR [8] with fewer parameters. Deep dense SR (DDSR) [10] was proposed for the SR of medical images, which uses densely connected hidden layers to obtain informative high-level features.

However, it remains a challenge for deep neural networks to go deeper because of the various difficulties associated with training, such as gradient vanishing/exploding problems. Residual learning was proposed to solve these problems. The deep residual network (ResNet) [11] is a representative model, which achieves a remarkable performance based on residual learning. Tai et al. [12] used residual learning and recursive learning to realize a very deep network without an enormous amount of parameters. The SR network using dense skip connections (SRDenseNet) [13] is another representative model based on residual learning. It bypasses all previous features to latter layers in blocks and densely concatenates all blocks. The enhanced deep SR network (EDSR) [14] proposed by Lim et al. removes the use of batch normalization (BN), which is harmful to the final performance in SR tasks. EDSR also employs a pretraining strategy and residual scaling techniques to improve the final performance. On the basis of residual learning, for the SR of three-dimensional (3D) brain MRI images, Pham et al. [15] proposed a deep 3D CNN.

Above classical SR methods are all feedforward SR methods; low-frequency information is directly passed to the following layer or bypassed to the latter layers through skip connections. The feedback mechanism enables the previous layers to perceive high-frequency information from the latter layers, but no new parameters are introduced. It is widely used in the domain of computer vision [16–19]. Recently, Haris et al. [20] proposed error feedback for image SR, which was used in two back-projection units. Thereafter, the SR feedback network (SRFBN) [21] was proposed, which contains a feedback block that functions in a self-feedback manner. For the SR of medical images, the feedback adaptive weighted dense network (FAWDN) [22] was proposed based on an adaptive weighted dense block and feedback connection.

Although the feedback mechanism is used in some SR methods, it is rarely used in lightweight SR methods. The feedback mechanism enables the previous layers to perceive high-frequency information from latter layers, but no new parameters are introduced. Therefore, it is very applicable for lightweight networks. Moreover, most medical devices are resource-constrained, so lightweight feedback SR networks with good performance are desired. In order to meet the demand, a lightweight dual mutual-feedback network (DMFN) is proposed for artificial intelligence in medical image super-resolution. The DMFN feeds the HR features generated by the up-projection unit back into the down-projection unit, and feeds LR features generated by the down-projection unit back into the up-projection unit, which forms a dual mutual-feedback architecture, as shown in Figure 1. Our method that was trained using natural images is named DMFN, and our method that was trained using medical images is named DMFN+. They were tested on MRI13 from [22] and compared with other state-of-the-art SR methods, as shown in Figure 2. Our method performs very well with little computational cost.

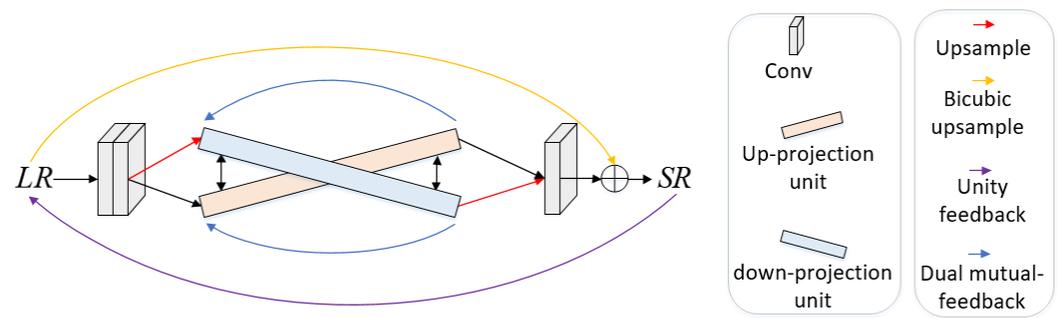


Figure 1. The structure of DMFN.

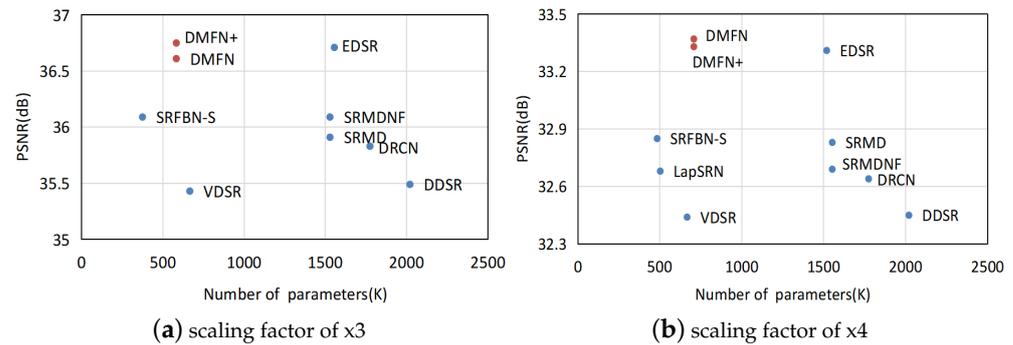


Figure 2. PSNR vs. parameters on MRI13 in [22].

Our contributions are summarized as follows:

- To better perceive the high-level information from each other, we designed a dual mutual-feedback structure. The HR features generated by the up-projection unit are fed back into the down-projection unit, and the LR features generated by the down-projection unit are fed back into the up-projection unit.
- To boost the expressive ability of the network, we propose a contrast-enhanced residual block (CRB) for use as each cell block in the projection units. CRB uses the contrast-enhanced channel and spatial attention within residual learning. The contrast-enhanced channel attention module learns the pixel contrast of each feature map to restore the textures, structures, and edges of images. The contrast-enhanced spatial attention module learns the pixel contrast in the same spatial location along the channel dimension to infer finer spatial-wise information.
- To narrow the search domain of the SR function, we designed a unity feedback. We down-sampled the SR result to LR image as the inverse process of SR. We then compared it with the input LR to calculate the unity feedback loss. The proposed unity feedback is helpful in terms of learning a better SR function with very few introduced parameters, which can be applied as a module to other SR networks.

2. Related Work

In this study, we designed a feedback network, which is inspired by SRFBN [21]. Moreover, inspired by [20], we used two back-projection units working in a dual mutual-feedback manner. Furthermore, we propose an attention-based module CRB for use as each cell block in the two back-projection units.

2.1. Attention Mechanism

The attention mechanism helps the networks perceive more informative features. Previously, the attention mechanism was used for image classification tasks [23] in RNN. Recently, inspired by the non-local means method, [24] learned the relationship between pixels with weighted sum t using long-range dependencies acquisition. Then, Hu et al. [25] learned the dependencies between channels with very little computational cost. The

residual channel attention block (RCAB) proposed in [26] first used channel attention within the residual block. The convolutional block attention module (CBAM) [27] enhanced the discriminate learning ability of the network with the help of channel and spatial attention. Hui et al. [28] proposed the contrast-enhanced channel attention (CCA) and argued that channel attention with standard deviation can better learn the interdependencies between feature channels.

Inspired by [28], we designed contrast-enhanced spatial attention, which learns the contrast of pixels in the spatial dimension to infer finer spatial-wise information in feature maps. Then, we used contrast-enhanced channel attention and spatial attention successively within the residual block, which is named the contrast-enhanced residual block (CRB).

2.2. Back-Projection

Irani et al. [29] used back-projection for image enhancements, which confirmed that iterative updates and down-sampling can minimize reconstruction error. Dai et al. [30] proposed bilateral back-projection for SR networks with a single LR input. Then Dong et al. [31] used iterative back-projection and incorporated non-local information to improve reconstruction performance. Timofte et al. [32] enhanced the reconstruction capabilities of learning-based SISR with the refinement of back-projection. Hairs et al. [20] learned the errors after up- and down-sampling to refine the intermediate features, which was used to realize up-projection and down-projection. The up- and down-projection units were then learned iteratively to further improve reconstruction performance.

Inspired by [20], we argue that mutual learning between two back-projection units will improve their performance, as it enhances the information exchange between the two. Further experimental results indicate that mutual learning between two back-projection units performs better than the existing independent learning methods.

2.3. Feedback Mechanism

In feedforward SR methods, the low-frequency information is directly passed to the following layer or is bypassed to the latter layers through skip connections. The feedback mechanism enables the previous layers to perceive the high-level information of latter layers, which is widely used in the domain of computer vision [16–19]. Recently, Hairs et al. [20] used error feedback in back-projection units to correct intermediate features. Then, Han et al. [33] designed a dual-state structure with delayed feedback to exchange signals between states. SRFBN [21] is a feedback network with a feedback block, which iteratively feeds the output features back to itself as the input.

Inspired by the above feedback methods, we used dual mutual feedback on two back-projection units, which feeds the HR features generated by the up-projection unit back into the down-projection unit, and feeds the LR features generated by the down-projection unit back into the up-projection unit. Our dual mutual feedback performs better than dual self-feedback and single feedback manners.

3. Method or Methodology

In this section, we present the overall architecture of DMFN, including the dual mutual-feedback component, the contrast-enhanced residual block (CRB) that is used as each cell block in the dual mutual-feedback component, and the loss function.

3.1. Architecture of DMFN

Similar to SRFBN [21], our DMFN can be unfolded into several iterations because of the feedback manner, and the iteration t is set from 1 to T . The back-projection units feed back their output results to each other iteratively in a dual mutual-feedback manner. As shown in Figure 3, two convolutional layers are firstly used to obtain shallow features, which are then up-sampled. Then, the shallow features and the up-sampled shallow features are learned by the dual mutual-feedback component. In the dual mutual-feedback component, the HR features generated by the up-projection unit are fed back into the

down-projection unit, and the LR features generated by the down-projection unit are fed back into the up-projection unit in the next iteration, which forms a dual mutual-feedback structure. Then, the outputs of dual mutual feedback from all iterations are concatenated for image reconstruction by fusing them with the bicubic interpolated results. Finally, we down-sampled the SR results to LR images as the inverse process of SR in the unity feedback component. Then, we compared it with the input LR to calculate the unity feedback loss.

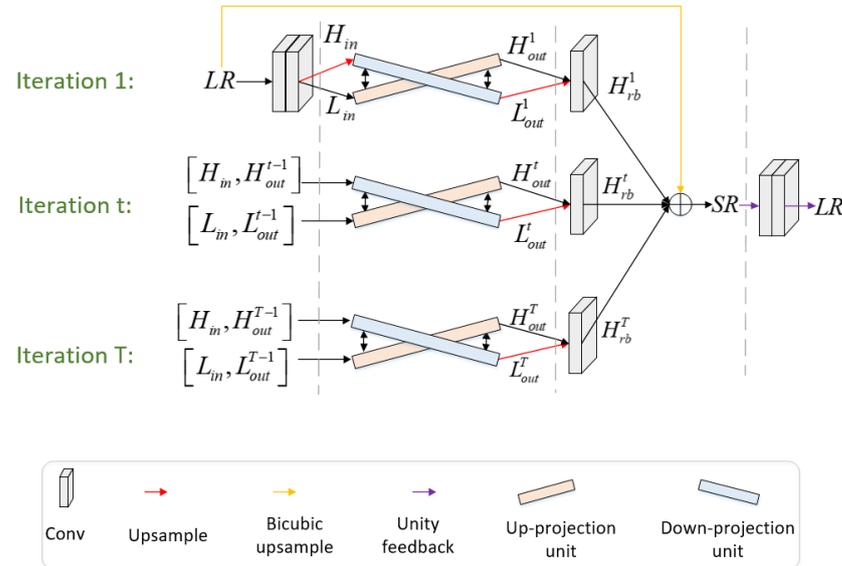


Figure 3. The unfolded DMFN.

We define L_{in} and H_{in} as the shallow features learned by the first component, which can be obtained by

$$\left\{ \begin{array}{l} L_{in} = f_c(LR) \\ H_{in} = f_{up}(L_{in}) \end{array} \right\}, \tag{1}$$

where f_c contains two convolutional layers to obtain shallow LR features. f_{up} is a deconvolutional upsampling operation.

In the dual mutual-feedback component of the t -th iteration, we use L_{out}^t to represent the LR features generated by the down-projection unit, and H_{out}^t to represent the HR features generated by the up-projection unit. The functions are as follows:

$$L_{out}^t = \left\{ \begin{array}{ll} f_{d-p}(H_{in}) & t = 1 \\ f_{d-p}([H_{in}, H_{out}^{t-1}]) & t \geq 2 \end{array} \right\}, \tag{2}$$

$$H_{out}^t = \left\{ \begin{array}{ll} f_{u-p}(L_{in}) & t = 1 \\ f_{u-p}([L_{in}, L_{out}^{t-1}]) & t \geq 2 \end{array} \right\}, \tag{3}$$

where f_{d-p} are the operations of the down-projection unit, which contains some features from the up-projection unit because of the mutual learning between the two back-projection units. f_{u-p} is the operations of the up-projection unit, which also contains some features from the down-projection unit. $[]$ is the concat function.

For reconstruction, we up-scale the LR features generated by the down-projection unit, which are then fused with the HR features generated by the up-projection unit. We define the final HR feature results of the t -th iteration as follows:

$$H_{rb}^t = H_{out}^t + f_{up}(L_{out}^t). \tag{4}$$

Since the final HR features of all iterations are fused and then added to the bicubic interpolated result of LR input, the final SR result is as follows:

$$SR = f_{cm}([H_{rb}^1, H_{rb}^2, \dots, H_{rb}^T]) + f_{BC}(LR), \tag{5}$$

where f_{cm} is a conv-3 compression layer, and f_{BC} represents the bicubic up-sample function.

Finally, we down sample the SR result to the LR image named LR' by the down-sampling function f_{down} , which contains two convolutional layers for down sampling and channel transformation. The unity feedback loss calculated by LR' and LR is used to narrow the search domain of the SR function.

$$LR' = f_{down}(SR), \tag{6}$$

3.2. Dual Mutual-Feedback Component

The dual mutual-feedback component of the t -th iteration is shown in Figure 4. Pink represents the up-projection unit, and blue represents the down-projection unit. Then, we unfold the two back-projection units. The upward arrows represent the up-sampling operation, and the downward arrows represent the down-sampling operation. The pink arrows connect to an up-projection unit, and the blue arrows connect to a down-projection unit. Then, we use mutual learning (black arrows) between the two back-projection units to exchange information. Finally, the outputs of the two units are fed back into each other in the next iteration to realize dual mutual feedback.

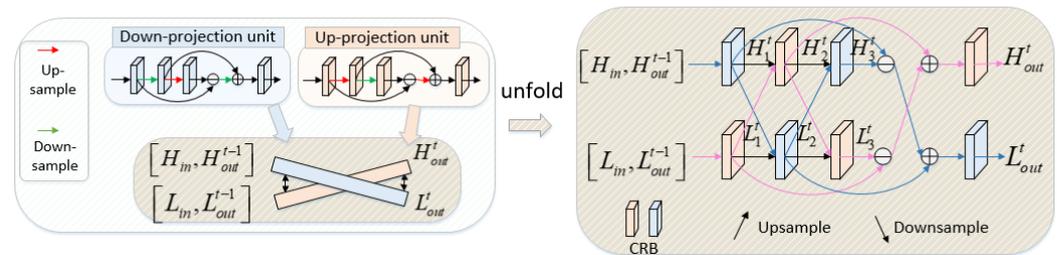


Figure 4. Dual mutual-feedback component of the t -th iteration in DMFN.

In the dual mutual-feedback component of the t -th iteration, we define the HR features as H_1^t, H_2^t and H_3^t , and the LR features as L_1^t, L_2^t and L_3^t . We use f_{CRB} to represent the operations of CRB. The dual mutual-feedback procedure is as follows:

$$L_1^t = \begin{cases} f_{CRB}(L_{in}) & t = 1 \\ f_{CRB}([L_{in}^t, L_{out}^{t-1}]) & t \geq 2 \end{cases}, \tag{7}$$

$$H_1^t = \begin{cases} f_{CRB}(H_{in}) & t = 1 \\ f_{CRB}([H_{in}^t, H_{out}^{t-1}]) & t \geq 2 \end{cases}, \tag{8}$$

$$\begin{cases} L_2^t = f_{CRB}([L_1^t, f_{down}(H_1^t)]) \\ H_2^t = f_{CRB}([H_1^t, f_{up}(L_1^t)]) \end{cases}, \tag{9}$$

$$\begin{cases} L_3^t = f_{CRB}([L_2^t, f_{down}(H_2^t)]) \\ H_3^t = f_{CRB}([H_2^t, f_{up}(L_2^t)]) \end{cases}, \tag{10}$$

$$\begin{cases} L_{out}^t = f_{CRB}(L_3^t + f_{down}(H_3^t - H_1^t)) \\ H_{out}^t = f_{CRB}(H_3^t + f_{up}(L_3^t - L_1^t)) \end{cases}. \tag{11}$$

3.3. Contrast-Enhanced Residual Block (CRB)

To further boost the expressive ability of our network, we propose a contrast-enhanced residual block (CRB), which is used as each cell block of the dual mutual-feedback component, as shown in Figure 5. CRB uses contrast-enhanced channel attention and spatial attention within the residual block. Contrast-enhanced channel attention assigns different weights to channels, and contrast-enhanced spatial attention assigns different weights to spatial locations. Therefore, the feature learning ability of residual blocks is enhanced. As shown in Figure 5, the input features F_{in} are learned by a multi-layer perceptron f_{mlp}

(Conv-ReLU-Conv) and then are learned by contrast-enhanced channel attention. The input of contrast-enhanced channel attention X is shown below:

$$X = f_{mlp}(F_{in}). \tag{12}$$

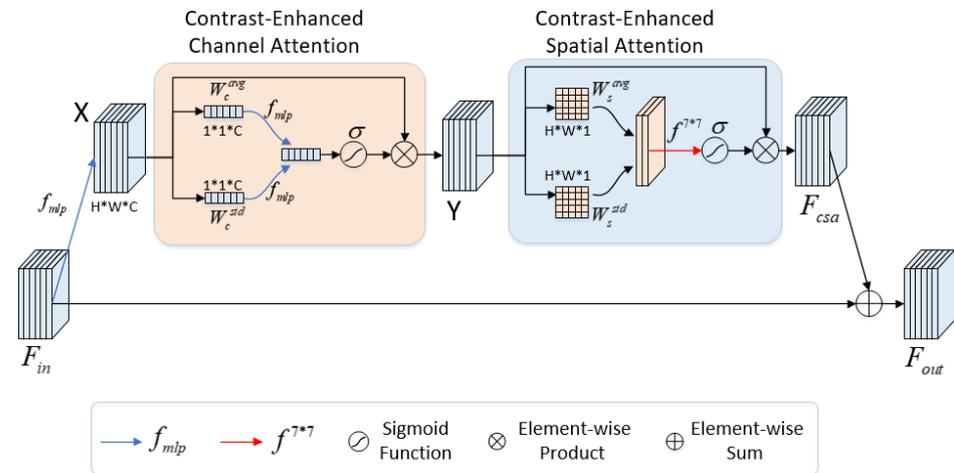


Figure 5. Contrast-enhanced residual block (CRB).

3.3.1. Contrast-Enhanced Channel Attention

As is the case for CCA [28], both standard deviation and average pooling are used to describe the context of each channel. Standard deviation enables the network to perceive more channels with a greater pixel contrast, as it represents image details related to structures, textures, and edges. Average pooling enables the network to perceive more informative channels. The size of feature maps X is $H \times W \times C$ and we use $c \in (1, \dots, C)$ to represent the channel number. We use $i \in (1, \dots, H)$ and $j \in (1, \dots, W)$ to represent the pixel location in each feature map. The weights of each channel calculated by average pooling and standard deviation are shown below:

$$w_c^{avg} = \frac{1}{HW} \sum_{(i,j) \in x_c} x_c^{i,j}, \tag{13}$$

$$w_c^{std} = \sqrt{\frac{1}{HW} \sum_{(i,j) \in x_c} (x_c^{i,j} - \frac{1}{HW} \sum_{(i,j) \in x_c} x_c^{i,j})^2}, \tag{14}$$

Then, we use $W_c^{avg} \in R^{1 \times 1 \times C}$ and $W_c^{std} \in R^{1 \times 1 \times C}$ to represent the average-pooled and standard deviation results of X on all channels. They are learned by f_{mlp} , and normalized with the application of the sigmoid function. Finally, the input feature maps X are rescaled by the element-wise product. The features learned by contrast-enhanced channel attention are shown below:

$$Y = X * \sigma(f_{mlp}(W_c^{avg}) + f_{mlp}(W_c^{std})). \tag{15}$$

3.3.2. Contrast-Enhanced Spatial Attention

We argue that the standard deviation value in the spatial dimension indicates the pixel contrast in the same spatial location along the channel dimension. The pixels with a higher standard deviation value must have a higher information value in some channels, which should be given more attention. Average pooling enables the network to perceive more informative spatial locations along the channel dimension. Therefore, both standard deviation and average pooling are used to describe the pixel weights in the spatial dimension, which enhances the image details. The size of feature maps Y is $H \times W \times C$ and we use $c \in (1, \dots, C)$ to represent the channel number. We use $i \in (1, \dots, H)$ and

$j \in (1, \dots, W)$ to represent the pixel location in each feature map. The weights of each spatial location calculated by average pooling and standard deviation are shown below:

$$w_{i,j}^{avg} = \frac{1}{C} \sum_{c=1}^C y_c^{i,j}. \quad (16)$$

$$w_{i,j}^{std} = \sqrt{\frac{1}{C} \sum_{c=1}^C (y_c^{i,j} - \frac{1}{C} \sum_{c=1}^C y_c^{i,j})^2}. \quad (17)$$

Then, we use $W_s^{avg} \in R^{H \times W \times 1}$ and $W_s^{std} \in R^{H \times W \times 1}$ to represent the average-pooled and standard deviation results of Y across the channel. They are learned using $f^{7 \times 7}$ (conv-7 layer), which is helpful to identify important spatial locations, and normalized with the application of the sigmoid function. Finally, it rescales the input Y using the element-wise product. The features learned by contrast-enhanced spatial attention are shown below:

$$F_{csa} = Y * \sigma(f^{7 \times 7}[W_s^{avg}, W_s^{std}]) \quad (18)$$

Finally, because CRB is a residual block, F_{out} , as the output of CRB, can be obtained by

$$F_{out} = F_{csa} + F_{in}. \quad (19)$$

3.4. Loss Function

We designed a unity feedback method that down-samples the SR result SR to LR images LR' , as the inverse process of SR. Then, we compared it with the input LR to obtain the unity feedback loss, which can be used to narrow the search domain of the SR function. We chose the L1 loss function and used w to represent the weight of the unity feedback loss. Accordingly, our loss function is as follows:

$$Loss = \|SR, HR\|_1 + w \|LR', LR\|_1, \quad (20)$$

4. Experimental Results

In this section, we first introduce the setting of our experiments. Then, we present the experiments and analyze the results to prove the effectiveness of our methods, which include unity feedback, dual mutual-feedback feedback, mutual learning, the concat function for SR reconstruction, and CRB.

4.1. Setting

4.1.1. Datasets

First, we trained our DMFN with the DIV2k [34] dataset and validated it with Set5 [35]. This was used to compare it with other SR methods trained using natural images. Moreover, the corresponding ablation models were trained with the DIV2k [34] dataset and validated with Set5 [35]. Then, as is the case with FAWDN [22], we trained our network with the medical image dataset MRIMP and validated it with MRI13, which was used in [22]. This was named DMFN+. Finally, all the comparison results were tested with three medical image datasets: the MRI13 dataset in [22], ADNI100 [36] dataset and OASIS100 [37] dataset.

The DIV2k [34] dataset contains 800 training images, which have a resolution of 2K. We increased the number of images 10-fold through rotation and cropping. The medical image dataset MRIMP in FAWDN [22] contains 1444 training images. These were obtained using GoogleMR by crawling the keywords IXI [38], ADNI [36], KneeMR [39], and LSMRI [40]. LR images come from the bicubic down-sampling of HR images. All experimental results were obtained on a GPU under the PyTorch framework.

4.1.2. Implementation

The Adam optimizer was employed to train our network. We set the initial learning rate (lr) to 0.0005, the epochs to 1000 (halved every 200 epochs), the batch size to 16, and the base filter number to 32. As is the case for SRFBN-S [21], we performed the dual mutual-feedback procedure four times.

4.2. Effectiveness of Unity Feedback

In our study, we designed a unity feedback to narrow the search domain of the SR function, so our loss function contains two parts, as shown in Equation (20). We used the SR loss and unity feedback loss to train our DMFN, and we used w to represent the weight assigned to the unity feedback loss. In this experiment, we increase the weight of unity feedback loss w from 0 to 1 to obtain the best trade-off. Then, we compare the SR results of DMFN with different weights in Table 1. The unity feedback improved the performance of our DMFN, and performed best when $w = 0.1$. Therefore, we set the weight of unity feedback loss to 0.1 to supervise the training of our methods.

Table 1. Comparison of different weights assigned to unity feedback loss on DMFN.

w	Scale	Params	MRI13	ADNI100	OASIS100
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
0	×3	583 k	36.58/0.9480	30.13/0.8716	34.22/0.9475
0.01		583 k	36.61/0.9482	30.15/0.8718	34.36/0.9479
0.1		583 k	36.72/0.9485	30.17/0.8719	34.35/0.9478
1		583 k	36.54/0.9480	30.14/0.8715	34.33/0.9477

4.3. Effectiveness of Dual Mutual Feedback

For evaluating the effectiveness of dual mutual feedback, we performed ablation studies for dual feedback, mutual feedback, and feedback. The ablation architectures are shown in Figure 6. First, to realize the ablation of dual feedback, we used feedback with down sampling, as shown in Figure 6b. Therefore, it performed two more up-sampling steps and one more down-sampling step than the proposed DMFN in every feedback session. Second, to realize the ablation of mutual feedback, the architecture has one more up-sampling and down-sampling step than the proposed DMFN in each feedback, as shown in Figure 6c. Finally, to realize the ablation of feedback, we used a down-sampling operation after the up-projection unit to build the feedforward architecture, and up-sampled all features at the last layer, as shown in Figure 6d. All the ablation architectures have more parameters and a higher computational complexity than the proposed DMFN.

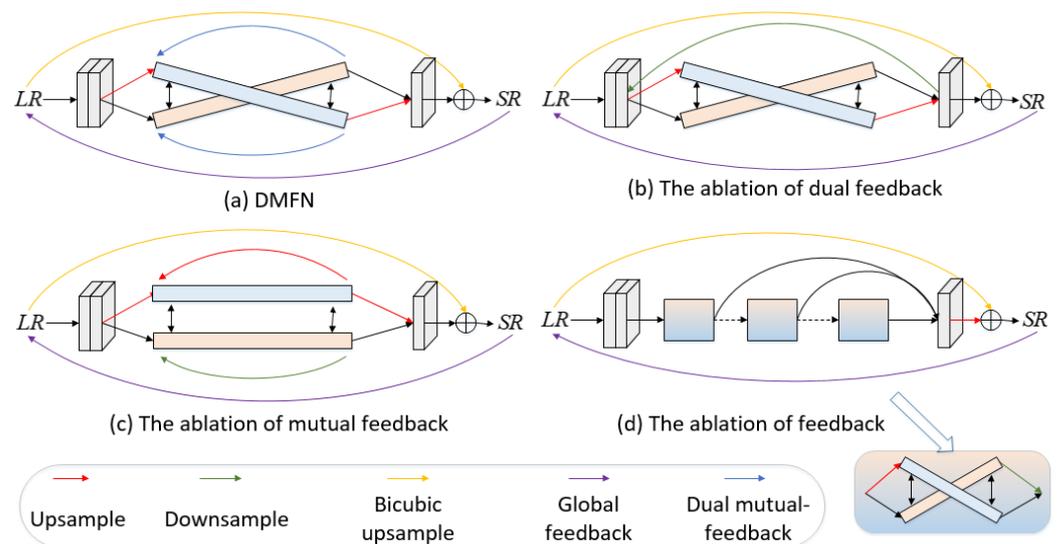


Figure 6. The ablation architectures. (a) is the DMFN. (b) represents the ablation of dual feedback. (c) represents the ablation of mutual feedback. (d) represents the ablation of feedback.

The experimental results are shown in Table 2. First, the ablation of dual feedback did not perform well. This is because the input of the two units is so similar that the information exchange between them cannot function adequately. Second, the ablation

of mutual feedback also exhibited a poorer performance than DMFN. This is because it has less information exchange for its self-feedback architecture, and the additional up sampling and down sampling are not directly used for SR reconstruction. Finally, DMFN with feedback manner has fewer parameters but performs better than the feedforward manner. This is because the feedback mechanism enables the previous layers to perceive high-level information of latter layers.

Table 2. The ablation studies of dual feedback, mutual feedback, and feedback on DMFN.

Dual	Mutual	Feedback	Scale	Params	MRI13	ADNI100	OASIS100
					PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
✓	✓	✓	×3	583 k	36.72/0.9485	30.17/0.8719	34.35/0.9478
×	✓	✓		631 k	36.49/0.9479	30.13/0.8716	34.24/0.9475
✓	×	✓		687 k	36.58/0.9480	30.14/0.8716	34.31/0.9477
✓	✓	×		633 k	36.53/0.9479	30.14/0.8714	34.35/0.9477

4.4. Ablation Study of Mutual Learning between Two Back-Projection Units

Inspired by [20], we argue that mutual learning between the two back-projection units will improve their performance, as it facilitates information exchange. We performed an ablation study for the mutual-learning method, in which two back-projection units were learned independently, as shown in Figure 7. As illustrated in Table 3, the mutual-learning method performed better than the independent-learning method.

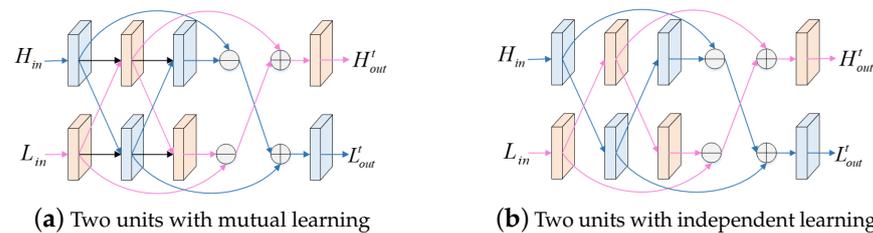


Figure 7. The comparison of our mutual-learning method and the independent-learning method between the two back-projection units. The black arrows show the mutual learning between two units, which are not used in independent-learning.

Table 3. Comparisons of mutual learning and independent learning on DMFN.

Architecture	Scale	Params	MRI13	ADNI100	OASIS100
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
DMFN with mutual learning	×3	583 k	36.72/0.9485	30.17/0.8719	34.35/0.9478
DMFN with independent-learning		574 k	36.55/0.9479	30.13/0.8713	34.31/0.9476

4.5. Ablation Study of the Concat Function for SR Reconstruction

Certain multi-branch methods [7,9,21] reconstruct the SR image using multi-prediction, such as SRFBN [21], which generates a prediction in each feedback procedure. However, we argue that previous feedback procedures cannot produce a meaningful prediction as a result of their very shallow HR features. Accordingly, we concatenated the HR features of all feedback procedures to obtain the final SR result. We compared our concat function and the multi-prediction method using the DMFN, as shown in Figure 8. As illustrated in Table 4, the concat function performed better than the multi-prediction method on the DMFN. Therefore, the concat function was shown to be effective and applicable to feedback networks.

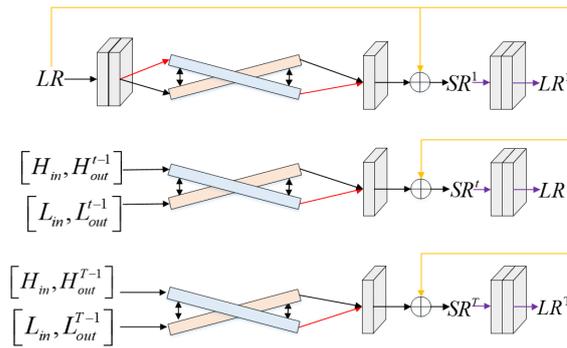


Figure 8. DMFN without the concat function, which degrades into the multi-prediction method.

Table 4. Comparison of the concat function and multi-prediction used on DMFN.

Reconstruction Method	Scale	Params	MRI13	ADNI100	OASIS100
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
concat function	×3	583 k	36.72/0.9485	30.17/0.8719	34.35/0.9478
multi-prediction		582 k	36.24/0.9468	30.08/0.8708	34.07/0.9465

4.6. Improvement of CRB

A contrast-enhanced residual block (CRB) was designed by fusing both the contrast-enhanced channel and spatial attention within residual learning, which is used in the dual mutual-feedback component as each cell block. In this experiment, we replaced the CRB with several attention-based modules used in existing methods to evaluate their effectiveness, such as CBAM [27], RCAB [26], and CCA [28]. Figure 9 shows a comparison of these attention models. Our CRB and CBAM [27] contain both channel and spatial attention, while CCA [28] and RCAB [26] are models based on the channel attention. We used the above attention models on our DMFN to compare their performance. They are denoted as DMFN-CBAM, DMFN-RCAB and DMFN-CCA. As illustrated in Table 5, our CRB performed better than the above attention-based models. Therefore, our CRB was shown to be efficient and to improve SR performance.

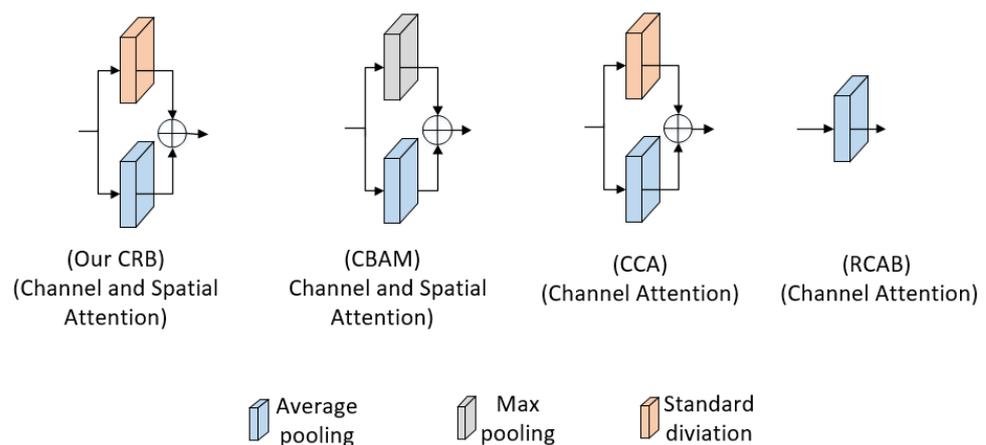


Figure 9. Existing attention models [26–28].

Table 5. Comparison of existing attention models and our CRB on DMFN.

Attention Module	Scale	Params	MRI13	ADNI100	OASIS100
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
DMFN	×3	583 k	36.72/0.9485	30.17/0.8719	34.35/0.9478
DMFN-CBAM		583 k	36.40/0.9476	30.13/0.8721	34.21/0.9472
DMFN-RCAB		582 k	36.47/0.9478	30.13/0.8718	34.24/0.9473
DMFN-CCA		582 k	36.56/0.9479	30.14/0.8717	34.29/0.9475

4.7. Comparison with Classical SR Methods

Our DMFN was trained using DIV2K [34]. Therefore, we compared it with other classical SR methods trained using natural images, such as SRCNN [2], FSRCNN [5], VDSR [8], DRCN [9], LapSRN [7], SRDenseNet [13], DDSR [10], EDSR [14], SRMD [41], SRMDNF [41], SRFBN-S [21] and FAWDN [22]. Moreover, we compared our DMFN+ with FAWDN+ [22], which was trained using the images of MRIMP and part of DIV2K [34], as the network suffered from overfitting when trained with MRIMP. Our DMFN+ was trained with MRIMP, and there was no overfitting, which demonstrates the stability of our method. As illustrated in Table 6, we compared their PSNR and SSIM values, and our DMFN was shown to perform better than the other natural image SR methods with fewer parameters. Furthermore, our DMFN+ demonstrated a better performance than the medical image SR methods with fewer parameters.

Finally, a visual comparison of SR medical images was performed, as shown in Figure 10. For ADNI100 [36], the performance of DMFN+ was the best, followed by EDSR [14]. For OASIS100 [37], the performance of DMFN+ was the best, followed by DMFN. However, DMFN+ requires fewer than half the parameters of EDSR [14], so our DMFN and DMFN+ provide a better trade-off. In summary, our methods recover image details and textures better than most of the other methods.

Table 6. Comparison of PSNR/SSIM for different scale factors on the MRI13 [22], ADNI100 [36], and OASIS100 [37] datasets. The red and blue represent the best and second-best results, respectively.

Methods	Scale	Params	MRI13	ADNI100	OASIS100
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic		-	37.95/0.9677	30.74/0.8943	33.54/0.9585
SRCNN [2]		8 K	40.76/0.9820	32.49/0.9258	35.52/0.9717
FSRCNN [5]		13 K	40.90/0.9809	40.90/0.9809	34.47/0.7778
VDSR [8]		666 K	41.78/0.9835	33.09/0.9328	36.70/0.9759
DRCN [9]		1774 K	42.15/0.9838	33.22/0.9336	37.19/0.9779
LapSRN [7]		251 K	41.99/0.9840	32.96/0.9317	36.98/0.9772
SRDenseNet [13]		7160 K	42.97/0.9846	33.33/0.9348	37.69/0.9788
DDSR [12]	×2	2020 K	41.92/0.9833	33.00/0.9318	36.97/0.9766
EDSR [14]		1370 K	43.61/0.9853	33.50/0.9359	38.18/0.9797
SRMD [41]		1511 K	42.26/0.9841	33.15/0.9335	37.27/0.9780
SRMDNF [41]		1511 K	42.76/0.9848	33.30/0.9348	37.69/0.9794
SRFBN-S [21]		282 K	42.77/0.9843	33.29/0.9342	38.88/0.9806
FAWDN [22]		7170 K	43.35/0.9850	33.41/0.9352	37.91/0.9791
FAWDN+ [22]		7170 K	43.59/0.9851	33.87/0.9400	38.10/0.9798
DMFN(ours)		475 K	43.38/0.9850	33.41/0.9353	39.27/0.9813
DMFN+(ours)		475 K	43.57/0.9851	33.84/0.9397	39.43/0.9818

Table 6. Cont.

Methods	Scale	Params	MRI13	ADNI100	OASIS100
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic		-	32.70/0.9126	27.95/0.8063	28.97/0.8919
SRCNN [2]		8 K	35.03/0.9393	29.29/0.8527	31.27/0.9225
FSRCNN [5]		13 K	35.35/0.9406	29.48/0.8562	30.91/0.7568
VDSR [8]		666 K	35.43/0.9449	29.83/0.8661	31.60/0.9308
DRCN [9]		1774 K	35.83/0.9449	29.92/0.8672	32.18/0.9371
SRDenseNet [13]		7160 K	36.25/0.9469	30.06/0.8704	32.64/0.9477
DDSR [12]	×3	202 0K	35.49/0.9433	29.68/0.8638	31.72/0.9330
EDSR [14]		1555 K	36.71/0.9484	30.27/0.8735	33.26/0.9444
SRMD [41]		1528 K	35.91/0.9456	29.91/0.8676	32.29/0.9385
SRMDNF [41]		1528 K	36.09/0.9465	30.02/0.8698	32.52/0.9405
SRFBN-S [21]		375 K	36.09/0.9459	29.99/0.8691	33.79/0.9453
FAWDN [22]		7170 K	36.60/0.9481	30.16/0.8719	33.00/0.9429
FAWDN+ [22]		7170 K	36.73/0.9479	30.75/0.8839	33.19/0.9450
DMFN(ours)		583 K	36.72/0.9485	30.17/0.8719	34.35/0.9478
DMFN+(ours)		583 K	36.75/0.9482	30.76/0.8843	34.41/0.9493
<hr/>					
Bicubic		-	29.90/0.8591	26.37/0.7298	26.37/0.7298
SRCNN [2]		8 K	31.75/0.8914	27.49/0.7838	28.47/0.8621
FSRCNN [5]		13 K	32.17/0.8934	27.66/0.7876	28.49/0.6900
VDSR [8]		666 K	32.44/0.9027	28.00/0.8028	28.96/0.8748
DRCN [9]		1774 K	32.64/0.9034	28.04/0.8029	29.29/0.8830
LapSRN [7]		502 K	32.68/0.9072	27.99/0.8027	29.26/0.8837
SRDenseNet [13]		7160 K	32.97/0.9075	28.18/0.8079	29.65/0.8900
DDSR [12]	×4	2020 K	32.45/0.9029	27.91/0.8011	29.15/0.8798
EDSR [14]		1518 K	33.31/0.9107	28.42/0.8136	30.48/0.9003
SRMD [41]		1552 K	32.83/0.9060	28.05/0.8044	29.66/0.8884
SRMDNF [41]		1552 K	32.69/0.9062	28.13/0.8080	28.13/0.8080
SRFBN-S [21]		483 K	32.85/0.9069	28.20/0.8086	31.01/0.9006
FAWDN [22]		7170 K	33.22/0.9098	28.30/0.8117	30.05/0.8957
FAWDN+ [22]		7170 K	33.21/0.9086	28.81/0.8259	30.38/0.8895
DMFN(ours)		707 K	33.37/0.9104	28.30/0.8117	31.47/0.9051
DMFN+(ours)		707 K	33.33/0.9094	28.94/0.8310	31.54/0.9079

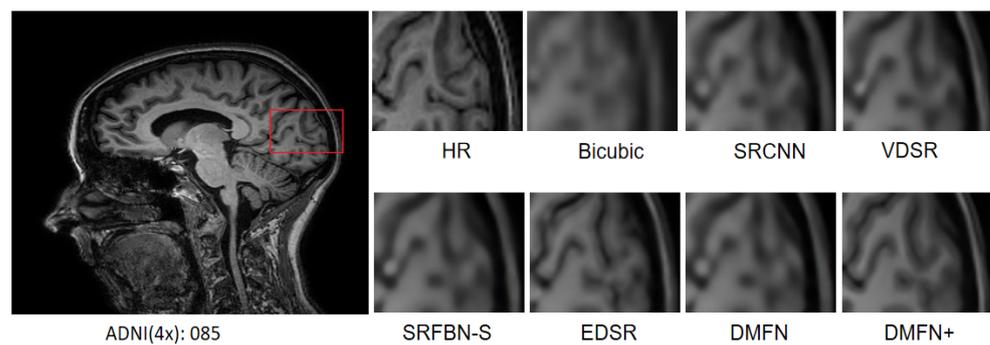


Figure 10. Cont.

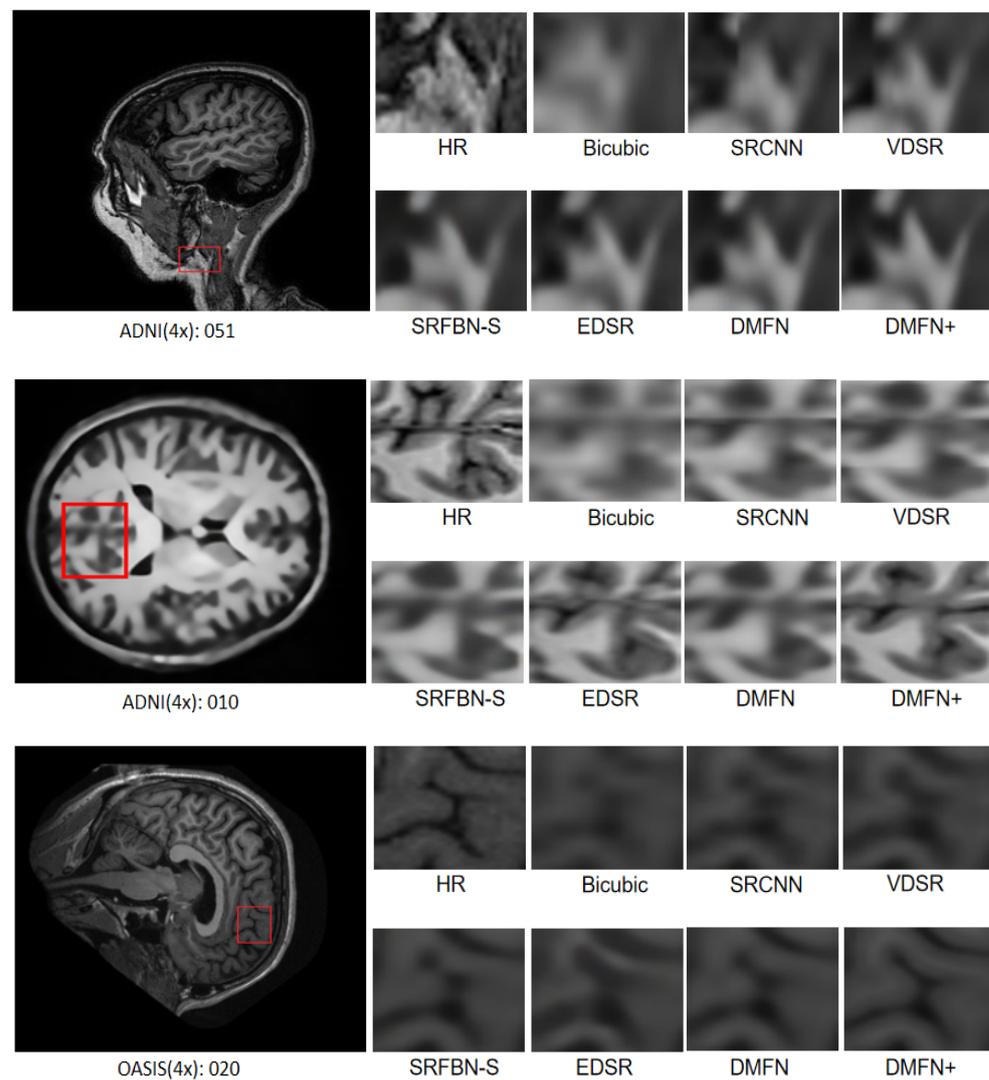


Figure 10. Comparison of visualization results on ADNI100 [36] and OASIS100 [37] datasets. Images on the right are the recovery details of the red box.

5. Conclusions

In this paper, a lightweight dual mutual-feedback network (DMFN) is proposed for use in artificial intelligence in medical image super-resolution. It contains two back-projection units working in a dual mutual-feedback manner. We propose a contrast-enhanced residual block (CRB), which is used in the back-projection units as each cell block. The CRB uses the contrast-enhanced channel and spatial attention within residual learning to enhance its ability to express details. We used the concat function for SR image reconstruction. Finally, a unity feedback method was designed to supervise the process of SR, which down-sampled the SR result to LR images as the inverse process of SR. As illustrated in the experimental results, our DMFN outperformed the other methods with very little computational cost. Accordingly, our method can help doctors to make accurate diagnoses by improving the resolution of medical images. The DMFN introduces a feedback mechanism into medical image SR and was shown to exhibit good performance on synthetic datasets. However, we are not sure whether it will perform well in real-world medical image SR, as the degradation of real-world images is complicated. In the future, we will focus our attention on real-world medical image SR, as it is possible that the feedback mechanism can also be used to improve performance in this scenario.

Author Contributions: B.W.: Conceptualization, Methodology, Software, Writing—Original draft; B.Y.: Supervision, Validation; G.J.: Supervision, Formal analysis; C.L.: Data Curation, Resources; X.Y.: Writing—Review, Editing, Funding acquisition; Z.Z.: Investigation, Visualization. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Sichuan University and Yibin Municipal People’s Government University and City strategic cooperation special fund project (Grant No. 2020CDYB-29); Science and Technology plan transfer payment project of Sichuan province (2021ZYSF007); The Key Research and Development Program of Science and Technology Department of Sichuan Province (No. 2020YFS0575, No.2021KJT0012-2021YFS0067); The funding from Science Foundation of Sichuan Science and Technology Department (2021YFH0119); The funding from Sichuan University (Grant No. 2020SCUNG205) and National Natural Science Foundation of China under Grant No. 62201370.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Dmitry Ulyanov, A.V.; Lempitsky, V. Deep image prior. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
2. 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 (ECCV), Zurich, Switzerland, 6–12 September 2014.
3. 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.
4. Fu, S.; Lu, L.; Li, H.; Li, Z.; Wu, W.; Paul, A.; Jeon, G.; Yang, X. A Real-Time Super-Resolution Method Based on Convolutional Neural Networks. *Circuits Syst. Signal Process.* **2020**, *39*, 805–817. [[CrossRef](#)]
5. Dong, C.; Loy, C.C.; Tang, X. Accelerating the Super-Resolution Convolutional Neural Network. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 8–16 October 2016.
6. 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 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
7. 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 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
8. Kim, J.; Lee, J.; Lee, K. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
9. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-Recursive Convolutional Network for Image Super-Resolution. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
10. Wei, S.; Wu, W.; Jeon, G.; Ahmad, A.; Yang, X. Improving resolution of medical images with deep dense convolutional neural network. *Concurr. Comput. Pract. Exp.* **2018**, *32*, e5084. [[CrossRef](#)]
11. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
12. Tai, Y.; Yang, J.; Liu, X. Image Super-Resolution via Deep Recursive Residual Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
13. Tong, T.; Li, G.; Liu, X.; Gao, Q. Image Super-Resolution Using Dense Skip Connections. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
14. Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
15. Pham, C.H.; Ducournau, A.; Fablet, R.; Rousseau, F. Brain MRI super-resolution using deep 3D convolutional networks. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI), Melbourne, Australia, 18–21 April 2017.
16. Gilbert, C.D.; Sigman, M. Brain States: Top-Down Influences in Sensory Processing. *Neuron* **2007**, *54*, 677–969. [[CrossRef](#)] [[PubMed](#)]
17. Stollenga, M.; Masci, J.; Gomez, F.; Schmidhuber, J. Deep Networks with Internal Selective Attention through Feedback Connections. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014.
18. Zamir, A.R.; Wu, T.L.; Sun, L.; Shen, W.; Shi, B.E.; Malik, J.; Savarese, S. Feedback Networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.

19. Carreira, J.; Agrawal, P.; Fragkiadaki, K.; Malik, J. Human Pose Estimation with Iterative Error Feedback. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
20. Haris, M.; Shakhnarovich, G.; Ukita, N. Deep Back-Projection Networks for Super-Resolution. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
21. Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; Wu, W. Feedback Network for Image Super-Resolution. In Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
22. Chen, L.; Yang, X.; Jeon, G.; Anisetti, M.; Liu, K. A Trusted Medical Image Super-Resolution Method based on Feedback Adaptive Weighted Dense Network. *Artif. Intell. Med.* **2020**, *106*, 101857. [[CrossRef](#)] [[PubMed](#)]
23. Mnih, V.; Heess, N.; Graves, A.; Kavukcuoglu, K. Recurrent models of visual attention. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Montreal, QC, Canada, 8–13 December 2014.
24. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local Neural Networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
25. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
26. 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.
27. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I. CBAM: Convolutional Block Attention Module In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
28. 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.
29. Irani, M.; Peleg, S. Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transparency. *J. Vis. Commun. Image Represent.* **2002**, *4*. [[CrossRef](#)]
30. Dai, S.; Han, M.; Wu, Y.; Gong, Y. Bilateral Back-Projection for Single Image Super Resolution. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Beijing, China, 2–5 July 2007.
31. Dong, W.; Zhang, L.; Shi, G.; Wu, X. Nonlocal back-projection for adaptive image enlargement. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP), Cairo, Egypt, 7–10 November 2009.
32. Timofte, R.; Rothe, R.; Van Gool, L. Seven Ways to Improve Example-Based Single Image Super Resolution. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
33. Han, W.; Chang, S.; Liu, D.; Yu, M.; Witbrock, M.; Huang, T.S. Image Super-Resolution via Dual-State Recurrent Networks. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.
34. Agustsson, E.; Timofte, R. NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.
35. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Albiro-Morel, M.L. Low-Complexity Single Image Super-Resolution Based on Nonnegative Neighbor Embedding. In Proceedings of the BMVC 2012 - Electronic Proceedings of the British Machine Vision Conference, Bristol, UK, 3–7 September 2012.
36. Adni. Available online: <http://adni.loni.usc.edu/> (accessed on 22 May 2019).
37. Oasis. Available online: <https://www.oasis-brains.org/> (accessed on 15 May 2019).
38. Ixi-Dataset. Available online: <http://brain-development.org/ixi-dataset/> (accessed on 16 May 2019).
39. Bien, N.; Rajpurkar, P.; Ball, R.; Irvin, J.; Park, A.; Jones, E.; Bereket, M.; Patel, B.; Yeom, K.; Shpanskaya, K.; et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: Development and retrospective validation of MRNet. *PLoS Med.* **2018**, *15*, e1002699. [[CrossRef](#)] [[PubMed](#)]
40. Sudirman, S.; Al Kafri, A.; Natalia, F.; Meidia, H.; Afriliana, N.; Al-Rashdan, W.; Bashtawi, M.; Al-Jumaily, M. Lumbar spine mri dataset. *Mendeley Data* **2019**, *2*. [[CrossRef](#)]
41. Zhang, K.; Zuo, W.; Zhang, L. Learning a Single Convolutional Super-Resolution Network for Multiple Degradations. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–22 June 2018.