Medical Image Super-Resolution via Diagnosis-Guided Attention

Jingwei Wang
Anhui Provincial International Joint Research Center for Advanced Technology in Medical Imaging, School of Computer Science and Technology, Anhui University
Hefei, China
e20201141@stu.ahu.edu.cn

Peng Zhou
Anhui Provincial International Joint Research Center for Advanced Technology in Medical Imaging, School of Computer Science and Technology, Anhui University
Hefei, China
zhoupeng@ahu.edu.cn

Xianjun Han
Anhui Provincial International Joint Research Center for Advanced Technology in Medical Imaging, School of Computer Science and Technology, Anhui University
Hefei, China
hxj@ahu.edu.cn

Yanming Chen
School of Computer Science and Technology, Anhui University
Hefei, China
cym@ahu.edu.cn

Abstract—**Medical image super-resolution (SR) is an important medical image processing task and is often helpful for downstream medical analysis tasks. Most of the conventional SR methods tried to generate visually more convincing images whereas ignoring the following downstream tasks. In this paper, we take the Alzheimer's disease diagnosis as the downstream task and propose a novel diagnosis-guided medical image SR network, which can make the SR and diagnosis be boosted by each other. The method contains two sub-networks, i.e., the SR network and the diagnosis network. To achieve better diagnosis performance, in the SR network, we apply the deformable convolution to capture the regions of interest (ROIs) with different and irregular sizes and shapes, which are important for diagnosis. Moreover, to integrate the two tasks, i.e., SR and diagnosis, more profoundly, we design a novel diagnosis-guided attention module, which makes the key regions for diagnosis can be reconstructed more clearly by the SR network. The extensive experiments on medical image data sets show that the proposed method often outperforms other state-of-the-art SR methods, which demonstrates its effectiveness. The codes of this paper are released in https://github.com/WJingwei/SRDA.**

Index Terms—**Medical image processing, super-resolution, diagnosis-guided attention module**

I. Introduction

Medical image analysis, such as diagnosis and segmentation, is an important problem in image processing. However, due to the hardware devices or image acquisition limitations, the obtained medical images often have a low resolution. For example, acquisitions with a high-resolution Magnetic Resonance Image (MRI) need a long imaging time, which may increase the potential of subject motion [1]; in Positron Emission Tomography (PET) image acquisitions, to decrease the patient radiation exposure, we have to decrease the radiation dosage, often leading to a low-resolution PET image [2].

One natural way to tackle this problem is to apply super-resolution methods to obtain high-resolution images before the medical image analysis. Super-resolution (SR) has been widely studied in natural image processing [3]–[8]. For example, Dong et al. applied convolutional neural networks (CNN) to SR, which was the first deep learning work for SR [3]; Liang et al. adopted swin transformer to SR [8]. Due to the promising performance of these super-resolution methods, they are also widely used in medical image processing [1], [2], [9]–[11]. For example, Sui et al. developed a deep neural network in an adversarial scheme for SR of MRI [1]; Chen et al. proposed a joint spatial-wavelet dual stream network for medical image SR [10].

However, despite the promising performance, since there are still some essential differences between natural image processing and medical image processing, the existing SR methods for natural images may be inappropriate for medical images. For example, in natural image SR methods, they wish to generate visually more convincing images [12]. However, in medical image super-resolution, SR is not our final goal and is often followed by some downstream medical image analysis tasks such as diagnosis. Therefore, when we apply SR to medical images, we should fully consider the downstream tasks. In this paper, we take the Alzheimer's disease diagnosis task as an example of the downstream task, and it is easily applied to other disease diagnosis tasks. Unfortunately, in
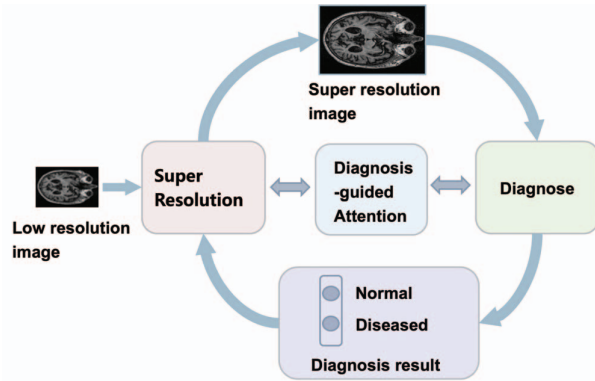
Fig. 1. The overview of SRDA.

practice, since SR is an ill-posed problem, i.e., the same low-resolution image can obtain different potential high-solution images, SR may introduce some artifacts which may mislead the following diagnosis.

To address this issue, in this paper, we propose a novel medical image Super-Resolution via Diagnosis-guided Attention (SRDA). The goal is that the generated SR image should be helpful to or at least should not mislead the following diagnosis task. To achieve this, we propose a multi-task framework for SR and diagnosis, which contains two sub-networks: the SR network and the diagnosis network. The overview of the proposed method is shown in Figure 1. We feed the low-resolution images into the SR network to obtain the SR images and immediately feed the SR images into the diagnosis network. The diagnosis network will affect the SR network in two-fold: firstly, since this is in an end-to-end way, the diagnosis loss can in turn guide the parameters learning in the SR network via backpropagation. Secondly, we design a diagnosis-guided attention module to make the SR network pay more attention to the important regions in the diagnosis.

In more detail, when designing the SR network, we fully consider the diagnosis task. In practice, there are some regions of interest (ROIs) which are important for the diagnosis. For example, according to previous literature [13], [14], some regions, such as the middle temporal gyrus and hippocampal formation, are key regions for Alzheimer's disease diagnosis. These regions often have different and irregular sizes and shapes, which are hard to be captured by the fixed-size convolution kernels used in conventional CNNs. To tackle this problem, our SR network applies the deformable convolution [15] to extract features. Due to the deformable convolution, it can capture ROIs with various scales and shapes, which is helpful to the diagnosis. Besides, the attention mechanism is another important tool for diagnosis which can make the model focus on the ROIs. In our method, we propose a novel diagnosis-guided attention for both SR and diagnosis. The basic idea is that the ROIs which are important for diagnosis are also worth being paid attention to when doing SR. For example, since the hippocampal formation is a key region for

Alzheimer's disease diagnosis, the SR network should also pay more attention to the hippocampal formation. To achieve this, we impose two same-structured attention networks on the SR and diagnosis networks, and force the attention maps obtained by the two attention networks to be close to each other, and thus the SR network and diagnosis network can pay attention to the similar regions. Due to the diagnosis-guided attention, the SR network and diagnosis network are seamlessly integrated into one unified framework. Notice that, different from other conventional SR methods which apply SR to enhance diagnosis, such as [11], which just combines the SR and diagnosis and lets the SR be affected by diagnosis only through backpropagation, in the proposed method, the diagnosis can guide the SR process more profoundly. In the proposed method, we carefully design and adjust the SR network to fit the diagnosis by introducing the deformable convolution and designing the diagnosis-guided attention, which makes the two tasks boosted by each other. On one hand, the ROIs in the SR images obtained by the SR network can be reconstructed more clearly which is more helpful for the diagnosis task; on the other hand, the diagnosis network finds the important ROIs which can in turn guide the SR network to obtain clearer and more useful SR images.

The main contributions are summarized as follows:

- We propose a novel framework for medical image SR, which applies the diagnosis to guide the SR. It can make the SR images more appropriate for the following diagnosis task.
- We carefully design a diagnosis-guided attention module that can make the SR and diagnosis network focus on similar regions. It makes sure that the key regions for diagnosis can be reconstructed more clearly.
- The experimental results on some medical image data sets show the effectiveness and superiority of the proposed method.

II. METHOD

In this section, we introduce our SRDA in more detail. Figure 2 shows the architecture of SRDA. It consists of two sub-networks: the SR network to obtain the high-resolution images and the diagnosis network to classify the high-resolution images. To integrate the two sub-networks into a unified framework, we also propose the diagnosis-guided attention module to characterize the ROIs. In this model, the SR and diagnosis can be boosted by each other. Notice that the influence from SR to diagnosis is in the forward propagation along the network, and the influence from diagnosis to SR is in the backpropagation of diagnosis loss. In the following, we will introduce the two sub-networks and the attention module in more detail. Here we take $\times 4$ super-resolution as an example, and it can easily be extended to other magnifications.

A. SR Network

In the SR network, we extract the features from the low-resolution images and apply the upsampling to obtain the high-resolution images. Since residual networks (ResNet) [16] has
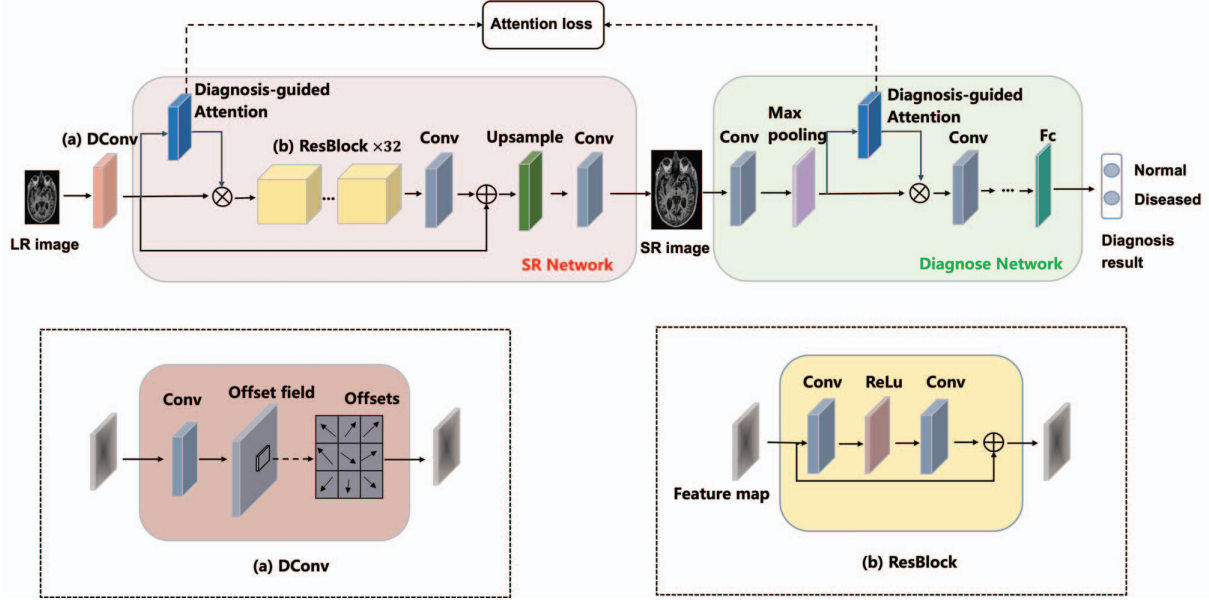
Fig. 2. The architecture of SRDA. It contains the SR and diagnosis networks. The influence from SR to diagnosis is in the forward propagation along the network, and the influence from diagnosis to SR is in the backpropagation of diagnosis loss.

demonstrated promising performance in SR tasks [6], [17], [18], we also apply ResNet to extract features in SR network as shown in Figure 2.

Different from natural image SR tasks, in real medical image processing tasks, SR is often followed by a downstream diagnosis task, and thus when extracting features, we should also consider the diagnosis task. In diagnosis tasks, we often pay more attention to some important ROIs, e.g. the hippocampal formation in Alzheimer's disease diagnosis. Therefore, these ROIs should also be specially handled in the SR network. Moreover, in real medical image analysis tasks, we observe that the ROIs often have different and irregular sizes and shapes. To tackle this problem, different from the conventional ResNet-based SR methods, which apply standard convolution operation to extract features, we adopt the deformable convolution [15] to extract features. Deformable convolution, as shown in Figure 2(a), augments the spatial sampling regions with learnable additional offsets, and thus can capture the ROIs with different and irregular sizes and shapes. Denoting $\mathbf{I}_{lr} \in \mathbb{R}^{C \times H \times W}$ as the images of the input low resolutions where C, H, and W are the number of channels, height, and width, respectively, and $f_{dconv}(\cdot)$ as the deformable convolution operation, we obtain the feature maps $\mathbf{F}_{dconv} \in \mathbb{R}^{C \times H \times W}$ as $\mathbf{F}_{dconv} = f_{dconv}(\mathbf{I}_{lr})$. The detailed structure of deformable convolution can be found in [15].

Then, we design a diagnosis-guided attention module to assign a weight to each pixel in the feature map $\mathbf{F}_{dconv}$. This attention is a kind of spatial attention and the output attention map is a weight matrix $\mathbf{W}_{sr} \in \mathbb{R}^{H \times W}$. In Section II-C we will introduce how to obtain it in more detail. Here we suppose we have already obtain such $\mathbf{W}_{sr}$, and then we obtain the weighted feature map $\mathbf{F}_{atten} \in \mathbb{R}^{C \times H \times W}$ as:

$$\mathbf{F}_{atten} = \mathbf{F}_{dconv} * \mathbf{W}_{sr}, \tag{1}$$

where $*$ is the element-wise production in each channel.

After obtaining the weighted feature map $\mathbf{F}_{atten}$, we feed it into 32 residual blocks as many other ResNet-based methods did. The structure of the residual block is shown in Figure 2(b). Then we obtain the high-resolution images with a $\times 4$ upsample layer with pixel shuffle, followed by a convolution layer which compresses the feature maps to three-channel or one-channel super-resolution images $\mathbf{I}_{sr} \in \mathbb{R}^{C \times 4H \times 4W}$.

B. Diagnosis Network

After obtaining the SR images $\mathbf{I}_{sr}$, we feed them into the diagnosis network. The diagnosis network is a classification network. Here we use ResNet50 [16] as its backbone.

After the first 7×7 convolution layer and the max pooling layer of the ResNet50, we obtain the feature maps $\mathbf{F}_{conv1} \in \mathbb{R}^{C_1 \times H \times W}$ where C_1 is the current number of channels, H and W are the same with those of the original low-resolution images. Different from the vanilla ResNet50, which directly feeds $\mathbf{F}_{conv1}$ into the next 1×1 convolution layer, we also impose the diagnosis-guided attention on such feature map $\mathbf{F}_{conv1}$. The attention network has the same structure as that in the SR network but has different network parameters, which will be introduced in Section II-C in more detail. Denote $\mathbf{W}_{diag} \in \mathbb{R}^{H \times W}$ as the output attention map of the attention module. Then, we obtain the weighted feature map $\mathbf{F}_{conv1}^{atten} \in \mathbb{R}^{C' \times H \times W}$ as: $\mathbf{F}_{conv1}^{atten} = \mathbf{F}_{conv1} * \mathbf{W}_{diag}$, which is similar to Eq.(1).

Then, we feed $\mathbf{F}_{conv1}^{atten}$ into the remaining parts of ResNet50 for diagnosis.

C. Diagnosis-guided Attention Module

To make the SR network and the diagnosis network pay attention to similar regions, we propose a diagnosis-guided attention module. This attention module is a kind of spatial attention and is used in both the SR network and the diagnosis network. The attention networks in the two sub-networks have the same structure but different network parameters, whose architecture is shown in Figure 3. Given a feature map $\mathbf{F} \in \mathbb{R}^{C_2 \times H \times W}$ (i.e., $\mathbf{F} = \mathbf{F}_{dconv}$ in the SR network and $\mathbf{F} = \mathbf{F}_{conv1}$ in the diagnosis network), similar to [19], we first apply the average pooling and the max pooling on $\mathbf{F}$ across the channel to obtain $\mathbf{F}_{avg} \in \mathbb{R}^{1 \times H \times W}$ and $\mathbf{F}_{max} \in \mathbb{R}^{1 \times H \times W}$, respectively. Then we concatenate $\mathbf{F}_{avg}$ and $\mathbf{F}_{max}$ along the channel to obtain $\mathbf{F}' = [\mathbf{F}_{avg}; \mathbf{F}_{max}] \in \mathbb{R}^{2 \times H \times W}$. After that, we apply a 7×7 convolution layer on $\mathbf{F}'$ to obtain new one-channel feature map $\mathbf{F}'' \in \mathbb{R}^{1 \times H \times W}$. At last we obtain the final attention map $\mathbf{W} \in \mathbb{R}^{H \times W}$ (i.e., $\mathbf{W}_{sr} = \mathbf{W}$ in SR network and $\mathbf{W}_{diag} = \mathbf{W}$ in diagnosis network) by feeding $\mathbf{F}''$ into a sigmoid function.
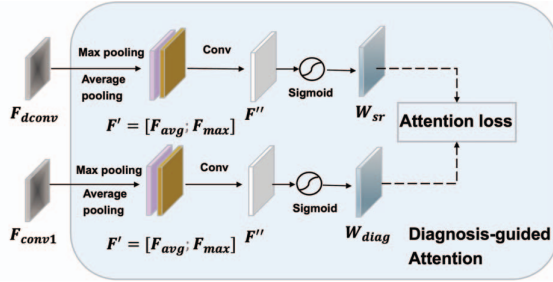

Fig. 3. The diagnosis-guided attention module.

Since we wish the key regions for diagnosis can be reconstructed more clearly by the SR network, we enforce the two attention networks to pay attention to similar regions. Therefore, we should make the outputs of the two attention networks (i.e., $\mathbf{W}_{sr}$ and $\mathbf{W}_{diag}$) be close to each other by minimizing the following diagnosis-guided attention loss:

$$\mathcal{L}_{Atten} = \|\mathbf{W}_{sr} - \mathbf{W}_{diag}\|_F^2 \qquad (2)$$

Due to this loss, although the SR network and the diagnosis network have different attention networks, they can still pay attention to similar regions, which can make the two tasks (i.e., SR and diagnosis) be boosted by each other.

D. Loss Function

The total loss function contains the SR loss, the diagnosis loss, and the diagnosis-guided attention loss. In the SR network, we compute the pixel-wise loss between the SR image $\mathbf{I}_{sr}$ and the ground truth high-resolution images $\mathbf{I}_{gt}$ with

$$\mathcal{L}_{SR} = \frac{1}{N} \sum_{i=1}^{N} \left\| \mathbf{I}_{sr}^{(i)} - \mathbf{I}_{gt}^{(i)} \right\|_1, \qquad (3)$$

where $\mathbf{I}_{sr}^{(i)}$ and $\mathbf{I}_{gt}^{(i)}$ are the SR image and the ground truth high-resolution image of the i-th image in the training set, respectively. N is the number of images in the training set.

In the diagnosis network, we adopt the cross-entropy loss between the predicted label (denoted as $\hat{y}_{ic}$) and the ground truth label (denoted as y_{ic}), where i denotes the i-th image and c denotes the c-th class. The diagnosis loss $\mathcal{L}_{Diag}$ is:

$$\mathcal{L}_{Diag} = \frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{M} y_{ic} log(\hat{y}_{ic}), \qquad (4)$$

where M is the number of classes of the data set.

The diagnosis-guided attention loss is defined as Eq.(2). To sum up, the total loss $\mathcal{L}_{total}$ is defined as:

$$\mathcal{L}_{total} = \mathcal{L}_{SR} + \lambda_1 \mathcal{L}_{Diag} + \lambda_2 \mathcal{L}_{Atten}, \qquad (5)$$

where λ_1 and λ_2 are two balancing hyperparameters.

III. EXPERIMENTS

A. Data Sets

In this paper, we focus on the SR for Alzheimer's disease diagnosis. We use three public data sets: ADNI-MRI, ADNI-PET, and Demented-MRI. ADNI-MRI and ADNI-PET contain MRI and PET images of brains for Alzheimer's Disease diagnosis collected by The Alzheimer's Disease Neuroimaging Initiative (ADNI) [1], respectively. The original images of ADNI-MRI and ADNI-PET are both 3D brain images, and we slice each 3D image into multiple 2D images. We slice them along the axial plane with 0.01 mm intervals. We randomly divide the data into the training, validation, and testing set in the ratio of 8:1:1. In more detail, for ADNI-MRI, the training set contains 4080 2D images, both the validation set and testing set contain 510 2D images. For ADNI-PET, the training set contains 9600 2D images, and the validation set and testing set contain 1200 2D images, respectively. Demented-MRI contains the 2D MRI images of brains for Alzheimer's disease diagnosis collected by Haywhale Community[2]. We use three classes (i.e., non-demented, mild-demented, and very-mild-demented) in Demented-MRI, where each class contains 890 MRI images. Therefore, it contains 2670 images in total. We randomly use 2136 images, 267 images, and 267 images as training data, validation data, and testing data, respectively.

B. Experimental Setup and Implementation Details

We use bicubic interpolation (denoted as Bicubic) as a baseline method. Moreover, we also compare with the following state-of-the-art SR methods: RCAN [20], IMDN [21], SWD [10], SwinIR [8], SRD [11], A^{2}F [22], ELAN [23]. We adopt Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) to evaluate the reconstruction quality. For both metrics, the larger values mean the better performance. Since most SR methods (including ours) need pairs of the low-resolution image and its ground-truth high-resolution image for training but the used data does not have such pairs, we use the original data as the ground-truth images

[1] https://adni.loni.usc.edu/
[2] https://www.heywhale.com/mw/dataset/6245d687e1d37c0017029c05/content

<div style="text-align:center">

TABLE I

QUANTITATIVE COMPARISON (AVERAGE PSNR AND SSIM) WITH
STATE-OF-THE-ART SR METHODS. THE BEST AND RESULTS ARE MARKED
IN **BOLD** FONTS.

</div>

Methods	ADNI-MRI		ADNI-PET		Demented-MRI	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	26.76	0.8153	27.55	0.8352	22.95	0.7954
RCAN [20]	34.02	0.9465	32.97	0.9437	30.30	0.9533
IMDN [21]	32.83	0.9338	31.23	0.9257	27.22	0.9216
A^{2}F [22]	34.39	0.9500	32.27	0.9384	28.89	0.9423
SWD [10]	31.70	0.9151	31.31	0.9272	27.15	0.9213
SwinIR [8]	32.74	0.9328	31.88	0.9331	28.83	0.9417
SRD [11]	32.83	0.9306	31.55	0.9305	27.19	0.9235
ELAN [23]	31.95	0.9231	31.00	0.9179	27.08	0.9192
SRDA	**35.56**	**0.9595**	**33.89**	**0.9487**	**30.69**	**0.9567**

and generate the low-resolution images by $\times 4$ downsampling from the original data.

The experiments are conducted with Pytorch on a PC with an NVIDIA GeForce RTX 3090 Ti GPU. In our method, we use Adam as the optimizer with a learning rate of 0.0001. We set the batch size as 128, 64, and 8 on the ADNI-MRI, ADNI-PET, and Demented-MRI, respectively. For all data sets, the number of epochs is fixed to 1000, and λ_1 and λ_2 are fixed as 0.0006 and 0.0001, respectively. For all compared methods, we use the codes released by the authors with the hyper-parameter setting suggested in the corresponding literature.

C. Experimental Results

Table I shows the PSNR and SSIM results of our method and other state-of-the-art SR methods on the three data sets. From Table I, we can see that our method achieves the best performance on all data sets. It well demonstrates the effectiveness of the proposed method.

Figures 4-6 show the qualitative comparison of all methods on all data sets. From these figures, we can find that the SR image obtained by our method contains more details compared with other methods. Such details may be important for diagnosis. Our method can reconstruct the details more clearly and thus the SR images of our method may be more helpful for the downstream diagnosis task.

D. Ablation Study

To show the effectiveness of each part in our network, we also conduct the ablation study. We use SR to denote the base model, which only contains the SR network without the deformable convolution, diagnosis-guided attention, and the diagnosis network. SR+Dconv denotes the SR network with deformable convolution but without diagnosis-guided attention and the diagnosis network. SR+Dconv+Diag denotes the SR network with deformable convolution and the diagnosis network but without the attention. SR+Dconv+Diag+Atten denotes the complete SRDA model with all parts.

Table II shows the results. Compared with the base model SR, deformable convolution can improve the performance to some extent. Then, by adding the diagnosis network, it achieves a better performance, which shows that the diagnosis can indeed guide the SR. Moreover, with diagnosis-guided
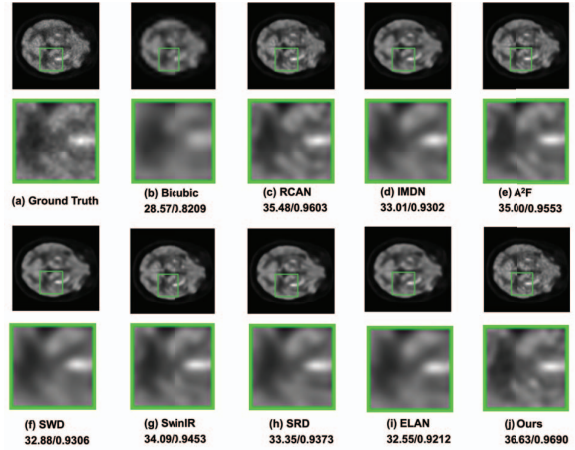

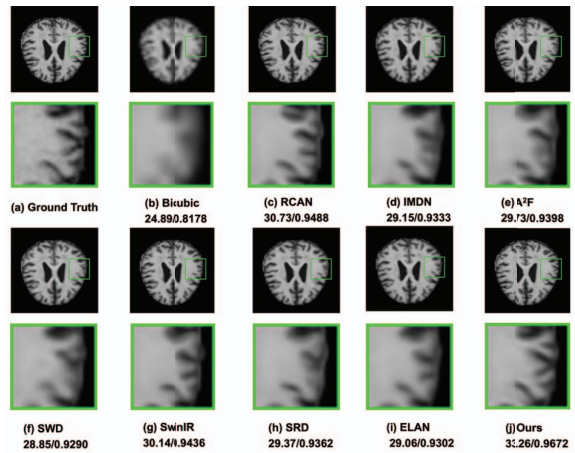Fig. 4. Qualitative comparison on ADNI-PET data set (PSNR/SSIM).

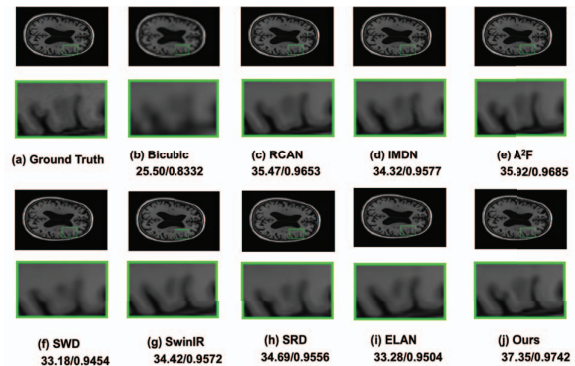Fig. 5. Qualitative comparison on Demented-MRI data set (PSNR/SSIM).

Fig. 6. Qualitative comparison on ADNI-MRI data set (PSNR/SSIM).

TABLE II
ABLATION STUDY.

Methods	ADNI-MRI		ADNI-PET		Demented-MRI	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
SR	32.42	0.9290	33.12	0.9444	30.14	0.9525
SR+Dconv	32.56	0.9321	33.16	0.9449	30.34	0.9540
SR+Dconv +Diag	34.17	0.9491	33.19	0.9452	30.44	0.9547
SR+Dconv +Diag+Atten	**35.56**	**0.9595**	**33.89**	**0.9487**	**30.69**	**0.9567**

TABLE III
TEST ACCURACY (ACC) FOR DIAGNOSIS TASK.

Methods	ADNI-MRI	ADNI-PET	Demented-MRI
LR	0.5863	0.6400	0.6515
RCAN [20]	0.7118	0.7158	0.7083
IMDN [21]	0.6039	0.6967	0.7083
A^2F [22]	0.6980	0.7433	0.6970
SWD [10]	0.6902	0.6692	0.6932
SwinIR [8]	0.7098	0.7558	0.6970
SRD [11]	0.6568	0.6225	0.6894
ELAN [23]	0.6902	0.7212	0.7008
SRDA	**0.8706**	**0.8292**	**0.7348**

attention, the performance is further improved, demonstrating the effectiveness of the attention module.

E. Diagnosis Experiments

To demonstrate that the proposed SRDA method is helpful for the diagnosis task, we also conduct the diagnosis experiments. In more detail, firstly, we apply SRDA and all other compared methods to generate the SR images respectively. Then, we train the ResNet50 [16] on each SR data set obtained by each SR method, and report the test accuracy (ACC) on the testing set in Table III. LR represents the classification results directly on the low-resolution images. We can see that the diagnosis accuracy on SR images is much higher than LR, which shows that SR is helpful for the diagnosis. Moreover, the accuracy of SRDA outperforms other SR methods significantly, which means that the SR images obtained by SRDA are easier to diagnose. It is consistent with our motivation that the SR images should be more helpful for diagnosis.

IV. CONCLUSION

In this paper, we proposed a novel diagnosis-guided SR method. To make the SR and diagnosis be boosted by each other, we carefully designed a diagnosis-guided attention module, which can make the SR network and the diagnosis network pay the attention to similar ROIs. Due to the diagnosis-guided attention module, the ROIs of diagnosis are reconstructed more clearly in the SR network, and thus the SR method is more appropriate for the following downstream diagnosis task. We conducted extensive experiments on image data sets for Alzheimer's disease diagnosis, and the experimental results demonstrated the effectiveness and superiority of the proposed method.

REFERENCES

[1] Yao Sui, Onur Afacan, Ali Gholipour, and Simon K. Warfield, "MRI super-resolution through generative degradation learning," in *MICCAI*, 2021, vol. 12906, pp. 430–440.

[2] John A. Kennedy, Ora Israel, Alex Frenkel, Rachel Bar-Shalom, and Haim Azhari, "Super-resolution in PET imaging," *IEEE Trans. Medical Imaging*, vol. 25, no. 2, pp. 137–147, 2006.

[3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a deep convolutional network for image super-resolution," in *ECCV*, 2014, vol. 8692, pp. 184–199.

[4] Chao Dong, Chen Change Loy, and Xiaoou Tang, "Accelerating the super-resolution convolutional neural network," in *ECCV*, 2016, vol. 9906, pp. 391–407.

[5] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *CVPR*, 2017, pp. 5835–5843.

[6] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *CVPR Workshops*, 2017, pp. 1132–1140.

[7] Shunta Maeda, "Unpaired image super-resolution using pseudo-supervision," in *CVPR*, 2020, pp. 288–297.

[8] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte, "Swinir: Image restoration using swin transformer," in *ICCV Workshop*, 2021, pp. 1833–1844.

[9] Yuchong Gu, Zitao Zeng, Haibin Chen, Jun Wei, Yaqin Zhang, Binghui Chen, Yingqin Li, Yujuan Qin, Qing Xie, Zhuoren Jiang, and Yao Lu, "Medsrgan: medical images super-resolution using generative adversarial networks," *Multim. Tools Appl.*, vol. 79, no. 29-30, pp. 21815–21840, 2020.

[10] Zhen Chen, Xiaoqing Guo, Chen Yang, Bulat Ibragimov, and Yixuan Yuan, "Joint spatial-wavelet dual-stream network for super-resolution," in *MICCAI*, 2020, vol. 12265, pp. 184–193.

[11] Zhen Chen, Xiaoqing Guo, Peter Y. M. Woo, and Yixuan Yuan, "Super-resolution enhanced medical image diagnosis with sample affinity interaction," *IEEE Trans. Medical Imaging*, vol. 40, no. 5, pp. 1377–1389, 2021.

[12] Mohammad Saeed Rad, Behzad Bozorgtabar, Urs-Viktor Marti, Max Basler, Hazim Kemal Ekenel, and Jean-Philippe Thiran, "SROBB: targeted perceptual loss for single image super-resolution," in *ICCV*, 2019, pp. 2710–2719.

[13] Nick C Fox and Jonathan M Schott, "Imaging cerebral atrophy: normal ageing to alzheimer's disease," *The Lancet*, vol. 363, no. 9406, pp. 392–394, 2004.

[14] Chandan Misra, Yong Fan, and Christos Davatzikos, "Baseline and longitudinal patterns of brain atrophy in mci patients, and their use in prediction of short-term conversion to ad: Results from adni," *NeuroImage*, vol. 44, no. 4, pp. 1415–1422, 2009.

[15] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei, "Deformable convolutional networks," in *ICCV*, 2017, pp. 764–773.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[17] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *CVPR*, 2016, pp. 1646–1654.

[18] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, 2017, pp. 105–114.

[19] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon, "CBAM: convolutional block attention module," in *ECCV*, 2018, vol. 11211, pp. 3–19.

[20] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," in *ECCV*, 2018, vol. 11211, pp. 294–310.

[21] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang, "Lightweight image super-resolution with information multi-distillation network," in *ACM MM*, 2019, pp. 2024–2032.

[22] Xuehui Wang, Qing Wang, Yuzhi Zhao, Junchi Yan, Lei Fan, and Long Chen, "Lightweight single-image super-resolution network with attentive auxiliary feature learning," in *ACCV*, 2020, vol. 12623, pp. 268–285.

[23] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang, "Efficient long-range attention network for image super-resolution," in *ECCV*, 2022, vol. 13677, pp. 649–667.