IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 10, 2019, accepted May 12, 2019, date of publication May 24, 2019, date of current version June 6, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2918926

Deep Generative Adversarial Networks for Thin-Section Infant MR Image Reconstruction

JIAQI GU[1], ZEJU LI[1], YUANYUAN WANG[1,3], (Senior Member, IEEE), HAOWEI YANG[2], ZHONGWEI QIAO[2], AND JINHUA YU[1,3], (Member, IEEE)

[1]School of Information Science and Technology, Fudan University, Shanghai 200433, China
[2]The Children's Hospital of Fudan University, Shanghai 201102, China
[3]Key Laboratory of Medical Imaging Computing and Computer Assisted Intervention of Shanghai, Department of Electronic Engineering, Institute of Functional and Molecular Medical Imaging, Fudan University, Shanghai 200433, China

Corresponding authors: Zhongwei Qiao (zqiao@fudan.edu.cn) and Jinhua Yu (jhyu@fudan.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61471125, in part by the National Basic Research Program of China under Grant 2015CB755500, and in part by the Shanghai Shenkang Hospital Development Center Clinical Auxiliary Capacity (Imaging Medicine) Construction Project SHDC22015031.

ABSTRACT Due to their high spatial resolution, thin-section magnetic resonance (MR) images serve as ideal medical images for brain structure investigation and brain surgery navigation. However, compared with the clinically widely used thick-section MR images, thin-section MR images are less available due to the imaging cost. Thin-section MR images of infants are even scarcer but are quite valuable for the study of human brain development. Therefore, we propose a method for the reconstruction of thin-section MR images from thick-section images. A two-stage reconstruction framework based on generative adversarial networks (GANs) and a convolutional neural network (CNN) is proposed to reconstruct thin-section MR images from thick-section images in the axial and sagittal planes. A 3D-Y-Net-GAN is first proposed to fuse MR images from the axial and sagittal planes and to achieve the first-stage thin-section reconstruction. A 3D-DenseU-Net followed by a stack of enhanced residual blocks is then proposed to provide further detail recalibrations and structural corrections in the sagittal plane. In this method, a comprehensive loss function is also proposed to help the networks capture more structural details. The reconstruction performance of the proposed method is compared with bicubic interpolation, sparse representation, and 3D-SRU-Net. Cross-validation based on 35 cases and independent testing based on two datasets with totally 114 cases reveal that, compared with the other three methods, the proposed method provides an average 23.5% improvement in peak signal-to-noise ratio (PSNR), 90.5% improvement in structural similarity (SSIM), and 21.5% improvement in normalized mutual information (NMI). The quantitative evaluation and visual inspection demonstrate that our proposed method outperforms those methods by reconstructing more realistic results with better structural details.

INDEX TERMS Deep learning, infant magnetic resonance (MR) images, super-resolution reconstruction, thick-section, thin-section.

I. INTRODUCTION

Thin-section head magnetic resonance (MR) images typically have a slice thickness of 1 mm and a spacing gap of zero. The high spatial resolution of thin-section head MR images is ideal for brain structure analysis, volumetric measurement, and surgery navigation. Thin-section head MR images, however, are not always available. Clinically routine head MR images are typically thick-section images with a slice thickness of 4 mm to 6 mm and a spacing gap of 0.4 mm to 1 mm. The higher section thickness leads to a lower spatial resolution, which limits the usage of thick-section MR images in brain-related research.

Compared with imaging data for adults, brain MR images of infants are even more valuable because these images provide great insight into human brain development after birth. The acquisition of infant brain MR images, however, is even more difficult since MR imaging, let alone thin-section imaging, is rarely performed on infants without sufficient reasons. This situation inspired us to develop a method that can provide a spatial resolution comparable to thin-section MR images by using available thick-section images.

The associate editor coordinating the review of this manuscript and approving it for publication was Mohsin Jamil.

A thin-section MR image reconstruction method is thus proposed in this paper.

This reconstruction method can also be used to normalize the image layer spacing. In a multi-center, multi-device scenario, proposed method can be used to normalize MR images obtained at different layer spacings to a uniform layer spacing, which is very beneficial for data-driven researches, such as human brain development statistics based on image big data.

Thin-section MR image reconstruction was considered a multiplanar MR image registration problem. For example, Mahmoudzadeh and Kashou [1] applied traditional interpolation algorithms to thick-section MR images in all three planes and combined them with the iterative registration algorithm optimized automatic image registration (OAIR) [2] with the guidance of a pixelwise loss function. The reconstruction results of this algorithm are visually improved but focus only on adult head MR images with limited consideration of structural similarity (SSIM) among human brains. Moreover, thin-section MR image reconstruction can be handled as a frame interpolation task. As proposed in [3], a decomposition-reconstruction method based on the rules of organ consistency is adopted to obtain a higher inter-slice resolution. Thin-section image reconstruction can also be considered a super-resolution problem. Yang *et al.* [4] proposed a trainable method to reconstruct high-resolution images by utilizing the same sparse representation between low-resolution image patches and their high-resolution counterparts. With the development of deep learning (DL) techniques, convolutional neural networks (CNN) and generative adversarial networks (GANs) have gained momentum recently, especially in the image super-resolution field. Accordingly, thin-section MR image reconstruction, if considered as a nonisotropic super-resolution problem, will benefit a great deal from the powerful modeling capacity of deep neural networks. For example, Heinrich *et al.* [5] recently applied a 3D-SRU-Net for isotropic super-resolution from nonisotropic three-dimensional (3-D) electron microscopy. Our group [6] proposed a residual-network-based 3D-SRGAN to reconstruct adult thin-section MR images from thick-section MR images, however only the reconstruction in the axial plane was considered. In addition, CNNs and GANs have been widely utilized to improve the resolution of MR images [7], [8]. Compared to traditional algorithms, DL algorithms show superior potential in thin-section MR image reconstruction by not only increasing reconstruction performance but also reducing the reconstruction time to seconds.

In this paper, the task is to combine the multiplanar feature fusion and 3-D nonisotropic super-resolution problems. Our proposed framework is inspired by several state-of-the-art DL architectures. First, U-Net [9], as it performs well in the biomedical segmentation field, distinguishes itself in feature fusion problems through multiscale convolution and upscaling. Super-resolution generative adversarial neural networks (SRGANs) [10] are empirically proven to have remarkable performance in the super-resolution field, as they extract both low- and high-frequency information from images. In addition, enhanced deep residual networks (EDSR) [11], a new residual architecture that won first prize in the NTIRE 2017, provide an efficient approach to recovering high-resolution images. Inspired by the above state-of-the-art models, we propose a two-stage reconstruction framework to apply the mapping from thick-section MR images in the axial and sagittal planes to their axial thin-section counterparts. Specifically, the first stage is a least-squares GAN (LSGAN) [12] with a newly proposed 3D-Y-Net generator, which is designed to fuse axial and sagittal thick-section MR images and map them onto the thin-section image space. The second stage is a cascade connection of 3D-DenseU-Net and enhanced residual blocks, designed to increase statistical metrics and eliminate artifacts via further detail refinement. A 3-D gradient correction loss and a self-adaptive Charbonnier loss are proposed to concentrate the generator's optimization attention and capture high-frequency differential information. We then evaluate the performance of the proposed two-stage framework by comparing the reconstruction results with the ground truth, showing that our proposed method is more effective than three representative methods, comprising bicubic interpolation [13], sparse representation [4], and 3D-SRU-Net [5]. We also undertake two experiments to further validate the contribution of multiplanar image fusion and our proposed comprehensive loss function. Finally, the conclusion summarizes the paper.

II. PROPOSED METHOD

A. OVERVIEW

CNNs have outperformed many traditional algorithms in the image super-resolution field. Via hierarchical spatial convolution and optional nonlinearity, CNNs can learn the prior knowledge from low-level and high-level features extracted from images and accordingly recover super-resolved images through upsampling operations such as fractionally-strided convolution [14] and sub-pixel convolution [15]. Recently, with an increasing number of state-of-the-art CNN models, e.g., EDSR, SRCNN [16], and VDSR [17], GANs are becoming gradually integrated with these popular CNN models in order to preserve high-frequency information. Under the supervision of the discriminator, the generator is driven to maximize the distribution similarity between the generated data and the real data, thus generating results that are more realistic. However, recently proposed super-resolution models mainly seek to upscale both dimensions of two-dimensional (2-D) images by the same factor [7], [15], [18], [19], [20], [21]. Even if several models are extended to handle 3-D images [5], [6], low-resolution images in multiple planes barely make a collaborative contribution in a single framework. In this study, we propose a two-stage reconstruction framework based on axial and sagittal thick-section MR images to reconstruct corresponding axial thin-section MR images with an upscaling factor of 8, as shown in Fig. 1. In our framework, multiplanar thick-section MR images are fully fused by our proposed 3D-Y-Net-GAN and

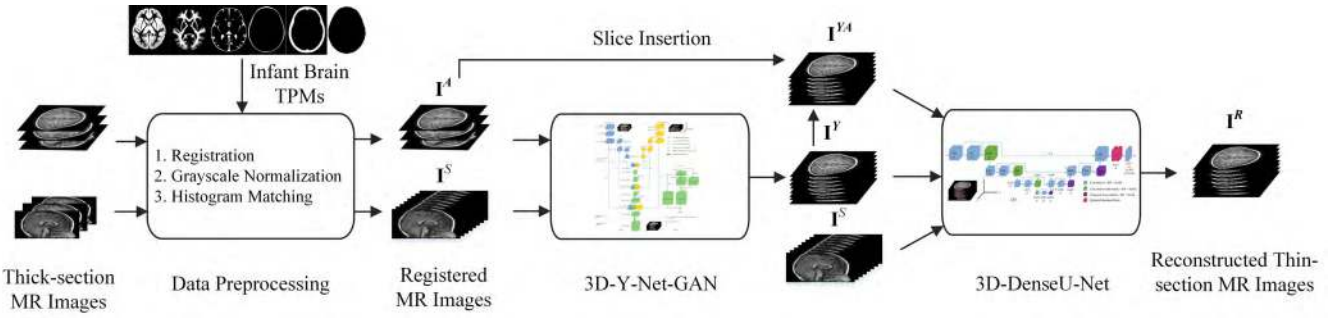


FIGURE 1. The proposed two-stage framework for thin-section MR image reconstruction. The first stage is 3D-Y-Net-GAN, and the second stage is 3D-DenseU-Net. TPMs represent tissue probability maps, which will be discussed in later sections.

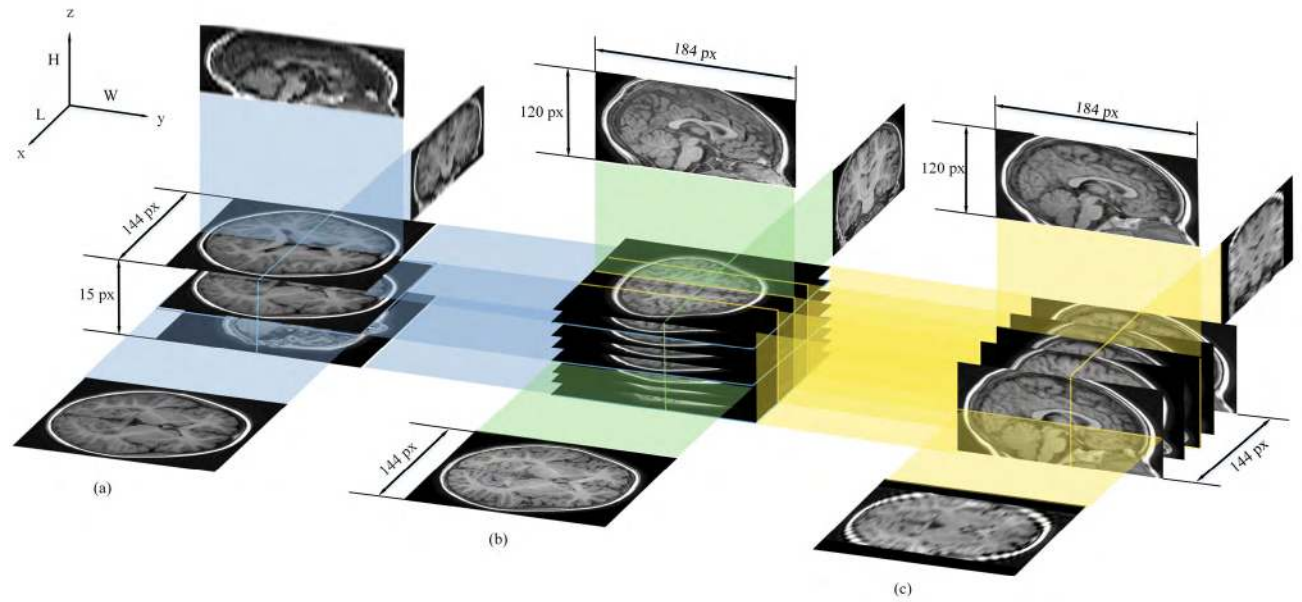


FIGURE 2. (a) 15 slices of normalized axial thick-section MR images, (b) 120 slices of normalized axial thin-section MR images, (c) 120 slices of normalized sagittal thick-section MR images. *x*, *y*, and *z* represent three axes in the coordinate system we use to describe the volumes. *L*, *W*, and *H* represent image sizes of MR images along *x*, *y*, and *z* axes, respectively. The yellow and blue lines illustrate their relative spatial locations. *px* is short for pixel.

3D-DenseU-Net to recover thin-section images collaboratively. In the following sections, we demonstrate the details of the proposed two-stage reconstruction framework and our proposed comprehensive loss function. To better demonstrate the task, relative spatial locations of thick-section and thin-section MR images are shown in Fig. 2.

B. NETWORK ARCHITECTURE

In this section, we introduce our proposed two-stage reconstruction framework. The first stage is a 3D-Y-Net-GAN consisting of a 3D-Y-Net generator and a conditional discriminator, which produces primary thin-section MR images for subsequent detail correction. The second stage is a 3D-DenseU-Net followed by a stack of enhanced residual blocks for final detail recalibration. The inputs are registered axial thick-section MR images, denoted as $\mathbf{I}^A$ with size $L \times W \times H$, and registered sagittal thick-section MR images, denoted as $\mathbf{I}^S$ with size $L \times W \times rH$ where r represents the upscaling factor along the z-axis. The outputs are thin-section MR images, denoted as $\mathbf{I}^R$ with size $L \times W \times rH$. Note that L, W, and H represent spatial sizes along x, y, and z axes respectively.

1) 3D-Y-NET-GAN

As the first stage of the whole framework, a 3D-Y-Net-GAN is proposed to take $\mathbf{I}^A$ and $\mathbf{I}^S$ as inputs and reconstruct thin-section MR images with an upscaling factor of r, denoted as $\mathbf{I}^Y$. The generator consists of three branches: 1) feature extraction (FE), 2) feature fusion (FF), and 3) reconstruction. The detailed network structure of the generator is illustrated in Fig. 3(a). In our case, r is set to 8, and we adopt a patch-based training strategy to reduce computational cost. Specifically, at the first stage, the size of the patches for $\mathbf{I}^A$ is $32 \times 32 \times 15$ and the size of the patches for $\mathbf{I}^S$ and $\mathbf{I}^Y$

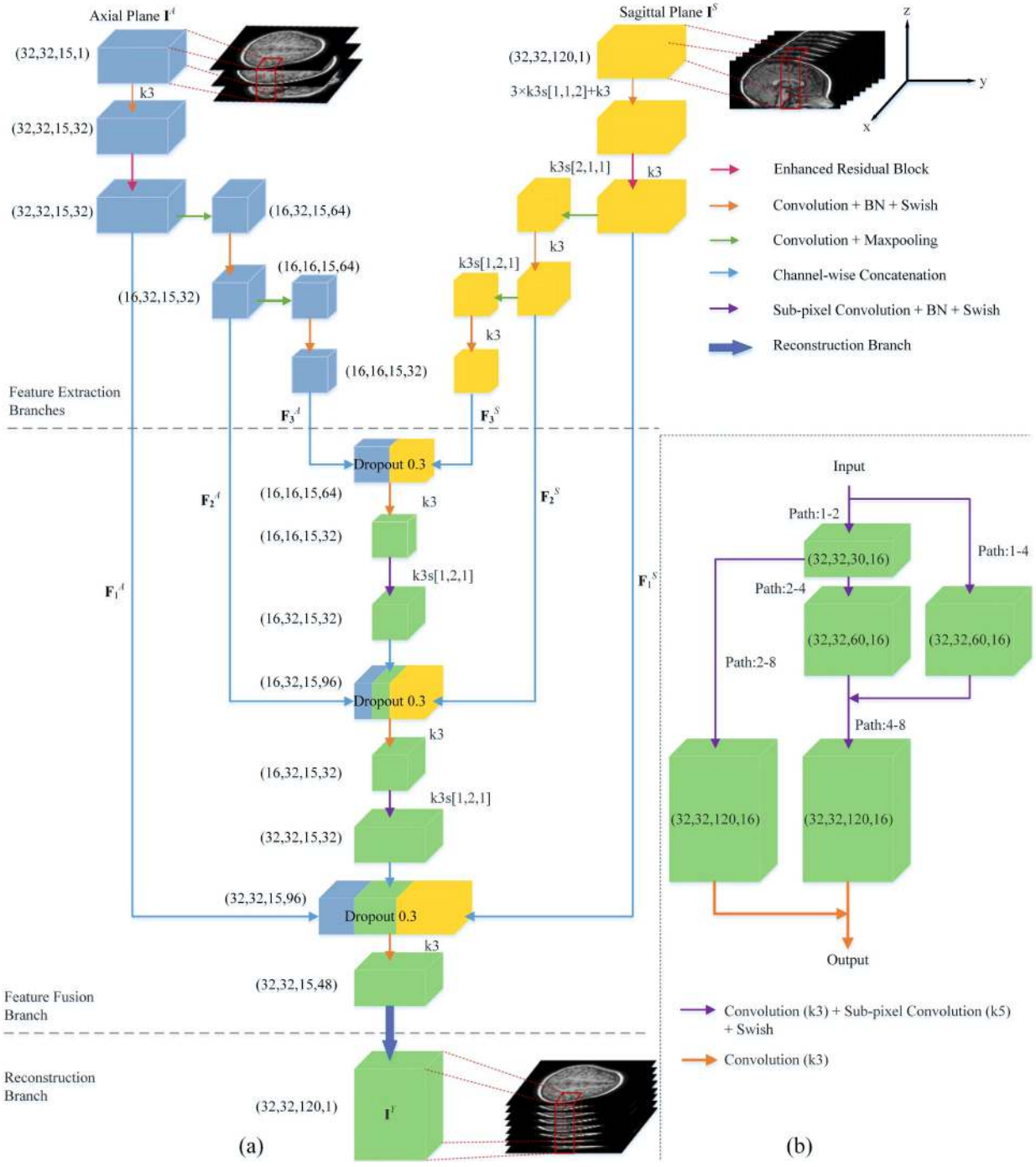


FIGURE 3. (a) Shows the network structure of 3D-Y-Net, e.g., (32,32,15,64) represents 64-channel feature maps with a spatial size of 32×32×15, e.g., k3s [1,2,1] represents a convolution kernel size of 3×3×3 with strides of [1,2,1]. Unless specified, kernel sizes, strides, and feature map shapes are identical between axial and sagittal branches, thus most parameters are only shown in either branch. *Dropout 0.3* represents the dropout operation with a drop rate of 0.3. Red frameworks represent patches for training; (b) shows the structure of the reconstruction branches. *Path* represents the upscaling process, e.g., *Path 1-4* means upsampling images with sizes of $L \times W \times H$ to images with sizes of $L \times W \times 4H$. The intersection of two arrows represents channel concatenation before convolution.

is 32 × 32 × 120. Note that, for inference, instead of image patches, full-size MR images are used as inputs.

a: FEATURE EXTRACTION BRANCHES

For the axial FE branch, 3-D convolutional layers are adopted to extract features from input images, and maxpooling layers with unbalanced strides of [1, 2, 1] or [2, 1, 1] are adopted to generate differently sized feature maps at different levels. Notably, maxpooling layers can ignore certain minute structural discrepancies, such that the negative impact induced by misalignment after registration would be mitigated to some degree. To clarify, the 3-D convolutional layer is

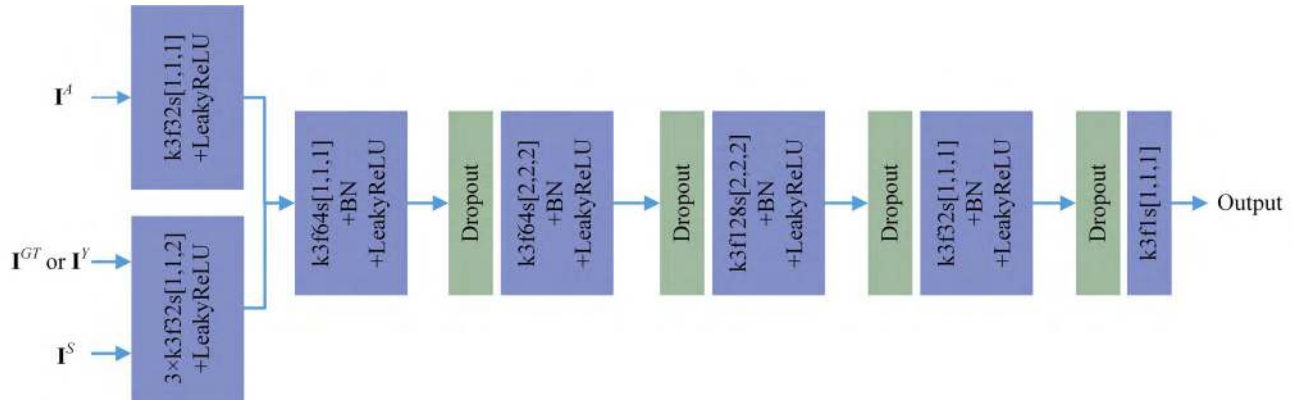


FIGURE 4. Network structure of the conditional discriminator. Note that the slope of the negative part of the Leaky ReLU is set to 0.2. $\mathbf{I}^{GT}$ is the real sample, $\mathbf{I}^{Y}$ is the fake sample, $\mathbf{I}^{A}$ and $\mathbf{I}^{S}$ are inputs of the generator. *k* represents the kernel size, and *f* represents the number of filters. All the dropout rates are set to 0.3.

Convolution + Batch Normalization + Swish. Specifically, Swish [22] is a new activation function that overcomes the dead-neuron problems caused by ReLU. In our framework, we set the untrainable parameter in Swish to be 1. The outputs of the axial FE branch are feature maps in three scales: $\mathbf{F}_1^A$ ($L \times W \times H$), $\mathbf{F}_2^A$ ($L \times W/2 \times H$), and $\mathbf{F}_3^A$ ($L/2 \times W/2 \times H$). The sagittal FE branch is generally of the same structure as the axial FE branch, which generates similar outputs $\mathbf{F}_1^S$ ($L \times W \times H$), $\mathbf{F}_2^S$ ($L \times W/2 \times H$), and $\mathbf{F}_3^S$ ($L/2 \times W/2 \times H$). However, given the size discrepancy between $\mathbf{I}^A$ and $\mathbf{I}^S$, an extra preprocessing module consisting of 3 convolutional layers with strides of [1, 1, 2] is appended to its entry.

b: FEATURE FUSION BRANCHES

The FF branch is a topological inversion of FE branches. At each level, the FF branch upsamples multiscale feature maps through sub-pixel convolution. Concretely, sub-pixel convolution [15] is a normal convolution followed by a pixel shuffler, which is an efficient substitution for transpose convolution. The means by which FE and FF branches are connected at three levels is a design inspired by the U-Net structure, which fully fuses multiscale features, guarantees the structural alignment, and avoids the gradient-vanishing problem.

c: RECONSTRUCTION BRANCHES

The detailed network structure of the reconstruction branch is shown in Fig. 3(b). This branch is specially designed for a large upscaling factor of 8. Instead of a sequential connection of 3 upsampling layers with an upscaling factor of 2, which might stretch the images and generate severe artifacts for lack of adequate information forwarding or feature reuse, we adopt a multipath upscaling strategy to mitigate such artifacts. Specifically, the outputs of *Path 2-4* and *Path 1-4* are concatenated as the inputs of *Path 4-8*; the outputs of *Path 4-8* and *Path 2-8* are concatenated as the inputs of the final convolution. We use $\mathbf{I}^Y$ to represent the output of the reconstruction branch, which is also the final output of the 3D-Y-Net generator.

d: DISCRIMINATOR

Given that the unsupervised GAN model is adopted here to solve a supervised regression problem, the original discriminator that gives high scores to realistic samples is not theoretically applicable for this supervised regression problem because our generator does not sample prior vectors from random noise. Instead, our discriminator is designed to be of a conditional structure [23], [24]. Specifically, the discriminator can recognize the input of the generator such that it can classify a reconstruction mapping from thick-section images to thin-section images as "real" or "fake." The detailed network structure is shown in Fig. 4. This structure takes $\mathbf{I}^A$, $\mathbf{I}^S$, and $\mathbf{I}^Y$ as fake inputs and $\mathbf{I}^A$, $\mathbf{I}^S$, and $\mathbf{I}^{GT}$ (ground-truth images) as real inputs and outputs a score tensor for later computation of loss functions.

2) 3D-DENSEU-NET

As the second stage of the whole framework, a 3D-Dense U-Net followed by a stack of 2 enhanced residual blocks is proposed for detail recalibration, whose network structures are shown in Fig. 5. The key point of detail recalibration lies in information reuse. To reuse axial thick-section images, we simply insert $\mathbf{I}^A$ into $\mathbf{I}^Y$ according to their corresponding spatial positions, which is denoted as $\mathbf{I}^{YA}$. Through this way, axial thick-section images can be easily used to correct axial images. When reusing sagittal thick-section images, slice insertion is not applied. The main reason is that we would like to reuse all slices in $\mathbf{I}^S$. But if $\mathbf{I}^S$ and sagittal-slice-inserted $\mathbf{I}^Y$ are both used as inputs of the second stage network, more information of sagittal slices than that of axial slices will be introduced into 3D-DenseU-Net, which could decrease the image quality in the axial plane. Based on the above consideration, We set $\mathbf{I}^Y$, $\mathbf{I}^S$, and $\mathbf{I}^{YA}$ as inputs of the 3D-DenseU-Net and let $\mathbf{I}^R$ denote the final output thin-section MR images. Notably, the dense architecture we adopt allows the output of the previous convolutional layers to be passed down to several convolutional layers, which, according to [21], [25], can fully leverage low-level and high-level features.

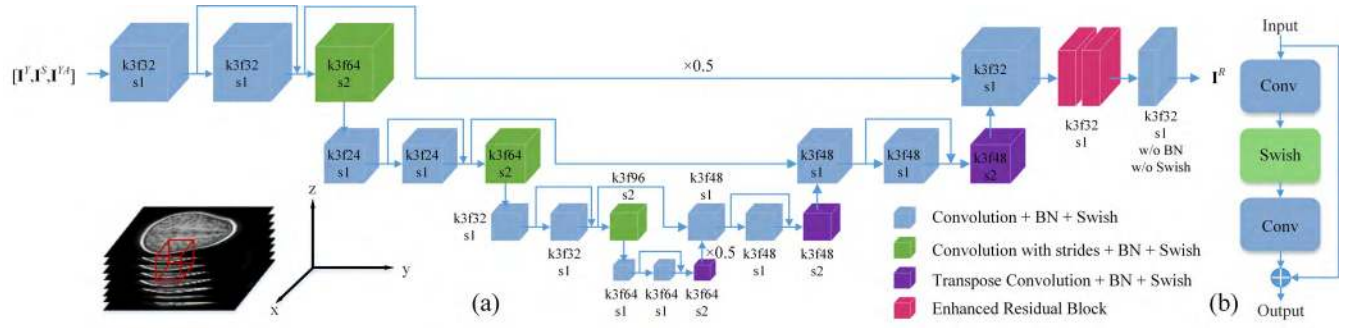


FIGURE 5. Network structure of the second stage. (a) is 3D-DenseU-Net; (b) is the enhanced residual block. Red framework represents patches for training. *×0.5* represents value decay by a factor of 0.5.

Additionally, to prevent blurriness and structural distortion caused by top-level and bottom-level skip connections, we apply value decay before channelwise concatenation to balance feature maps at different levels. Moreover, the tail enhanced residual blocks are also designed for similar consideration, since shallow features passing through the top-level skip connection could corrupt the final outputs. Without traditional batch normalization layers, enhanced residual blocks are also preferred as they cut down GPU RAM usage, thus allowing larger batch size in training phase.

Given limited GPU capacity, there is a trade-off between convergence rate and receptive field. Specifically, relatively larger patch size leads to larger receptive field, thus more useful information can be seen by convolution kernels. But it also reduces the maximum batch size we could use, which could harm the convergence rate especially when batch size is already small. After hyperparameter search, we train 3D-DenseU-Net based on randomly sampled patches with size of $48 \times 48 \times 48$ to strike a balance between convergence and receptive field.

It is worth noting that 3D-DenseU-Net and 3D-Y-Net-GAN are trained separately instead of end-to-end. Two major reasons can account for this. First, separate training could guarantee the functionality of initial reconstruction as designed for 3D-Y-Net-GAN, and also decouple the functionality of two stages. Second, end-to-end training of two 3-D DL models are currently not feasible on our GPU resources if using acceptable batch size.

3) LOSS FUNCTION

To train the 3D-Y-Net-GAN to learn the mapping from $\mathbf{I}^A$ and $\mathbf{I}^S$ to $\mathbf{I}^Y$, we need to search the set of network parameters $\boldsymbol{\theta}_G$ and obtain the optimal parameters $\hat{\boldsymbol{\theta}}_G$ that minimize the generator's loss function L_G, which is described as in (1), where G is taken as the generator and $\mathbf{I}^{GT}$ is taken as the ground truth.

$$\hat{\boldsymbol{\theta}}_G = \underset{\theta_G}{\operatorname{argmin}} L_G\left(G\left(\mathbf{I}^A, \mathbf{I}^S\right), \mathbf{I}^{GT}\right) \tag{1}$$

To find a loss function that evaluates the difference between the generated images and the ground-truth images, we design a loss function that consists of a self-adaptive Charbonnier loss, a 3-D gradient correction loss, an adversarial loss, and an l_2 weight regularization term:

$$\mathrm{L_G} = \mathrm{L_{SC}^G} + \lambda_1 \mathrm{L_{GC}^G} + \lambda_2 \mathrm{L_{AD}^G} + \lambda_3 \mathrm{L_{WR}^G} \tag{2}$$

where λ_1, λ_2, and λ_3 represent the respective terms' weights. The above four components will be further discussed in the following paragraphs. Given the second stage 3D-DenseU-Net is not based on adversarial learning, we train it with the same loss function L_G as used for 3D-Y-Net-GAN except that λ_2 is set to 0.

a: SELF-ADAPTIVE CHARBONNIER LOSS

In supervised regression problems, the ℓ_1 and ℓ_2 norms are widely used because pixelwise restriction is practically important to guarantee the basic SSIM. However, the ℓ_2 norm often leads to overly smooth results, and the ℓ_1 norm penalizes the deviation of the prediction from the ground truth indiscriminately. In one study [20], a Charbonnier loss, a differentiable variant of the ℓ_1 norm, showed better performance and higher robustness than an ℓ_1 and ℓ_2 norm. In another study [5], a cubic-weighted mean square error (MSE) loss was introduced to emphasize the performance in ''difficult'' areas, which represent areas with relatively large pixelwise differences between generated images and ground truth. However, the difference between the ground truth and upsampled images through bicubic interpolation will not always be a good indicator of the actual difficult areas along the training process and is even worse when facing a large upscaling factor. Therefore, we propose the use of the dynamic coefficients calculated by the difference between the current generated images and the ground truth to weigh the robust Charbonnier loss:

$$\mathrm{L_{SC}^G} = \frac{1}{rLWH} \sum_{x,y,z=1,1,1}^{L,W,rH} \sqrt{\left(\mathbf{I}_{x,y,z}^{GT} - \mathbf{I}_{x,y,z}^{Y}\right)^2 + \varepsilon} \cdot \left(\frac{1}{2} + \frac{\left(\mathbf{I}_{x,y,z}^{GT} - \mathbf{I}_{x,y,z}^{Y}\right)^2}{2\max\left(\left(\mathbf{I}^{GT} - \mathbf{I}^{Y}\right)^2\right)}\right) \tag{3}$$

where ε is a small value, which is set to 10^{-6}, $(\cdot)^2$ performs element-wise square if input is a tensor, and max$(\cdot)$ function calculates the global maximum element of a tensor, which outputs a scalar value.

b: 3-D GRADIENT CORRECTION LOSS

A Charbonnier loss merely addresses the pixelwise difference, which may lead to inadequate attention to the second-order differential information. This being the case, we adopt a 3-D gradient correction loss to explicitly exert a second-order constraint between adjacent pixels along the x,y, and z-axes, which can help our model generate sharper edges:

$$\begin{aligned} L_{GC}^{G} = &\mathbb{E}\left[\left(\nabla_x \mathbf{I}_{x,y,z}^{GT} - \nabla_x \mathbf{I}_{x,y,z}^{Y}\right)^2\right] \\ &+ \mathbb{E}\left[\left(\nabla_y \mathbf{I}_{x,y,z}^{GT} - \nabla_y \mathbf{I}_{x,y,z}^{Y}\right)^2\right] \\ &+ \mathbb{E}\left[\left(\nabla_z \mathbf{I}_{x,y,z}^{GT} - \nabla_z \mathbf{I}_{x,y,z}^{Y}\right)^2\right] \end{aligned} \tag{4}$$

c: ADVERSARIAL LOSS

To make the generated images more realistic, we utilize a conditional discriminator to supervise the learning process of the generator. Taking into account robustness and implementation efficiency, we use the LSGAN loss as the adversarial loss. For the conditional discriminator, its loss function is defined as follows:

$$\mathrm{L}^{D} = \frac{1}{2}\mathbb{E}\left[\left(\mathrm{D}\left(\mathbf{I}^{GT}, \mathbf{I}^{A}, \mathbf{I}^{S}\right) - 1\right)^2 + \left(\mathrm{D}\left(\mathbf{I}^{Y}, \mathbf{I}^{A}, \mathbf{I}^{S}\right) - 0\right)^2\right] \tag{5}$$

where D represents the discriminator and $\mathbb{E}$ represents the mathematical expectation, which practically calculates the mean value of the output tensor. To make it clear, the discriminator tries to make the score of ground truth close to 1 and that of fake inputs close to 0.

The generator tries to fool the conditional discriminator by increasing the score of the fake samples. Accordingly, the adversarial loss for the generator is shown below:

$$\mathrm{L}_{AD}^{G} = \mathbb{E}\left[\left(\mathrm{D}\left(\mathbf{I}^{Y}, \mathbf{I}^{A}, \mathbf{I}^{S}\right) - 1\right)^2\right] \tag{6}$$

Notably, the balance between the generator and the discriminator is crucial when training GANs, which means we need to strike a balance between the adversarial loss and the Charbonnier loss. Therefore, we consider this adversarial loss an auxiliary term in the generator's loss function and set a small value for its weight λ_2. The hyperparameter setting of λ_1, λ_2, and λ_3 will be further discussed in the Experimental Results section.

d: ℓ_2 WEIGHT REGULARIZATION LOSS

Theoretically, parameters with smaller norms lead to lower model complexity, which is indicative of a decreased likelihood of encountering the overfitting problem. Thus, we adopt an ℓ_2 weight regularization loss to mitigate overfitting problem in this study:

$$\mathrm{L}_{WR}^{G} = \sum \|\mathbf{W}_G\|_2^2 \tag{7}$$

where $\mathbf{W}_G$ represents all the kernel weights of the generator, and $\|\bullet\|_2$ represents the ℓ_2 norm.

III. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of multiplanar MR image fusion, we conduct an ablation experiment among 3 cases: 1) our full framework with axial and sagittal images as inputs (Ours Full), 2) a partial version of our method with only axial images as input (Ours Partial Axial), and 3) a partial version of our method with only sagittal images as input (Ours Partial Sagittal). Specifically, for Ours Partial Axial and Ours Partial Sagittal, we modify the 3D-Y-Net generator to have two FE or FF branches and discard the input $\mathbf{I}^{S}$ or $\mathbf{I}^{YA}$, respectively, at the second stage. After the above network modifications, we have two partial versions of our proposed framework, which only leverage thick-section MR images in a single plane.

To validate our proposed comprehensive loss function, we conduct another ablation experiment among four cases: 1) $\ell_1 norm + \mathrm{L}_{GC} + \mathrm{L}_{AD} + \mathrm{L}_{WR}$, 2)$\mathrm{L}_{SC} + \mathrm{L}_{GC} + \mathrm{L}_{WR}$, 3)$\mathrm{L}_{SC} + \mathrm{L}_{AD} + \mathrm{L}_{WR}$, *and* 4)$\mathrm{L}_{SC} + \mathrm{L}_{GC} + \mathrm{L}_{AD} + \mathrm{L}_{WR}$.

To evaluate our proposed reconstruction method, three representative methods and our proposed first-stage network are used for comparison: 1) traditional bicubic interpolation [13], 2) sparse representation (SR) [4], 3) 3D-SRU-Net [5], and our proposed first stage 3D-Y-Net-GAN. Each of those comparison methods will be detailed below.

Traditional bicubic interpolation [13] is an untrainable algorithm that predicts a certain pixel with adjacent 16 pixels. SR [4] is a trainable method that seeks a SR for each patch of the low-resolution image and then uses the coefficients of this representation to generate its high-resolution counterpart. Specifically, we train the coupled dictionaries based on the 2-D slices in the sagittal plane. As described in the introduction, 3D-SRU-Net has been proposed for isotropic super-resolution reconstruction from nonisotropic 3-D electron microscopy. In [5], low-resolution and high-resolution images are jointly leveraged when training this variant of the original U-Net to predict high-resolution images from their blurred counterparts. In our paper, we increase the depth of its network and append 3 convolutional layers with strides of [2,1,1] to its entry, similar to that in our second-stage framework, such that its upscaling factor is extended to 8 and can take the same inputs as our proposed first-stage networks. We also show the reconstruction results of our proposed first stage network 3D-Y-Net-GAN in order to validate the effectiveness of the second stage network.

For quantitative evaluation, we adopt metrics, including peak signal-to-noise ratio (PSNR), SSIM, and normalized mutual information (NMI), for image quality assessment. Note that we clip the pixels that are out of the valid dynamic range $[-1,1]$ and cast the generated MR images and the

TABLE 1. Imaging parameters of our dataset.

Imaging Parameters	Axial Thin-section Images	Axial Thick-section Images	Sagittal Thick-section Images
Imaging Pulse Sequence	3D T1 BRAVO	T1 FLAIR	T1 FLAIR
Voxel Size (mm)	0.5×0.5×1	0.4687×0.4687×6.5	6.5×0.4687×0.4687
Number of Slices	126-153	19	19
Repetition Time (ms)	8.17	1450	2291.30
Echo Time (ms)	3.17	25.10	25.30
Inversion Time (ms)	450	627.84	749.77
Flip Angle (degree)	12	111	111

ground truth to an 8-bit grayscale. PSNR is defined as follows:

$$PSNR{=}20 \cdot \log_{10}\left(\frac{MAX_I}{\sqrt{\frac{1}{rLWH}\sum_{x,y,z}\left(\mathbf{I}^{R}_{x,y,z}-\mathbf{I}^{GT}_{x,y,z}\right)^2}}\right) \tag{8}$$

where MAX_I represents the maximum pixel value, which is 255 in this case; and r represents the upscaling factor, which is 8 in this case. L, W, and H represent the spatial size of the generated MR images, which are 144, 184, and 120, respectively, in this case. SSIM measures the structural similarity between two images by calculating their cross-correlation, which is defined as follows:

$$SSIM = \frac{(2\mu_a\mu_b + c_1)(2\sigma_{ab} + c_2)}{(\mu_a^2 + \mu_b^2 + c_1)(\sigma_a^2 + \sigma_b^2 + c_2)} \tag{9}$$

where μ_a and μ_b represent the respective mean values of two images; σ_a^2 and σ_b^2 represent the respective variances; σ_{ab} represents the covariance of the two images; and $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ are two constants that prevent the dominator from being 0, where k_1 and k_2 are typically set to be 0.01 and 0.03, respectively; and L represents the dynamic range of the pixel values, which is set to be 255 in our case. NMI measures the mutual dependence between two variables, which is defined as follows:

$$\begin{cases} H(X) = -\sum_{x_i \in X} p(x_i) \log p(x_i) \\ H(X,Y) = -\sum_{y_j \in Y}\sum_{x_i \in X} p(x_i, y_j) \log p(x_i, y_j) \\ NMI(X,Y) = 2\dfrac{H(X) + H(Y) - H(X,Y)}{H(X) + H(Y)} \end{cases} \tag{10}$$

where $H(X)$ is the entropy of variable X, $H(X,Y)$ is the joint entropy of X and Y,$p(x_i)$ is the marginal probability distribution function of x_i, and $p(x_i, y_i)$ is the joint probability distribution function of x_i and y_i. Higher PSNR, SSIM, and NMI mean that the generated MR images are much closer to the ground truth.

A. DATA AND PREPROCESSING

We validate our two-stage framework on the reconstruction of thin-section infant head MR images. Thick-section and thin-section MR images of 154 infants aged 2 to 5 years old were collected from the Children's Hospital of Fudan University, Shanghai, China. For each individual infant, we collected axial thick-section, sagittal thick-section, and axial thin-section MR images with the specific imaging parameters listed in Table 1. We randomly selected 40 samples for the cross-validation dataset, another 65 samples as the independent testing dataset 1, and the rest 49 samples as the independent testing dataset 2. Note that the collection time interval of two independent testing sets is half a year. We applied spatial normalization, grayscale normalization, and histogram matching to our raw MR image data for data preprocessing. We use preprocessed thick-section images as inputs of our model, and preprocessed thin-section images as the ground truth during the training phase.

Given different imaging parameters (e.g. field of view) and various intensities between thin-section and thick-section MR images, we observed spatial misalignment and intensity imbalance in raw image-domain MRI data, for which raw MR images in DICOM format can not be directly used in our experiments. Thus we preprocess all raw MR images as followings. For registration, we apply unified spatial normalization to all the MR images using MATLAB tools SPM12 [26], to mitigate spatial misalignment between thin-section and thick-section MR images. We firstly transform MR images from Digital Imaging and Communications in Medicine (DICOM) format to Neuroimaging Informatics Technology Initiative (NIfTI) format. Secondly, we segment the infant brain atlas [27] to generate the full version of tissue probability maps (TPMs) which contain probability maps of various tissues in the image data, including gray matter (GM), white matter (WM), cerebrospinal fluid (CSF), skull, scalp, and air mask. Thirdly, SPM12 estimates nonlinear deformation field that best aligns the generated TPMs to the individual's MR images. Then, MR images are warped according to their own estimated deformation field. Finally, we obtain $\mathbf{I}^A$ of size 144 × 184 × 15, $\mathbf{I}^S$ and $\mathbf{I}^{GT}$ of size 144 × 184 × 120. For detailed configuration of registration, we set voxel size of thin-section images to 1 × 1 × 1 mm ^{3}, axial thick-section images to 1 × 1 × 8 mm ^{3}, and sagittal thick-section images to 1×1 × 1 mm ^{3}. Besides, we use ICBM Asian brain template in affine regularization and adopt appropriate bounding box such that registered MR images have the exact spatial size as illustrated in Fig. 2. Other configurations are kept default. After registration, possible misalignment due to various spatial positions and head shapes is minimized. Note that the field of view of sagittal

TABLE 2. Quantitative evaluation of thin-section MR image reconstruction methods using different input data: PSNR, SSIM, and NMI.

Dataset	Input plane(s)	PSNR (dB)		SSIM		NMI	
		Mean (std.)	Med.	Mean (std.)	Med.	Mean(std.)	Med.
Cross-validation Dataset	Axial Only	19.53(0.79)	19.60	0.63(0.04)	0.64	0.20(0.01)	0.20
	Sagittal Only	19.65(0.99)	19.57	0.67(0.05)	0.69	**0.21(0.02)**	**0.21**
	Axial and Sagittal	**19.75(0.85)**	**19.69**	**0.69(0.05)**	**0.71**	**0.21(0.02)**	**0.21**
Independent Testing Dataset	Axial Only	18.77(0.96)	18.84	0.62(0.05)	0.63	0.19(0.01)	0.19
	Sagittal Only	19.06(0.75)	19.03	0.64(0.03)	0.65	0.19(0.01)	0.20
	Axial and Sagittal	**19.12(0.85)**	**18.98**	**0.66(0.03)**	**0.66**	**0.20(0.01)**	**0.20**

Bold text indicates the best performance.

thick-section MR images is smaller than that of thin-section images, thus there are uncovered head areas at each side in $\mathbf{I}^S$, as shown in Fig. 2(c). To avoid the structural incompleteness of $\mathbf{I}^S$, we upsample corresponding areas in $\mathbf{I}^A$, and simply use them to fill the uncovered areas in $\mathbf{I}^S$. Since SPM12 can not guarantee successful registration on all samples, we actually found 5 poorly-registered samples in the cross-validation dataset, which were accordingly excluded from it. Thus, 35 samples comprise the actual cross-validation dataset for all experiments.

Also, given that registered MR images have a 16-bit grayscale with various intensities among different subjects, we normalize intensities of all MR images into [−1,1], using simple linear transformation. Then, we apply a histogram-matching algorithm to all MR images with a fixed sample as reference to eliminate histogram imbalance.

In order to enlarge our training dataset and mitigate the overfitting problem for the data-driven DL model, we adopt data augmentation by applying radial transformation [28] and mirror reflection to our training dataset.

B. EXPERIMENTAL SETTINGS

We adopt 5-fold cross-validation on the cross-validation dataset to evaluate our framework. For fold s, we divide the cross-validation dataset of 35 samples randomly into 2 parts, with 7 samples as the validation data and the other 28 as the training data. For data augmentation, we apply radial transformation and mirror reflection to the training data, such that it is enlarged to 336 samples at the first stage and 56 samples at the second stage. All of the validation procedures are applied to 5 iterations. To further validate the generalization of our proposed model, we select a certain model with the best performance in the cross-validation and evaluate it on independent testing dataset 1 of 65 samples, and independent testing dataset 2 of 49 samples, whose collection time interval is half a year.

For 3D-Y-Net-GAN, we randomly sample 12 patches per volume with a size of $32\times32\times15$ for $\mathbf{I}^A$ and $32\times32\times120$ for $\mathbf{I}^S$ and $\mathbf{I}^{GT}$. The mini-batch size and epoch are set to 16 and 200, respectively. For the generator, we use the Adam optimizer [29] with the momentum parameter $\beta_1 = 0.9$ and adopt a stepwise, exponential-decay learning rate schedule with initial value $= 5 \times 10^{-4}$, decay step $= 252$, and decay rate $= 0.989$. We use the same optimizer and learning rate schedule for the discriminator. We initialize the generator and the discriminator with an He normal initializer [30]. We set λ_1, λ_2, and λ_3 in L_G to be 0.2, 0.02, and 0.1, respectively.

For 3D-DenseU-Net, we randomly sample 80 patches per volume with a size of $48 \times 48 \times 48$. The mini-batch size and epoch are set to 12 and 300, respectively. We use the Adam optimizer with $\beta_1 = 0.9$ and adopt a stepwise, exponential-decay learning rate schedule with initial value $= 5 \times 10^{-4}$, decay step $= 373$, and decay rate $=0.989$. We initialize it with the He normal initializer and set λ_1 and λ_3 in its loss function to 1 and 0.001, respectively. Note that not like training, inference is not patch-based. On the contrary, it is based on whole MR images. Therefore, no special post-processing is needed in our method.

For the SR [4] method, we set appropriate parameters for coupled dictionary training. Concretely, we set dictionary size $= 512$, patch number $= 100{,}000$, patch size $= 13 \times 13$, sparsity regularization $= 0.15$, and overlap $= 12$. Notably, for bicubic interpolation and SR methods, we only utilize axial thick-section MR images given their limitations.

For 3D-SRU-Net, we choose appropriate hyper-parameters to guarantee its best performance while maintaining good comparability. Concretely, we consider the patch size of $32 \times 32 \times 15$ for $\mathbf{I}^A$ and $32 \times 32 \times 120$ for $\mathbf{I}^S$ and $\mathbf{I}^{GT}$. We set the mini-batch size and epoch to 32 and 300, respectively. We adopt the Adam optimizer with a parameter of $\beta_1 = 0.9$, initial learning rate $= 5 \times 10^{-4}$, and the bicubic-weighted MSE loss function as adopted in [5].

The SR method was implemented in MATLAB2017a. The training process took approximately 10 hours, while the reconstruction process took approximately 2 hours per sample. All the DL methods were implemented with Python3.6.2 and TensorFlow1.3, running on a NVIDIA Titan Xp GPU with 12 GB of RAM. Our 3D-Y-Net-GAN took approximately 20 hours for training, 3D-DenseU-Net took approximately 20 hours for training, and 3D-SRU-Net took approximately 11 hours for training.

C. ABLATION EXPERIMENT ON INPUT DATA

In this section, we design an experiment on the three aforementioned cases to demonstrate the impact of different input data. The reconstruction results of the three cases

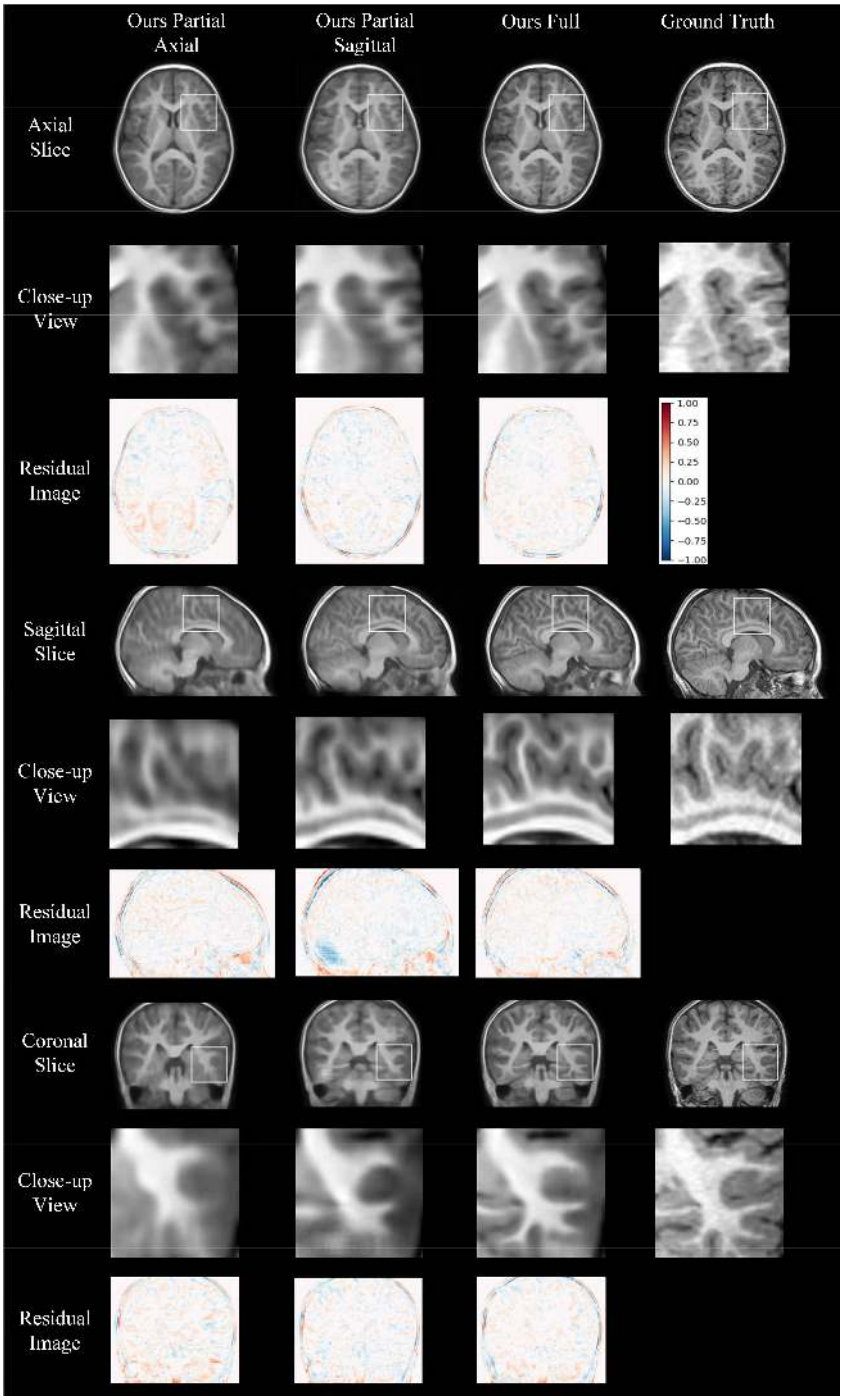


FIGURE 6. Visual comparison to show the contribution of different input data.

are visualized in Fig. 6. We can see that reconstructed thin-section MR images based on images in the axial and sagittal planes have more structural details and less distortion compared to images generated from single-plane thick-section images. This is because multiplanar thick-section MR images could be fused and thus contribute collaboratively to the reconstruction task. Their quantitative evaluation is summarized in Table 2, which shows that the reconstruction method with multiplanar MR image fusion can generate thin-section images of higher similarity with ground-truth images.

D. ABLATION EXPERIMENT ON LOSS FUNCTION

In this section, to validate the contribution of each term in our proposed comprehensive loss function, we set three comparison experiments to show the effectiveness of self-adaptive

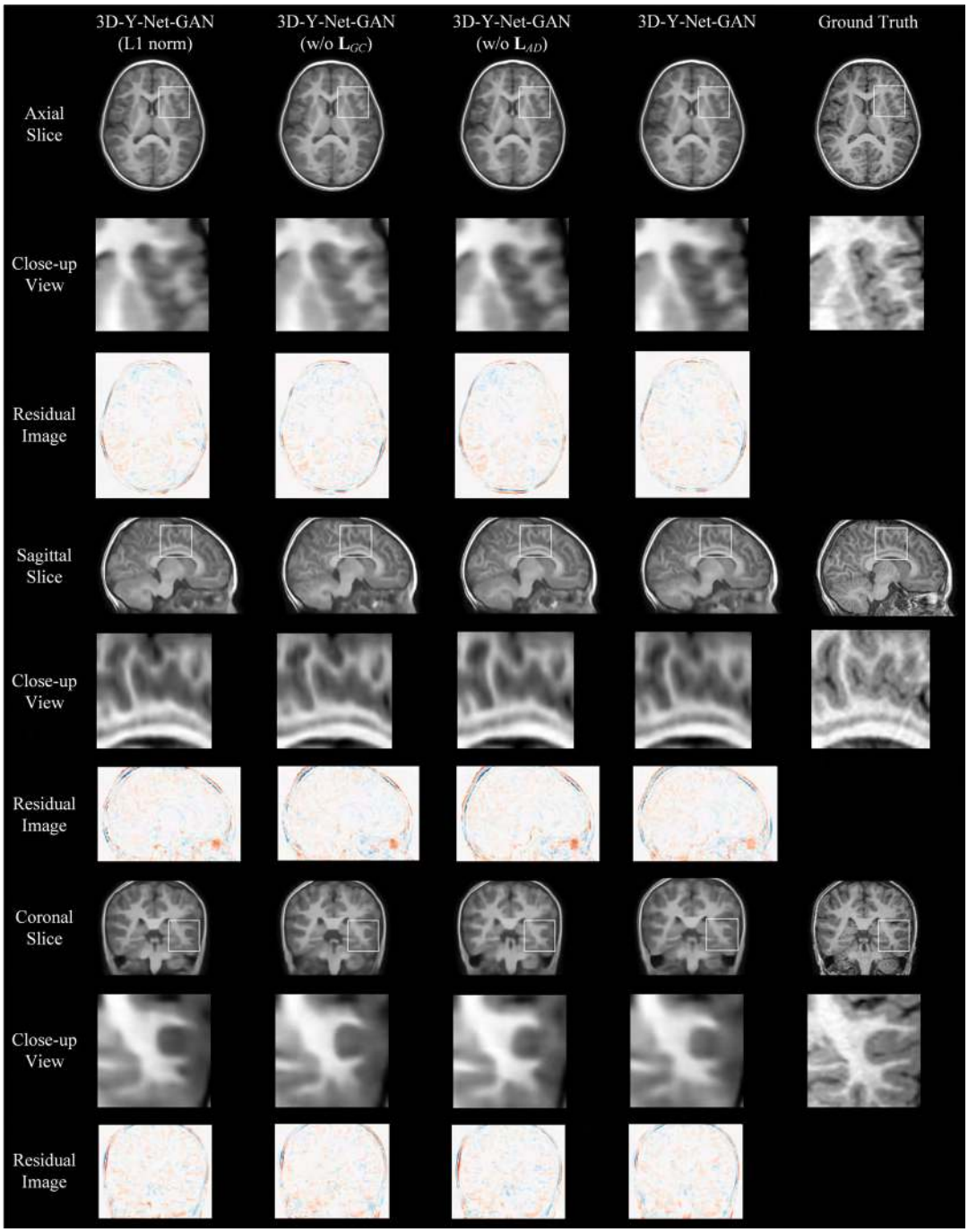


FIGURE 7. Visual comparison to show the effectiveness of our proposed comprehensive loss function.

Charbonnier loss, gradient correction loss, and adversarial loss. Note that this ablation experiment is based on our proposed 3D-Y-Net-GAN, and we do not conduct 5-fold cross-validation here. Their reconstruction results are shown as Fig. 7. From the visualization comparison, we can see that the ℓ_1 norm generates blurry images compared to self-adaptive Charbonnier loss. The results based on a loss function without gradient correction loss show less sharp edges compared to our proposed loss function. A loss function without adversarial loss generates less realistic images than our proposed loss function. The quantitative evaluation shown in Table 3 further validates the contribution of our proposed loss function.

E. COMPARISON WITH OTHER METHODS

In this section, we design a comparison experiment to evaluate our proposed method by comparing it with three existing methods, traditional bicubic interpolation [13], sparse

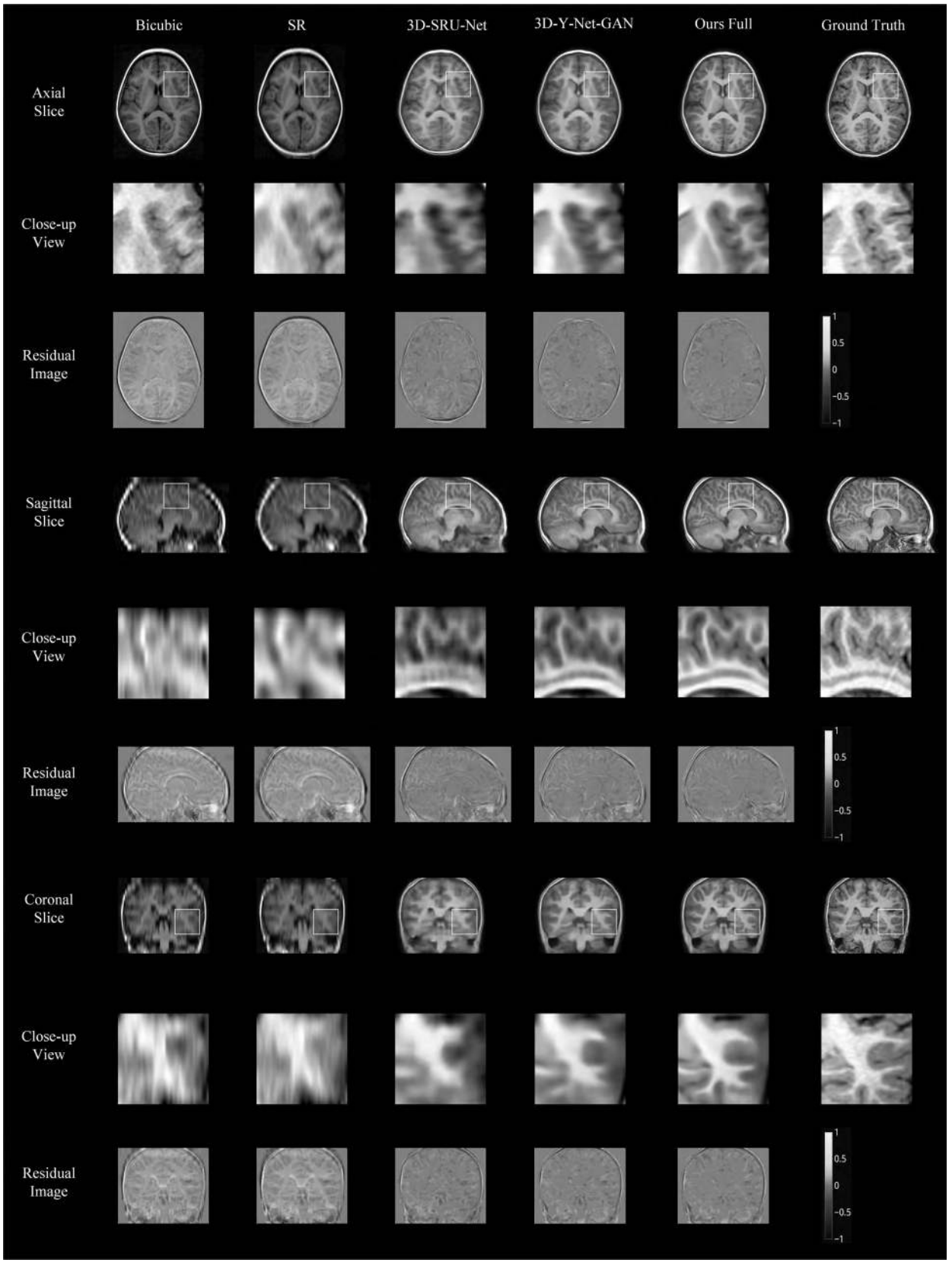


FIGURE 8. Visual comparison among four reconstruction methods. Color bars illustrate the intensity range of residual images. The first, fourth, seventh rows illustrate the axial, sagittal, and coronal views of the reconstructed thin-section MR images by using four different methods, respectively. The second, fifth, and eighth rows illustrate the local enlarged views. The third, sixth, and ninth rows illustrate the error maps.

TABLE 3. Quantitative evaluation of thin-section MR image reconstruction methods using different loss functions: PSNR, SSIM, and NMI.

ℓ_1 norm	L_{SC}	L_{GC}	L_{AD}	PSNR (dB)		SSIM		NMI	
				Mean (std.)	Med.	Mean (std.)	Med.	Mean(std.)	Med.
×		×	×	19.96(0.57)	20.09	0.70(0.03)	0.71	0.21(0.01)	0.21
	×	×		19.92(0.63)	19.75	0.70(0.03)	0.71	0.21(0.01)	0.21
	×		×	19.94(0.56)	20.03	0.70(0.03)	0.71	0.21(0.01)	0.21
	×	×	×	**20.03(0.61)**	**20.06**	**0.70(0.03)**	**0.71**	**0.21(0.01)**	**0.21**

× represents involvement of the corresponding term in the loss function. Bold text indicates the best performance.

TABLE 4. Quantitative evaluation of thin-section MR image reconstruction methods: PSNR, SSIM, NMI, and MAE.

Dataset	Methods	Input plane(s)	PSNR (dB)		SSIM		NMI		MAE	
			Mean (std.)	Med.	Mean (std.)	Med.	Mean(std.)	Med.	Mean(std.)	Med.
Cross-validation Dataset	Bicubic [13]	Axial Only	14.59(0.44)	14.58	0.31(0.04)	0.30	0.17(0.01)	0.17	32.96(1.69)	33.11
	SR [4]	Axial Only	14.65(0.52)	14.72	0.30(0.05)	0.30	0.15(0.01)	0.16	32.58(2.58)	32.48
	3D-SRU-Net [5]	Axial and Sagittal	19.07(0.69)	19.17	0.61(0.05)	0.62	0.19(0.01)	0.20	15.16(0.33)	15.15
	3D-Y-Net-GAN	Axial and Sagittal	19.65(0.81)	19.58	0.68(0.04)	0.69	**0.21(0.01)**	**0.21**	14.16(1.77)	13.65
	Ours Full	Axial and Sagittal	**19.75(0.85)**	**19.69**	**0.69(0.05)**	**0.71**	**0.21(0.02)**	**0.21**	**13.90(1.79)**	**13.40**
Independent Testing Dataset 1	Bicubic [13]	Axial Only	13.59(0.92)	13.74	0.27(0.06)	0.28	0.16(0.01)	0.16	35.32(3.88)	34.25
	SR [4]	Axial Only	13.83(0.95)	13.96	0.28(0.06)	0.29	0.15(0.01)	0.15	35.26(3.87)	34.13
	3D-SRU-Net [5]	Axial and Sagittal	17.63(1.27)	17.47	0.57(0.06)	0.57	0.17(0.02)	0.17	18.29(2.77)	18.34
	3D-Y-Net-GAN	Axial and Sagittal	18.00(1.41)	17.63	0.60(0.07)	0.60	0.18(0.02)	0.18	17.00(2.96)	17.15
	Ours Full	Axial and Sagittal	**18.18(1.44)**	**17.86**	**0.61(0.08)**	**0.61**	**0.19(0.02)**	**0.19**	**16.93(2.95)**	**16.98**
Independent Testing Dataset 2	Bicubic [13]	Axial Only	13.76(1.20)	14.26	0.27(0.07)	0.28	0.16(0.01)	0.16	34.16(4.95)	32.03
	SR [4]	Axial Only	13.66(1.03)	14.04	0.23(0.05)	0.23	0.14(0.01)	0.14	35.33(4.33)	33.18
	3D-SRU-Net [5]	Axial and Sagittal	18.14(1.55)	18.84	0.56(0.08)	0.58	0.19(0.02)	0.19	17.25(3.27)	15.95
	3D-Y-Net-GAN	Axial and Sagittal	18.41(1.71)	19.01	0.59(0.09)	0.62	0.19(0.02)	**0.20**	**16.32(3.48)**	15.13
	Ours Full	Axial and Sagittal	**18.56(1.72)**	**19.17**	**0.61(0.09)**	**0.64**	**0.20(0.03)**	**0.20**	16.40(3.47)	**15.07**

Bold text indicates the best performance. We show an extra metric mean absolute error (MAE) for quantitative evaluation of residual images. Note that all the metrics are calculated based on 8-bit grayscale images.

representation [4], and 3D-SRU-Net [5]. In addition, we also illustrated the result of the first-stage 3D-Y-Net-GAN, to validate the effectiveness of the second-stage network. The reconstruction results of a certain slice in the center of the sampling intervals are visualized in Fig. 8. Compared to the other three methods, our proposed reconstruction framework generates the most realistic MR images, which is closer to the ground truth on the rightmost column of Fig. 8.

The traditional bicubic interpolation method shows blurry reconstructed results and suffers from severe detail distortion as well as artifacts, partly due to its limited receptive fields, untrainable structure, and lack of sagittal information.

The sparse representation method generates smoother results with relatively better tissue coherency than bicubic interpolation but still outputs poor results in the sagittal and coronal planes for its 2-D receptive field and limited modeling capacity.

While 3D-SRU-Net reconstructed thin-section MR images with less artifacts, it provided worse results than our proposed framework. Two factors can account for its worse performance. First, given its single-stage architecture, 3D-SRU-Net suffers from inevitable insufficiency in modeling capacity and thus cannot provide a balance among feature fusion, upsampling, and detail preservation, which leads to a poor performance in the sagittal plane reconstruction. Second, an upscaling *Path 1-8* based on shallow features passes through the top-level skip connection. Potential downsides of this design are that features with severe artifacts caused by fractionally-strided convolution with a small kernel size and a very large upscaling factor is directly passed to last several layers through the top-level connection, which harms the reconstruction results.

In the close-up views, we note that our framework reconstructs realistic images that are spatially closer to the ground truth after the first-stage reconstruction and recovers more tissue details in sagittal and coronal planes after the detail recalibration of the second stage, which reflects the effectiveness of our proposed two-stage reconstruction framework.

The overall experimental results are summarized in Table 4, in which we compare the mean values, standard deviations, and median values of the above metrics. We show the experimental results on three different datasets to illustrate the generalizability and robustness of our proposed method.

Our method outperforms existing methods on all three datasets, with higher PSNR, SSIM, NMI, and mean absolute error (MAE). Specifically, in contrast to the untrainable bicubic interpolation method, our method can learn from training samples to generate images with better tissue coherency. Compared to the SR method, our method can utilize 3-D receptive fields and greater modeling capacity to recover more realistic thin-section images. Also, notice that SR method has worse statistical results on the independent testing dataset 2 than on independent testing dataset 1, which shows that our model has better robustness than SR

method when dealing with different datasets. Compared to 3D-SRU-Net, our full framework can better learn mapping from the thick-section MR images to corresponding thin-section MR images, which means dealing with feature fusion, upsampling, and detail recalibration separately and successively will assign a clear task to the neural networks given their limited modeling capacity. This improvement in the final results further confirms the superiority of our proposed method. Note that our model not only shows better performance on cross-validation dataset, but also shows better reconstruction quality on two more testing datasets. Also, since the number of testing data samples in our experiments is around 4 times as many as the training data samples we used, our proposed reconstruction framework shows good generalizability and robustness to be applied to larger database.

IV. CONCLUSION

We proposed a two-stage reconstruction framework to reconstruct thin-section infant head MR images from thick-section images in the axial and sagittal planes. Our proposed 3D-Y-Net-GAN, trained on paired patches of thick-section MR images, reconstructed preliminary thin-section MR images for subsequent refinement. Then, based on the output of the first stage and original thick-section images, our proposed 3D-DenseU-Net was trained for further detail refinement and performance improvement. Moreover, we proposed a comprehensive loss function composed of a self-adaptive Charbonnier loss, a 3-D gradient correction loss, an adversarial loss, and an ℓ_2 weight regularization loss for more effective and more realistic reconstruction.

Two ablation experiments on different input data and our proposed loss function have been conducted. The visualization and quantitative evaluation demonstrated that our proposed multiplanar image fusion and comprehensive loss function could contribute to performance improvement in reconstruction. A comparison experiment with three existing methods was conducted based on a cross-validation dataset and two independent testing datasets. The quantitative evaluation revealed that our proposed method is able to reconstruct thin-section MR images with higher PSNR, SSIM, and NMI compared to the other three methods, including traditional bicubic interpolation [13], sparse representation [4], and 3D-SRU-Net [5]. Note that we show mean absolute error to demonstrate that our reconstruction results have lower residues by average, where we use 8-bit grayscale for evaluation. Even though MAE of Ours Full is a little bit worse than that of our first stage network on independent testing dataset 2, our full model still shows overall better reconstruction details because its loss function focuses on penalizing outlier pixel predictions to generate more realistic images. In addition, we illustrated visualized results generated from the above four methods to bolster the superiority of our method's performance compared to other methods. Although the objective of our proposed method is the reconstruction of thin-section infant head MR images from thick-section images in axial and sagittal planes, it can be easily extended to other application contexts, such as three-plane reconstruction or adult head MR image reconstruction. Furthermore, this reconstruction method can also be used to normalize image layer spacing, to benefit data-driven researches based on image big data.

Data preprocessing is an important factor to guarantee the applicability of our proposed reconstruction framework. We apply unified spatial normalization, histogram matching, and grayscale normalization to all MR images to mitigate the impacts caused by their various intensities and contrast ranges. We also adopt data augmentation to enlarge our training dataset. In future work, we will generalize our reconstruction method and perform validation on more data of different categories.

REFERENCES

[1] A. P. Mahmoudzadeh and N. H. Kashou, "Interpolation-based super-resolution reconstruction: Effects of slice thickness," *J. Med. Imag.*, vol. 1, no. 3, Dec. 2014, Art. no. 034007.

[2] A. P. Mahmoudzadeh and N. H. Kashou, "Evaluation of interpolation effects on upsampling and accuracy of cost functions-based optimized automatic image registration," *Int. J. Biomed. Imag.*, vol. 2013, May 2013, Art. no. 395915.

[3] Q. Lin, Q. Zhang, and L. Tongbin, "Slice interpolation in MRI using a decomposition-reconstruction method," in *Proc. ICISCE*, Changsha, China, Jul. 2017, pp. 678–681.

[4] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[5] L. Heinrich, J. A. Bogovic, and S. Saalfeld, "Deep learning for Isotropic Super-resolution from non-isotropic 3D electron microscopy," in *Proc. MICCAI*, Montreal, QC, Canada, 2017, pp. 135–143.

[6] Z. Li, Y. Wang, and J. Yu, "Reconstruction of thin-slice medical images using generative adversarial network," in *Proc. MICCAI*, Montreal, QC, Canada, 2017, pp. 325–333.

[7] S. McDonagh, B. Hou, A. Alansary, O. Oktay, K. Kamnitsas, M. Rutherford, J. V. Hajnal, and B. Kainz, "Context-sensitive super-resolution for fast fetal magnetic resonance imaging," in *Proc. MICCAI RAMBO*, Montreal, QC, Canada, 2017, pp. 116–126.

[8] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Med. Imag.*, vol. 38, no. 1, pp. 167–179, Jan. 2019.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, Munich, Germany, 2015, pp. 234–241.

[10] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 105–114.

[11] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 1132–1140.

[12] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, Z. Wang, and S. P. Smolley, "Least squares generative adversarial networks," in *Proc. ICCV*, Venice, Italy, Oct. 2017, pp. 2813–2821.

[13] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 29, no. 6, pp. 1153–1160, Dec. 1981.

[14] M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 2018–2025.

[15] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 1874–1883.

[16] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.

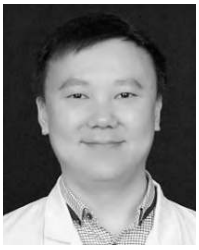
[17] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 1646–1654.
[18] S. Zhang, G. Liang, S. Pan, and L. Zheng, "A fast medical image super resolution method based on deep learning network," *IEEE Access*, vol. 7, pp. 12319–12327, 2018.
[19] H. Lu, Y. Li, S. Nakashima, H. Kim, and S. Serikawa, "Underwater image super-resolution by descattering and fusion," *IEEE Access*, vol. 5, pp. 670–679, 2017.
[20] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 5835–5843.
[21] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," Feb. 2018, *arXiv:1802.08797*. [Online]. Available: https://arxiv.org/abs/1802.08797
[22] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions," Oct. 2017, *arXiv:1710.05941*. [Online]. Available: https://arxiv.org/abs/1710.05941
[23] M. Mirza and S. Osindero, "Conditional generative adversarial nets," Nov. 2014, *arXiv:1411.1784*. [Online]. Available: https://arxiv.org/abs/1411.1784
[24] G.-J. Qi, "Loss-sensitive generative adversarial networks on Lipschitz densities," Mar. 2018, *arXiv:1701.06264*. [Online]. Available: https://arxiv.org/abs/1701.06264
[25] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269.
[26] D. T. Chard, G. J. M. Parker, C. M. B. Griffin, A. J. Thompson, and D. H. Miller, "The reproducibility and sensitivity of brain tissue volume measurements derived from an SPM-based segmentation methodology," *J. Magn. Reson. Imag.*, vol. 15, no. 3, pp. 259–267, Mar. 2002.
[27] F. Shi, P.-T. Yap, G. Wu, H. Jia, J. H. Gilmore, W. Lin, and D. Shen, "Infant brain atlases from neonates to 1- and 2-year-olds," *PLoS ONE*, vol. 6, no. 4, Apr. 2011, Art. no. e18746.
[28] H. Salehinejad, S. Valaee, T. Dowdell, and J. Barfett, "Image augmentation using radial transform for training deep neural networks," Aug. 2017, *arXiv:1708.04347*. [Online]. Available: https://arxiv.org/abs/1708.04347
[29] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, Vancouver, BC, Canada, 2015, pp. 1–13.
[30] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. ICCV*, Santiago, Chile, Dec. 2015, pp. 1026–1034.

JIAQI GU received the B.Sc. degree in microelectronic science and engineering from Fudan University, Shanghai, China, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, The University of Texas at Austin. His current research interests include medical image processing and machine-learning applications.

ZEJU LI received the B.Sc. and M.S. degrees in electronic engineering from Fudan University, China, in 2018. He is currently pursuing the Ph.D. degree with the Department of Computing, Imperial College London. His current research interests include biomedical image processing and deep-learning applications.

YUANYUAN WANG received the B.Sc., M.Sc., and Ph.D. degrees in electronic engineering from Fudan University, Shanghai, China, in 1990, 1992, and 1994, respectively.

From 1994 to 1996, he was a Postdoctoral Research Fellow of the School of Electronic Engineering and Computer Science, University of Wales, Bangor, U.K. In 1996, he went back to the Department of Electronic Engineering, Fudan University, as an Associate Professor. He was then promoted to a Full Professor, in 1998. He is currently the Director of the Biomedical Engineering Center, Fudan University. He has authored or coauthored over six books and 500 research papers. His research interests include medical ultrasound techniques and medical image processing.



HAOWEI YANG received the Master of Medicine degree from Fudan University, Shanghai, China, in 2011. Since 2009, he has been an attending Doctor with the Radiology Department of Children's Hospital of Fudan University. During the master's degree, he was involved in the DTI analysis of temporal lobe epilepsy and continue with further study of imaging epilepsy.

ZHONGWEI QIAO received the M.D. degree from Shanghai Jiao Tong University, Shanghai, China. He was a Postdoctoral Fellow of The University of Hong Kong University. He is currently the Chief Radiologist and the Director of the Department of Radiology of Children's Hospital of Fudan University, and the Principle Investigator in pediatrics Institute of Functional and Molecular Medical Imaging, Fudan University. His research interests include pediatric radiology and AI in radiologic diagnosis.

JINHUA YU received the Ph.D. degree in electronic engineering from Fudan University, Shanghai, China, in 2008.

From 2008 to 2010, she was a Postdoctoral Fellow of the Department of Bioengineering, University of Missouri, Columbia, MO, USA. She is currently a Full Professor with the Electronic Engineering Department, Fudan University. Her current research interests include medical signal analysis, radiomics, and ultrasound imaging.