Received: 8 April 2021 | Revised: 1 December 2021 | Accepted: 7 December 2021

DOI: 10.1002/mp.15438

RESEARCH ARTICLE

MEDICAL PHYSICS

Three-dimensional self super-resolution for pelvic floor MRI using a convolutional neural network with multi-orientation data training

Fei Feng[1] | **James A. Ashton-Miller**[2] | **John O.L. DeLancey**[3] | **Jiajia Luo**[4]

[1] University of Michigan-Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai, China

[2] Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan, USA

[3] Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, Michigan, USA

[4] Biomedical Engineering Department, Peking University, Beijing, China

Correspondence
Jiajia Luo, Biomedical Engineering Department, Peking University, Beijing 100191, China.
Email: jiajia.luo@pku.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 31870942; Peking University Clinical Medicine Plus X - Young Scholars Project, Grant/Award Numbers: PKU2020LCXQ017, PKU2021LCXQ028; PKU-Baidu Fund, Grant/Award Number: 2020BD039; U.S. Public Health Service, Grant/Award Number: R01 HD038665; HHS, Grant/Award Number: P50 HD044406

Abstract

Purpose: High-resolution pelvic magnetic resonance (MR) imaging is important for the high-resolution and high-precision evaluation of pelvic floor disorders (PFDs), but the data acquisition time is long. Because high-resolution three-dimensional (3D) MR data of the pelvic floor are difficult to obtain, MR images are usually obtained in three orthogonal planes: axial, sagittal, and coronal. The in-plane resolution of the MR data in each plane is high, but the through-plane resolution is low. Thus, we aimed to achieve 3D super-resolution using a convolutional neural network (CNN) approach to capture the intrinsic similarity of low-resolution 3D MR data from three orientations.

Methods: We used a two-dimensional (2D) super-resolution CNN model to solve the 3D super-resolution problem. The residual-in-residual dense block network (RRDBNet) was used as our CNN backbone. For a given set of low through-plane resolution pelvic floor MR data in the axial or coronal or sagittal scan plane, we applied the RRDBNet sequentially to perform super-resolution on its two projected low-resolution views. Three datasets were used in the experiments, including two private datasets and one public dataset. In the first dataset (dataset 1), MR data acquired from 34 subjects in three planes were used to train our super-resolution model, and low-resolution MR data from nine subjects were used for testing. The second dataset (dataset 2) included a sequence of relatively high-resolution MR data acquired in the coronal plane. The public MR dataset (dataset 3) was used to demonstrate the generalization ability of our model. To show the effectiveness of RRDBNet, we used datasets 1 and 2 to compare RRDBNet with interpolation and enhanced deep super-resolution (EDSR) methods in terms of peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index. As 3D MR data from one view have two projected low-resolution views, different super-resolution orders were compared in terms of PSNR and SSIM. Finally, to demonstrate the impact of super-resolution on the image analysis task, we used datasets 2 and 3 to compare the performance of our method with interpolation on the 3D geometric model reconstruction of the urinary bladder.

Results: A CNN-based method was used to learn the intrinsic similarity among MR acquisitions from different scan planes. Through-plane super-resolution for pelvic MR images was achieved without using high-resolution 3D data, which is useful for the analysis of PFDs.

KEYWORDS
3D super-resolution, deep learning, MRI

1 | INTRODUCTION

MR imaging is an important modality for medical image analysis. Compared with ultrasound (US) imaging, it provides better image quality and tissue contrast. Therefore, it is suitable for soft tissue imaging and is widely used for the evaluation of pelvic floor disorder (PFD), such as pelvic organ prolapse. Three-dimensional (3D) MR images are commonly used for pelvic organ segmentation,[1–3] pelvic floor evaluation,[4] computer simulation of pelvic organ prolapse,[5,6] and evaluation of tissue material properties.[7] High-resolution MR images are necessary for high-precision analysis of the above tasks. However, acquiring high-resolution 3D MR data is both expensive and time-consuming. Moreover, artifacts due to human movement, breathing, or organ contraction may be introduced when acquiring high-resolution 3D MR images. In addition, it is difficult to maintain the same pose for a long time, such as in maximal Valsalva maneuver. Therefore, it is a common practice to use a stack of 2D slices instead of 3D scans. For convenience, we will use the terms in-plane resolution to refer to the resolution of the 2D slices, and through-plane resolution to indicate the resolution between neighboring 2D slices. The in-plane resolution is usually less than 1 mm, whereas the through-plane resolution is not less than 5 mm. In this way, it increases the spacing between two slices or decreases the through-plane resolution while maintaining the high in-plane resolution characteristics for each 2D slice. This approach reduces the scanning time, but the deterioration of through-plane resolution limits the precision of downstream analysis tasks, such as 3D segmentation, reconstruction, and prolapse evaluation.

Some digital techniques can improve the through-plane resolution when hardware updating is not available. An intuitive solution is the interpolation method, such as bilinear interpolation and spline interpolation. Compared with bilinear interpolation, spline interpolation can produce smoother results. However, interpolation methods cannot consider the semantic and structural information of MR images, so they may cause artifacts. In contrast, the learning-based method uses the structural information between slices to obtain results with better fidelity. When learning, both low- and high-resolution pairs are required. This method usually downscales the high-resolution images to obtain the corresponding low-resolution images, which can be called "self super-resolution." For example, Timofte et al.[8] proposed the anchored neighborhood regression (ANR) method for natural image super-resolution. Schulter et al.[9] used random forest method for local image regression to achieve super-resolution. More recently, methods based on deep convolutional neural networks (CNNs) outperformed the previous methods and produced new state-of-the-art results in image super-resolution (SRCNN and EDSR)[10,11] because of the powerful representation ability of CNNs. However, these methods were designed for natural image super-resolution, and therefore, had some differences when applied to medical image super-resolution, especially for 3D medical image data. Therefore, Peng et al.[12] proposed a spatially aware interpolation network for 3D CT super-resolution. However, such approach requires high-resolution data in the training phase, which is not easily available for pelvic floor MR imaging. Jog et al.[13] used ANR and Fourier burst accumulation (FBA) to achieve neuroimaging super-resolution, and Zhao et al.[14] proposed an improved method for brain MRI based on the EDSR method. Zhao et al.[15] later applied this technique to neural, cardiac, and tongue MR images super-resolution.

In this work, we designed a CNN-based algorithm to achieve super-resolution of 3D pelvic MR images based on only low-resolution 3D MR acquisitions from three orientations. Our contribution can be summarized in three aspects. First, as shown in Figure 1, MR data from three views (coronal, sagittal, and axial) were used to train a 2D super-resolution model. For convenience, we used the terms "high-resolution view" and "low-resolution view" for the 2D MR images with high resolution in both dimensions and for images with low resolution in either dimension, respectively. For example, Figure 1(a) shows a high-resolution MR image, and Figures 1(d) and (g) present its corresponding low-resolution projected images. The three-view data training ensured that the model had the ability to achieve super-resolution on different views, thereby avoiding the use of high-resolution 3D MR images data for training. Second, an advanced deep CNN backbone, RRDBNet[16] was used. As RRDBNet was already shown to have better performance for natural image super-resolution compared with other CNN models, it was used in this work. Third, the 2D super-resolution model was applied sequentially on two low-resolution views to improve the super-resolution performance. Subsequently, we validated the performance of our method in three areas. First, a group of holdout high-resolution MR sequences were used to validate the true super-resolution performance. Second, to show the generalization ability, we applied our method to another, public MR dataset without training a new model. Third, to demonstrate the advantages of our method, we compared the 3D reconstruction results from our method and from the interpolation-based method. In summary, we first demonstrated the 3D self super-resolution of pelvic MR images using the deep CNN method, which means that 3D MR images super-resolution is achieved without any high-resolution 3D MR data.

2 | METHODS AND EXPERIMENTS

The conceptual framework of our method is shown in Figure 2(a). We named the high-resolution 3D MR data as $I_{(x,y,z)}$, where x, y, and z are scanning directions

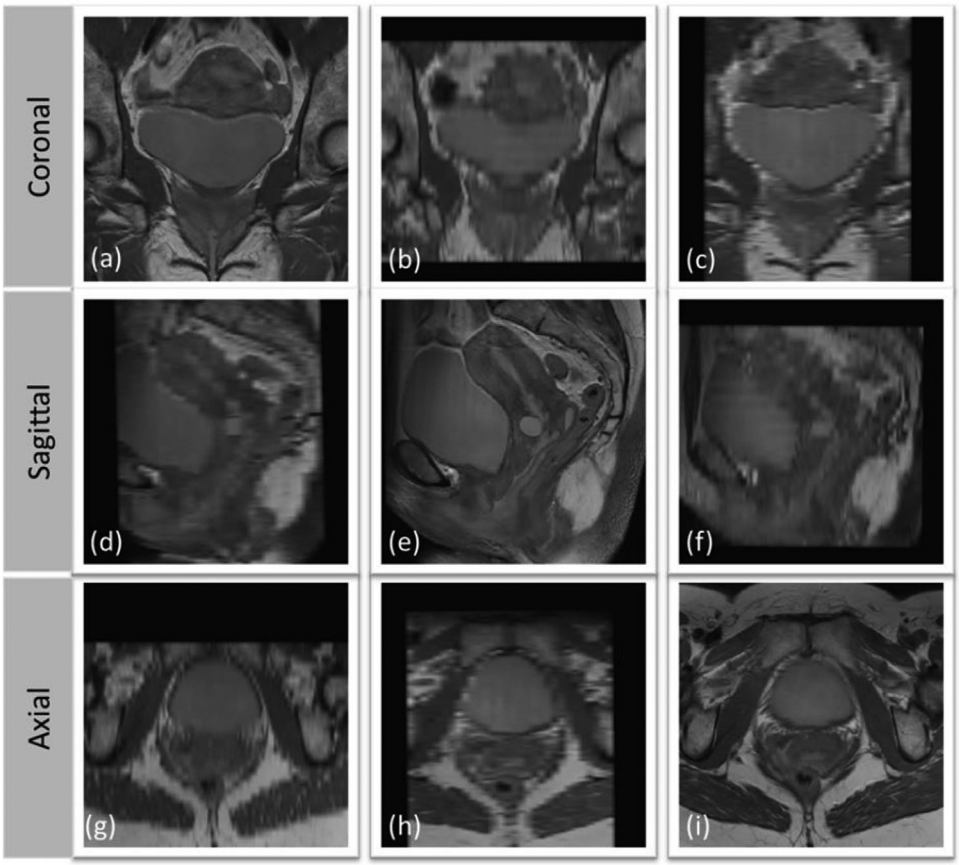

FIGURE 1 Three-view pelvic MR images. (a), (e), and (i) are the scanned high-resolution coronal, sagittal, and axial images, respectively. (d) and (g) are low-resolution projected from (a), (b), and (h) are projected from (e), and (c) and (f) are projected from (i)

for coronal, sagittal, and axial MR data, respectively. For example, for a super-resolution task, we had low-resolution 3D MR data scanned in the coronal view, denoted as $I_{(\hat{x},y,z)}$. Therefore, $I_{(x,y,z)}$ was then expected to be reconstructed from $I_{(\hat{x},y,z)}$. We adopted a 2D approach to address this problem. First, we performed isotropic analytic interpolation of $I_{(\hat{x},y,z)}$ with spline interpolation algorithm. This process ensured that all three dimensions have the same resolution. Next, we sectioned it from the z-axis (axial view) and applied the 2D super-resolution model on all slices. As we estimated $I_{(x,y)}$ from $I_{(\hat{x},y)}$, x-axis resolution was improved. Therefore, we achieved the 3D super-resolution after traversing all the axial view slices and stacking them. As the MR data were reconstructed from the z-axis, we denoted it as $I^{SR-z}_{(x,y,z)}$. Similarly, starting from $I^{SR-z}_{(x,y,z)}$, we continued to apply the same procedures on the y-axis. After that, we obtained the final 3D super-resolution result, which was denoted as $I^{SR-z-y}_{(x,y,z)}$. However, if we changed the order of the super-resolution axis in the process, that is, if we achieved super-resolution from the y-axis before the z-axis, we obtained $I^{SR-y-z}_{(x,y,z)}$. We compared their difference in the Section 3. In our method, no 3D high-resolution MR data were used, and the super-resolution task was simplified as a 2D super-resolution problem that required the 2D CNN model that could achieve super-resolution on multiple views. Therefore, we used MR data from three views for training. As for the CNN model, we used RRDBNet (Figure 2(b)). After fully optimizing the model, the model was applied for the 3D super-resolution. The model training process is introduced in the following subsection.

2.1 | RRDBNet training

Two key points for training the model are the training data and the model structure. To train the CNN model, pairs of low- and high-resolution image data are needed. With all three-view high-resolution 2D MR images that we acquired, we downscaled the high-resolution data in one dimension to create the corresponding low-resolution MR images. As we used three-view MR data to train the model, it ensured that the CNN model had the ability to recover three-view images. Three-view high-resolution 2D MR images have the same image size (256×256), so it ensures that the model training can be performed without resizing. As the obtained low-resolution images are downsampled, cubic interpolation

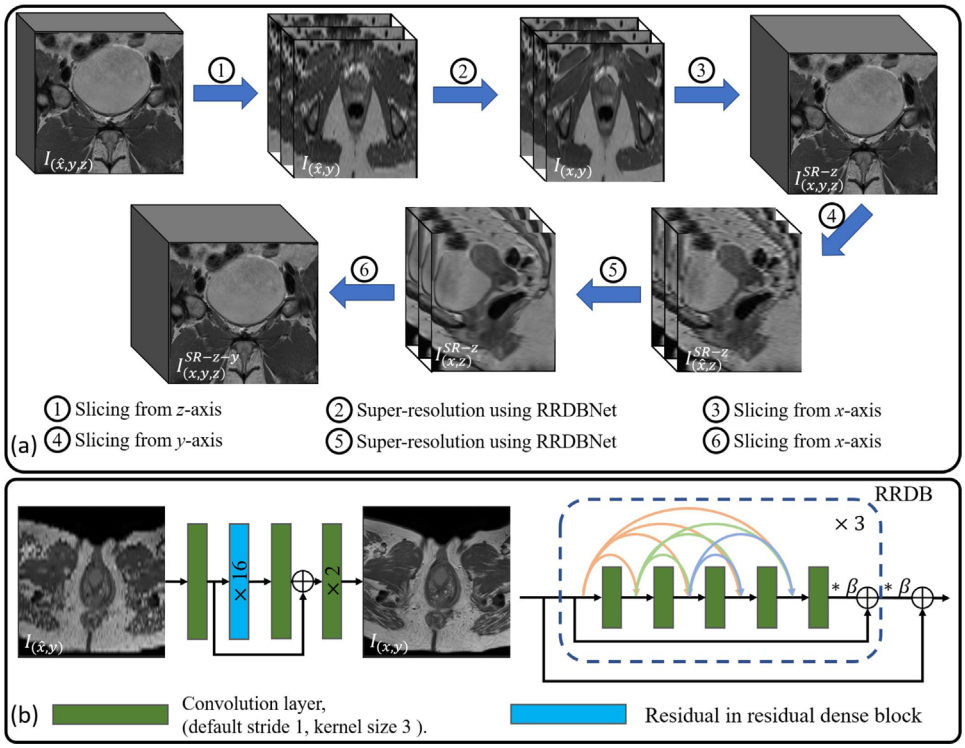

FIGURE 2 The pipeline of our method. (a) Through-plane super-resolution data flow. (b) RRDBNet model structure. × 16 means 16 repetitions. $* \beta$ means the output feature is multiplied by β, where β is equal to 0.2

was used to ensure that the low-resolution images have the same size as high-resolution images.

2.2 | RRDBNet model structure

As shown in Figure 2(b), RRDBNet consists of 16 RRDB modules. Each RRDB module consists of three residual dense blocks (RDB), and there are five densely connected convolutional layers for each RDB.[16] Dense connections ensure that each CNN layer receives the outputs from all previous CNN layers, which promote efficient feature reusing and avoid overfitting. There is a residual connection outside of three RDBs to connect the input and output of RRDB. Residual scaling[17] was used to avoid the training instability and the scaling factor β was set to 0.2 empirically in our experiment. As RRDBNet is a fully convolutional network, different sized inputs are allowed during testing to handle the different sizes of three-view slices that may occur. The fully convolutional networks have been successfully applied to super-resolution, receiving inputs of different sizes.[10] Since pooling is not used in RRDBNet, it manages to retain the maximum information of the input, that is, the input and output images have the same size. Therefore, when images of different sizes are used for testing, the batch size should be set to one. There is no doubt that the size of the input should be larger than the maximum filter size in the network (3×3) and satisfy the maximum memory limit of the processor.

2.3 | Loss function and metrics

In addition, the loss function and evaluation metrics are important for the model training and model evaluation, respectively. The L1 loss is used as the loss function, defined as follows:

$$Loss = \frac{1}{MN} \sum_{m,n}^{M,N} |g_{mn} - p_{mn}|, \quad (1)$$

where M and N are length and width, respectively. g_{mn} and p_{mn} are the pixel values for the ground truth and prediction, respectively. As L1 is the pixelwise evaluation between two images, we used the peak signal-to-noise ratio (PSNR) to evaluate the similarity of two images from the image level. PSNR was defined as follows:

$$PSNR = 20 \log\left(\frac{255}{\sqrt{MSE}}\right), \quad (2)$$

$$MSE = \frac{1}{MN} \sum_{m,n}^{M,N} (g_{mn} - p_{mn})^2, \quad (3)$$

where MSE is the mean square error between the ground truth and prediction. However, PSNR could not guarantee the structural similarity (SSIM) between two images. Previous studies have shown that two images with same MSE can have very different SSIM indices. The image with a larger SSIM has a better visual result.[18,19] Therefore, it is used as a complementary metric to evaluate super-resolution from a macroscopic perspective. SSIM is defined as follows:

$$SSIM = \frac{(2\mu_g\mu_p + 2.55^2)(2\sigma_{g,p} + 7.65^2)}{(\mu_g{}^2 + \mu_p{}^2 + 2.55^2)(\sigma_g{}^2 + \sigma_p{}^2 + 7.65^2)}, \quad (4)$$

where μ_g and μ_p are the average of the ground truth and prediction, respectively. σ_g and σ_p are the standard deviation of the ground truth and prediction, respectively, and $\sigma_{g,p}$ is the covariance between the ground truth and prediction.

To evaluate the overlap of two geometric reconstructions, relative absolute volume difference (RAVD) is defined as follows:

$$RAVD = \frac{|V_1 - V_2|}{V_1} \times 100\%, \quad (5)$$

where V_1 is the reference volume and V_2 is the evaluated volume.

2.4 | Experiments

Three experiments were designed to validate the effectiveness of our method using three datasets. The first dataset, called the dataset 1, consisted of MR data from 43 subjects. Each subject's data included T2 MR data of coronal-, sagittal-, and axial-plane acquisitions. Each 3D MR sequence had an in-plane resolution of 0.78 mm × 0.78 mm and a through-plane resolution of 5.0 mm. The second dataset (dataset 2) consisted of a coronal view 3D MR sequence to quantitatively validate the super-resolution performance. It included 65 images with a through-plane resolution of 2.2 mm and an in-plane resolution of 0.63 mm × 0.63 mm. Both the dataset 1 and dataset 2 were taken from the Michigan Pelvic Floor collection with the approval from the institutional ethics review board. The third dataset (dataset 3) was selected from a public dataset (from the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC))[20,21] to validate the generalization capability of our method. Dataset 3 also had three-view scans with each scan having one high-resolution view. It had a through-plane resolution of 5.2 mm for three-view scans, and its in-plane resolutions varied from 0.78 to 0.94 mm. More imaging parameters for the three datasets are attached in Table S1.

In the first experiment, we split the dataset into a training set and a testing set, containing 34 and 9 subjects' MR data, respectively. There were 3037 images in the training set and 796 images in the validation set. There were 990 coronal, 1020 sagittal, and 1027 axial MR images in the training set. As discussed in Section 2.1, we downsampled the high-resolution images to create their corresponding low-resolution images. As the projected image has only one low-resolution dimension, we downsampled the row or column direction to mimic the projected image. As three-view scans were used for training, it could accommodate the differences among different scans during projection. We set three levels of the downsampling ratios, 2:1, 4:1, and 6:1, respectively. Some examples of the training data are shown in Figure 3. We compared our method with both the spline interpolation method and EDSR method.[11] Note that the EDSR model was trained with all 3037 training images. We also investigated the improvement after training with three-view MR data over training with single-view MR data. We first trained an RRDBNet using 3037 images, which was named $\text{RRDBNet}_{\text{all}}$. We also trained another RRDBNet using only coronal-plane MR images of all training subjects (990 images), and named this model $\text{RRDBNet}_{\text{c}}$. Similarly, we also trained $\text{RRDBNet}_{\text{s}}$ (1020 images) and $\text{RRDBNet}_{\text{a}}$ (1027 images). Since the number of training images for $\text{RRDBNet}_{\text{all}}$ was almost three times higher than for $\text{RRDBNet}_{\text{c}}$, $\text{RRDBNet}_{\text{s}}$, and $\text{RRDBNet}_{\text{a}}$, we used 12 subjects' MR data from three planes (998 images) to train another model, which was named $\text{RRDBNet}_{\text{partial}}$ for comparison.

We used Adam optimizer and an NVIDIA TITAN RTX graphics card with 24 GB of computation memory. RRDBNet was trained for 10^6 batches with a batch size of 4 and a learning rate of 0.0002. After the deep learning model was well optimized, we tested its performance on the testing dataset.

In the second experiment, we used the dataset 2 to validate the 3D super-resolution performance quantitatively. As the original MR data had a relatively high through-plane resolution of 2.2 mm, we evaluated the performance of 3D super-resolution performance on this basis. We extracted half of the slices to generate data with a through-plane resolution of 4.4 mm as a model input, and used the remaining half slices as the ground truth for evaluation. The super-resolution performance was then evaluated from three areas. First, we evaluated the 2D super-resolution performance from the sagittal and axial views. In this step, the spline interpolation and EDSR methods were used for comparison. Second, we obtained the 3D super-resolution results using RRDBNet. We evaluated the super-resolution performance on the hidden slices. The interpolation and FBA[22] method were used for comparison. When applying the model sequentially on the two projection views, there were two variants, which were distinguished

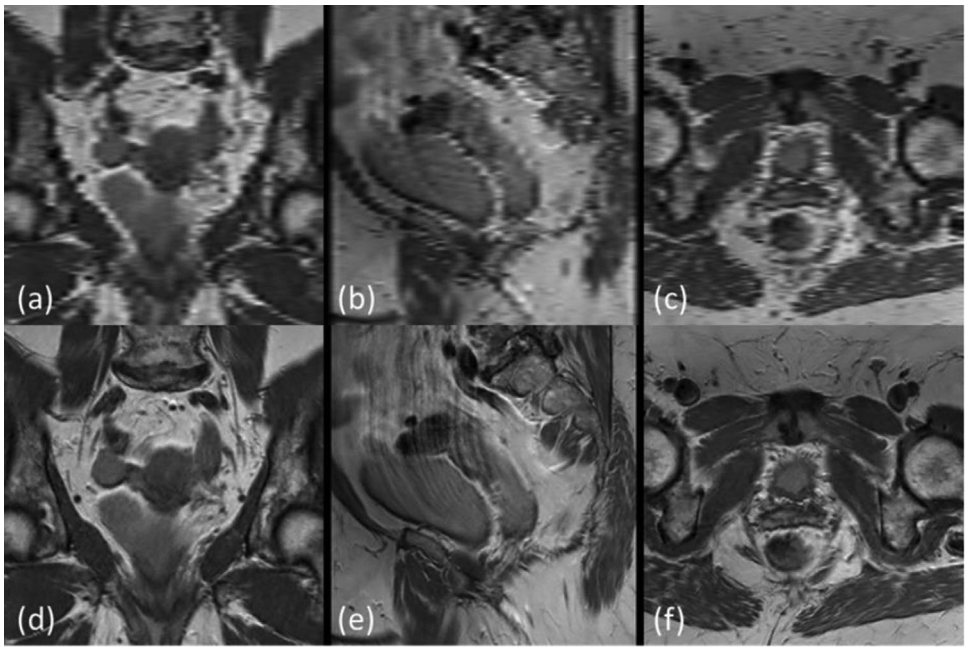

FIGURE 3 Examples of training images from different views. (a), (b), and (c) are downsampled images from coronal(d), sagittal (e), and axial (f) images, respectively

as RRDBNet^{SR-z-y} and RRDBNet^{SR-y-z}. In addition, we also tested the single-view super-resolution variants, which were RRDBNet^{SR-y} and RRDBNet^{SR-z}. Third, we reconstructed the geometrical model of the urinary bladder based on segmentation results from the interpolation method and our method for mutual comparison.

In the third experiment, we directly applied our method to the dataset 3 without training on it. Similarly, the super-resolution performance was also demonstrated in three areas. First, we showed the super-resolution results on the low-resolution views. Second, we showed the super-resolution results on the high-resolution views. Third, the geometrical model reconstruction results were compared. In these comparisons, the spline interpolation method was used as the baseline method.

3 | RESULTS

3.1 | Validation on the testing set of the dataset 1

The super-resolution results of different methods for the testing set of the dataset 1 are summarized in Table 1. Both EDSR and RRDBNet outperformed the interpolation method. However, for CNN-based methods, models based on RRDBNet had higher PSNR and SSIM than EDSR. RRDBNet$_{partial}$ also outperformed the EDSR model. Moreover, PSNR and SSIM values of RRDBNet$_{partial}$ were higher than those of RRDBNet$_c$, RRDBNet$_s$, and RRDBNet$_a$, which were trained using single-view MR data. In addition, RRDBNet$_s$ had better performance than RRDBNet$_c$ and RRDBNet$_a$ and their *p*-values comparison is attached in Table S2.

TABLE 1 Super-resolution performance on the testing set of the dataset 1. *p*-Values were calculated using RRDBNet$_{all}$ as a reference

Methods	PSNR (dB)	*p*-value	SSIM	*p*-value
Interpolation	26.84	<0.001	0.7664	<0.001
EDSR	28.41	<0.001	0.8101	<0.001
RRDBNet$_c$	28.23	<0.001	0.8101	<0.001
RRDBNet$_s$	29.26	<0.001	0.8249	<0.001
RRDBNet$_a$	28.37	<0.001	0.8168	<0.001
RRDBNet$_{partial}$	**29.94**	<0.001	**0.8453**	<0.001
RRDBNet$_{all}$	**30.44**	–	**0.8549**	–

The top two performances were highlighted in bold.

After super-resolution, the through-plane resolution was improved in-plane resolution so that the nominal resolution for dataset 1's super-resolution results is 0.78×0.78×0.78 mm^3. We compared their super-resolution results on the low-resolution views, as shown in Figures 4 and 5. The results obtained by the spline interpolation method have jagged edges, whereas the results by CNN methods are smoother and more faithful. Besides, compared with the EDSR model, RRDBNet obtained better results in terms of image smoothness and fidelity. The high-resolution view images were compared by using rigid registration for reference. In Figure 5, it shows that RRDBNet's results are smoother

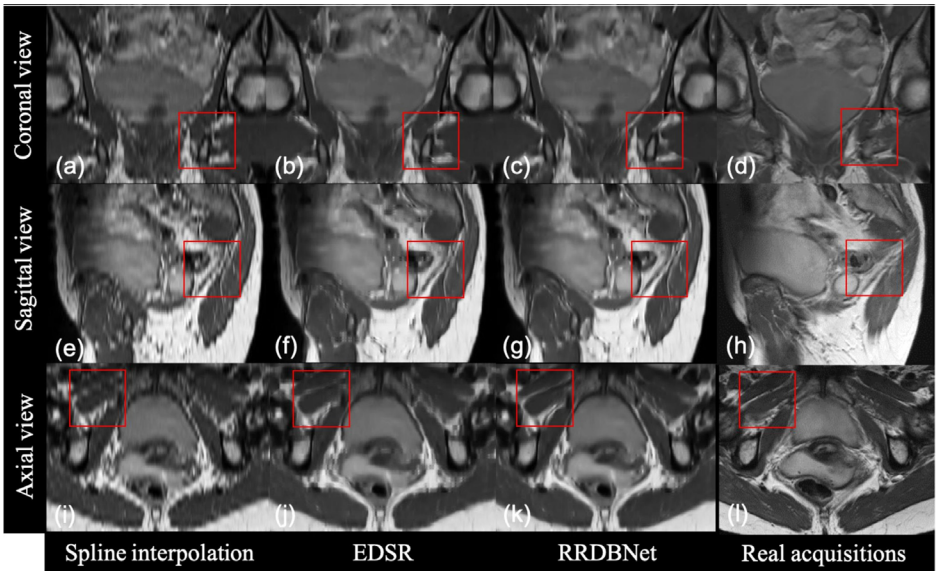

FIGURE 4 Comparison of super-resolution results for the projection view. (a), (e), and (i) are obtained by spline interpolation (order = 3). (b), (f), and (j) are obtained with the EDSR model. (c), (g), and (k) are obtained with RRDBNet. (d), (h), and (l) are reference images from real acquisitions using registration. Regions in red boxes were zoomed in for comparison in Figure 5

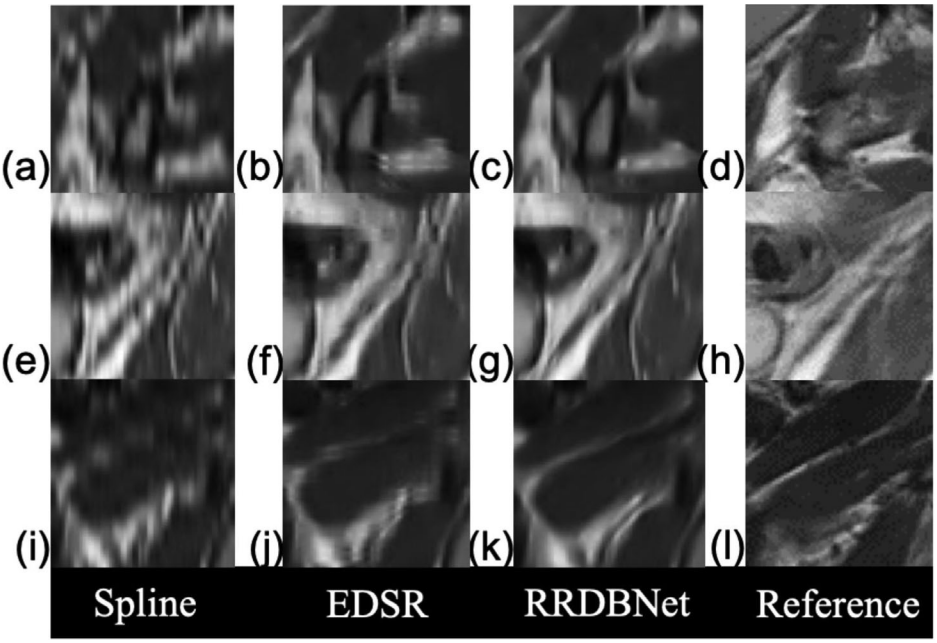

FIGURE 5 Comparison of super-resolution results for local regions. (a)–(l) correspond to the regions in red boxes of (a)–(l) in Figure 4. Spline refers to spline interpolation

than EDSR and interpolation results. Moreover, RRDBNet reconstructed results have high similarity with the reference images but are not exactly matched. Subsequently, we compared their scan-plane results, as shown in Figure S1. Those results show that there are artifacts in the interpolation results. However, the results of CNN-based methods show fewer artifacts and better smoothness. In addition, compared with the EDSR model on the scanned view images, the results of RRDBNet also show smoother edges with fewer artifacts. To demonstrate that the CNN results have more continuous variations, we compared their results in Figure 6. The bladder for RRDBNet has a larger size at + 0 position, whereas it has a smaller size at + 5 position that reflects that the urinary bladder has more continuous changes in the RRDBNet results than in the interpolation results. The original MR images can be found in Figure S2.

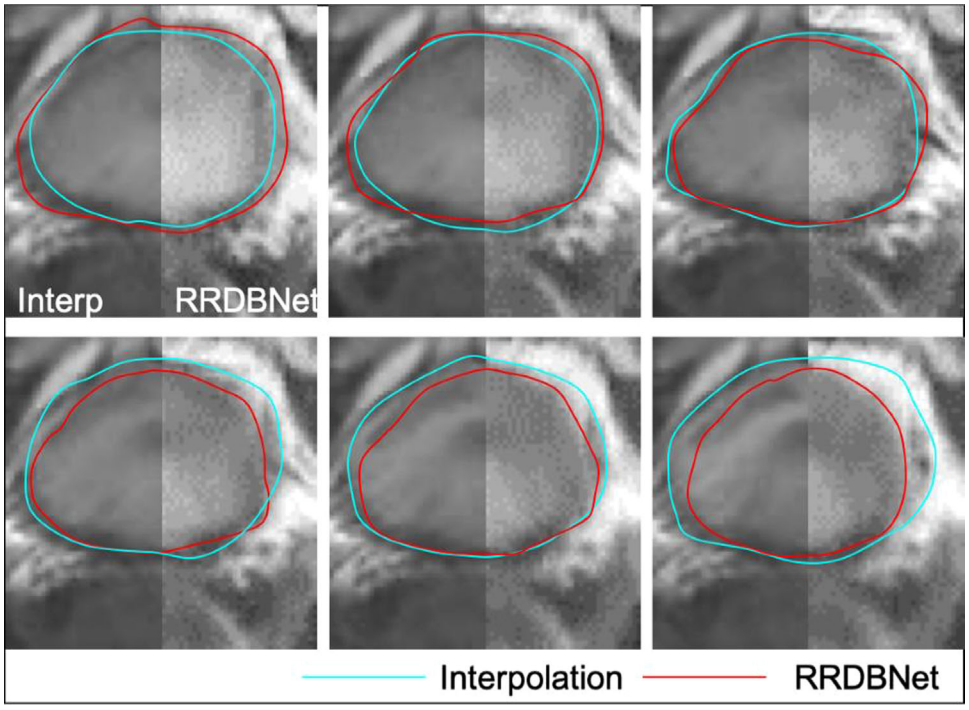

FIGURE 6 Comparison of successive changes in the urinary bladder. In each subimage, the left half is the result of interpolation and the right half is the result of RRDBNet. The urinary bladders are segmented into different colors

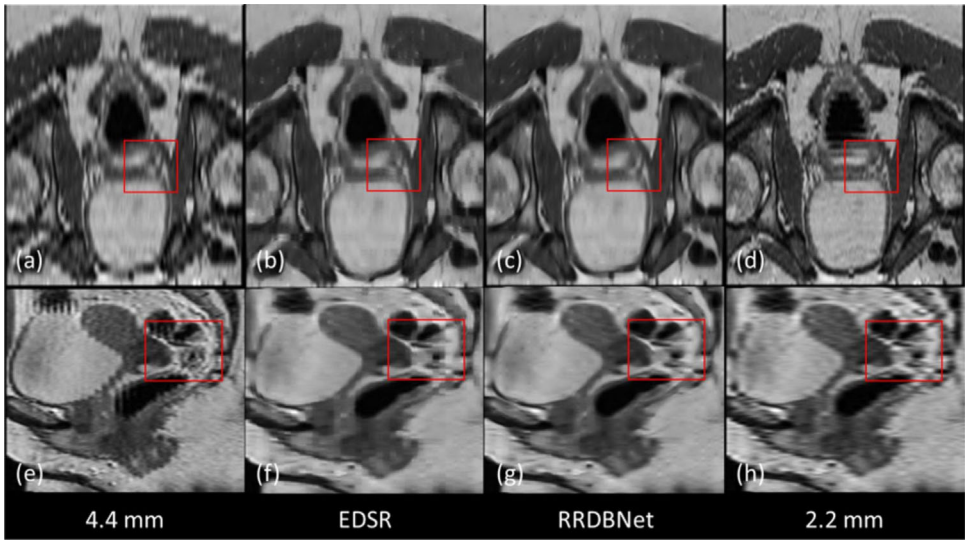

FIGURE 7 Projection view super-resolution performance of the dataset 2. (a) and (e) are the images obtained with spline interpolation (order = 3). The raw through-plane resolution is 4.4 mm. (b) and (f) are the super-resolution results of (a) from the EDSR. (c) and (g) are the super-resolution results of (a) from RRDBNet. (d) and (h) are the reference image data of (a) and (e) with a through-plane resolution of 2.2 mm, respectively

3.2 | Validation on the dataset 2

The nominal resolution of the super-resolution reconstruction of dataset 2 is $0.63\times0.63\times0.63$ mm^3. Some super-resolution examples are shown in Figure 7 for visual comparison. Figures 8(a) and 8(e) show blurred edges of the interpolation results, whereas both CNNs (EDSR and RRDBNet) result in better image smoothness and fidelity. It also shows that the results of RRDBNet (Figures 7(c) and 7(g)) are even smoother than the 2.2 mm reference data (Figures 7(d) and 7(h)). The quantitative results of super-resolution on the projection view (average values for axial and sagittal) are shown in Table 2. The PSNR and SSIM values obtained by

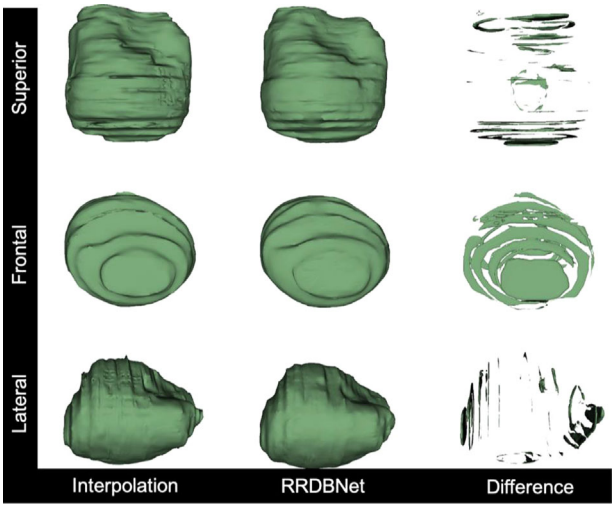

FIGURE 8 Comparison of the geometric model reconstructions of the urinary bladder from three viewpoints. "Difference" means the difference between the reconstructions of the interpolation method and RRDBNet. Geometrical models were smoothed with the same parameters when reconstruction

TABLE 2 Comparison of the super-resolution performance of the projected views of dataset 2. *p*-Values were calculated using $RRDBNet_{all}$ as a reference

Methods	PSNR (dB)	*p*-value	SSIM	*p*-value
Interpolation	26.55	<0.001	0.8083	<0.001
EDSR	26.98	<0.001	0.8192	<0.001
$RRDBNet_{all}$	**27.32**	–	**0.8292**	–

The best performance was highlighted in bold.

TABLE 3 Comparison of scan-plane super-resolution performance for the dataset 2. *p*-Values were calculated using $RRDBNet^{SR-z-y}$ as a reference

Methods	PSNR (dB)	*p*-value	SSIM	*p*-value
Interpolation	21.50	<0.001	0.4809	<0.001
FBA	22.96	<0.001	0.5799	<0.001
$RRDBNet^{SR-y}$	22.40	<0.001	0.5774	<0.001
$RRDBNet^{SR-y-z}$	22.60	<0.001	0.5965	<0.001
$RRDBNet^{SR-z}$	**23.40**	<0.001	**0.6237**	0.156
$RRDBNet^{SR-z-y}$	**23.27**	–	**0.6255**	–

The top two performances were highlighted in bold.

RRDBNet are higher than those obtained by the EDSR model and interpolation method. Therefore, we used the RRDBNet in the following high-resolution view comparisons. The quantitative evaluation results of PSNR and SSIM are summarized in Table 3. Namely, the CNN methods substantially outperformed the interpolation method in both PSNR and SSIM. Besides, both $RRDBNet^{SR-y-z}$ and $RRDBNet^{SR-z-y}$ obtained higher SSIM than the FBA method. In addition, $RRDBNet^{SR-z}$ had better results than $RRDBNet^{SR-y}$ (p-value < 0.001 for both PSNR and SSIM). The scan-plane results are provided in Figure S3. It shows that interpolation results have some ghosting patterns (Figure S3(d)), whereas the RRDBNet results have fewer artifacts. Finally, Figure 8 shows the results of 3D urinary bladder reconstruction by the interpolation method and RRDBNet method. Geometrical models were smoothed under the same configuration during reconstruction. The volume obtained with interpolation results is 412.2 mm^3 and the volume obtained with the RRDBNet result is 409.8 mm^3. The RVAD between them is 0.58%. The "difference" results (Figure 8, column (3)) show the differences between the results of RRDBNet and those of the interpolation method.

3.3 | Generalization testing on the dataset 3

The generalization ability of our method was evaluated using dataset 3. The in-plane resolution of MRI in dataset 3 is 0.78×0.78 or 0.94×0.94 mm^2. After super-resolution, the through-plane resolution was improved to the in-plane resolution, so the nominal resolution of the super-resolution reconstruction of dataset 3 is 0.78×0.78×0.78 or 0.94×0.94×0.94 mm^3 depending on the original in-plane resolution of the 2D high-resolution MR images. Figure 9 shows the super-resolution results in low-resolution views. The MR data from the public dataset have a different appearance from our training dataset. However, the RRDBNet results are sharper and smoother than those obtained using the spline interpolation. Scan-plane results are also provided in Figure S4. It shows the super-resolution results in high-resolution views and the results of RRDBNet have fewer artifacts than interpolation results. Similar to Section 3.1, we selected the urinary bladder as the region of interest and built 3D reconstruction models to evaluate the impact of the super-resolution results on the subsequent reconstruction task. We also used the same smoothing parameters for all geometrical models during reconstruction. The volume obtained with interpolation result is 20.6 mm^3 and the volume obtained with the RRDBNet result is 24.0 mm^3. The RVAD between them is 14.1%. As shown in Figure 10, the shape continuity and surface smoothness of the 3D bladder model obtained by our method are superior to those of the interpolation method. The "difference" results (Figure 1, column (3)) show that there are evident differences between the two reconstructions.

4 | DISCUSSION

We developed a novel CNN-based method for super-resolution of 3D pelvic MR data using only low-resolution 3D data with RRDBNet. There are three novel

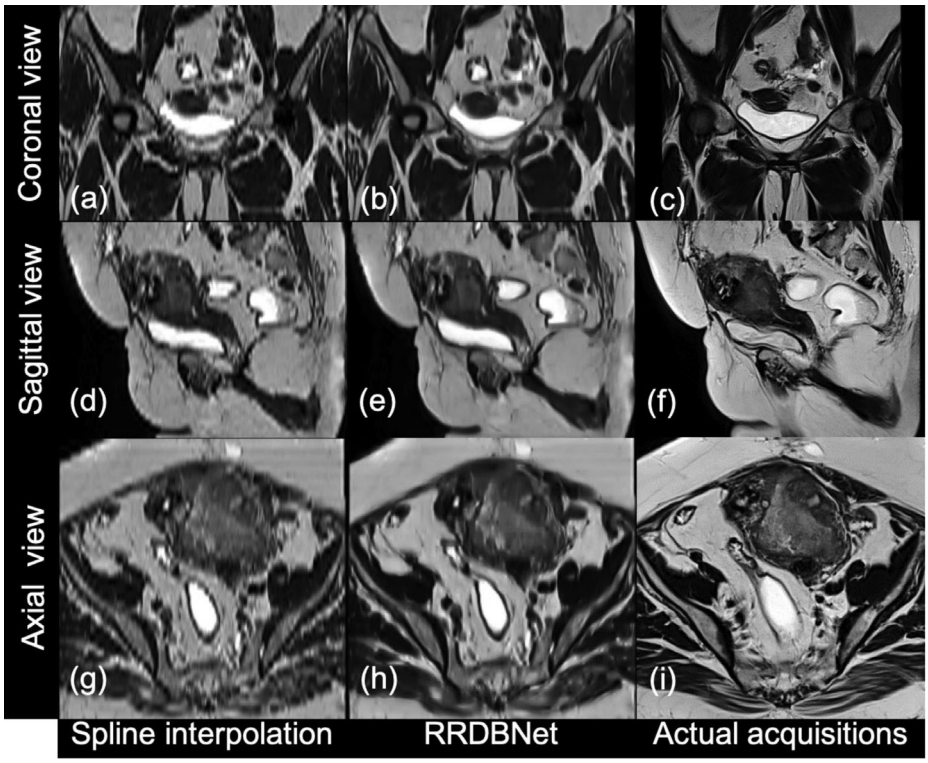

FIGURE 9 Projection views super-resolution results for the dataset 3. (a), (c), and (e) are the results of spline interpolation (order = 3). (b), (d), and (f) are the super-resolution results of RRDBNet

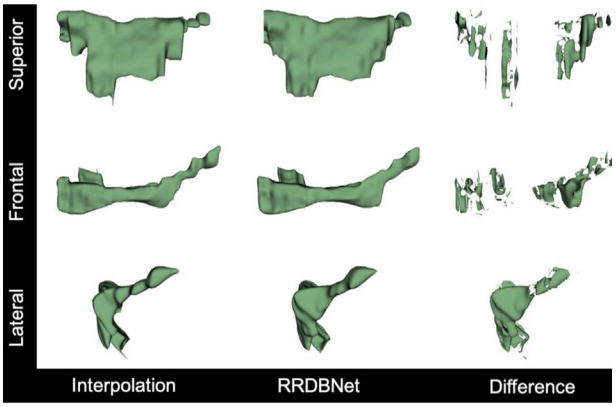

FIGURE 10 Comparison of the geometric model reconstructions of the urinary bladder for the dataset 3 from three viewpoints. "Difference" means the difference between interpolation's and RRDBNet's reconstructions. The geometrical models were smoothed with the same parameters during reconstruction

aspects to this work. First, it represents a new application of 3D self super-resolution of pelvic MR images. We exploited the intrinsic similarity of MR images from three MR views to avoid using 3D high-resolution MR training data, and solved the 3D super-resolution problem using a 2D approach. Second, we established that three-view data could improve the model performance compared with single-view data, even for the same number of images. Third, we demonstrated the advantages of our method with three datasets, proving its effectiveness on MR images super-resolution of different views and 3D geometric model reconstruction.

Super-resolution is crucial for high-resolution and high-precision medical image analysis. Some related works focused on the brain,[14,15,23,24] cardiac,[15] tongue,[15] musculoskeletal,[25] kidney,[12] and knee applications.[26] Compared to them, our pelvic floor imaging study has some important differences. First, we are concerned about the improvement of the through-plane resolution. As for PFD analysis, the through-plane resolution is always the limitation. Some researchers also investigated the through-plane resolution problem by using deep learning methods for other body regions.[14,15,25,27–30] However, most of them still require high-resolution MR images during training.[25,27–29] Second, the pelvic floor has a complex structure and large variability in the shape and size of different organs. The shape and size of some pelvic organs, such as the urinary bladder and uterus, may change due to abdominal pressure and prolapse, whereas other organs in the body, such as the brain, usually have less variation. Third, we used low-resolution MR data from three views for 3D super-resolution, whereas previous studies used the paired training data of low- and high-resolution to train the super-resolution model. Compared with brain imaging, high-resolution 3D pelvic MR data are usually not available, the pelvic floor imaging process is long and costly due to the large pelvic area, and patients

cannot remain in the same position for a long period of time, especially in Valsalva maneuver. Therefore, direct training of a 3D super-resolution model (3D CNN) is not a feasible solution. Previously, Zhao et al.[14,15,30] also investigated the 3D super-resolution problem using CNNs, which does not require 3D MR data for training. In contrast to their approaches, the proposed method takes advantage of three-view training data, so it can learn the view-specific characteristics. In addition, we implemented the super-resolution of projection views sequentially instead of FBA. Moreover, the proposed method used RRDBNet, which showed better super-resolution performance than EDSR. Natural image super-resolution and medical image super-resolution are also closely related. SRCNN,[10] EDSR,[11] and RRDBNet[16] were first used for the natural image 2D super-resolution, but they can also be transferred for medical image super-resolution. Some generative adversarial networks (GANs)[16,31] have been proposed to avoid oversmoothing and to obtain more photorealistic results. However, one challenge of GANs is their unstable training and some efforts have been made to improve the stability of trained GANs. Different medical applications of GANs in super-resolution have also been investigated to produce photorealistic results.[32,33]

As in the first experiment, we used a low-resolution 3D MR dataset to train the CNN models. The results show that the CNN methods have higher PSNR and SSIM than the interpolation method, indicating that the CNN methods have higher image quality and better SSIM with the ground truth data. This is because CNN methods are data-driven methods that make better use of a large amount of training data to capture the structural patterns behind the training data. Therefore, CNN methods can provide results with better smoothness and image fidelity. The downsampling ratios during training were set to 2:1, 4:1, and 6:1, but we further tested the super-resolution performance for data with a downsampling ratio of 7:1, as shown in Table S3. This shows that the CNN method works well when the downsampling ratio does not match the ratio in the training data. However, larger downsampling ratios will make super-resolution more difficult because there is less information available. Next, RRDBNet provided better results than the EDSR model, which means that RRDBNet is more powerful and better suited for this task. Another important question, whether using three views data have better performance than using single-view data, was also investigated and answered. It shows that training with three-view data provided better performance than training with only a single-view data even with almost the same number of images. This finding is relevant for pelvic MR images super-resolution because three-view MR data can be scanned instead of high-resolution MR data from only one view. In this way, three views of MR data triple the number of training images, thereby further improving the super-resolution performance. It also benefits from the fact that each of the three scan planes has complementary strengths and weaknesses based on the angle at which they intersect a structure. One region may be clear on an axial scan but fuzzy in sagittal, whereas another would be the reverse, which can cause the scanning difference among different views. In Figure 5, RRDBNet results are highly similar to the reference images but not fully matched. Besides, the movement during multiple times scanning is also the reason. Moreover, it shows that different results were obtained when using different view data for training. The training results with the sagittal view data are better than those training with the coronal and axial views data. We believe that reflects larger image variation in the sagittal view compared with the other two views, as discussed in a previous study on pelvic organ segmentation.[3] Hence, as more variances were learned by the model, it became more powerful.

Training using 2D high-resolution MR images of three views is featured in this method to avoid using high-resolution 3D MR images for training. As three-view data are not scanned simultaneously, there may be slight differences between the datasets due to motion and breathing. However, it does not affect the proposed approach. Although the 2D MR images are acquired at different times, the model will only use the paired high-resolution images and downsampled low-resolution images from the same view for training. It does not require the information among different views, so it does affect the training process. As for testing, super-resolution model will be applied to a single acquisition of MR data, so these differences between different views will not affect it either. However, the inconsistency between different views resulted in our inability to register the imaging volume of different view scans for training purposes. Training with such registered data would lead to oversmoothed results due to the mismatch between input and output. We only used the image registration method to generate reference data to test the super-resolution performance of $\text{RRDBNet}_{\text{all}}$ trained with downsampled data. As the original MR images have a low through-plane resolution, we aligned the MR images scanned from two planes by image registration. For example, if the sagittal MR image is registered with the coronal MR image, the super-resolution performance of the sagittal projection view acquired from the coronal acquisition can be evaluated with the registered sagittal images. Then, RRDBNet was compared with the spline interpolation method on the projection views using the registered data for PSNR and SSIM, as shown in Table S4 for dataset 1 and Table S5 for dataset 3. These results demonstrated the actual super-resolution capability of RRDBNet.

Next, we quantitatively proved the effectiveness of our method with dataset 2. We hid half of the slices to generate the low-resolution data and used our method to achieve super-resolution. When compared with the

high-resolution data on projection views, RRDBNet had a significant improvement over the EDSR model and interpolation model in terms of PSNR and SSIM, which was consistent with the visual results (Figure 7). We then evaluated the super-resolution performance for the scanned views. Overall, the results of the proposed method are better than those from the interpolation method. The interpolation results have some ghosting patterns (Figure S3(a)), whereas the RRDBNet results also have some artifacts, but these are different from the interpolation ones. The ghosting pattern from the interpolation method is because it does not consider the semantic continuity of the data. And we think that the artifacts in CNN results can be explained in two ways. First, the 3D super-resolution results were achieved using a 2D approach to avoid using high-resolution MR data, which may sacrifice some 3D continuity. Second, original MR data were also acquired slice by slice, which may also introduce some artifacts due to movement and breathing. We found that RRDBNet^{SR-z} outperformed RRDBNet^{SR-y} (Table 3). We think that this may be due to the difference of variance among the three views. Since sagittal view images present a larger variance, it is more difficult to reconstruct from this view. Besides, it shows the differences in achieving super-resolution under different orders. We investigated whether the processing order mattered for the other two datasets. As high-resolution 3D MR data were not available for dataset 1 and dataset 3, we tested the super-resolution performance on projection views with the acquisition of the corresponding planes as reference, using both rigid and nonrigid registration methods. Results for the test data of dataset 1 are shown in Table S4. It shows that for the super-resolution of the coronal view data, the processing from the axial view data outperforms the processing from the sagittal view when both registration methods are used, which is consistent with the results of dataset 2. In addition, for the sagittal view super-resolution of dataset 1, processing from the coronal view is slightly better than processing from the axial view. For the axial view super-resolution of dataset 1, processing from the coronal view produces better results. Similarly, evaluation results for the data of dataset 3 are shown in Table S5. In dataset 3, it is shown that for the coronal view super-resolution, processing from the sagittal view produces better PSNR and SSIM than processing from the axial view when different registration methods are used. For the sagittal view super-resolution of dataset 3, processing from the sagittal view has better SSIM but the difference in PSNR is not significant. For the axial view super-resolution of dataset 3, processing from the sagittal view produces higher PSNR and SSIM. Dataset 1 and dataset 2 are from the same data source and have similar imaging parameters, but they are different from dataset 3, so the differences may be due to the scanning parameters. In the experiments of dataset 2, we evaluated the super-resolution performance on hidden slices, but whether the original inputs changed during this process has not been tested yet. Then we compared the original high-resolution scan plane images in Table S6, and it shows that both spline interpolation and CNN methods introduced small changes in input slices. In terms of PSNR and SSIM, there is no significant difference between the results of the two methods (p-value > 0.05). As the raw MR images were scanned slice by slice, this may lead to discontinuities between slices, which can cause small changes during super-resolution to consider the 3D semantic continuity. Finally, the visual comparison of the 3D geometrical models of the urinary bladder (Figure 8) shows that our reconstructed bladder model has a smoother surface than the interpolation results, especially in regions with dramatic shape changes as indicated from the "difference" results.

Finally, we validated our method on the dataset from a different source, obtained from different scanners and different operators. The results showed that our method yielded high-quality super-resolution results. Similarly, we also compared the reconstruction results of the geometric model of the urinary bladder (Figure 10). We found that the reconstruction results of our method were more faithful in terms of surface smoothness and shape continuity compared to the results of the interpolation method. From the comparison between Figures 8 and 10, the RVAD in Figure 10 is larger than that of Figure 8 because the volume in Figure 8 is larger. When a bladder has a larger volume, the change in shape is flatter, so downsampling has less effect on it. Otherwise, the difference between RRDBNet and the interpolation method is more obvious, which means that the super-resolution is more significant for small features.

In this work, there are some limitations. First, we did not have 3D pelvic floor MR images with high through-plane resolution, so we could not comprehensively assess the 3D super-resolution performance especially in low-resolution views. However, the visual improvement could prove the effectiveness and advantage of our method qualitatively. Second, stress MR images, images made while an individual is straining down, which are used for PFD evaluation, are not included in the current work. Prolapse can be better observed in stress images, where low through-plane resolution exists due to the difficulty of maintaining the maneuver for long periods of time under large abdominal pressure. Therefore, super-resolution in stress MR images is of interest and can be explored in future work. Third, the training images are not sufficient for model training because a deep CNN usually requires "big data" for training. Since we found that RRDBNet$_{all}$ had better performance compared to RRDBNet$_{partial}$, we deduced that more training data could further improve the model performance. However, there are usually limited training sequences for a single hospital or medical center. If

we can utilize data from different sources to train the model, such as the dataset 1 and dataset 3, it may further improve the model performance. Therefore, using data from different sources to improve the performance and generalization of the model is another meaningful direction for further research.

5 | CONCLUSION

We proposed a CNN-based framework to achieve 3D super-resolution for pelvic MR images, while using only low-resolution 3D MR data. Our approach takes advantage of the intrinsic similarity between data from different scan planes for training to achieve 3D super-resolution from projection views. By evaluating low-resolution data, high-resolution data, and unseen data, the effectiveness and good generalization of our method compared with interpolation and EDSR methods were demonstrated. The comparison of 3D urinary bladder geometric model reconstruction results demonstrates that our method could be beneficial for the image analysis and may be useful for high-resolution and high-precision PFD evaluation.

ACKNOWLEDGMENTS

The authors thank for support from NSFC General Program grant 31870942, Peking University Clinical Medicine Plus X-Young Scholars Project PKU2020LCXQ017 and PKU2021LCXQ028, PKU-Baidu Fund 2020BD039, NIH R01 HD038665, and P50 HD044406.

CONFLICT OF INTEREST

The authors have no relevant conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

Author elects to not share data.

REFERENCES

1. Hoyte L, Ye W, Brubaker L, et al. Segmentations of MRI images of the female pelvic floor: a study of inter- and intra-reader reliability. *J Magn Reson Imaging*. 2011;33:684-691.
2. Akhondi-Asl A, Hoyte L, Lockhart ME, Warfield SK. A logarithmic opinion pool based staple algorithm for the fusion of segmentations with associated reliability weights. *IEEE Trans Med Imaging*. 2014;33:1997-2009.
3. Feng F, Ashton-Miller JA, DeLancey JOL, Luo J. Convolutional neural network-based pelvic floor structure segmentation using magnetic resonance imaging in pelvic organ prolapse. *Med Phys*. 2020;47:4281-4293.
4. Larson KA, Luo JJ, Guire KE, Chen LY, Ashton-Miller JA, DeLancey JOL. 3D analysis of cystoceles using magnetic resonance imaging assessing midline, paravaginal, and apical defects. *Int Urogynecol J*. 2012;23:285-293.
5. Chen L, Ashton-Miller JA, DeLancey JOL. A 3D finite element model of anterior vaginal wall support to evaluate mechanisms underlying cystocele formation. *J Biomech*. 2009;42:1371-1377.
6. Luo J, Chen L, Fenner DE, Ashton-Miller JA, DeLancey JOL. A multi-compartment 3-D finite element model of rectocele and its interaction with cystocele. *J Biomech*. 2015;48:1580-1586.
7. Luo J, Smith TM, Ashton-Miller JA, DeLancey JOL. In vivo properties of uterine suspensory tissue in pelvic organ prolapse. *J Biomech Eng*. 2014;136:021016-1-021016–6.
8. Timofte R, De V, Gool LV. Anchored neighborhood regression for fast example-based super-resolution. In: *2013 IEEE International Conference on Computer Vision*; 2013:1920-1927.
9. Schulter S, Leistner C, Bischof H. Fast and accurate image upscaling with super-resolution forests. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2015:3791-3799.
10. Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. *IEEE Trans Pattern Anal Mach Intell*. 2016;38:295-307.
11. Lim B, Son S, Kim H, Nah S, Lee KM. Enhanced deep residual networks for single image super-resolution. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; 2017:1132-1140.
12. Peng C, Lin WA, Liao H, Chellappa R, Zhou SK. SAINT: spatially aware interpolation network for medical slice synthesis. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; 2020:7747-7756.
13. Jog A, Carass A, Prince JL. Self super-resolution for magnetic resonance images. In: Ourselin S, Joskowicz L, Sabuncu MR, Unal G, Wells W, eds. *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016*; Springer International Publishing; 2016:553-560.
14. Zhao C, Carass A, Dewey BE, et al. A deep learning based anti-aliasing self super-resolution algorithm for MRI. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, eds. *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018*; Springer International Publishing; 2018:100-108.
15. Zhao C, Shao M, Carass A, et al. Applications of a deep learning method for anti-aliasing and super-resolution in MRI. *Magn Reson Imaging*. 2019;64:132-141.
16. Wang X, Yu K, Wu S, et al. ESRGAN: enhanced super-resolution generative adversarial networks. In: Leal-Taixé L, Roth S, eds. *Computer Vision - ECCV 2018 Workshops*, Springer International Publishing; 2018:100-108.
17. Szegedy C, Ioffe S, Vanhoucke V, Alemi AA. Aaai. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: *31st Aaai Conference on Artificial Intelligence, Assoc Advancement Artificial Intelligence*, Palo Alto; 2017.
18. Zhou W, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-612.
19. Wang Z, Bovik AC. Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE Signal Process Mag*. 2009;26:98-117.
20. Clark K, Vendt B, Smith K, et al. the cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26:1045-1057.
21. "National cancer institute clinical proteomic tumor analysis consortium (CPTAC)". Radiology Data from the Clinical Proteomic Tumor Analysis Consortium Uterine Corpus Endometrial Carcinoma [CPTAC-UCEC] Collection [Data set]. The Cancer Imaging Archive; 2018.
22. Delbracio M, Sapiro G. Burst deblurring: removing camera shake through Fourier burst accumulation. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2015:2385-2393.
23. Chen Y, Xie Y, Zhou Z, Shi F, Christodoulou AG, Li D. Brain MRI super resolution using 3D deep densely connected neural networks. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*; 2018:739-742.

24. Pham C, Ducournau A, Fablet R, Rousseau F. Brain MRI super-resolution using deep 3D convolutional networks. In: *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*; 2017:197–200.
25. Chaudhari AS, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med*. 2018;80:2139-2154.
26. Neubert A, Bourgeat P, Wood J, et al. Simultaneous super-resolution and contrast synthesis of routine clinical magnetic resonance images of the knee for improving automatic segmentation of joint cartilage: data from the Osteoarthritis Initiative. *Med Phys*. 2020;47:4939-4948.
27. Sood RR, Shao W, Kunder C, et al. 3D Registration of pre-surgical prostate MRI and histopathology images via super-resolution volume reconstruction. *Med Image Anal*. 2021;69:101957.
28. Du J, He Z, Wang L, et al. Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network. *Neurocomputing*. 2020;392:209-220.
29. Georgescu M-I, Ionescu RT, Verga N. Convolutional neural networks with intermediate loss for 3D super-resolution of CT and MRI scans. *IEEE Access*. 2020;8:49112-49124.
30. Zhao C, Carass A, Dewey BE, Prince JL, Self super-resolution for magnetic resonance images using deep networks. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*; 2018:365-368.
31. Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017: 105-114.
32. Chen Y, Shi F, Christodoulou AG, Xie Y, Zhou Z, Li D. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Springer International Publishing; 2018:91-99.
33. You C, Li G, Zhang Y, et al. CT super-resolution GAN constrained by the identical, residual, cycle learning ensemble (GAN-CIRCLE). *IEEE Trans Med Imaging*. 2020;39:188-203.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Feng F, Ashton-Miller JA, DeLancey JOL, Luo J. Three-dimensional self super-resolution for pelvic floor MRI using a convolutional neural network with multi-orientation data training. *Med Phys*. 2022;49:1083–1096.
https://doi.org/10.1002/mp.15438