2D CNN-Based Slices-to-Volume Superresolution Reconstruction

ZHANG SIYUAN\(^1\), DONG JINGXIAN\(^1\), JIANG CAIWEN\(^3\), HOU WENGUANG\(^1\),
AND DENG XIANBO\(^2\)

\(^1\)College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
\(^2\)Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China
\(^3\)School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China

Corresponding authors: Hou Wenguang (houwenguang99@163.com) and Deng Xianbo (1951547885@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 41611530696.

**ABSTRACT** In the case of slice-to-volume reconstruction, the usual method is to head up the slices according to their relative positions. However, it is common that the space between two scanned slices is more than the interval of the pixels on each slice, leading to the result that the reconstructed volume is anisotropic. To obtain isotropic reconstruction, traditional methods attempt to interpolate missing slices based on interpolation algorithms. However, the resolutions of the reconstructed volume are still different in different directions since the slices are highly related. Moreover, this may lead to misplacement among the slices in the nonscanning directions of the volume. As such, we intend to overcome the above two obstacles on the basis of the 2D CNN superresolution strategy. In the proposed method, the superresolution is conducted for the nonscanned slices. Moreover, the samples for training the CNN are specifically designed, which have tomographic features similar to those of the nonscanned slices. Additionally, a multichannel and dense connected CNN is borrowed to perform slice superresolution considering that each slice is actually related to its neighbors. Finally, we conduct slice-to-volume reconstruction after performing superresolution for slices. Experiments in the case of single-channel and multichannel CNNs are conducted to perform validation. The proposed method is robust and can obtain high accuracy.

**INDEX TERMS** Slices-to-volume reconstruction, CNN, isotropic, superresolution.

**I. INTRODUCTION AND MOTIVATION**

In the case of slice-to-volume reconstruction (SVR), such as computed tomography (CT) and ultrasound computed tomography (UCT), the usual method is to heap the slices according to their theoretical locations. In this way, the reconstructed volume is usually anisotropic because the space between the slices is usually larger than the interval of pixels on each slice. The resolutions of the reconstructed volume along the scanning direction should be higher than those of the other coordinate directions. Then, the traditional methods attempt to perform interpolation among the original slices to make the voxel’s size the same in three coordinated axes [1]. Additionally, some two-dimensional (2D) convolutional neural network (CNN)-based methods are borrowed to perform superresolution (SR) for each interpolated slice [2]. They treat the volume as a series of 2D images and combine the SR results based on their spatial positions. This kind of method is simple and efficient. However, the reconstructed volume is still anisotropic since the interpolated images are highly related to the original slices. Moreover, SR for the interpolated slices may lead to misalignment among the slices in the sagittal and coronal planes (the scanning plane is taken as the horizontal plane). This is because the continuous relationship among the neighboring slices is not considered in the process of SVR. Additionally, some three-dimensional (3D) CNNs have been proposed to perform the SVR in [3], [4]. Generally, the 3D CNN contains many parameters, and a large number of samples are needed in training. However, 3D medical samples are difficult to obtain. As such, more focus is still on 2D CNN-based SVR. However, there are two main questions in the usual 2D CNN-based SVR. The first is that the volume is not actually isotropic, and the second is that SR may increase the misalignment among the nonscanned slices.

This paper proposes a solution by introducing the deeply recursive convolutional network (DRCN) [5]. In our application, we interpolate an initial volume as preprocessing. Then, we use the trained DRCN to perform SR for the 2D
nonscanned slices of the initial reconstructed volume. The input of the single-channel DRCN is a low resolution (LR) slice, and the output is the related high resolution (HR) version. The inputs of the multichannel DRCN are several LR slices centered at the considered slice, and the output is the HR version of the considered slice. Finally, we use these HR images to fulfill the SVR. What is important is that two strategies are adopted. The first is the method for generating the training samples, and the other is that the context of each slice is considered in image SR. In generating the LR version from an HR medical image, we first sample the HR image by row/column interval. It should be mentioned that the usual method is to perform downsampling in both directions. Then, the LR version is upsampled to the original size by interpolating the unsampled rows/columns. As such, the original HR and the interpolated LR image are the sample pair. This method makes the relationship between the input and output of DRCN close to the 3D scanning process. That is, this method makes samples have similar tomographic features as those of the nonscanned slices. The other strategy considers the context information of each slice in the process of slice SR by designing a multichannel DRCN, which reduces the misalignment along the nonscan directions on the volume. The single-channel DRCN only enters a single image into the network each time, and it does not consider the relationship among the slices in the neighbor. By comparison, the multichannel DRCN can obtain better results, in which the inputs are $N$ consecutive images and the output is the weighted average of $N$ results. This method takes the spatial continuity of the volume as a precondition and realizes SR reconstruction using only a small number of samples. In addition, the computation is moderate.

II. RELATED WORKS

In image processing, SR refers to the process of synthesizing a new image with richer details from one or some LR images [6]. Specifically, SR refers to the process of reconstructing the HR image from one or more LR images. It is used in solving the problems of low image quality and blurred target area caused by the limitation of the image acquisition environment and system. Generally, SR methods can be divided into three types: interpolation-, reconstruction- and learning-based methods. By comparison, the learning-based method can obtain higher accuracy since it introduces high-frequency information from the samples. As such, the learning-based image SR represents the main stream, which uses the multilayer network to establish the relationship between LR and HR images.

Deep learning was first applied to 2D image SR in 2014 as a superresolution convolutional neural network (SRCNN) [7]. The LR image is upsampled to the desired size via bicubic interpolation before it is input to the SRCNN. The input of the network is the interpolated version, while the output is the related HR image, which establishes a three-layer CNN to map the LR image to the high one. Superresolution using very deep convolutional networks (VDSR) considers a large image context by using a large receptive field. Moreover, residual learning and extremely high learning rates are applied to speed up training [8]. Enhanced deep residual networks for single image superresolution (EDSR) remove unnecessary modules from conventional residual networks (ResNets) and employs residual scaling techniques to stably train large models [9]. This method was extended to the SVR by performing SR for each slice of the volume or directly establishing the 3D CNN [10]. In 2017, Pham applied the convolution network to 3D medical SR reconstruction and officially proposed the superresolution convolutional neural network (SRCNN3D) [11], which replaces the 2D convolution kernel in the convolutional layer with a 3D convolution kernel. Inspired by the SRCNN, Oktay [12] applied residual learning to a 3D CNN [13]. Additionally, they added skip connection points to the network to prevent the gradient vanishing in training and greatly deepened the number of layers in the network. Based on a deep 3D CNN, Pham et al. proposed a multiscale 3D SR reconstruction network recurrent convolutional neural network (ReCNN) [11]. In this way, the LR nuclear magnetic resonance (NMR) image is first interpolated to the target size through upsampling. Then, the network extracts the features of the input image through multiple 3D convolutional layers. In addition, the rectified linear unit (ReLU) is taken as the activation function to obtain the residual image [15]. Finally, the residual image and the input image are added to calculate the final output by skip connection points. ReCNN uses the idea of ResNet to set skip connection points [16]. As such, it increases the convergence speed and avoids the vanishing gradient problem and overfitting [17]. To reuse convolutional layers, Chen proposed a 3D dense skip-connected network that can reconstruct the high-frequency information of MRI brain images [4]. In addition, they also proposed a multilevel densely connected network (MDCSRNN) using generative adversarial networks [19] for NMR image SR reconstruction. The network combines perceptual loss and adversarial loss function to improve the authenticity of the reconstructed image. In 2018, Du proposed the dense denoised superresolution (DDSR) by introducing dense blocks to the application of 3D medical image reconstruction [3], [20]. The DDSR directly takes 3D MRI brain tissue images without interpolation as input. Low-level features are first extracted through a 3D convolutional layer, followed by a series of dense blocks to extract high-level features. Each dense block has a gated unit to fuse different levels of features. There is a skip connection point following each gating unit, which directly connects the result of the gated unit to the final deconvolution layer. Finally, the multilayer information is fused and upsampled by the deconvolution layer to obtain the final HR output. SRCNN3D is combined with a regular space shifting mechanism in [21], in which the quality of the restored images is increased by applying a regular shifting model.

Compared with the SRCNN, ReCNN obtains a larger receptive field by deepening the network structure. With the idea of ResNet, ReCNN uses a residual learning strategy
to learn only the residuals between the input and output to speed up the network convergence. However, the deep ResNet requires longer in training, and each layer is only connected to the previous two or three layers. DDSR uses dense blocks to input the features of each layer to all subsequent layers, and the features of all layers are linked. As such, the final decision is made on the basis of more multilevel information. The defect is that memory consumption is extremely high in training. In SVR of medical images, the above methods can theoretically directly reconstruct the volume using the 3D CNN. However, there are some limitations. First, training a 3D CNN requires a large number of samples, but the volume is difficult to obtain, especially for the HR CNN. Second, the parameters of the 3DCNN are too many, resulting in high time cost and high hardware requirements. To solve these problems, we transfer the 2D DRCN to SVR by designing the single-channel DRCN and multichannel DRCN [22], [23].

III. METHODOLOGY

A. THE BASIC IDEA

As mentioned, training a 3D CNN is relatively difficult to some extent. As such, we still use the 2D CNN to perform SR for each slice and then reconstruct the volume using the processed slices because there are many effective 2D CNNs and the 2D medical samples are easily accessed. The main process of the proposed method is as follows. We preliminarily reconstruct the volume on the basis of the original scanned slices by interpolating the missing slices, i.e., the intervals of the reconstructed volume in three coordinate directions seem to be the same. However, it is actually anisotropic since the interpolated slices are linearly related to the original slices. Then, a multichannel CNN is established whose inputs are several adjoining slices (centered at the considered slice), and the output is the SR result of the center slice. Finally, we assemble these processed slices to finish the SVR. It should be mentioned that the SR is conducted along the sagittal or coronal plane rather than along the horizontal plane, which is one of the characteristics of the proposed method considering that there may be misalignment among the original scanned slices, which leads to a jagged phenomenon when SR is still conducted along the horizontal direction. Figure 1 shows the slices along the coronal plane of an initially reconstructed volume. In each slice, the solid line corresponds to the original scanning layer whose resolution is high. The dotted line is interpolated according to the original ones. The solid line and the dotted line appear alternately in the slices. It should be mentioned that this is just an example; there may be two or more dotted lines between two solid rows.

Before establishing the multichannel CNN, we train a signal channel CNN. First, we generate the training samples using some 2D medical images. The method for generating the samples is shown in Figure 2. The solid line is a row of the original HR pixels, and the dotted line is a row of the LR pixels obtained through interpolation. For an HR medical image, we extract some rows/columns to form the anisotropy LR image. Then, we interpolate the anisotropic LR image to its original size through bilinear interpolation. As such, we actually obtain LR samples that have the same tomographic characteristics as those of the original slices. Then, the LR image is taken as the input, and the HR image is taken as the label to form the sample pair. As such, the network trained by this kind of sample will have the ability to reconstruct HR images in a similar situation. In application, the preliminary reconstructed volume will be decomposed into a series of 2D slices along the coronal/sagittal plane. Then, each slice will be input into the trained DRCN network to obtain an HR image. To some extent, a single-channel DRCN can fulfill the SVR. However, single-channel DRCN reconstructs these 2D slices one by one without considering the relationship among 2D slices. That is, single-channel DRCN does not take the spatial continuity of the volume as a training element. Therefore, it may lead to sawtooth or even abrupt changes in the reconstructed volume.

In response to the above problems, we connect multiple single-channel DRCNs in parallel and add a final output layer to form a multichannel DRCN whose structure is shown in Figure 3. It keeps the parameters of the single-channel DRCN unchanged. To obtain these weights, only a small number of 3D samples is needed (the signal channel DRCN is trained using the 2D samples as mentioned above). For the multichannel DRCN, the inputs are several neighboring slices centered at the considered slice, and the output is the SR result of the center slice. In the application, we input $N'$ (always an odd number) 2D medical slices centered at the considered slice, and the SR result is the weighted average of $N'$ results. In the multichannel DRCN training, only the weights of the output layer need to be optimized, which greatly simplifies the computational complexity. In this way, the SR considers the neighboring slices and then improves the reconstruction quality, especially the spatial continuity.

B. CNN

1) NETWORK STRUCTURE

DRCN applies a recursive neural network to deepen the network and makes the network have a larger receptive field. Additionally, it uses the idea of residual learning to set the skip connection in the network contributing to fine output. Its structure is shown in Figure 4, which is mainly composed of three parts: embedding, inference and reconstructed networks. They are used for extracting features, nonlinearly...
mapping features and outputting HR images. Moreover, the data passes through the inference network multiple times. The inference network is equivalent to multiple convolution layers, H1 to HD, using the same parameters in series, as shown in Figure 5. These convolutional layers are directly connected to the reconstruction layer. The outputs of all layers are added to obtain the output. All recursions are supervised and trained at the same time.

When the network has more layers, gradient vanishing or gradient explosion problems easily occur [24]. To avoid them, two solutions are used: recursive supervision and skip connection [25]. Recursive supervision refers to all recursive networks in the inferential network being supervised. In the inference network part, D convolutional layers share the same set of parameters. The reconstruction network computes the weighted average of these D results to obtain the final result. During training, each weight \( W_i \) is optimized to achieve recursive supervision. Skip connection uses the idea of residual learning, which adds the original input image to the output of each channel. Skip connections can speed up network convergence, avoiding gradient explosion.

2) PARAMETER OPTIMIZATION

DRCN training is used to find the best set of parameters so that the network can obtain the best prediction. The label \( \hat{y}_i \) is the HR image, and \( x^i \) is the interpolated image obtained by degenerating the original image, as shown in Figure 2. The loss function is defined as the minimum variance, which is expressed as:

\[
l_1 (\theta) = \sum_{d=1}^{D} \sum_{i=1}^{N} \frac{1}{2DN} \| y^i - \hat{y}_d^i \|^2, \tag{1}
\]

\[
l_2 (\theta) = \sum_{d=1}^{D} \frac{1}{2N} \| y^i - \sum_{d=1}^{D} w_d * \hat{y}_d^i \|^2, \tag{2}
\]

\[
L(\theta) = \alpha l_1 (\theta) + (1 - \alpha) l_2 (\theta) + \beta \| \theta \|^2. \tag{3}
\]
where $l_1(\theta)$ is the error of the recursive layer’s output, $\hat{y}^d_i$ is the output set of $D$ recursive layers, and $y^i$ is the prediction. $l_2(\theta)$ is the error of the total output. The final objective function $L(\theta)$ of the network needs to optimize the error of each recursive layer’s output and the error of the total output, where $\beta$ is the weight attenuation. The initial value of $\alpha$ is set to be higher to make the training process stable. As the training continues, $\alpha$ gradually decays to improve the performance of the final output. The training of the network is conducted under the TensorFlow framework. Adam instead of the stochastic gradient descent algorithm is used to optimize the objective function to speed up the training process [26], [27].

\[
\text{PSNR} = 10 \log_{10} \frac{255^2 * W * H}{\sum_{i=1}^{W} \sum_{j=1}^{H} [y(i,j) - \hat{y}(i,j)]^2} \quad (4)
\]

\[
\text{SSIM} = \frac{(2u_xu_{x'} + C_1)(2\sigma_{xx'} + C_2)}{(u_x^2 + u_{x'}^2 + C_1)(\sigma_x^2 + \sigma_{x'}^2 + C_2)} \quad (5)
\]

### C. ERROR METRICS

Structural similarity (SSIM) and peak signal-to-noise ratio (PSNR) are adopted here to evaluate the accuracy of SVR. They can be calculated according to Equations 4 and 5, $H$, $W$ are the height and width of the image. $u_x$ and $u_{x'}$ are the mean values of $x$ and $x'$, $\sigma_x$, $\sigma_{x'}$ are their related standard deviations. $\sigma_{xx'}$ is the covariance between $x$ and $x'$. $C_1$ and $C_2$ are two constants [28]. PSNR reflects the difference between the corresponding pixels of two images. Its advantage is that it does not rely on the subjective visual perception of the human eyes since human eyes are more sensitive to luminance contrast than chroma and more sensitive to the contrast of low spatial frequency. In addition, human perception is affected by the environment. SSIM evaluates the differences between two images in terms of structure, brightness and contrast. From the viewpoint of image composition, SSIM reflects the structure of objects in the scene, which is independent of brightness and contrast. At the same time, distortion is modeled as a combination of brightness, structure and contrast.

As mentioned above, we perform SR for the slices along the coronal/sagittal plane. To evaluate the accuracy of the reconstructed volume, we can decompose the reconstructed volume along three directions: the scanning, the sagittal and coronal planes. Afterward, these slices are evaluated with SSIM and PSNR so that we can obtain the objective indexes about the spatial continuity of the reconstructed volume. Additionally, to intuitively reflect the spatial continuity, we can select a row of pixels perpendicular to the coronal plane and draw a pixel value distribution map, as shown in Figure 6. If the reconstructed volume has better spatial continuity, the row pixel distribution is continuous, and there...
are fewer jumps. That is, there are fewer local maximum or minimum points, and vice versa.

IV. EXPERIMENTS
To validate the effectiveness of our strategies, single-channel DRCN and multichannel DRCN are all tested under different conditions. Single-channel DRCN is trained using samples obtained in Figure 2. In multichannel DRCN training, the parameters in the last layer are optimized with a small amount of 3D data. It should be pointed out that the parameters in the single DRCN are fixed in multi DRCN training. All CNNs are trained under the TensorFlow framework, and Adam is used to optimize the network.

A. DATASET
The test dataset contains 457 MRI brain 2D images with different sizes collected from many sources. Approximately 50 CT and MRI volumes were captured from the Union Hospital, Wuhan, China. Their sizes are $256 \times 256 \times 220$ and $652 \times 768 \times 127$. We divided the MRI images into two groups: MRI1 and MRI2. Figure 7 shows one example of each kind of image. These samples are used to train and test the single-channel and the multichannel DRCN. It should be mentioned that the error values represent the average of all the testing volumes. According to the demands of our experiments, they are processed into the following datasets.

2D MR: We randomly order the 457 2D MR brain images to ensure that there is no relationship between the neighboring images to make unbiased training. Meanwhile, data enhancement is taken to increase the number of images. The enhancement operations are mirror and rotation. Then, we obtain 1,828 2D MR images with different sizes. Subsequently, we generate the samples for training a single-channel DRCN according to the method in Figure 2.

3D CT and MRI Images: By convention, we name the scanning direction of the CT or MRI brain image in the horizontal section. The size of the 3D CT brain image is $256 \times 256 \times 220$. As such, we can obtain 256 coronal images with a size of $256 \times 220$. Similarly, 768 2D medical images can be obtained with a size of $652 \times 127$ from each volume of MRI1 and MRI2. MRI2 is used for training the multichannel DRCN, and CT and MRI1 are used to test the SVR methods.

B. SINGLE-CHANNEL SUPERRESOLUTION
In generating the samples for training the DRCN, we randomly extract some of rows from the HR image, and the sampled image is upsampled to be the original size. This means that the obtained image is the related interpolated version and has similar tomographic characteristics with the original slices of the volume along the sagittal and coronal directions. Different from most degradation methods, this method actually performs alternate sampling. After obtaining the samples, the DRCN will be trained by taking the interpolated image as input and the HR image as its label. To limit the training cost, the number of DRCN recursive network layers is set to 9, the weight attenuation is set to 0.0001, the momentum parameter is set to 0.9 and all convolutional layers contain 96 filters of $3 \times 3$. We initialize weights of the nonrecursive layers using the method of [29], and all weights and biases of recursive layers are initialized to zero except the weights connected to the next layer. The learning rate is initially set to 0.01. If the verification error does not decrease within 5 cycles, the learning rate is reduced by 10 times. The training stops when the learning rate is less than $10^{-6}$.

1) DIFFERENT SAMPLING RATIOS
Theoretically, the SVR accuracy is inversely proportional to the interval between the scanned slices. Then, the cases of different intervals between the original slices are considered. We degenerate 3D images to the LR volumes according to Figure 2 and then decompose HR and LR volumes into coronal slices. Then, each interpolated coronal slice is taken to perform SR. The results are shown in Table 1, in which the sampling ratio means the percent of original slices extracted in generating the LR volume. In addition, the evaluation value represents the average of that of all slices of the testing
FIGURE 7. The sample data in this article, where a is a two-dimensional MR brain image, b is a three-dimensional CT brain image, c and d are three-dimensional MR brain images.

TABLE 1. Coronal and sagittal slices SR results at different sampling ratios of single-channel DRCN.

| Axis  | Dataset | Error Metric | 1/5 | 1/2 | 2/3 |
|-------|---------|--------------|-----|-----|-----|
|       |         | PSNR         | 40.05 | 38.51 | 37.17 |
|       |         | SSIM         | 0.959 | 0.951 | 0.949 |
|       | CT      | PSNR         | 38.12 | 37.22 | 36.37 |
|       | MRI1    | SSIM         | 0.950 | 0.945 | 0.942 |

FIGURE 8. Weights $W_1$ and $W_3$ with increasing iterations during training.

1) DIFFERENT SAMPLING RATIOS

Table 2 shows the results when MRI2 is used as training data in training the multichannel DRCN; CT and MRI1 are taken for testing. The inputs of the multi-DRCN are 5 consecutive 2D slices centered at the considered slice. Moreover, the final SR result of the considered slice is the weighted average of $N$ outputs of the single-channel DRCNs. In the training of the multichannel DRCN, the parameters of each single DRCN are frozen, i.e., only the weights $W_n$ in the last layer will be optimized. Assume that the input is $X_n$, the output of DRCN is $\hat{Y}_n$, the target image is $Y$ and the loss function is defined as the minimum variance. The learning rate is initialized to 0.0001, and the number of iterations is set to 1,000. The loss function $L$ can be expressed as:

$$L(\theta) = \sum_{i=1}^{D} \frac{1}{2D} \| Y^i - \sum_{n=1}^{N} W_n * i * \hat{Y}_n \|^2$$

2) EVALUATION IN THE OTHER TWO DIRECTIONS

We want to obtain an HR volume, so it is not suitable to evaluate the reconstruction volume only from the coronal plane. To fully evaluate the results of SVR, we decompose the reconstructed volume into a series of 2D slices along the sagittal and horizontal planes. Then, these 2D slices are evaluated. The results on the sagittal planes are shown in Table 1. The error of the slice on the sagittal plane is much worse than that on the coronal plane. This is because SR is conducted for the slices along the coronal plane, and the relationship among the slices is not considered in SR, which then leads to the misalignment in the other two directions. That is, the reconstructed 3D volume only shows well in the reconstructed direction (coronal plane), but the pixel values may be discontinuous in the sagittal plane and the horizontal plane. To overcome this problem, we further propose the multichannel DRCN.

C. MULTICHANNEL SUPERRESOLUTION

The multichannel DRCN is established by combining several single-channel DRCNs in parallel. Then, the inputs of the multichannel DRCN are the $N$ ($N$ is an odd number) consecutive 2D slices centered at the considered slice. Moreover, the final SR result of the considered slice is the weighted average of $N$ outputs of the single-channel DRCNs. In the training of the multichannel DRCN, the parameters of each single DRCN are frozen, i.e., only the weights $W_n$ in the last layer will be optimized. Assume that the input is $X_n$, the output of DRCN is $\hat{Y}_n$, the target image is $Y$ and the loss function is defined as the minimum variance. The learning rate is initialized to 0.0001, and the number of iterations is set to 1,000. The loss function $L$ can be expressed as:

$$L(\theta) = \sum_{i=1}^{D} \frac{1}{2D} \| Y^i - \sum_{n=1}^{N} W_n * i * \hat{Y}_n \|^2$$
2) CASES OF DIFFERENT INPUT NUMBERS

To consider the cases of different input numbers, MRI2 is used as training data, CT and MRI1 are used as test data, and 1/2 row pixels are extracted in generating the degenerate volume. The results are shown in Table 3. It can be seen that the multichannel DRCN improved the reconstruction quality compared with the single-channel DRCN ($N = 1$), and the reconstruction quality also becomes better when $N$ increases.

The improvement is because the more consecutive 2D slices are input to the CNN, the more information can be referenced when reconstructing the HR image. When the input number $N$ is 3, the weight parameters in the output layer are shown in Figure 8 with the training iteration. During the training process, the curves of $W_1$ and $W_3$ are almost the same, which indicates that the images adjacent to the target image have almost approximate contributions to the SR reconstruction of the target image in a continuous 2D image sequence.

3) EVALUATION IN THE OTHER TWO DIRECTIONS

Similar to the single-channel DRCN, we also evaluate the spatial continuity of the reconstructed volume. We decompose the reconstructed volume into a series of slices along the sagittal plane and horizontal plane. Then, these 2D slices are taken to evaluate the reconstruction quality on the basis of the metrics of PSNR and SSIM. The results on the sagittal plane are shown in Table 3 (the results along the horizontal plane are close). By comparison, the reconstruction in the sagittal direction has a significant improvement with the increase in the input slices. It shows that the more slices are input to the multichannel DRCN, the better the spatial continuity of the reconstructed volume will be. To intuitively reflect the spatial continuity of the reconstructed volume, we drew the gray value distribution of some rows on a randomly chosen sagittal slice. Figure 9 shows the comparisons. The original slice is basically continuous, and the reconstructed volume of the single-channel DRCN ($N = 1$) shows an obvious jump phenomenon. When $N = 7$, the jump phenomenon of
In multichannel testing, the number of inputs sampling ratio is 1/2 in generating the degenerate volumes. VDSR and EDSR methods, considering that they are classic deep CNNs. However, we still conduct comparisons with samples are needed in DRCN training compared with other methods can be taken as single-channel CNNs. DRCN is an effective SR method since it deepens the network without greatly increasing the number of parameters. Then, fewer parameters are used in DRCN training compared with other deep CNNs. However, we still conduct comparisons with some classic SR methods, including the bicubic, SRCNN, VDSR and EDSR methods, considering that they are classic. CT and MRI are still used as the testing set, and the sampling ratio is 1/2 in generating the degenerate volumes. In multichannel testing, the number of inputs \( N \) is 5, and the parameters in generating the degenerate volume are the same as those in the single-channel case. The results are shown in Table 4. They show that in the same situation, the SR reconstruction of the multichannel DRCN is better than other methods.

The multichannel DRCN intends to overcome the defect that the SR in one direction may lead to the misalignment in the other directions. As such, some related slices centered at the considered slice are referenced in performing SR. According to the experiments, the SVR can be effectively performed in the metrics of continuity or smoothness. Additionally, it should be mentioned that the 2D CNN-based SR can be conducted along three directions, and the final result should be their average, which can ensure continuity in any direction.

### D. COMPARISON WITH OTHER ALGORITHMS

As mentioned above, our contributions lie in that we propose a specialized way to generate the samples and establish a multichannel SVR framework. Theoretically, all kinds of SR methods can be taken as single-channel CNNs. DRCN is an effective SR method since it deepens the network without greatly increasing the number of parameters. Then, fewer parameters are used in DRCN training compared with other deep CNNs. However, we still conduct comparisons with some classic SR methods, including the bicubic, SRCNN, VDSR and EDSR methods, considering that they are classic. CT and MRI are still used as the testing set, and the sampling ratio is 1/2 in generating the degenerate volumes. In multichannel testing, the number of inputs \( N \) is 5, and the parameters in generating the degenerate volume are the same as those in the single-channel case. The results are shown in Table 4. They show that in the same situation, the SR reconstruction of the multichannel DRCN is better than other methods.

The multichannel DRCN intends to overcome the defect that the SR in one direction may lead to the misalignment in the other directions. As such, some related slices centered at the considered slice are referenced in performing SR. According to the experiments, the SVR can be effectively performed in the metrics of continuity or smoothness. Additionally, it should be mentioned that the 2D CNN-based SR can be conducted along three directions, and the final result should be their average, which can ensure continuity in any direction.

### V. CONCLUSION

In applications of CT and UCT, it is common that the 3D model is reconstructed on the basis of scanned slices. Generally, the distance between the original slices is more than the interval between pixels on the scanned slice. As such, the reconstructed volume on the basis of the interpolation algorithms is not actually isotropic. To add more details to the interpolated slices, the SR has been applied. The traditional methods are mainly based on the 2D CNN considering that the 3D CNN is difficult to train. However, these traditional 2D CNN-based SVRs cannot obtain the balance between adding details and ensuring continuity among the neighboring slices at the same time. As such, we perform the 2D SR for non-scanned slices on the basis of the specialized samples that are obtained by simulating the process of slice scanning. This method improves the accuracy of the reconstruction. Additionally, the multichannel DRCN considers the information among the neighbors of the considered slice, which further improves the robustness and accuracy. In addition, there are still some issues to be considered in the future. The first is that more CNN structures should be adapted to perform more comparisons. Moreover, the method for accelerating the training process should also be detected. The second is how to consider the continuity among the neighboring slices along three directions at the same time, which may further improve the accuracy and robustness. In addition, how to evaluate the SVR accuracy in the volume metric should be studied in the future.

### REFERENCES

[1] R. Keys, “Cubic convolution interpolation for digital image processing,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASP-29, no. 6, pp. 1153–1160, Dec. 1981.

[2] Y. Jia, A. Gholipour, Z. He, and S. K. Warfield, “A new sparse representation framework for reconstruction of an isotropic high spatial resolution MR volume from orthogonal anisotropic resolution scans,” *IEEE Trans. Med. Imag.*, vol. 36, no. 5, pp. 1182–1193, May 2017.

[3] J. Du, L. Wang, A. Gholipour, Z. He, and Y. Jia, “Accelerated super-resolution MR image reconstruction via a 3D densely connected deep convolutional neural network,” in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 349–355.

[4] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li, “Brain MRI super resolution using 3D deep densely connected neural networks,” in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 739–742.

[5] J. Kim, J. K. Lee, and K. M. Lee, “Deeply-recursive convolutional network for image super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1637–1645.

[6] R. Tsai and T. Huang, “Multiframe image restoration and registration,” *Adv. Comput. Vis. Image Process.*, vol. 12, pp. 317–339, Jan. 1984.

[7] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *Proc. 13th Eur. Conf. Comput. Vis.* (ECCV), Zürich, Switzerland, vol. 8692, Sep. 2014, pp. 184–199.

[8] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.

[9] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.

[10] E. Plenge, D. H. J. Poot, W. J. Niessen, and E. Meijering, “Super-resolution reconstruction using cross-scale self-similarity in multi-slice MRI,” in *Proc. Int. Conf. Med. Image Comput. Assist. Intervent.*, in Lecture Notes in Computer Science, vol. 9151, 2015, pp. 123–130.

[11] C.-H. Pham, A. Ducournau, R. Fablet, and F. Rousseau, “Brain MRI super-resolution using deep 3D convolutional networks,” in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Melbourne, VIC, Australia, Apr. 2017, pp. 197–200.

[12] O. Oktay, W. Bai, M. Lee, R. Guerrero, K. Kamnitsas, J. Caballero, A. de Marvao, S. Cook, D. O’Regan, and D. Rueckert, “Multi-input cardiac image super-resolution using convolutional neural networks,” in *Medical Image Computing and Computer-Assisted Intervention*. Athens, Greece: Springer, 2016, pp. 246–254.

**TABLE 4. The SVR results of different methods.**

| Dataset | Error Metric | Bicubic | SRCNN | VDSR | EDSR | Single-Channel DRCN | Multichannel DRCN |
|---------|-------------|--------|------|------|------|---------------------|-------------------|
| CT      | PSNR        | 35.44  | 37.52| 38.42| 38.71| 38.51               | 38.92             |
|         | SSIM        | 0.933  | 0.948| 0.950| 0.956| 0.951               | 0.956             |
| MRI     | PSNR        | 34.05  | 37.03| 37.53| 38.61| 37.22               | 37.73             |
|         | SSIM        | 0.927  | 0.939| 0.944| 0.947| 0.945               | 0.950             |
ZHANG SIYUAN was born in Heilongjiang, China, in 1995. He received the B.S. degree in biomedical engineering from the Huazhong University of Science and Technology, in 2018, where he is currently pursuing the master’s degree in biomedical engineering. His research interests include image super-resolution reconstruction and 3D image display.

DONG JINGXIAN was born in Shandong, China, in 1997. She is currently pursuing the bachelor’s degree in biomedical engineering with the Huazhong University of Science and Technology. Her research interests include image super-resolution reconstruction and image processing.

JIANG CAIWEN was born in Hubei, China, in 1997. He received the B.S. degree in bio-medical engineering from the Huazhong University of Science and Technology, in 2019. He is currently pursuing the master’s degree in biomedical engineering from ShanghaiTech University. His research interests include image super-resolution reconstruction and image recognition.

HOU WENGUANG was born in Shaxi, China, in 1974. He received the B.S. and M.S. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, in 1997, 2003, and 2006. From 2009 to 2010, he was a Research Assistant with The Hong Kong PolyTechnic University. Since 2013, he has been an Associate Professor with the Bio-Medical Engineering Department, Huazhong University of Science and Technology. He has published more than 35 articles and ten inventors. His research interests include image processing, 3D reconstruction, and rendering.

DENG XIANBO was born in Hubei, China, in 1974. He received the B.S. and Ph.D. degrees from the Tongji Medical College, Huazhong University of Science and Technology, in 1997 and 2005, respectively. Since 2005, he has been a Radiologist with the Union hospital, Huazhong University of Science and Technology. He majors the medical image processing and analysis. He has been the one of the main investigators in many research projections and has published more than 20 articles.