Noise-Robust MRI Upsampling Using Adaptive Local Steering Kernel

JING HU[1], (Member, IEEE), XINYAN LI[1], XIAOLONG WANG[2], YAN LI[3], AND XI WU[1], (Member, IEEE)

[1]Department of Computer Science, Chengdu University of Information Technology, Chengdu 610225, China
[2]Sichuan Diwu Surveying and Mapping Company Ltd., Chengdu 610036, China
[3]Chongqing Cigarette Industry Company Ltd., Chongqing 400060, China

Corresponding author: Jing Hu (jing_hu09@163.com)

ABSTRACT Upsampling and denoising of magnetic resonance images are conventionally performed separately, which would introduce unwanted artifacts such as blurring. To address this problem, we propose an innovative adaptive interpolation framework to achieve simultaneous image upsampling, denoising, and detail enhancement. First, local steering kernel (LSK) function is leveraged to adapt the interpolation weights according to geometric structures in the magnetic resonance (MR) images. An adaptive sharpening of the LSK weight matrix and a Rician bias correction are then adopted to remove Rician noise and enhance fine details. In this regard, the adaptive LSK extends the zero-order point estimation framework to higher orders of regression, and therefore facilitates edge preservation and detail reconstruction. The post-processing Rician correction avoids the bias caused by the asymmetry of Rician noise distributions. Experimental results using both real and synthetic clinical MR cranial images (with and without noise) demonstrated that our algorithm produced better reconstruction results than several traditional interpolation-based upsampling methods, including nearest neighbor (NN), non-local means (NLM), self-learning super resolution (SLSR), Gaussian process regression (GPR), and even comparable to four deep-learning-based methods but with less data requirements and computational complexity. The proposed technique resulted in PSNR and SSIM values were ~3%–16% higher than any of the other traditional algorithms tested, and our method recovered more clear textures from noisy images compared with deep-learning-based methods. As such, the presented technique is a viable new approach to MR upsampling, particularly for noisy images.

INDEX TERMS Adaptive sharpening, image upsampling, local steering kernel, magnetic resonance imaging, noise robust.

I. INTRODUCTION

One of the primary objectives of medical imaging is the automated extraction and modeling of 3D anatomical regions of interest (ROIs) within the human body [1]. A variety of imaging modalities have been developed to achieve this goal, including computed tomography and magnetic resonance imaging (MRI). MRI is capable of achieving excellent soft-tissue contrast and is thereby effective for viewing differences between normal and diseased tissue. Higher res-

olution images provide a more comprehensive understanding of anatomy at the cost of reduced signal-to-noise ratio (SNR) and increased imaging time [2]. However, clinically, an MRI scan is often fast because long scan time increases costs, leads to patient discomfort, and induces motion artifacts in images [2]. As a result, the resolution of a clinical magnetic resonance (MR) image is limited.

Conventional interpolation methods, such as spline interpolation, have been used to increase MR image resolution. More specifically, a voxel in a high-resolution (HR) image is estimated using a weighted averaging of several sampled voxels. The coefficient for each sampled voxel reflects the

---

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

similarity between sampled and target voxels. Commonly, sampled voxels are selected within a spatial neighborhood of the target voxel, and thus coefficients are typically fixed as a function of the spatial distance between target and sampled voxels. Due to the simplicity, images reconstructed using these traditional interpolation methods often suffer from blurred edges and stair-casing artifacts.

To solve these two problems, several researchers have proposed methods that performed adaptive interpolation [3]–[10]. In the pioneering work of Manjón *et al.* [3], sampled voxels were selected from a large cube centered on the target voxel. Interpolation coefficients were then determined by intensity distances between two 3D image patches located around the target and sampled voxels. Rather than using fixed coefficients as conventional interpolation methods do, Manjón's method used adaptive coefficients that were more flexible to image structures and thereby avoid blurring edges in the interpolated results. Several variant techniques were subsequently proposed, which all focused on coefficient refinement. For example, Plenge *et al.* [7] took advantage of in-plane patches having a higher spatial resolution than those resulting from slice-selection. They calculated coefficients by measuring the 3D patch similarity in the in-plane directions, so as to estimate voxels in slice-selection. Within a similar notion, high-resolution images acquired with other MRI modalities have also been leveraged in several approaches to calculate interpolation weights so as to better reflect image structure [8], [9].

Besides image interpolation, deep learning has witnessed a tremendous amount of attention over the last few years in the field of image processing, because of the high representational capacity. With extensive parameters and a good learning process, deep-learning-based models have the ability of fitting on a large number of training data and exploiting the underlying structures of natural images. Dong *et al.* [22] introduced CNN into the SR task and proposed the SRCNN model that was composed of a three-layer network to learn the mapping from LR images to HR images. This model achieved much better performance compared with the traditional algorithms. Kim *et al.* [24] proposed the VDSR model that used a very deep network with 20 layers which produced improved performance compared with SRCNN. A main contribution of this method was to employ residual learning which encouraged a fast convergence rate in the training process. Tai *et al.* introduced recursive blocks in DRRN [25] and memory block in Memnet [31] for deeper networks. Haris *et al.* [32] developed a novel architecture which was named as DBPN. This model exploited iterative up and down sampling layers, providing an error feedback mechanism for project errors at each stage. DBPN improved super-resolution performance, yielding superior results and in particular establishing new state-of-the-art performance for large scale factors such as ×8 on multiple datasets. Zhang *et al.* [33] proposed the channel attention mechanism to build a deep model called RCAN to further improve the performance of SR. An image super-resolution feedback network (SRFBN) [34] was pro-

posed to redefine the low-level representation with high-level information, which was realized by the hidden state in the constrained recurrent neural network (RNN).

Although their performance is state-of-the-art, these deep-learning-based methods also have some disadvantages. Firstly, the deep-learning-based methods often require a large number of training data sets, while a large MRI dataset is hard to obtain. Second, the generalization capabilities of deep-learning-based methods are also very limited, and they require a close distribution of training data and testing data. On the contrary, the conventional image processing methods often have less data requirements and wide range of applications. Last but not least, there is a need for pre-denoising before upsampling noisy images, so the edges of final results are blurry or smooth. In this regard, this paper focuses on the conventional method in MRI image upsampling.

For interpolation methods, they cannot consistently reconstruct high-frequency details from low-resolution (LR) images. This is because interpolation-based algorithms all belong to the framework of zero-order regression estimation [10]. To this end, a regression-inspired upsampling method using second-order polynomials was proposed in our previous study [10]. However, both adaptive interpolation methods and our previous work fail to account for the noise present in MR images, mainly produced by echo planar imaging [11]. Image denoising is typically used in these methods to remove noise beforehand. It has been demonstrated that a simultaneous interpolation and denoising eventually outperforms an asynchronous framework [10]. Motivated by this, we propose to develop a new unified MR image upsampling framework that simultaneously performs HR image reconstruction, denoising, and high-frequency detail enhancement, and thereby leads to more accurate treatment contours. Unlike the abovementioned adaptive interpolation methods using intensity distance to calculate interpolation coefficients, the proposed method refined the adaptive interpolation coefficient by comparing patches' structure difference. Moreover, although the proposed method was also a zero-order regression-based method, the incorporated adaptive weight sharpening strategy enabled detail preservation.

The main contributions of the proposed method are the following four aspects:

1) With the assumption that input image is clean, most image upsampling methods have to include an additional step to remove image noise. But this denoising step blurs fine image details as well. To this end, the proposed method presents a unified framework that simultaneously carries out upsampling, denoising, and detail sharpening.

2) Compared with deep-learning-based methods, the proposed method recovers more clear textures from noisy images with less data requirements and computational complexity.

3) Local steering kernel (LSK) is adopted to help interpolation weights be sensitive to local structure of images.

4) An adaptive sharpening technique is proposed to extend the traditional weighted-averaging interpolation framework (zero-order estimation) to a two-order estimation framework.

Besides, using adaptive weight sharpening enforces the distribution of interpolation weights to be consistent with local image structure, which helps denoise background and sharpen fine details.

5) Rician noise correction is leveraged to correct the bias caused by the asymmetry of Rician distributions.

Our method is described in the following three subsections. Specifically, details of the LSK and adaptive sharpening methods are elaborated in section II-A and II-B. In section II-C, a Rician noise correction step is introduced.

The remainder of this paper is structured as follows. Section II details the proposed algorithm. Experimental results for several MR phantoms and real data are provided in Section III, together with visual and quantitative comparisons with other methods. Finally, the conclusions are described in Section IV.

II. METHODS

We have noticed that both image denoising and interpolation could be realized by the weighted averaging framework:

$$\tilde{y}_i = \frac{1}{Z} \sum_{k \in \Omega(i)} w(y_k, y_i) y_k \qquad (1)$$

where y_i is the intensity value of voxel i that needs to be estimated, Ω is a restricted search volume surrounding the voxel i, and y_k is the sampled voxel within the search zone, w measures the similarity between voxels k and i respectively. Z is a normalization constant. Intuitively, has a determining influence on this intensity estimation.

As for denoising, averaging voxels with a similar structure can reduce noise without compromising details [12]–[14]. More specifically, we have found in our pilot study [14] that when denoising voxels in flat regions, the weights used for averaging should be nearly isotropic since using sampled voxels in all directions facilitates noise removal. Meanwhile, when denoising voxels in textured regions, weights used for averaging should be anisotropic so that details are preserved by using sampled voxels along a particular direction. In other word, weights should be adapted to image structures so that noise removal and detail preservation are balanced in the context of denoising.

On the other hand, coefficient computation is also a fundamental issue for interpolation-based MR image upsampling. Previous studies [8], [10] have advocated that using high-order patch statistics in deriving interpolation weights can be a big help for reconstructing high frequency details. In other words, weights should be adapted to high-order patch features so that details are preserved in the context of interpolation.

Based on the above analysis, we propose a modified weighting strategy to unify MR image upsampling and denoising under the framework of weighted averaging. More specifically, the coefficients should be sensitive to image structure and reflect the similarity of high-order patch features. To this end, a specific kernel based on image structure tensor was adopted in the proposed method to compare high-order patch structure. An adaptive sharpening scheme was

also used to ensure the coefficients responded appropriately to image structures. Finally, a Rician correction step, which was conventionally used in several MR image denoising methods [11], was employed. Fig. 1 displays a flow chart for the proposed technique.

A. INTERPOLATION WEIGHTING USING LOCAL STEERING KERNEL (LSK)

LSK [26] is an image feature that describes the local structure information of the image and has good robustness to noise. The basic idea of LSK is to measure both geometric and photometric local similarity of a pixel to its neighbors. It analyzes the pixel value differences based on estimated gradients in a local window. It can robustly estimate the intrinsic image structure and address pixel-level image noise and ambiguity [27]. For example, patches containing a flat region, textural clutter, or structural part can behave very differently for LSK-based descriptors [26]. In addition, extensive experiments have shown that feature descriptors using LSKs are robust to brightness variation and noise interference through estimation of the local intrinsic structure. The effectiveness of the LSK has been verified in various applications, e.g., image reconstruction [28], image deblurring [15], object detection [29] and single image super-resolution [30].

By analyzing the intensity differences voxel-wisely on estimated gradients, LSK captures structure information and uses it to determine the shape and size of a canonical Gaussian kernel. Mathematically, this kernel is represented as:

$$K_h(x - x_i) = \sqrt{\frac{\det(C_i)}{2\pi h^2}} \exp\left\{-\frac{(x - x_i)^T C_i (x - x_i)}{2h^2}\right\}, \qquad (2)$$

where C_i denotes a 2×2 symmetric covariance matrix centered at voxel x_i, which is estimated from a collection of first derivatives along the vertical and horizontal directions. The term $x - x_i$ represents the spatial distance between sampled (x_i) and target voxels (x); h is a smoothing parameter that controls the decay of the Gaussian function. It is important to note that the matrix C_i is an image structure tensor. The specific formation of this matrix is described as:

$$C_i = \sum_{m_k \in N(x_i)} \left[\begin{pmatrix} z_1(m_k) z_1(m_k) & z_1(m_k) z_2(m_k) \\ z_1(m_k) z_2(m_k) & z_2(m_k) z_2(m_k) \end{pmatrix} \right], \qquad (3)$$

where $N(x_i)$ represents a search window centered at x_i and m_k is the voxel located inside $N(x_i)$. The terms $z_1(*)$ and $z_2(*)$ denote first derivatives along the vertical and horizontal directions, respectively.

LSKs are able to reflect local structure because they are computed using local statistics. Fig. 2 illustrates LSK shape for neighboring sampled voxels for a variety of image structures (i.e., texture, flat, strong and weak edge) in a T1 image with different noise levels (0%, 5%, and 9% of maximum intensity). The yellow dot inside each red box denotes the target voxel and the rest voxels inside the red box are sampled voxels. The blue box depicts the LSK values of the sampled
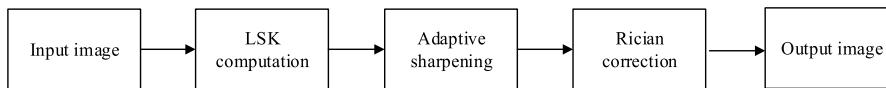
FIGURE 1. A flowchart for the proposed method.

(a) Noise-free image   (b) noisy image (5% Rician noise)
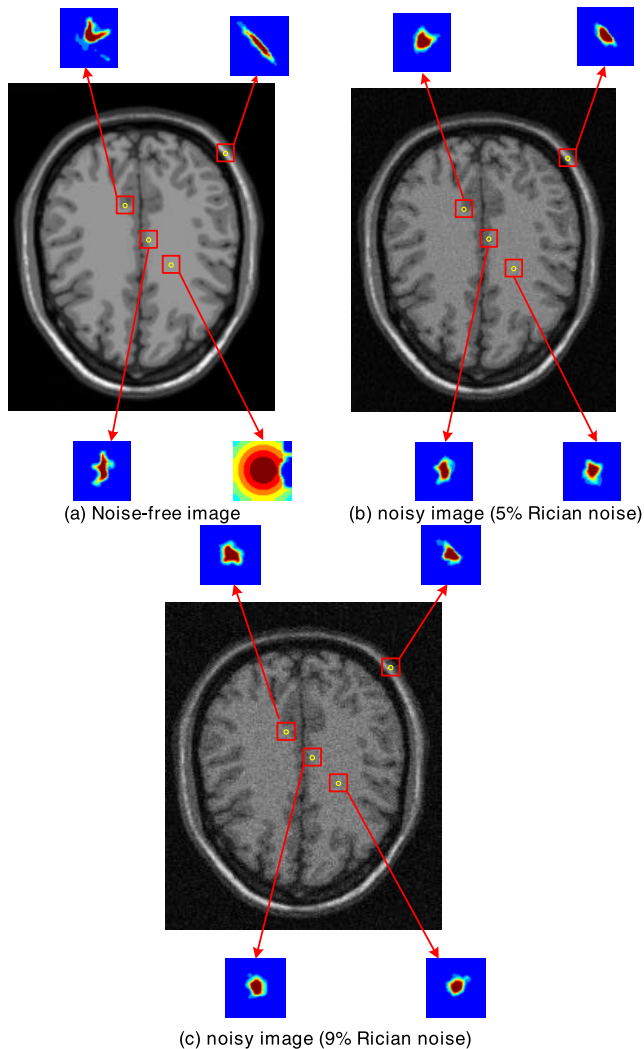
(c) noisy image (9% Rician noise)

FIGURE 2. The LSK weights (Wi) in various noisy conditions. The yellow dot denotes the target voxel. Sampled voxels are contained inside the red box, and their corresponding LSK are depicted in the blue box.

voxels. In the blue box, the dark red color indicates that sampled voxels in this location have a comparatively large weight (near 1) in the weighted averaging, while the blue color represents corresponding sampled voxel with a smaller weight (near 0). As is clearly shown in Fig. 2, LSKs reliably capture local data structures for a clean image even in regions of complex texture.

However, it is also important to note that, in the noisy case, the shape and orientation of LSKs are quite deviated from the noiseless case. This is possible due to the degradation of image structures in the presence of noise. Therefore, for noisy

images, LSKs must be refined by certain strategies to reflect the latent image structures that have been corrupted by noise.

B. ADAPTIVE WEIGHT SHARPENING

Although LSK is not capable of reflecting latent image structures for noisy images, it, as shown in Fig.2, still appears to be anisotropic in textured regions and isotropic in flat regions among these images. But compared to the distributions of LSK for a clean image, we find that such distribution tendency is actually weakened in noisy images. Motivated by these two findings, we propose reinforcing the tendency of LSK distribution on noisy images by adaptive sharpening in order to mimic the LSK distribution on clean image. In other words, for LSKs in flat regions, the spread of LSKs becomes wider and essentially isotropic; for LSKs in textured regions, the spread of LSKs shrinks to the texture outlines.

In order to achieve this adaptation, adaptive sharpening was applied to LSK values. More specifically, after collecting LSKs into a weighting matrix W_i, an adaptive sharpening was carried out on W_i by adding or subtracting a fraction of the high-pass filtered weight matrix back to the same matrix again:

$$\hat{W}_i = W_i + \lambda \left(W_i \otimes H \right). \tag{4}$$

Here the matrix H represents a high-pass filter and λ is a scalar used to modulate sharpness. As indicated in [17], the sign λ of determines whether the input (W_i) is sharpened (positive) or smoothed (negative). Since a smoothed W_i is achieved by taking out the high frequency information of W_i, the spread of its corresponding LSKs will be more isotropic than the original one; similarly, a sharpened W_i implies that the spread of its corresponding LSKs will be biased towards a particular direction. Therefore, according to the previous analysis, for estimating voxels in flat regions, W_i ought to be smoothened; while for estimating voxels in textured regions, W_i ought to be sharpened. In this way, the adaptively sharpened weight matrix adapts better to the latent image structures than the original one does.

At first, before performing adaptive sharpening on weight matrix, each image voxel ought to be classified according to the type of its neighboring patch structure (i.e., flat, weak, and strong texture). In this paper, image structure tensor (C_i in Eq. (3)) was used to discriminate textured voxels from flat voxels. The relative discrepancies between two eigenvalues of C_i reflect how strongly the distribution of gradients in an image patch is biased towards a particular direction [13]. In other words, smaller eigenvalue differences exist for voxels in smooth regions, while larger differences exist for voxels in

textured regions. Voxel classification can thereby be achieved by examining eigenvalue differences for each voxel. More specifically, if s_1 and $s_2 (s_1 \geq s_2)$ denote two eigenvalues of C_i, we use $T = s_1 - s_2$ to reflect the degree of texture for every voxel. Voxels with different texture magnitudes can be classified by analyzing the cumulative histogram of in T an MR image slice:

$$(i, j) \in \begin{cases} c_1 & T(i,j) > t_1 \\ c_2 & t_2 < T(i,j) \leqslant t_1 \\ c_3 & T(i,j) \leqslant t_2 \end{cases} \tag{5}$$

Here $T(i, j)$ is the T value of the voxel at location (i, j); t_1 and t_2 are the bin values corresponding to 80% and 50% in the cumulative T histogram, respectively. The terms c_1, c_2, and c_3 represent strongly-textured, weakly-textured, and smooth regions, respectively. Fig. 3 presents the results of voxel classification for a T1 image using Eq. (5). It is evident that voxel classification is consistently the same in all images, regardless of whether they are heavily contaminated. This result indicates that image structure tensors provide a reliable classification system in the presence of noise.
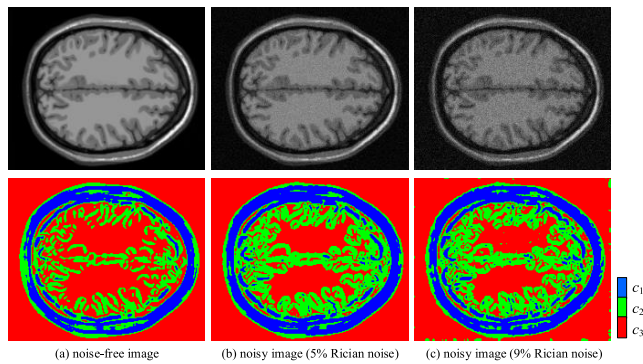

FIGURE 3. Voxel classification results for various noisy images. The first row: a typical MR slice. The second row: corresponding voxel classification results. c1, c2, and c3 represent strongly-textured, weakly-textured, and smooth regions, respectively.

Second, λ in Eq. (4) should be carefully designed in our method to adaptively enhance weight matrix. Based on the previous analysis, this parameter is empirically selected in accordance with region type. Besides this, the problem of voxel misclassification, which is inevitable in highly corrupted images (for example, in Fig. 3(c) there are green dots present in the red background, which indicates that the noisy background is misclassified into the texture), is also considered in the process of λ selection. More specifically, we used a positive λ in classes c_1 and c_2 to preserve and sharpen details and a negative λ in class c_3 to remove background noise. Additionally, a sigmoid function was used to refine λ in c_2 to remedy misclassification. With Q_i $(i = 1, 2, 3)$ denoting the median value of T in each class type, the specific form of

λ is derived as follows:

$$\lambda(n, m)$$
$$= \begin{cases} 0.2\, l_1 & if\, (n, m) \in c_1 \\ 0.15\, l_2 \cdot f\left(T^*\right) & if\, (n, m) \in c_2, l_i = \sqrt{\dfrac{Q_i}{10}}, i = 1, 2, 3 \\ -0.15\, l_3 & if\, (n, m) \in c_3. \end{cases} \tag{6}$$

where $f(*)$ is a typical sigmoid function that returns a small value when the input is low:

$$f(x) = \frac{1}{1 + \exp(-10x + 5)}, \quad x \in [0, 1] \tag{7}$$

In Eq. (6), T^* is a normalized T for voxels classified as c_2. We observed that the misclassified voxels in c_2 all have small T values near t_2. As such, we used a sigmoid function to assign these voxels to λ values near 0. This was done to prevent improper sharpening or smoothing caused by misclassification.

An illustration of the adaptive sharpened LSK (Eqs. (4)–(7)) for the same regions in Fig. 2 is presented in Fig. 4. These refined LSKs coincide better with local image structures and are more adaptable than those of Fig. 2. For example, in the flat region, the LSKs are wide and essentially isotropic. In the edge regions, the LSK shrinks to accurately depict the edge outline. Using the refined LSKs facilitates noise removal and edge preservation, since samples with similar geometric structure are more likely to be leveraged in a weighted averaging framework.

C. RICIAN CORRECTION
It is worth noting that the weighted averaging framework is designed for additive Gaussian noise. Nevertheless, MRI image is commonly believed to be governed by noise with Rician distribution [18]. Directly applying this framework to MRI datasets would cause biasing because of the asymmetry of Rician distributions [11]–[18]. To this end, a post-processing bias correction introduced in [11]–[18] was carried out in the proposed method. The final voxel intensity estimation was calculated as:

$$z(x_i) = \sqrt{\max\left(\hat{W}_i y / sum\left(\hat{W}_i\right) - 2\sigma^2, 0\right)}, \quad y \in N(x_i) \tag{8}$$

where $z(x_i)$ represents the estimated intensity value of a voxel x_i in the SR image, $y \in N(x_i)$ denotes x_i's neighboring sampled voxels, and $\hat{W}_i$ is a matrix comprised of the sampled voxels' corresponding adaptively sharpened LSK values, $sum(*)$ is an operator denoting the summation of matrix elements, and σ is the standard deviation of MR image noise.

III. EXPERIMENTS AND DISCUSSIONS
The proposed algorithm was compared with other state-of-the-art upsampling methods using both synthetic and clinical MRI datasets. Two T1 MR phantoms from Brainweb (simulated images, http://brainweb.bic.mni.mcgill.ca/brainweb/
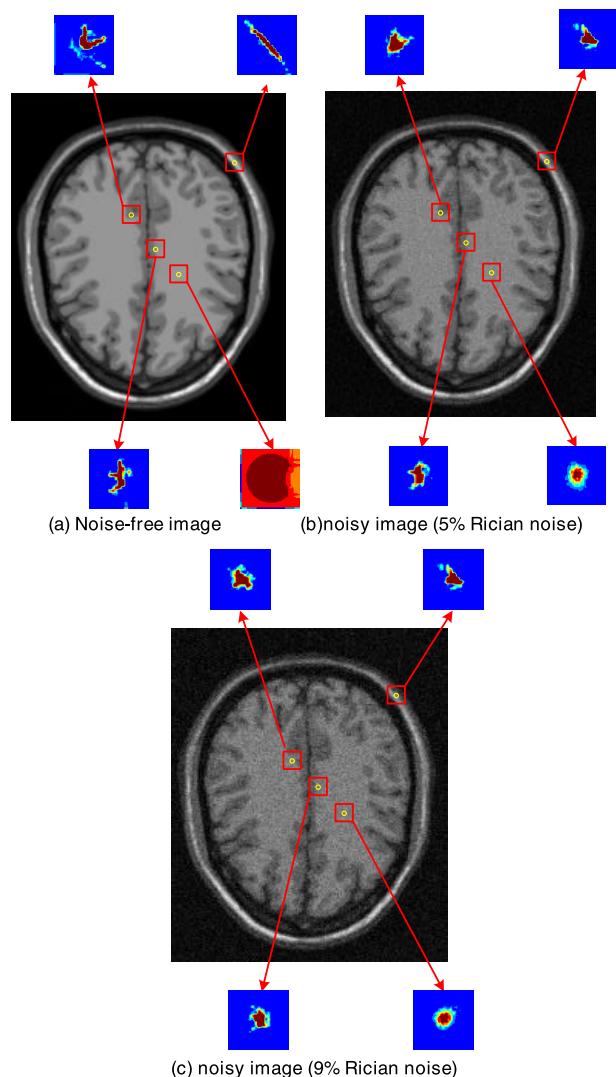
FIGURE 4. Adaptive sharpened LSK weights (Wi) in various noisy conditions. The yellow dot denotes the target voxel. Sampled voxels are contained inside the red box, and their corresponding LSK are depicted in the blue box.

selection _normal.html) and the Human Connectome Project (HCP, real images, https://db.humanconnectome.org/) were utilized to evaluate reconstruction performance for the proposed technique in both cases with and without noise. Moreover, a real clinical MRI dataset was also used to evaluate the robustness of the proposed method for a realistic scenario, where subjects gave informed consent to participate and recordings were used for study purposes. Note that the proposed method made no prior assumption of the image, it could thereby be applied to other anatomical regions such as knees, liver and heart. The LR volumes used in this study were generated by blurring and downsampling. More specifically, the blurred volumes were generated by convolving an HR volume with a 3D Gaussian kernel of standard deviation 0.8 (in voxel space) along each dimension. The blurred volume

was subsequently downsampled to a lower voxel resolution at $2 \times 2 \times 2$ mm^3 or $3 \times 3 \times 3$ mm^3.

A. ALL COMPETING MODELS

In this study, nearest neighbor (NN) interpolation, non-local means (NLM) based upsampling [3], self-learning super resolution (SLSR) [10], Gaussian process regression (GPR) based upsampling [19], deep-learning-based methods (SCN [22], VDSR [23], DRRN [24], and SRFBN [34]) were employed for performance comparisons. GPR method, NLM method and SLSR method are both interpolation-based methods. The former two are zero-order based and the latter is second-order based. As a pioneer CNN model for SR, super-resolution convolutional neural network (SCN) predicts the nonlinear LR-HR mapping via a fully convolutional network, and significantly outperforms classical non-DL methods. Kim [24] *et al.* proposed a super-resolution method using very deep networks (VDSR). This work use residual-learning and extremely high learning rates to optimize a very deep network that fast maximizing the convergence speed. In DRRN [25], an enhanced residual unit structure is recursively learned in a recursive block, and several recursive blocks are stacked to learn the residual image between the HR and LR images, and then added to the input LR image from a global identity branch to estimate the HR image. SRFBN [34] introduced a feedback module that shared weights by using recurrent network. Besides, a curricular approach was introduced to handle different tasks. Variable parameters used in the latter nine methods were selected per author suggestions. The only parameters used in the proposed method are the size of 3D patches, in which the sampled voxels are determined, as well as the bandwidth in Eq. (2) that is used to calculate the LSK: the 3D patch size was fixed to be $3 \times 3 \times 3$ and the bandwidth was selected according to noise level.

In general, NLM, GPR, and the proposed method belong to a zero-order estimation framework. SLSR belongs to a second-order estimation framework, which has been shown to be beneficial for detail preservation [10]. In addition, GPR and the proposed method require no additional denoising, while NN, NLM, SLSR, SCN, VDSR, DRRN, and SRFBN require a pre-denoising step. As a result, they were used to perform on the denoised LR volumes. This was achieved by filtering the original noisy LR volumes with the APW-NLM method [14], which employs an adaptive bandwidth and patch size.

The performance of the proposed method was evaluated by comparing reconstructed HR volumes with the original HR volumes, as well as those reconstructed using other methods. Peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [23] were used to quantify agreement between images. A high PSNR score indicated that a recovered MR image contained little distortion and low noise. An SSIM value near 1 implied reconstructed MR images were close to the ground truth.

The utility and novelty of the proposed algorithm lie in its ability to process corrupted MR images without additional

TABLE 1. Investigations of Racian correction and adaptive sharpening. These upsampling results are from images with noise level at 3% of the maximum intensity.

Rician correction	✘	✔	✘	✔
Adaptive sharpening	✘	✘	✔	✔
PSNR(2x)	26.63	26.62	28.68	28.69

TABLE 2. A comparison of PSNR values [dB] and SSIM values (PSNR/SSIM) for noisy MR image upsampling.

Noise level	3%	5%	7%	9%
NN*	25.02/0.71	23.82/0.67	22.93/0.65	21.70/0.63
The proposed	**28.94**/0.78	**26.50**/0.75	**25.02**/0.72	23.24/0.70
NLM	25.88/0.74	24.48/0.71	23.55/0.68	22.19/0.66
GPR	26.21/0.74	24.65/0.70	24.00/0.68	22.14/0.65
SLSR	27.06/0.76	25.30/0.73	24.13/0.70	22.70/0.68
SCN	28.02/0.74	25.48/0.70	24.43/0.68	22.07/0.65
VDSR	26.78/**0.90**	25.49/**0.88**	24.54/**0.86**	23.27/**0.84**
DRRN	26.80/**0.90**	25.50/**0.88**	24.54/**0.86**	23.28/**0.84**
SRFBN	26.84/**0.90**	25.52/**0.88**	24.61/**0.86**	**23.33**/**0.84**

*Note that all methods except GPR and the proposed are performed on the denoised LR images.

denoising. In section III-B, an ablation study was carried out to demonstrate the effectiveness of the combined adaptive sharpening and Rician correction. In section III-C, an experiment was conducted with corrupted MR images to test how well the proposed approach performed given different noise levels. In addition, an experiment on noise-free MR images was also carried out to demonstrate the detail preservation capabilities of the proposed method (i.e. a sharp contour is obtained using the proposed method), and it is described in the following subsection. Last, an additional experiment on real clinical datasets is described section III-E.

B. ABLATION STUDY
In order to investigate the effectiveness of the combined adaptive sharpening and Rician correction, an ablation study is carried out where Rician correction and adaptive sharpening are respectively removed.

Images with noise level at 3% of the maximum intensity is used and the upscaling factor is set to 2. As shown in Table 1, the adaptive sharpening technique is a key factor in the performance improvement. Moreover, it is important to note that post Rician correction (the second column in Table 1) has a lower PSNR value than LSK-based upscaling framework (the first column in Table 1). This is possibly due to the reason that LSK is not capable of reflecting latent image structures for noisy images, and therefore leads to an inaccurate weight matrix. In this regard, the subsequent Rician correction would worsen the weight bias.

C. MR IMAGES WITH DIFFERING NOISE LEVELS
Four different noisy T1 volumes (voxel resolution 1 mm³, $180 \times 216 \times 180$ voxels) from a Brainweb phantom were used in this experiment. They were corrupted by Rician noise

at levels of 3%, 5%, 7%, and 9% of the maximum intensity respectively. The simulated LR data (voxel resolution 2 mm³) in these four noisy volumes were respectively upsampled to 1 mm isotropic resolution using the proposed technique and various comparison methods. Fig. 5 compares the reconstruction results obtained using these different methods for each noise level. Local regions of interest are also presented to provide a better comparison. Table 2 summarizes the quantitative comparisons.
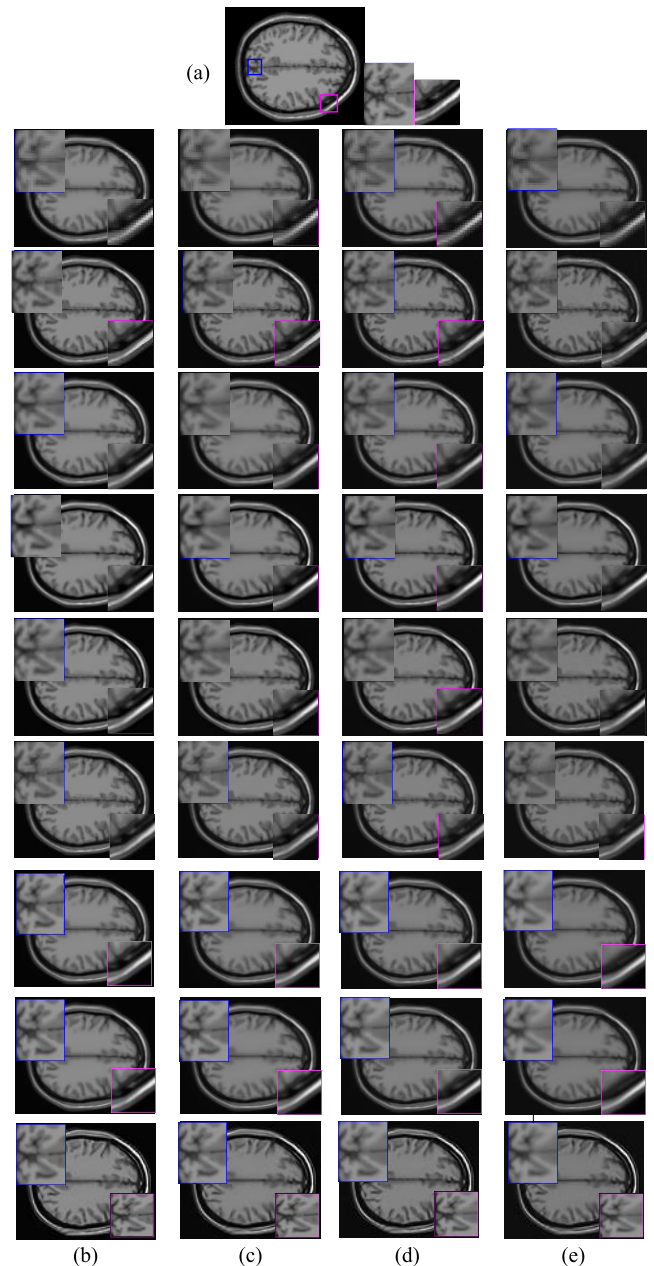

FIGURE 5. Comparison of upsampling results for images with different noise levels: (a) the ground truth, (b) noise level at 3% of the maximum intensity, (c) noise level at 5% of the maximum intensity, (d) noise level at 7% of the maximum intensity, (e) noise level at 9% of the maximum intensity. From top to bottom: results produced by the NN method, the proposed method, NLM, SLSR, GPR, SCN, VDSR, DRRN, and SRFBN.

We observe from Fig. 5 that the proposed method reconstructed more vivid details across all noisy images. Although it was directly applied to noisy LR images, the proposed method achieved a denoising effect comparable to that of comparison methods processing denoised LR images. On the other hand, fine details of the cortex and tissue boundaries were more apparent with the proposed method, while comparison methods which relied on pre-denoising tended to produce blurry or smooth edges. Moreover, the proposed technique also produced better interhemispheric fissures than the noise-robust methods like GPR method. It is important to note that although SCN, VDSR, DRRN, and SRFBN were able to reconstruct high-frequency image details and achieved state-of-the-art performance in several image upsampling scenarios, they are not suitable for super-resolving noisy images. For instance, fine details they reconstructed were blurred by the subsequent denoising method, so the overall performance of these deep learning-based methods (denoising+SCN, denoising+VDSR, denoising+DRRN, and denoising+SRFBN) deteriorate and even becomes visually comparable to GPR, a noise-robust zero-order based interpolation method. The degradation of these deep-learning-based methods on noisy images implies that it is important to incorporate image upsampling and denoising, so as to obtain fine details.

Table 2 demonstrates that the proposed method also achieved higher quantitative values than the other algorithms we tested. Specifically, its average PSNR gain was 0.92 dB, 1.00dB and 0.48dB higher than that of the second-best method with noise level at 3%, 5%, 7% of the maximum intensity, respectively. And for noise level at 9% of the maximum intensity, our method achieves slightly lower PSNR than the compared deep-learning-based methods (VDSR, DRRN, and SRFBN). Regarding SSIM, the proposed method also achieved the highest SSIM scores comparing with all conventional methods, which indicates that our method faithfully recovered fine details. But the proposed method has lower SSIM score than the compared deep-learning-based methods. It is not surprising that our method, an interpolation-based method, gives slightly inferior performance over the deep-learning-based methods, since their models are deep enough to approximate the non-linear mapping between LR and HR images. However, these deep-learning based methods require much larger dataset e.g., DIV2K contains 1000 RGB images with 2K resolution, more complex training processing and more time-consuming than our methods. On the contrary, our method uses the input LR image only and could be generalized to any image.

D. MR IMAGES WITHOUT NOISE

We have observed in Fig. 5 that the upsampling results from other compared methods are blurred, but this degradation could be possibly caused by the pre-denoising step. Hence, in order to make a fair comparison of detail preservation ability for these upsampling techniques, we devised an additional experiment in which noise-free MR images were used.

These MR images without noise were downloaded from the Brainweb phantom and corresponding LR images (downsampled to $3 \times 3 \times 3$ mm3) were generated using the described workflow in section III-A.
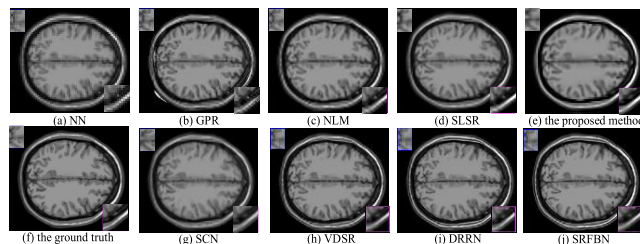

FIGURE 6. Upsampling results ($1 \times 1 \times 1$ mm^3) for simulated data with a $3 \times 3 \times 3$ mm^3 resolution. The PSNR and SSIM values for each method are: (a) 26.18/0.89, (b) 24.90/0.91, (c) 27.80/0.93, (d) 28.52/0.93, (e) 28.25/0.91, (g) 30.03/0.96, (h) 30.97/0.94, (i) 30.06/0.94, (j) 31.48/0.95.

As shown in Fig. 6, both GPR and NLM resulted in sever ghost artifacts along edges. This was caused by limitations of the zero-order regression-based framework, which cannot consistently reconstruct high-frequency details [10]. Conversely, the SLSR method, a second-order regression-based framework, and SCN, the model in which was trained from thousands of image patches via neural network, achieved visually superior performance near edges. Although the proposed method includes a zero-order regression-based framework, the resulting contours were comparable to that of SLSR. This confirms the adaptive weight sharpening strategy embedded in the proposed technique not only helps to remove noise, but also facilitates detail preservation.

E. REAL DATA

The proposed method was applied to two real clinical data sets and its performance was evaluated both quantitatively and qualitatively. The first was a T1 volume ($260 \times 311 \times 260$ voxels, $1 \times 1 \times 1$ mm^3 resolution) from the HCP dataset, which contained apparent noise. It was acquired on a Siemens 3T scanner and processed by a structural pre-processing pipeline that included spatial noise removal, surface generation, cross-modal registration, and alignment in standard space [21]. The second was a T1 volume ($256 \times 256 \times 78$ voxels, $2 \times 2 \times 2$ mm^3 resolution) acquired on a GE MR750 scanner, which contained little noise.

The first T1 volume was downsampled to a voxel resolution of $2 \times 2 \times 2$ mm^3 and then upsampled to $1 \times 1 \times 1$ mm^3 using the proposed technique and other conventional methods. Fig. 7 compares the reconstructed results for a sample slice processed using these methods. We observe the ground truth still exhibits considerable noise despite the pre-processing step in generation of this dataset, which indicates noise cannot be neglected in MR image upsampling. Although the blurring procedure involved in LR image generation may somewhat mitigate noise effects (i.e., the NN result included less noise than the ground truth), the other interpolation algorithms tended to amplify background noise.
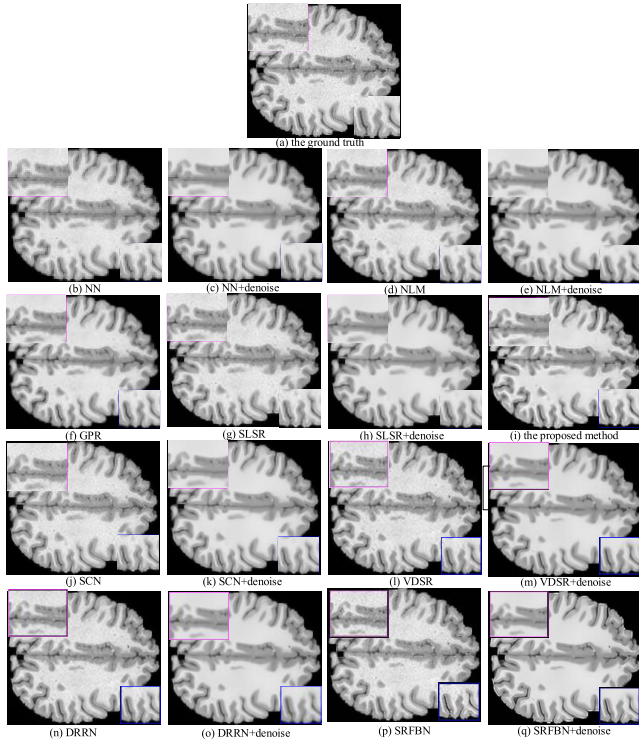
FIGURE 7. Upsampling of the HCP dataset subject using different methods.

FIGURE 8. Upsampling results for an adult image.

is O(N). For Rician noise correction, it takes O(N) to obtain the final voxel estimation. Therefore, the overall computational complexity of the proposed method is approximately $O(Nr^2)$.

FIGURE 9. Upsampling results using different techniques. (a) The LR image (enlarged by NN method for display), (b) and (c) are the proposed method with and without adaptive sharpening on LSK.

This excluded GPR and the proposed method, which reduced noise. Using an extra denoising step, NLM, SLSR and SCN could effectively remove noise, but fine details were also washed out. On the other hand, the proposed method achieved better detail preservation and edge reconstruction, due to higher contrast in these texture regions.

The second T1 image was directly upsampled to $1 \times 1 \times 1$ mm^3 using SCN, SLSR, and the proposed method. It is evident from Fig. 8 that the proposed method produced reconstruction results comparable to the other two methods (in terms of detail preservation) when dealing with MR images containing low noise.

F. COMPUTATIONAL COMPLEXITY

As outlined in Fig. 1, the major computational cost of the proposed method comes from these four parts: learning the local steering kernel, computing the weight matrix, adaptive weight sharpening, and Rician noise correction step. Given an image with N pixels and the local window radius r, it takes about $O(Nr^2)$ to learn the local steering kernel (including estimation of the gradient covariance matrix and computation of the weights according to Eq. (3) and (2) respectively). And then, it takes about O(N) to calculate the weight matrix. Regarding the adaptive weight sharpening process, two steps were involved: voxel classification according to the type of its neighboring patch structure, and adding or subtracting a fraction of the high-pass filtered weight matrix back to the same matrix. The computation complexity of these two steps
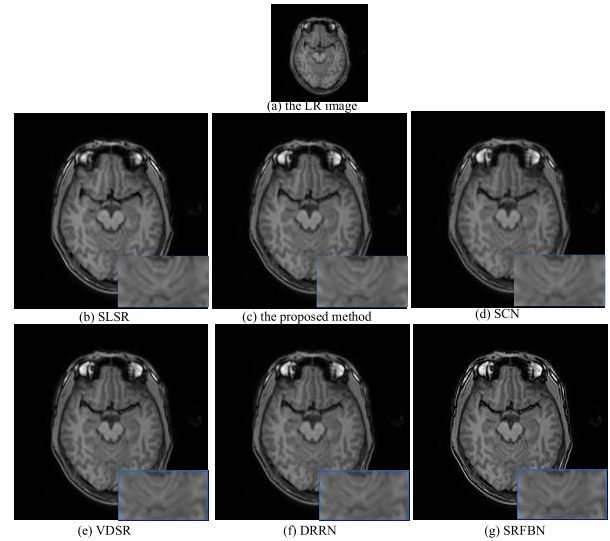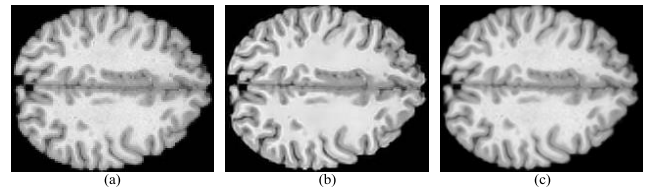
G. DISCUSSION

From the above experimental results, we can see that the proposed method achieves simultaneous MR image upsampling, denoising, and detail enhancement, and it outperforms other state-of-the-art image upsampling method when dealing images with noise. In general, our method contains several techniques, including LSK, adaptive sharpening on LSK and Rician correction. To discuss whether our noise robustness benefits from the proposed adaptive sharpening technique, the following experiment was carried out: the noisy LR image from HCP dataset (used in Fig. 7) was respectively processed by the proposed method with and without adaptive sharpening on LSK, and the corresponding results were shown at Figs. 9(b)-(c). It can be clearly observed that without LSK sharpening technique, the proposed method still produced noisy upsampling result. Moreover, without LSK sharpening, the resulting image appeared blurry near edges. These two findings suggested that the proposed adaptive sharpening technique contributes to the simultaneous denoising and detail enhancing in our method. As shown in Table 2, when processing noisy images, our proposed

method can get better results than the deep-learning-based method. For the deep-learning based methods, we firstly perform a denoising step on a noisy image, and then use a deep-learning-based method to upsample it. We can observe from Fig. 7 that using the two-step framework cause the edges of the reconstructed SR image blurred, and the details is not as good as the proposed method, which is a simultaneous image resolution enhancement and noise removal. However, Fig. 6 illustrates that when processing noise-free images, the deep-learning-based method can obtain higher PSNR. This make sense since deep-learning-based methods use much more high- resolution images to learn the relations between LR-HR image patches.

IV. CONCLUSION

In this paper, a unified interpolation-based SR method was proposed to reduce noise during MR image upsampling. By observing that both image interpolation and denoising essentially belong to a weighted averaging framework, we conducted simultaneous image upsampling, denoising, and detail enhancement by carefully devising the weights used in the averaging process. First, the LSK was employed in weight computation to implicitly capture local geometric structure. Additionally, an adaptive sharpening of the weight matrix was integrated to enhance detailed areas and smooth flat areas. Finally, by removing Rician bias from averaged squared intensities, we adapted the averaging process to data corrupted by Rician noise.

Experiments were conducted on both simulated and real cranial MR datasets and the upsampling results, achieved using various methods, were compared both quantitatively and qualitatively. Experiments with noisy MR images demonstrated that the proposed noise-robust interpolation method produced less blurry edges than a traditional two-step upsampling framework (upsampling and denoising), although the utilized upsampling method in this two-step framework was devised for detail preservation. Additionally, although the proposed method was a zero-order regression-based, it succeeded in recovering fine brain structure, comparable to SLSR, a second-order regression-based, in noise-free MR images. These two findings confirm the proposed method successfully adapted to spatial MR data structures and was highly robust in the presence of noise.

Currently, the proposed method is mainly developed for a single modality MR image upsampling. In the future, we plan to use information from high-resolution images acquired in other modalities to further refine interpolation coefficients, and thereby improve upsampling quality.

REFERENCES

[1] H. Greenspan, "Super-resolution in medical imaging," *Comput. J.*, vol. 52, no. 1, pp. 43–63, 2008.

[2] E. Plenge, D. H. J. Poot, M. Bernsen, G. Kotek, G. Houston, P. Wielopolski, L. van der Weerd, W. J. Niessen, and E. Meijering, "Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time?" *Magn. Reson. Med.*, vol. 68, no. 6, pp. 1983–1993, Dec. 2012.

[3] J. V. Manjón, P. Coupé, A. Buades, V. Fonov, D. L. Collins, and M. Robles, "Non-local MRI upsampling," *Med. Image Anal.*, vol. 14, no. 6, pp. 784–792, Dec. 2010.

[4] F. Rousseau, "A non-local approach for image super-resolution using intermodality priors," *Med. Image Anal.*, vol. 14, no. 4, pp. 594–605, Aug. 2010.

[5] J. V. Manjón, P. Coupé, A. Buades, D. L. Collins, and M. Robles, "MRI superresolution using self-similarity and image priors," *Int. J. Biomed. Imag.*, vol. 2010, Dec. 2010, Art. no. 425891.

[6] P. Coupé, J. V. Manjón, M. Chamberland, M. Descoteaux, and B. Hiba, "Collaborative patch-based super-resolution for diffusion-weighted images," *NeuroImage*, vol. 83, pp. 245–261, Dec. 2013.

[7] 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. MICCAI*, 2013, pp. 123–130.

[8] K. Jafari-Khouzani, "MRI upsampling using feature-based nonlocal means approach," *IEEE Trans. Med. Imag.*, vol. 33, no. 10, pp. 1969–1985, Oct. 2014.

[9] K. Jafari-Khouzani, E. Gerstner, B. Rosen, and J. Kalpathy-Cramer, "Upsampling dynamic contrast enhanced MRI," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2015, pp. 1032–1035.

[10] J. Hu, X. Wu, and J. Zhou, "Second-order regression-based MR image upsampling," *Comput. Math. Methods Med.*, vol. 2017, Jan. 2017, Art. no. 6462832.

[11] J. Hu, J. Zhou, and X. Wu, "Non-local MRI denoising using random sampling," *Magn. Reson. Imag.*, vol. 34, no. 7, pp. 990–999, Sep. 2016.

[12] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 60–65.

[13] W. L. Zeng and X. B. Lu, "Region-based non-local means algorithm for noise removal," *Electron. Lett.*, vol. 47, no. 20, pp. 1125–1127, Sep. 2011.

[14] J. Hu and Y.-P. Luo, "Non-local means algorithm with adaptive patch size and bandwidth," *Optik*, vol. 124, no. 22, pp. 5639–5645, 2013.

[15] H. Takeda, S. Farsiu, and P. Milanfar, "Deblurring using regularized locally adaptive kernel regression," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 550–563, Apr. 2008.

[16] H. Takeda, P. Milanfar, M. Protter, and M. Elad, "Super-resolution without explicit subpixel motion estimation," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 1958–1975, Sep. 2009.

[17] H. Wang, "Multiscale image sharpening adaptive to edge profile," *J. Electron. Imag.*, vol. 14, no. 1, Jan. 2005, Art. no. 013002.

[18] J. Manjon, J. Carbonellcaballero, J. Lull, G. Garciamarti, L. Martibonmati, and M. Robles, "MRI denoising using non-local means," *Med. Image Anal.*, vol. 12, no. 4, pp. 514–523, Aug. 2008.

[19] H. He and W.-C. Siu, "Single image super-resolution using Gaussian process regression," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 449–456.

[20] E. Lopez-Rubio, "Super-resolution from a single noisy image by the median filter transform," *SIAM J. Imag. Sci.*, vol. 9, no. 1, pp. 82–115, 2016.

[21] M. F. Glasser, S. N. Sotiropoulos, J. A. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. Van Essen, and M. Jenkinson, "The minimal preprocessing pipelines for the human connectome project," *NeuroImage*, vol. 80, pp. 105–124, Oct. 2013.

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

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[24] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. CVPR*, Jun. 2016, pp. 1646–1654.

[25] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. CVPR*, Jun. 2017, pp. 3147–3155.

[26] R. Lai, Y. Mo, Z. Liu, and J. Guan, "Local and nonlocal steering kernel weighted total variation model for image denoising," *Symmetry*, vol. 11, no. 3, p. 329, Mar. 2019.

[27] Y. Li and Y. Zhang, "Robust infrared small target detection using local steering kernel reconstruction," *Pattern Recognit.*, vol. 77, pp. 113–125, May 2018.

[28] H. Takeda, S. Farsiu, and P. Milanfar, "Kernel regression for image processing and reconstruction," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 349–366, Feb. 2007.

[29] H. Jong Seo and P. Milanfar, "Training-free, generic object detection using locally adaptive regression kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1688–1704, Sep. 2010.

[30] K. Zhang, X. Gao, D. Tao, and X. Li, "Single image super-resolution with non-local means and steering kernel regression," *IEEE Trans. Image Process.*, vol. 21, no. 11, pp. 4544–4556, Nov. 2012.

[31] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Oct. 2017, pp. 4539–4547.

[32] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 1664–1673.

[33] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.

[34] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3867–3876.

XINYAN LI received the B.E. degree in computer science and technology from the Chengdu University of Information Technology, China, in 2018. She is currently pursuing the master's degree with the Artificial Intelligence Laboratory of Computer Science, Chengdu University of Information Technology. Her current research interests include deep learning and image processing.

XIAOLONG WANG received the B.E. degree in surveying engineering from Jilin Jianzhu University, China, in 2009. He is currently working with Sichuan Diwu Surveying and Mapping Company Ltd. His current research interests include tilt photogrammetry and LiDAR mapping.

YAN LI received the B.E. degree in computer science and technology from the Department of Computer Science, Chongqing Three Gorges College, China, in 1997. She is currently working with Chongqing Cigarette Industry Company Ltd. Her current research interests include deep learning and image processing, and the practical application of this technology in manufacturing.

XI WU (Member, IEEE) received the B.S. degree in communication engineering from Sichuan University, Chengdu, China, in 2003, the M.S. degree in communication engineering from the University of Electronic Science and Technology of China, Chengdu, in 2006, and the Ph.D. degree in computer science from Sichuan University, in 2012. He is currently a Professor with the Department of Computer Science, Chengdu University of Information Technology, Chengdu. His current research interests include image processing and computer vision.

JING HU (Member, IEEE) received the bachelor's degree from the University of Electronic Science and Technology of China (UESTC), in 2009, and the Ph.D. degree from Tsinghua University, in 2015. She is currently an Associate Professor with the Chengdu University of Information Technology. Her current research interests include image processing, deep learning, reinforcement learning, and their applications.

• • •