Received: 27 June 2023 | Revised: 19 January 2024 | Accepted: 23 January 2024

DOI: 10.1002/mrm.30047

RESEARCH ARTICLE

Magnetic Resonance in Medicine

Self-supervised multicontrast super-resolution for diffusion-weighted prostate MRI

Batuhan Gundogdu | **Milica Medved** | **Aritrick Chatterjee** | **Roger Engelmann** | **Avery Rosado** | **Grace Lee** | **Nisa C. Oren** | **Aytekin Oto** | **Gregory S. Karczmar**

Department of Radiology, University of Chicago, Chicago, Illinois, USA

Correspondence
Batuhan Gundogdu, Department of Radiology, University of Chicago, Goldblatt Pavilion, A27 S Maryland Ave, Chicago, IL 60637, USA.
Email: gundogdu@uchicago.edu

Funding information
Sanford J. Grossman Charitable Trust; University of Chicago Medicine Comprehensive Cancer Center, Grant/Award Number: P30 CA014599-37; Foundation for the National Institutes of Health, Grant/Award Numbers: 1R41CA244056-01A1, 1S10OD018448-01, R01 CA17280, R01 CA227036

Abstract

Purpose: This study addresses the challenge of low resolution and signal-to-noise ratio (SNR) in diffusion-weighted images (DWI), which are pivotal for cancer detection. Traditional methods increase SNR at high *b*-values through multiple acquisitions, but this results in diminished image resolution due to motion-induced variations. Our research aims to enhance spatial resolution by exploiting the global structure within multicontrast DWI scans and millimetric motion between acquisitions.

Methods: We introduce a novel approach employing a "Perturbation Network" to learn subvoxel-size motions between scans, trained jointly with an implicit neural representation (INR) network. INR encodes the DWI as a continuous volumetric function, treating voxel intensities of low-resolution acquisitions as discrete samples. By evaluating this function with a finer grid, our model predicts higher-resolution signal intensities for intermediate voxel locations. The Perturbation Network's motion-correction efficacy was validated through experiments on biological phantoms and in vivo prostate scans.

Results: Quantitative analyses revealed significantly higher structural similarity measures of super-resolution images to ground truth high-resolution images compared to high-order interpolation ($p < 0.005$). In blind qualitative experiments, 96.1% of super-resolution images were assessed to have superior diagnostic quality compared to interpolated images.

Conclusion: High-resolution details in DWI can be obtained without the need for high-resolution training data. One notable advantage of the proposed method is that it does not require a super-resolution training set. This is important in clinical practice because the proposed method can easily be adapted to images with different scanner settings or body parts, whereas the supervised methods do not offer such an option.

KEYWORDS

DWI, implicit neural representation, prostate MRI, super-resolution for MRI

1 | INTRODUCTION

Diffusion-weighted images (DWI) are sensitive to the molecular displacement of water within each voxel. By acquiring DWI at varying diffusion-weighting values (*b*-values), we can calculate apparent diffusion coefficient (ADC) maps. These maps reveal tissue diffusivity levels and are particularly important because the increased cell density typically present in cancer significantly reduces the ADC values, which allows cancer detection. However, DWI often exhibits lower image resolution and signal-to-noise ratio (SNR) compared to other imaging sequences like T1 and T2-weighted images. To improve the SNR, multiple high-b DWI acquisitions are performed. These are then averaged to create a single image. However, this approach is susceptible to significant artifacts.[1] Artifacts in these images mainly arise from two types of motion or magnetic field fluctuations: (1) gross motion or field changes *between* acquisitions (e.g., due to breathing, rectal gas motions), and (2) localized motions or field changes *during* diffusion encoding gradients (e.g., arterial pulse or peristalsis). The nonrigid gross motion occurring between the separate signal acquisitions makes the image blurry, consequently reducing the resolution whereas the localized motion or B0 field fluctuations during the acquisition can reduce or completely suppress the signal.[2]

This presents a challenging situation with no clear solution. Specifically, when dealing with small voxel sizes, the common approach of signal averaging (obtaining multiple acquisitions and averaging them) can enhance the SNR but leads to decreased image quality and resolution due to motion-induced variations between acquisitions. In clinical practice, high SNR, reduced scanning times, and patient comfort typically outweigh the need for high resolution. Consequently, DWI images often have low resolution, affecting their specificity and the accuracy of targeted biopsies.[3] Thus, resolution is considered crucial to the "diagnostic image quality" in prostate MRI. This is measured with the PI-QUAL score, which reflects the extent to which the image assists in ruling in and ruling out clinically significant cancer.[4,5] Studies using the PI-QUAL score have shown that DWI image quality, including SNR and resolution, significantly impacts the accuracy of cancer detection.[6–9]

This study introduces a super-resolution model for diffusion-weighted prostate imaging that potentially reduces scan times and improves patient comfort, particularly for those with claustrophobia or Parkinson's disease. The method enables high-quality imaging from lower-resolution inputs, mitigating the need for a large number of averages needed for obtaining high-resolution images with adequate SNR. Our study aims to improve prostate DWI resolution through self-supervised deep learning, utilizing anatomical patterns across varying *b*-values and the nonlinear inter-acquisition motion. Its efficacy was assessed using in vivo prostate MR images with biopsy-confirmed cancers and carefully designed biological phantom experiments.

1.1 | Super-resolution in medical images

Super-resolution in medical images has been a topic of interest over the past two decades. Early methods concentrated on multi-image registration and stitching, utilizing shifted, orthogonal, anisotropic, or rotated scans.[10–14] However, the effectiveness of methods using linear subvoxel shifts has been debated, as they provide minimal or no in-plane improvement. This is because MRI is a Fourier imaging technique, here rigid spatial shifts only introduce a linear phase gradient without yielding a more densely sampled k-space.[15] Consequently, these techniques have only shown success in achieving through-plane super-resolution.

The success of deep learning-based super-resolution technologies in real-world imaging has inspired the application of similar supervised methods in medical imaging, drawing from techniques proven effective in digital photography.[16] Recent supervised deep learning methods assume that a low-resolution image is a "degraded" version of its high-resolution counterpart, aiming to reverse this "degradation process" to recover the high-resolution image.[17] While these methods are increasingly applied and show promise, it is crucial to note that, as with other medical deep learning applications, their effectiveness is often limited by the scarcity of available training data. Consequently, much of the super-resolution training in medical imaging depends on synthetically degraded input–output pairs.

However, it is important to approach these supervised methods cautiously. In medical imaging, high- and low-resolution image pairs are not readily available, and synthetically degrading images to create these pairs for training can lead to artifacts and hallucinations. The emerging self-supervised super-resolution paradigm presents a promising solution to overcome these limitations. Self-supervised learning utilizes the inherent information in the data, eliminating the need for explicit high- and low-resolution pairs. Table 1 offers a concise summary of the relevant literature in this domain.

This paper presents a self-supervised super-resolution model* for DWI utilizing the Implicit Neural Representation (INR) framework.[40] Unlike supervised methods that assume a low-resolution image is simply a degraded high-resolution image and seek to learn the inverse degradation function, our model adopts a more

TABLE 1 Literature review of super-resolution in medical imaging.

References	Year	Modality	Body part	Method
Peled et al.[10]	2001	DWI	Brain	Registration of shifted scans
Carmi et al.[11]	2006	T2W	Phantom	Registration of rotated scans
Kennedy et al.[12]	2007	PET/CT	Lung	Registration of shifted scans
Mayer et al.[13]	2007	T2W	Phantom	Registration of shifted scans
Lu et al.[14]	2010	T2W	Body	Registration of parallel scans from different coils
Scherrer et al.[15]	2011	DWI	Brain	Registration of rotated scans
Tieng et al.[18]	2011	T2W	Phantom	Registration of shifted scans
Van Reeth et al.[19]	2012	T2W	Brain	Registration of rotated/shifted scans (review article)
Scherrer et al.[20]	2012	DWI	Brain	Registration of orthogonal scans
Wu et al.[21]	2016	DWI	Brain	High-order singular value decomposition
Jurek et al.[22]	2017	T2W	Prostate	Registration of orthogonal scans
Alexander et al.[23]	2017	DWI	Brain	Patch regression
Chaudhari et al.[16]	2018	T2W	Knee	Supervised CNN
Zeng et al.[24]	2018	T1W	Brain	Supervised CNN
Chen et al.[25]	2018	T2W	Brain	Supervised CNN
Sood et al.[26]	2019	T2W	Prostate	Supervised GAN
Liu et al.[27]	2019	T2W	Prostate	Supervised CNN
Hong et al.[28]	2019	DWI	Brain	Supervised CNN
Dar et al.[29]	2019	T1, T2	Brain	Supervised GAN
He et al.[30]	2020	T2W	Prostate	Supervised deep attention networks
Zhou et al.[31]	2020	T1, T2, FLAIR	Brain	Supervised multimodal fusion-net
Park et al.[32]	2021	T2W	Lung	Supervised autoencoder-inspired CNN
Sood et al.[33]	2021	T2W	Prostate	Supervised GAN
Chatterjee et al.[34]	2021	DWI	Brain	Supervised Unet
Wu et al.[35]	2021	T1W	Brain	Self-supervised IRN
Jiang et al.[36]	2022	T1W	Brain	Supervised CNN with real 0.35T vs 3T scans
Molahasani et al.[37]	2022	T2W	Prostate	Supervised GAN
Mahaptra et al.[38]	2022	PD, T1 and T2	Brain	Supervised CNN and vision transformers
Guo et al.[39]	2023	DWI	Brain	Semi-supervised GAN
This work	2023	DWI	Prostate	Self-supervised INR

Abbreviations: CNN, convolutional neural networks; DWI, diffusion-weighted image; GAN, generative adversarial Networks; INR, implicit neural representation.

realistic assumption. It posits that the high-resolution image is a continuous function mapping the continuous domain of x-space to the range of signal intensities, and the low-resolution voxels are merely samples of this continuous function. Therefore, the objective is to learn this continuous function via the INR network, using the sparse samples (voxels) from the low-resolution image in the training set. In other words, this framework uses the voxel locations as input, and the acquired signal intensities within these voxels as the output.

While the INR framework was initially not designed for super-resolution, Wu et al. showed that using INR is an effective means of super-resolution on single-contrast T1W brain images.[35] We introduce two novelties to the INR approach and extend its application to DWI of the prostate: Firstly, we incorporate multiple degrees of diffusion weighting (quantified by the *b*-value) along with the standard three-dimensional spatial inputs. This modification transforms the INR into a multicontrast solution, thereby broadening its scope of application. Second, we

introduce the input Perturbation Network (PN) to correct voxel mismatches observed between different signal acquisitions and potentially to harvest important morphological details that might exist in individual acquisitions but can be lost within the large voxel sizes and during aggregation of different acquisitions. This PN learns to move the input voxel locations slightly for each acquisition while minimizing the reconstruction loss of the INR.

An analysis of the literature in Table 1 reveals that most super-resolution research focuses on brain imaging, with prostate MRI studies being relatively scarce, even though prostate cancer is the second deadliest cancer in men.[41] Moreover, the majority of super-resolution applications in MRI have been conducted using T1- or T2-weighted images. However, applying super-resolution to DWI which typically has lower resolution, could be more impactful. To our knowledge, existing super-resolution studies in DWI are limited to brain imaging, and the few prostate studies focus exclusively on T2-weighted images, with no attempts made on prostate DWI. Our study underscores the significance of high-resolution prostate DWI in (1) identifying extra-prostatic tumor extension, (2) pinpointing precise tumor locations, and (3) generating accurate coronal DWI views for enhanced cancer localization.

2 | METHODS

The foundational assumption of our model is that a full-resolution analog image can be represented as a continuous function, with low-resolution acquisitions being just discrete and noisy samples from this function. An INR network is used to approximate this continuous function. As previously mentioned, DWI are typically acquired at various b-values, including multiple acquisitions at high b-values. Each acquisition can be represented as a discrete three-dimensional volume:

$$S_k(b) \in \mathbb{R}^{M_x \times M_y \times M_z} \qquad k = 1 \ \ldots \ K, \tag{1}$$

where b is the diffusion-weighting value, M_x, M_y, and M_z denote the number of voxels along the x, y and z directions, respectively, with K representing the number of independent acquisitions. In standard practice, these acquisitions, denoted S_k are averaged to produce a single mean image for radiological examination:

$$S_{\text{mean}}(b) = \frac{1}{K}\sum_{k=1}^{K} S_k(b). \tag{2}$$

The INR network models the full-resolution image as a function $f : \mathbb{R}^4 \rightarrow \mathbb{R}$. This four-dimensional input space comprises the continuous three-dimensional voxel locations and the b-value, which serves as the fourth dimension. The network's output is the signal intensity corresponding to a specific voxel location and b-value. The INR is structured as a deep neural network with fully connected layers and sinusoidal activation functions, following the initialization guidelines from Sitzmann et al.[40] Training set inputs are quadruplets (m_x, m_y, m_z, b) representing the normalized voxel location and b-value, scaled to range between -1 and 1 along each axis. The model's output labels are the intensity values of the image observed at the input location. Thus, the INR is trained to minimize the error between the predicted diffusion values at input locations $\mathbf{x}_i = (m_x, m_y, m_z)$ and the measured diffusion values $S_k(\mathbf{x}_i, b)$:

$$\mathcal{L} = \sum_{i,k} \|S_k(\mathbf{x}_i, b) - \text{INR}(\mathbf{x}_i, b)\|^2. \tag{3}$$

Essentially, INR's objective is to estimate the signal for *any* given input location, even for those not included in the training set, that is, "intermediate voxels" with higher resolution. Take, for instance, our practically discrete input space: while input samples like $\mathbf{x}_1 = (0, 0, 0, 0)$ and $\mathbf{x}_2 = \left(\frac{1}{M_x/2}, 0, 0, 0\right)$ are part of the training set, an intermediate sample along the first axis, such as $\mathbf{x}_3 = \left(\frac{1}{M_x}, 0, 0, 0\right)$, is not.† Once trained with $\mathbf{x}_1, \mathbf{x}_2$ and other location indices, the INR is capable of predicting diffusion values for the "intermediate voxel locations," like $\mathbf{x}_3$. Importantly, it should be emphasized that the INR functions beyond the capabilities of a simple sinc interpolator. This distinction arises for three reasons: (1) INR integrates information across various b-values within shared layers (Equation 3), (2) the PN applies nonlinear motion correction across acquisitions and b-values (Section 2.1), and (3) the training process incorporates data-driven learning weights for each voxel's acquisitions, addressing inter-acquisition variability (Section 2.2).

2.1 | Input PN for motion correction

The motion-induced intra- and inter-acquisition variabilities are often nonlinear due to B0 inhomogeneities and bowel-enclosed gas. We represent the warping effect of millimetric motion as a perturbation to the INR input. We assume that the reason $S_{k1}(\mathbf{x}, b) \neq S_{k2}(\mathbf{x}, b)$ is the millimetric motion between acquisitions $k1$ and $k2$ and that $S_{k1}(\mathbf{x}, b) = S_{k2}(\mathbf{x} + \epsilon, b)$. This implies that when an acquisition's signal value differs from the network output, the method considers whether the discrepancy is due to the acquisition originating not from the designated voxel location but from a nearby subvoxel location, affected by small inter-acquisition motions. A one-dimensional example

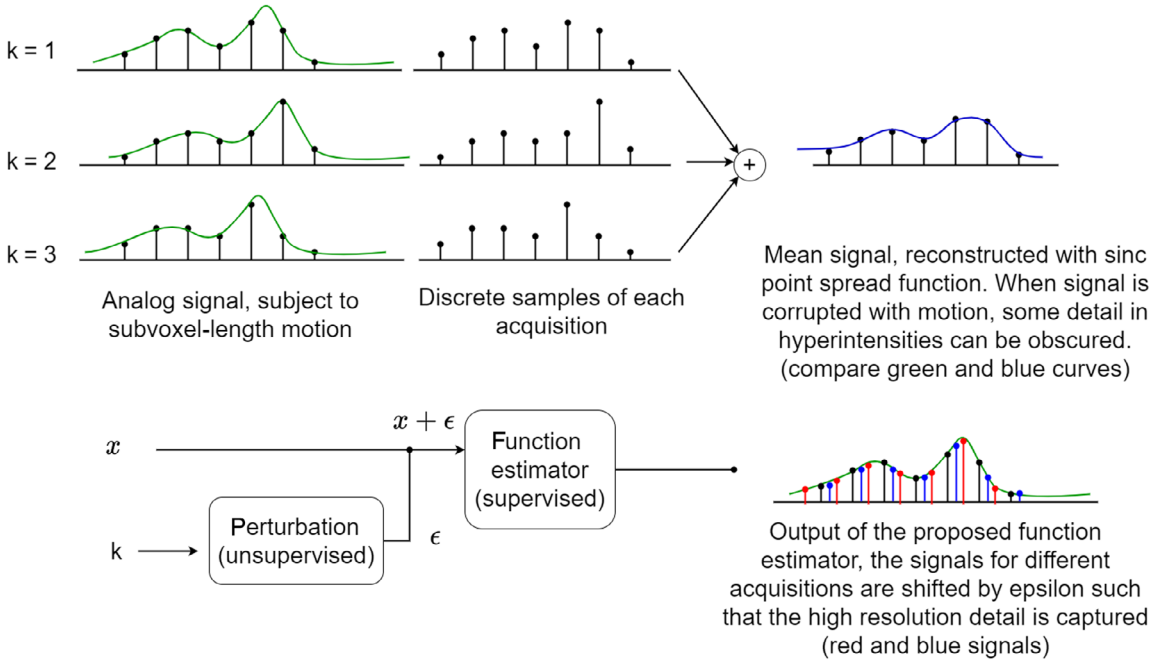

FIGURE 1 A one-dimensional example of the zero-shot super-resolution methodology. The supervised block uses the low-resolution voxel intensities as labels, hence effectively there is no need for high-resolution ground truth labels.

of the idea behind the proposed methodology is visualized in Figure 1. In this example, when the point spread function is applied on each voxel, very small motions that occur between acquisitions cause the sharp edge to be lost even when it is sampled at the Nyquist rate. In the proposed approach, the discrete input space is generalized as a continuous domain, and the input PN is trained in an unsupervised fashion to estimate a warping function for each acquisition to minimize the reconstruction loss at the output.

We propose that this perturbation value can be learned jointly with the INR parameters, on a separate neural network. The error that the signal for such an acquisition creates at the output is backpropagated to assign this value to a better location near the center of the input voxel such that the total loss is minimized for all voxels and all acquisitions. This warping is done via the PN which is modeled as another fully connected neural network that takes the acquisition number and the *b*-value as input, in addition to the voxel location $\mathbf{x}$. The network then estimates a perturbation value ϵ via a scaled `tanh` activation function such that the perturbation is constrained to vary within physically feasible motion limits by enforcing the following cost:

$$\mathcal{L} = \sum_{i,k} \|S_k(\mathbf{x}_i, b) - \mathrm{INR}(\mathbf{x}_i + \epsilon_{\max}\mathtt{tanh}(\mathrm{PN}(\mathbf{x}_i, b, k)), b)\|^2. \tag{4}$$

The perturbation upper-bound ($\epsilon_{\max}$) was set as half a voxel size along each direction, that is, $\epsilon_{\max} = \left(\frac{1}{M_x}, \frac{1}{M_y}, \frac{1}{M_z}\right)$.

Note that the acquisition number k and the *b*-value used as inputs to the PN take discriminative roles, and no temporal order along acquisitions is assumed. This creates an opportunity to capture sharp edges within the image, even if the voxel size is too large to capture it, and this edge is captured in different acquisitions for the same voxel due to the motion. The flowchart of the proposed network and the role of the PN is demonstrated in Figure 2.

2.2 | Weighted learning for suppressing anomalously low signals

Besides the inter-acquisition bulk motion of the prostate as discussed in Section 2.1, the disparity in $S_{k1}(\mathbf{x}, b)$ and $S_{k2}(\mathbf{x}, b)$ for $k1 \neq k2$ may also stem from the molecular motion of water or the fluctuation in local magnetic field *during* application of the diffusion-sensitizing gradients. These factors lead to variations that are too large to be explained by thermal Rician noise.[42] This type of variation is commonly referred to as physiological noise. Physiological noise typically leads to an anomalous reduction in signal intensity, resulting in a right-skewed per-voxel signal intensity distribution. This aspect is crucial, as conventional signal averaging tends to yield ADC values significantly higher than the actual ADC values, which are key for evaluating tissue composition and diagnosing cancers.

Representing a DWI S, derived from multiple acquisitions S_k, as a single INR (as proposed in this study),

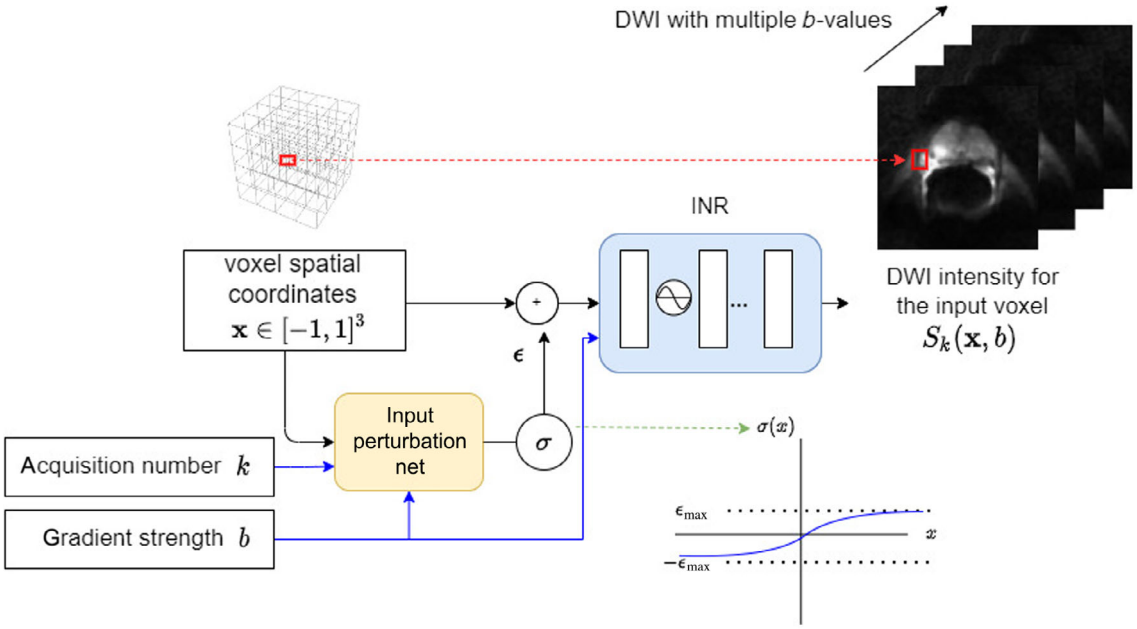

FIGURE 2 The flowchart of the implicit neural representation learning, augmented by the input perturbation network.

effectively addresses such signal mismatches. The INR framework's operation within continuous spaces for both domain and range grants it an awareness of proximal voxels. This characteristic allows the INR to potentially favor intensity values of the acquisitions that are more consistent with the neighboring voxels, while naïve-averaging merely calculates the mean of different signal intensities for each voxel.

The proposed methodology also presents an opportunity to assign weight values to each acquisition such that the motion-corrupted signals can be automatically suppressed during training. For this, we follow the recipe that was recently proposed by Gundogdu et al.[42] Accordingly, we alter the loss function in Equation (3) to incorporate learning weights (w_k) as thus:

$$\mathcal{L} = \sum_{i,k} w_k(\mathbf{x}_i, b) \| S_k(\mathbf{x}_i) - \mathrm{INR}(\mathbf{x}_i + \epsilon_{\max} \mathtt{tanh}(\mathrm{PN}(\mathbf{x}_i, b, k))) \|^2. \tag{5}$$

Detailed derivation and further interpretation of the learning weights are provided in Appendix S1.

2.3 | Evaluation of super-resolution performance and accuracy

In the inference phase, to generate the super-resolution image, the INR network employs a finer input grid than the resolution of the voxels used in training. This approach yields an image with enhanced detail and improved resolution. However, to be able to assert that the new details introduced by the network correspond to actual anatomical features, it is crucial to establish reliable ground truth images for evaluation. In the case of super-resolution examples in digital photography, such datasets are readily available.[43] On the other hand, for MRI, especially DWI, acquiring high-resolution images to this degree is not feasible. In vivo acquisitions with higher resolution would result in unacceptably low SNR. Furthermore, the long acquisition times would increase vulnerability to motion artifacts. Therefore, in addition to the in vivo prostate scans, we employed a biological phantom to evaluate our model's performance.

2.3.1 | Evaluation on phantom

This specifically designed phantom experiment aimed to test the unsupervised motion-correction efficacy of the PN. In this experiment, we emulated both the presence and absence of inter-acquisition motion in MRI scans of a biological phantom. Kiwifruit was selected as the biological phantom, as various studies have demonstrated its effectiveness as a proxy for prostate MRI, DWI included.[44,45]

DWIs of kiwifruit were acquired under three distinct settings. Firstly, we acquired images with a resolution of $1.5\,\mathrm{mm} \times 1.5\,\mathrm{mm} \times 3.0\,\mathrm{mm}$ in nine repetitions. To simulate rigid motion in the axial plane, we simulated transitive motion to each repetition by shifting the center of the field-of-view in-plane by −0.5, 0.0, or +0.5 mm in the readout or phase-encoding directions. In some cases,

the shifting of the center of the field-of-view is achieved in postprocessing by the scanner, rather than by changing the sequence parameters, and thus the underlying k-space data may not have changed due to the shift. However, our algorithm operates on the produced magnitude images without relying on the complex k-space information and the main objective was to assess if the PN corrects this motion with unsupervised training, thus this was deemed acceptable.

The second image was acquired as a ground truth with the same resolution and the same number of acquisitions as the first one, but with no motion. We recognize that the synthetically introduced motion is rigid and might not fully replicate the nonrigid warping often encountered in prostate diffusion imaging. However, we opted to apply motion during the MR scan rather than inducing artifacts in postprocessing, to more accurately mirror real-world conditions.

Additionally, we acquired a third image with a high-resolution of $0.75\,\text{mm} \times 0.75\,\text{mm} \times 3\,\text{mm}$, serving as a reference to evaluate the detail captured in the INR output and the first two images (with and without motion). All scans were conducted on a dStream Philips 3T Ingenia scanner with digital coil technology, a 70-cm bore, dual RF transmission, and maximum gradient strength of 45 mT/m, and a maximum slew rate of 200 T/m/s. DWI images were acquired at $b = 1000\,\text{s/mm}^2$, with TR/TE = 4000/90 ms, four EPI segments, in-plane resolution of $1.5 \times 1.5\,\text{mm}^2$ (acquired) and $1.4 \times 1.4\,\text{mm}^2$ (reconstructed), and slice thickness of 3 mm. The phase encoding direction is along the main axis of the kiwi fruit, parallel to its striated structural pattern.

To see the correcting impact of the PN, we trained one INR by setting $\epsilon_{\max} = 0$ and another by setting $\epsilon_{\max}$ to be half a voxel size. We used the scans with the motion as input to the network and compared the output to the no-motion ground truth, as well as the very high-resolution ground truth.

2.3.2 | Evaluations on in vivo prostate images

We evaluated the performance of the proposed method on diffusion-weighted prostate images of patients with biopsy-verified prostate cancers. This was conducted following the approval of the Institutional Review Board (IRB) (Approval number: IRB17-1694). We ran both qualitative and quantitative experiments. To evaluate the resolution improvement, we created a training set of low-resolution prostate images by applying low-pass filtering and subsampling to the original images. Another approach could have involved acquiring a separate set of low-resolution images with larger voxel sizes during a second scan of the patient. However, this would have introduced a different SNR, rendering the original images suboptimal as ground truth. No additional noise was introduced to the training set. The network was trained with DWIs with four *b*-values: 0, 150, 1000, and $1500\,\text{s/mm}^2$, so that it can yield a super-resolved and motion-corrected ADC map. Orthogonal gradients were in the readout, phase encoding, and slice directions. At high *b*-values multiple acquisitions at each orthogonal direction were used as independent acquisitions to train INR. The INR framework addresses the inter-acquisition motion directly via the PN, and intra-acquisition motion indirectly by suppressing the resulting low signals via weighted learning.

The network was trained with the voxels of the subsampled low-resolution image. Then the trained network was used to generate the 2× super-resolution images. It is noteworthy that the network has the potential to produce outputs at multiple resolutions, not limited to the 2× factor. Two fellowship-trained and board-certified attending radiologists, with 12+ and 6+ years of experience in prostate MRI (NCO and GL respectively), evaluated the 2× super-resolution images, 2× bicubic-interpolated images (serving as an approximation to sinc interpolation), and ground truth (GT) images. They were tasked with ranking these three sets of images—super-resolution, bicubic-interpolated, and GT—based on their perceptual quality. This experiment involved 65 cases with high-*b* images, and the order of presentation for the three test images was randomized for each subject during evaluation. Hence, the radiologists were unaware of the identity of the specific image they were evaluating.

The same set of 65 cases was utilized for a quantitative analysis, comparing the super-resolution and bicubic-interpolated images against their respective ground truth counterparts. For this analysis, we used established quantitative image similarity metrics common in digital image analysis, such as the Structural Similarity Index Measure (SSIM), Multiscale SSIM (MS-SSIM), Feature Similarity Index Measure (FSIM), and Spectral Residual-based Similarity Index Measure (SR-SIM).[46–48] Each metric was used to compare the 2× super-resolution and 2× bicubic-interpolated images with their corresponding high-resolution ground truth images. Statistical significance was evaluated using paired t-tests with $N = 65$.

3 | RESULTS

Model training was carried out utilizing the PyTorch software package on a workstation equipped with Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz, Tesla V100 GPU with 32 GB memory.

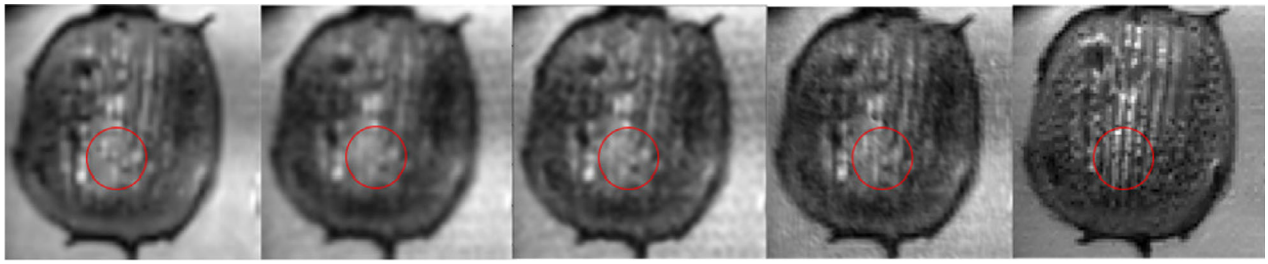

FIGURE 3 Proposed network on kiwifruit diffusion-weighted images (DWI). (A) image acquired without simulated motion, (B) image acquired with simulated motion, (C) model output with $\epsilon_{\max}$ is set to zero, that is, Perturbation Network (PN) is not in effect, (D) model output with Perturbation Network turned on and (E) very high-resolution "ground truth" image used for visual validation. The vertical detail enclosed in the red circle is lost as a result of motion as seen in (B). The model output at (D) does not only address the effects of motion, but it also provides a better image than the "no-motion" image shown in (A), as can be verified from the very high-res image in (E). (A) Image without motion; (B) Image with simulated motion: (C) INR output when the PN turned off ($\epsilon_{\max} = 0$); (D) INR output when the PN turned on ($\epsilon_{\max} = 0.7$ mm); (E) Very high-resolution scan.

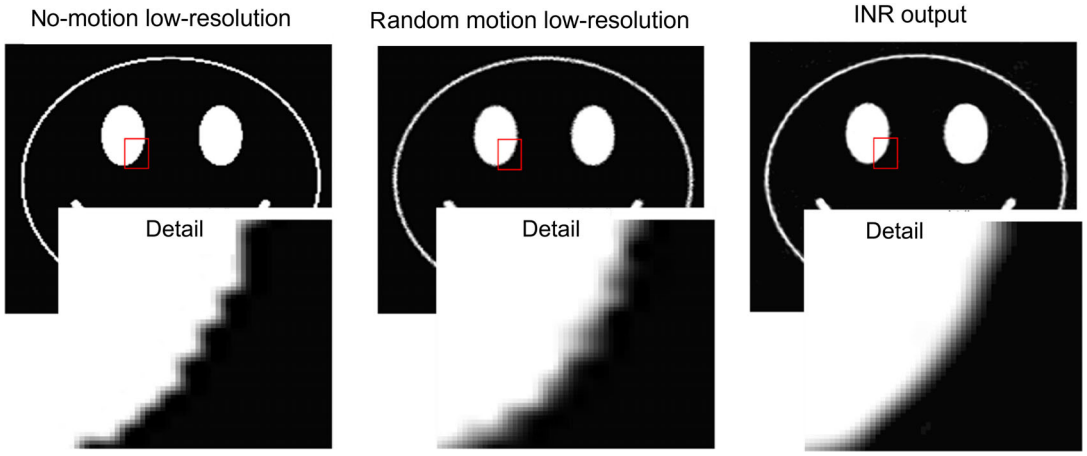

FIGURE 4 Comparison between the in silica image with no motion corruption (left), the motion-corrupted average image (middle), and the output of the implicit neural representation (INR) (right). The input to the INR model was the middle image.

3.1 | Unsupervised motion correction with the PN

Figure 3 displays representative images from the phantom experiment. Since MRI is a Fourier technique, the calculated shifts resulted in a linear phase change in k-space without contributing to increased resolution. The purpose of this experiment was to demonstrate that MRI scans can be modeled using an INR and that the PN can effectively mitigate motion effects. The PN successfully inferred the direction and magnitude of voxel movements ($\boldsymbol{\epsilon}$) for each acquisition and applied the necessary corrections.

In the kiwifruit experiment, the perturbation vector $\boldsymbol{\epsilon}$ remained consistent across all voxels but varied for each acquisition. To assess the PN's ability to decipher completely random perturbations for all voxels, we created a motion-corrupted image by slightly displacing each voxel in random directions. This was achieved by slightly perturbing the location of each voxel in each acquisition towards a random direction. This in silica experiment simulated a scenario significantly more severe than the actual motion effects typically observed in DWI scans. We conducted this test to determine whether the PN could learn such nonlinear motion effects despite being trained in an unsupervised manner. Figure 4 provides a comparison between the image without motion corruption, the motion-corrupted average image, and the INR output.

3.2 | Results on in vivo prostate images

In the observer study, 96.1% of the super-resolution images were rated as having superior perceptual quality compared to the interpolated images. Moreover, 40.7% of the super-resolution images were voted to have better quality than the ground truth images, despite the ground truth

images having double the resolution and quadruple the voxel count of the model's input. This latter finding is particularly noteworthy as it indicates that the PN significantly contributes to the improvement in image quality in these cases. The most likely reason for this outcome is that the model is trained by $K = 12$ individual acquisitions despite seeing only a quarter of the number of voxels as input (compared to ground truth). It is worth noting that the interpolated and ground truth images were also generated using the mean of the same number of acquisitions, implying no expected SNR gain with the algorithm. When the PN is not in effect, the main INR learns to average these acquisitions for the observed locations in the training set and attempts to predict the intermediate voxels based on the continuity of the INR function. However, considering that the model can produce outputs that visually outperform the high-resolution ground truth, we can argue that the PN learns to reposition certain features away from the voxel center, thereby providing higher resolution information and accounting for subvoxel motion. This effect is most prominent when dealing with sharp features that exhibit relatively high SNR.

In quantitative analyses, super-resolution INR outputs exhibited significantly higher similarity to the ground truth images compared to the bicubic-interpolated images ($p < 0.005$) across all quantitative similarity metrics (SSIM, MS-SSIM, FSIM, and SR-SIM). Naturally, image areas are smooth regions and consist of low-resolution components that are similar in super-resolution, bicubic-interpolated and ground truth images. Consequently, these measurements yield very high similarity values in the aforementioned measurements. However, our primary focus is on determining whether the high-resolution details introduced by the super-resolution model are present in the ground truth images. To investigate this, we applied a high-pass filter to all super-resolution, ground truth, and bicubic images, effectively removing the low spatial frequency components, and focused solely on the high spatial frequency details. This analysis revealed that most of the gain obtained by super resolution is in high spatial resolution components, for example, edges etc. Table 2 presents the similarity scores of the super-resolution and bicubic-interpolated mean images (baseline) relative to the high-resolution ground truth. Figure 5 showcases examples from this evaluation set. Additional details on the similarity scores and further image examples can be found in the Appendix S1.

3.3 | Super-resolution performance along slice-select direction

The proposed methodology treats the DWI as three-dimensional volumetric objects, aiming to enhance super-resolution not only in the axial view but also along the slice select direction. Typically, the z-axis slices are taken thicker than the axial resolution to improve SNR gain. Consequently, the coronal or sagittal views of DWI, despite their potential significance, are often overlooked in clinical practice. Reconstructing INR with finer details along the slice-select direction enables the visualization of coronal or sagittal DWI views, derived from axial slices. Figure 6 demonstrates the super-resolution reconstructed coronal view of a sample DWI alongside the corresponding slice of the T2W coronal image.

3.4 | Contrast enhancement via weighted training

To assess the contrast enhancement due to weighted training, we trained the network with the high-resolution GT image and applied the learning weights as detailed in the Appendix S1. ADC maps calculated using the model output are compared to the ADC maps of GT images. The analyses were done on a cohort of 12 patients with biopsy-verified cancers by measuring both the SNR on cancer and the cancer-to-healthy-prostate contrast ratio. Statistical analysis was conducted using paired t-tests. The contrast ratio significantly improved on average from 1.9 to

TABLE 2 Image similarity measures of super-resolution compared to bicubic interpolation.

	All bands				High-frequency bands			
Metric	Bicubic versus GT	SR versus GT	Gain (%)	*p*-Value	Bicubic versus GT	SR versus GT	Gain (%)	*p*-Value
SSIM	88.5	89.3	0.9	<0.001	52.5	56.3	7.5	<0.001
MS-SSIM	88.6	90.1	1.8	<0.001	72.0	75.8	5.4	<0.001
SR-SIM	87.8	89.0	1.5	<0.005	87.6	89.1	1.8	<0.005
FSIM	90.2	91.9	1.9	<0.001	88.5	90.5	2.3	<0.001

Abbreviations: FSIM, feature similarity index measure; GT, ground truth; MS-SSIM, multiscale SSIM; SR, super-resolution; SSIM, similarity index measure.

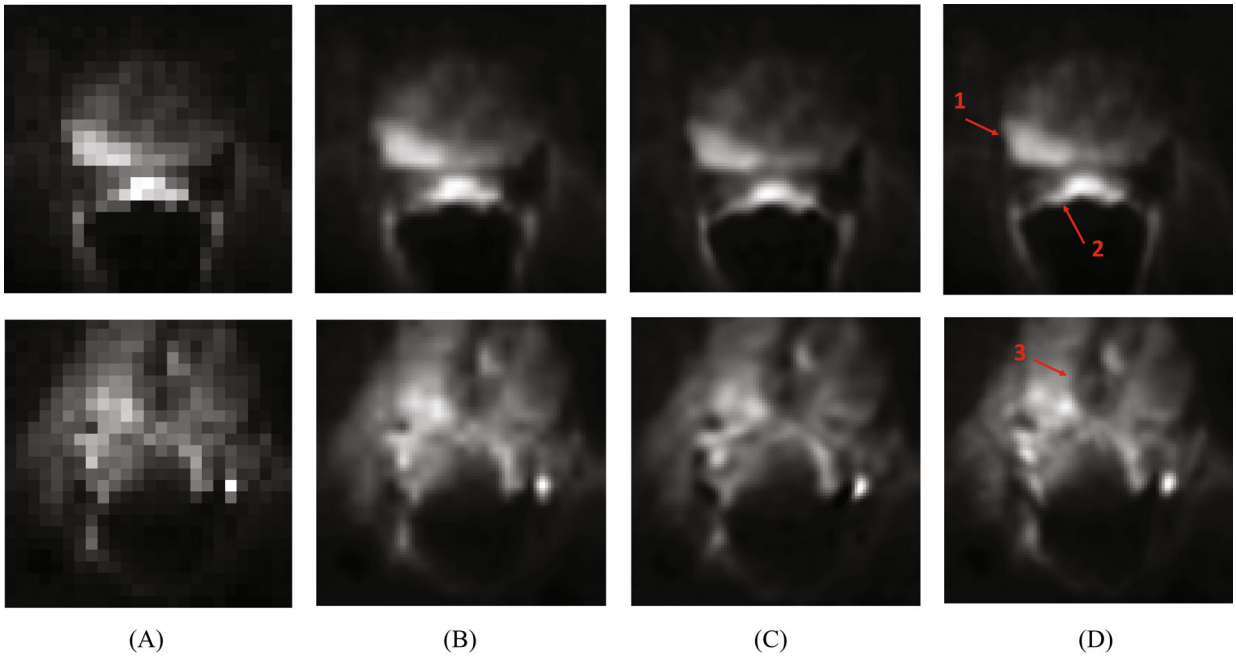

FIGURE 5 Performance of the proposed method on two cases: (A) LR image, (B) 2× bi-cubic interpolated image, (C) the output of the model with 2× super-resolution, and (D) the HR ground truth image. The red arrows show anatomic details that implicit neural representation enhances. Arrow 1 shows the enhancement in the border of the cancer along the margin of the prostate, arrow 2 shows the rectal wall, and arrow 3 shows that the boundaries of the urethra and an adjacent suspicious lesion can be seen better in the super-resolution image. (A) LR; (B) Bi-cubic interpolation; (C) INR output; (D) HR ground truth.

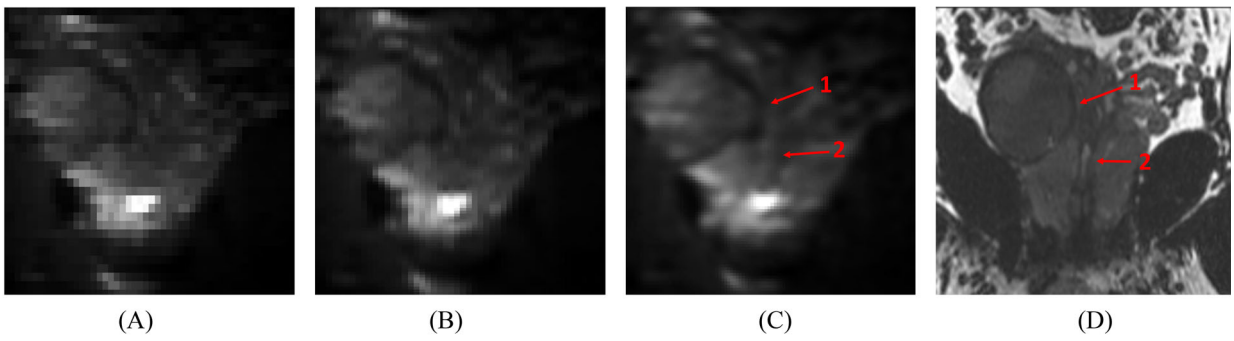

FIGURE 6 Super-resolution reconstructed coronal view of one sample image compared to the interpolated image. T2W coronal image that corresponds to the same slice is given on the right for reference. Arrow 1 shows the boundaries and extent of the cancer whereas Arrow 2 shows the prostatic urethra which was not visible in interpolation but visible in super-resolution image. (A) DWI coronal view: (B) Interpolation along z-axis; (C) INR output; (D) T2W coronal image.

2.4 ($p < 0.01$) when comparing the model output with the input image. Similarly, the SNR showed a significant average increase from 1.87 to 1.92 ($p < 0.01$), despite no noise suppression being applied.

4 | DISCUSSION AND CONCLUSIONS

This study validates the feasibility of zero-shot super-resolution in diffusion-weighted prostate imaging and underscores the potential of using individual acquisitions for self-supervised motion mitigation to some extent. Although introducing synthetic shifts and obtaining super-resolution is a well-established implementation, we provided a novel unsupervised method to utilize the underlying nonrigid motion to obtain super-resolution. We showed this by comparing the model output with the motion-free ground truth, using the settings with the PN turned on and off. The new detail introduced by the network was also compared with a very-high-resolution image, which was acquired in an idealistic setting obtained

solely for this experiment and would not have been practically possible in actual patient scans.

As discussed earlier, precalculated linear shifts during imaging provide negligible gains in spatial resolution. However, the "shifting" observed between subsequent in vivo acquisitions is nonrigid.[49] Moreover, local B0 inhomogeneities, arising from poor shimming, tissue boundaries, or proximity to bowel-enclosed gas, contribute additional nonlinear alterations to the MRI signal. Thus, we argue that additional information may, in fact, be present in the repeated acquisitions of the same image. Much of the existing literature on super-resolution to DWI focuses on brain imaging, where motion is typically minimal compared to prostate imaging, thus not leading to significant non-linear changes in the MRI signal.

It's important to note that phantom experiments (Figures 3 and 4) and in vivo experiments (Figures 5 and 6) are designed to showcase two distinct features of the proposed model. All kiwifruit images in Figure 3A–D share the same resolution and voxel count. This figure effectively demonstrates the motion-correction capability of our model, especially the role of the PN. The PN aims not only to correct signal mismatches in different acquisitions of inter- and intra-*b*-value scans but also to recover details potentially lost in naive averaging. Conversely, Figures 5 and 6 showcase the super-resolution capability of the proposed method highlighting how the model, trained exclusively with LR image voxels, compares with the HR image. The model's estimation of "intermediate voxels" is noticeably superior to bi-cubic interpolation, a difference attributed to our method functioning globally, in contrast to interpolation's local operation. The quantitative structural similarity results in Table 2 further substantiate this observation.

Naturally, this study has certain constraints and limitations that should be acknowledged. For example, we selected bi-cubic interpolation as a competitor for evaluation, whereas other studies employing synthetically generated low-resolution inputs could compare their approaches directly. However, it is important to note that the challenge we undertook in the context of "real image super-resolution" does not allow for a uniform or universally accepted ground truth. If we had acquired images with smaller voxel sizes, it would have resulted in lower SNR, or longer acquisition times at lower voxel sizes would have introduced variations in motion exposure. As a result, direct comparisons or supervised training with other CNN-based super-resolution methods would not have been feasible. In essence, we intentionally adopted this self-supervised setting to address these challenges. Key advantages of the proposed model over the supervised deep learning-based methods are enumerated below:

1. **Input/Output:** Supervised methods rely on processing low-resolution images as *input*, utilizing local kernel operations. At the output, these models expect the high-resolution versions of the images. In contrast, the INR method takes only the voxel locations (x,y,z) as input and emit the voxel intensity that corresponds to the input location.
2. **Resolution factor:** Supervised models are trained with predesigned low-/high-resolution pairs and learn to increase the resolution by the factor dictated in the difference of high- and low-resolution training images. INR models the continuous space and can provide improvement for any factor, *even noninteger ones*.
3. **Training set:** Supervised models typically require a large training set of input/output pairs. Deep models that are trained with fewer images than required possess the risk of hallucination, that is, carrying some information/detail seen on the training set to a test image. INR does not need a training set of images from other patients. It merely uses the existing voxel intensities for each scan to infer a motion-corrected and higher-resolution image.
4. **Risk of artifacts:** Acquiring low-/high-resolution training pairs for medical imaging is impractical, often leading to the synthetic generation of low-resolution samples. However, INR bypasses such complexities, as each image individually trains its own model, eliminating the risk of artifact generation due to incorrect modeling.
5. **Practicality:** Supervised models, trained for specific imaging settings (TE, TR, *b*-value, etc.), or data from a certain vendor (Phillips, GE, Siemens etc.) or images from a certain modality (T2W, T1W, DWI, etc.), require fine-tuning for new settings, posing challenges for inter-institutional operation and clinical practice. Conversely, INR is free from such constraints; it can train and super-resolve a new scan within minutes without a pre-existing model for inference.

A particularly promising implication of this study is the potential for reduced scan times, which is especially beneficial for patients experiencing claustrophobia or those with Parkinson's disease. By employing the proposed model for super-resolution, it becomes possible to obtain high-quality images from lower-resolution inputs, thus mitigating the need for a large number of averages needed for obtaining high-resolution images with adequate SNR. This has significant benefits in terms of patient comfort and convenience. In addition, shortened echo train durations will help reduce the spatial distortion often present in EPI-based images.

From the clinical perspective, reconstructed coronal DWI images of the prostate, as we proposed in this paper,

could be helpful in more accurate localization of the tumor to the appropriate level of apex, midgland, or base; and distinguishing a lesion location in the peripheral zone of the prostate from normal periprostatic tissue such as the surgical capsule or the periprostatic plexus by lesion location confirmation with an additional view, or distinguishing lesion location in the central gland, where BPH nodules could mimic cancer, from the peripheral zone, where malignant lesions are more prevalent.

CONFLICT OF INTEREST STATEMENT

Drs. Oto, Karczmar, and Chatterjee hold equity in QMIS LLC.

ENDNOTES

*https://github.com/batuhan-gundogdu/INR.

†Without loss of generality, this example assumes that M_x is an even number, which is in fact fair since FFT in DWI reconstruction typically calls for sizes with powers of 2.

ORCID

Batuhan Gundogdu https://orcid.org/0000-0002-9395-7519

Milica Medved https://orcid.org/0000-0003-2945-5511

Aritrick Chatterjee https://orcid.org/0000-0003-4022-7106

Gregory S. Karczmar https://orcid.org/0000-0002-2455-5858

REFERENCES

1. le Bihan D, Poupon C, Amadon A, Lethimonnier F. Artifacts and pitfalls in diffusion MRI. *J Magn Reson Imaging*. 2006;24:478-488.
2. Liau J, Lee J, Schroeder ME, Sirlin CB, Bydder M. Cardiac motion in diffusion-weighted MRI of the liver: artifact and a method of correction. *J Magn Reson Imaging*. 2012;35:318-327.
3. Baxter GC, Patterson AJ, Woitek R, Allajbeu I, Graves MJ, Gilbert F. Improving the image quality of DWI in breast cancer: comparison of multi-shot DWI using multiplexed sensitivity encoding to conventional single-shot echo-planar imaging DWI. *Br J Radiol*. 2021;94:20200427.
4. Giganti F, Kirkham A, Kasivisvanathan V, et al. Understanding PI-QUAL for prostate MRI quality: a practical primer for radiologists. *Insights Imaging*. 2021;12:1-19.
5. Francesco G, Alexander N, Aqua A, et al. Global variation in magnetic resonance imaging quality of the prostate. *Radiology*. 2023;309:e231130.
6. Dinneen E, Allen C, Strange T, et al. Negative mpMRI rules out extra-prostatic extension in prostate cancer before robot-assisted radical prostatectomy. *Diagnostics*. 2022;12:1057.
7. Karanasios E, Caglic I, Zawaideh JP, Barrett T. Prostate MRI quality: clinical impact of the PI-QUAL score in prostate cancer diagnostic work-up. *Br J Radiol*. 2022;95:20211372.
8. Windisch O, Benamran D, Dariane C, et al. Role of the prostate imaging quality PI-QUAL score for prostate magnetic resonance image quality in pathological upstaging after radical prostatectomy: a multicentre European study. *European Urology Open Science*. 2023;47:94-101.
9. Reijnen JS, Ryg U, Marthinsen JB, Schönhardt I, Seierstad T, Hole KH. Monoparametric high-resolution diffusion weighted MRI as a possible first step in an MRI-directed diagnostic pathway for men with suspicion of prostate cancer. *Front Oncol*. 2023;13:1102860.
10. Peled S, Yeshurun Y. Superresolution in MRI: application to human white matter fiber tract visualization by diffusion tensor imaging. *Magn Reson Med*. 2001;45:29-35.
11. Carmi E, Liu S, Alon N, Fiat A, Fiat D. Resolution enhancement in MRI. *Magn Reson Imaging*. 2006;24:133-154.
12. Kennedy JA, Israel O, Frenkel A, Bar-Shalom R, Azhari H. Improved image fusion in PET/CT using hybrid image reconstruction and super-resolution. *Int J Biomed Imaging*. 2007;2007:1-10.
13. Mayer GS, Vrscay ER. Measuring information gain for frequency-encoded super-resolution MRI. *Magn Reson Imaging*. 2007;25:1058-1069.
14. Lu Y, Yang R, Zhang J, Zhang C. Super resolution image reconstruction in parallel magnetic resonance imaging. Paper presented at: 2010 8th IEEE International Conference on. Control and Automation. IEEE. 2010; Xiamen, China:761-766.
15. Scherrer B, Gholipour A, Warfield SK. Super-resolution in diffusion-weighted imaging. Paper presented at: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011: 14th International Conference; September 18-22. 2011; Toronto, Canada:124-132.
16. Chaudhari A, Fang Z, Hyung LJ, Gold G, Hargreaves B. Deep learning super-resolution enables rapid simultaneous morphological and quantitative magnetic resonance imaging. Paper presented at: International Workshop on Machine Learning for Medical Image Reconstruction. 2018; Granada, Spain:3-11.
17. Yamac M, Ataman B, Nawaz A. Kernelnet: a blind super-resolution kernel estimation network. Paper presented at: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2021; Nashville, TN:453-462.
18. Tieng QM, Cowin GJ, Reutens DC, Galloway GJ, Vegh V. MRI resolution enhancement: how useful are shifted images obtained by changing the demodulation frequency? *Magn Reson Med*. 2011;65:664-672.
19. Van Reeth E, Tham IWK, Tan CH, Poh CL. Super-resolution in magnetic resonance imaging: a review. *Concepts Magn Reson Pt A*. 2012;40:306-325.
20. Scherrer B, Gholipour A, Warfield SK. Super-resolution reconstruction to increase the spatial resolution of diffusion weighted images from orthogonal anisotropic acquisitions. *Med Image Anal*. 2012;16:1465-1476.
21. Wu X, Yang Z, Hu J, Peng J, He P, Zhou J. Diffusion-weighted images superresolution using high-order SVD. *Comput Math Methods Med*. 2016;2016:1-9.
22. Jurek J, Kociński M, Materka A, et al. Reconstruction of high-resolution t2W mr images of the prostate using maximum a posteriori approach and Markov random field regularization. Paper presented at: 2017 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA). 2017; Poznan, Poland:96-99.

23. Alexander DC, Zikic D, Ghosh A, et al. Image quality transfer and applications in diffusion MRI. *Neuroimage*. 2017;152:283-298.
24. Zeng K, Zheng H, Cai C, Yang Y, Zhang K, Chen Z. Simultaneous single-and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. *Comput Biol Med*. 2018;99:133-141.
25. Chen Y, Xie Y, Zhou Z, Shi F, Christodoulou AG, Li D. Brain MRI super resolution using 3D deep densely connected neural networks. Paper presented at: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). 2018; Washington, D.C.:739-742.
26. Sood R, Rusu M. Anisotropic super resolution in prostate mri using super resolution generative adversarial networks. Paper presented at: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). 2019; Venice, Italy:1688-1691.
27. Liu K, Ma Y, Xiong H, et al. Medical image super-resolution method based on dense blended attention network. *arXiv preprint arXiv:1905.05084*. 2019.
28. Hong Y, Chen G, Yap P-T, Shen D. Reconstructing high-quality diffusion MRI data from orthogonal slice-undersampled data using graph convolutional neural networks. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2019; Shenzhen, China:529-537.
29. Dar SU, Yurt M, Karacan L, Erdem A, Erdem E, Cukur T. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. *IEEE Trans Med Imaging*. 2019;38:2375-2388.
30. He X, Lei Y, Fu Y, et al. Super-resolution magnetic resonance imaging reconstruction using deep attention networks. *Proc Vol*. 2020;11313:642-647.
31. Zhou T, Fu H, Chen G, Shen J, Shao L. Hi-net: hybrid-fusion network for multi-modal MR image synthesis. *IEEE Trans Med Imaging*. 2020;39:2772-2781.
32. Park S, Gach HM, Kim S, Lee SJ, Motai Y. Autoencoder-inspired convolutional network-based super-resolution method in MRI. *IEEE J Transl Eng Health Med*. 2021;9:1-13.
33. Sood RR, Shao W, Kunder C, et al. 3D registration of pre-surgical prostate MRI and histopathology images via super-resolution volume reconstruction. *Med Image Anal*. 2021;69:101957.
34. Chatterjee S, Sciarra A, Dünnwald M, et al. ShuffleUNet: super resolution of diffusion-weighted MRIs using deep learning. Paper presented at: 2021 29th European Signal Processing Conference (EUSIPCO). 2021; Dublin, Ireland:940-944.
35. Wu Q, Li Y, Xu L, et al. Irem: high-resolution magnetic resonance image reconstruction via implicit neural representation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention. 2021; Strasbourg, France:65-74.
36. Jiang J, Qi F, du H, et al. Super-resolution reconstruction of 3T-like images from 0.35 T MRI using a hybrid attention residual network. *IEEE Access*. 2022;10:32810-32821.
37. Molahasani Majdabadi M, Choi Y, Deivalakshmi S, Ko S. Capsule GAN for prostate MRI super-resolution. *Multimed Tools Appl*. 2022;81:4119-4141.
38. Mahapatra D. Improved super resolution of MR images using CNNs and vision transformers. *arXiv preprint arXiv:2207.11748*. 2022.
39. Guo H, Wang L, Gu Y, Zhang J, Zhu Y. Semi-supervised super-resolution of diffusion-weighted images based on multiple references. *NMR Biomed*. 2023;36:e4919.
40. Sitzmann V, Martel JN, Bergman AW, Lindell DB, Wetzstein G. Implicit neural representations with periodic activation functions. *Adv Neural Inform Process Syst*. 2020;33:7462-7473.
41. Giaquinto AN, Miller KD, Tossas KY, Winn RA, Jemal A, Siegel RL. Cancer statistics for African American/black people 2022. *CA Cancer J Clin*. 2022;72:202-229.
42. Gundogdu B, Pittman JM, Chatterjee A, et al. Directional and inter-acquisition variability in diffusion-weighted imaging and editing for restricted diffusion. *Magn Reson Med*. 2022;88:2298-2310.
43. Dong C, Loy CC, Tang X. Accelerating the super-resolution convolutional neural network. Paper presented at: European Conference on Computer Vision. 2016; Amsterdam, The Netherlands:391-407.
44. Mueller-Lisse UG, Murer S, Mueller-Lisse UL, Kuhn M, Scheidler J, Scherr M. Everyman's prostate phantom: kiwi-fruit substitute for human prostates at magnetic resonance imaging, diffusion-weighted imaging and magnetic resonance spectroscopy. *Eur Radiol*. 2017;27:3362-3371.
45. Tatsuya H, Shimpei Y, Shinya K, Toshimune I. Variability in contrast and apparent diffusion coefficient of kiwifruit used as prostate MRI phantom: 1-week validation. *Radiol Phys Technol*. 2022;15:1-6.
46. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-612.
47. Zhang L, Zhang L, Mou X, Zhang D. FSIM: a feature similarity index for image quality assessment. *IEEE Trans Image Process*. 2011;20:2378-2386.
48. Zhang L, Li H. SR-SIM: a fast and high performance IQA index based on spectral residual. Paper presented at: 2012 19th IEEE International Conference on Image Processing. 2012; Orlando, FL:1473-1476.
49. Sadinski M, Medved M, Karademir I, et al. Short-term reproducibility of apparent diffusion coefficient estimated from diffusion-weighted MRI of the prostate. *Abdom Imaging*. 2015;40:2523-2528.

SUPPORTING INFORMATION

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

Appendix S1. Supporting information.

How to cite this article: Gundogdu B, Medved M, Chatterjee A, et al. Self-supervised multicontrast super-resolution for diffusion-weighted prostate MRI. *Magn Reson Med*. 2024;1-13. doi: 10.1002/mrm.30047