

MRI reconstruction using deep Bayesian estimation

Guanxiong Luo¹ | Na Zhao² | Wenhao Jiang¹ | Edward S. Hui¹ | Peng Cao¹ 

¹Department of Diagnostic Radiology, The University of Hong Kong, Hong Kong, Hong Kong

²Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, Hong Kong

Correspondence

Peng Cao, Department of Diagnostic Radiology, The University of Hong Kong, Hong Kong, 5 Sassoon Road, Pok Fu Lum, Hong Kong, Hong Kong.
Email: caopeng1@hku.hk

Purpose: To develop a deep learning-based Bayesian estimation for MRI reconstruction.

Methods: We modeled the MRI reconstruction problem with Bayes's theorem, following the recently proposed PixelCNN++ method. The image reconstruction from incomplete k-space measurement was obtained by maximizing the posterior possibility. A generative network was utilized as the image prior, which was computationally tractable, and the k-space data fidelity was enforced by using an equality constraint. The stochastic backpropagation was utilized to calculate the descent gradient in the process of maximum a posterior, and a projected subgradient method was used to impose the equality constraint. In contrast to the other deep learning reconstruction methods, the proposed one used the likelihood of prior as the training loss and the objective function in reconstruction to improve the image quality.

Results: The proposed method showed an improved performance in preserving image details and reducing aliasing artifacts, compared with GRAPPA, ℓ_1 -ESPRiT, model-based deep learning architecture for inverse problems (MODL), and variational network (VN), last two were state-of-the-art deep learning reconstruction methods. The proposed method generally achieved more than 3 dB peak signal-to-noise ratio improvement for compressed sensing and parallel imaging reconstructions compared with the other methods.

Conclusions: The Bayesian estimation significantly improved the reconstruction performance, compared with the conventional ℓ_1 -sparsity prior in compressed sensing reconstruction tasks. More importantly, the proposed reconstruction framework can be generalized for most MRI reconstruction scenarios.

KEYWORDS

Bayesian estimation, compressed sensing, deep learning reconstruction, generative network, parallel imaging

1 | INTRODUCTION

In compressed sensing MRI reconstruction, the commonly used analytical regularization such as ℓ_1 regularization can

ensure the convergence of the iterative algorithm and improve MR image quality.¹ The conventional iterative reconstruction algorithm with analytical regularization has an explicit mathematical deduction in gradient descent, which ensures

the convergence of the algorithm to a local or global optimal and the generalizability. Besides, the dictionary learning is an extension of analytical regularization, providing an improvement over the ℓ_1 regularization in specific applications.² The study of analytically regularized reconstruction mainly focused on choosing the appropriate regularization function and parameters. As an extension of analytical regularization, the deep learning reconstruction was employed as an unrolled iterative algorithm for solving the regularized optimization^{3,4} or used as a substitute for analytic regularization.^{5,6} With the advances of deep learning methodology, research started shifting the paradigm to structured feature representation and recovery of MRI, such as cascade,³ iterative with fidelity term,^{4,6} and generative⁶ deep neural networks. Especially, the method proposed in³ recast the compressed sensing reconstruction into a specially designed neural network that still partly imitated the analytical data fidelity and regularization terms. In that study, the analytical regularization term was replaced with convolutional layers and a specially designed activation function.³ In a latter study, several data-consistency layers were embedded in a feed-forward convolutional network to keep the reconstructed image consistent with k-space data.⁵ In another study, a discriminator (from a generative adversarial network) was used to create an image manifold, ensuring the reconstruction can explore the feasible data space, along with the data consistency layers.⁶ In a recent study, following the well-known “unrolling” of iterative reconstruction approach, a semi-iterative convolutional network was also developed as a model-based deep learning architecture for inverse problems (MODL).⁴ In a more recent study, a sophisticated framework for generalized compressed sensing reconstruction was formulated as a variational network (VN) that was embedded in an unrolled gradient descent scheme.⁷ These deep learning methods may show improved performance in some predetermined acquisition settings or pretrained imaging tasks. However, they also lack flexibility when used with changes in MRI undersampling scheme, the number of radio-frequency coils, and matrix size or spatial resolution. Such restriction is caused by the embedment of k-space data fidelity and the regularization terms into neural network implementations. Therefore, it was preferable to separate the k-space data fidelity and neural network-based regularization for improving the flexibility in changing MRI acquisition configurations.

This study applied Bayesian estimation to model the MRI reconstruction problem, and the statistical representation of an MRI database was used as a prior model. In Bayesian estimation, the prior model is required to be computationally scalable and tractable.^{8,9} The scalability indicates that the prior model has an explicit probabilistic distribution function, which can be used as loss function for both network training and image reconstruction.⁸⁻¹⁰ The tractability of the

prior model means the probability of an given image can be calculated directly without any approximation, besides the gradient that facilitates the maximization of posterior distribution can be calculated by stochastic backpropagation.^{8,9} In such Bayesian estimation, the image to be reconstructed was referred to as the parameters of the Bayesian model, which was conditioned on the measured k-space data (as the posterior). Bayes's theorem expressed the posterior as a function of the k-space data likelihood and the image prior. Recently, a variational autoencoder (VAE) was applied as the deep density prior in MRI reconstruction, under a Bayesian estimation framework,¹¹ but the evidence lower bound (ELBO) used in VAE is an approximation density model. The ELBO is calculated via Monte Carlo sampling which leads to expensive computational cost. Moreover, the general challenge of VAE is to match the conditional distribution in latent space to the explicit distribution.¹² For the image prior, Refs. [13,14] proposed a generative deep learning model, providing a tractable and scalable likelihood. In those studies, the image prior model was written as the multiplication of the conditional probabilities those indicated pixel-wise dependencies of the input image. The k-space data likelihood described how the measured k-space data was computed from a given MR image. The relationship between k-space data and MR image can be described, using the well-known MRI encoding matrix in an equality constraint.¹⁵ With such computationally scalable and tractable prior model, the maximum a posterior can serve as an effective estimator⁹ for the high dimensional image reconstruction problem tackled in this study. To summarize, the Bayesian estimation for MRI reconstruction had 2 separate models: the k-space likelihood model that was used to encourage data consistency and the image prior model that was used to exploit knowledge learned from an MRI database.

This paper presented a generic and interpretable deep learning-based reconstruction framework, using Bayesian estimation. It employed a generative network as the MR image prior model. The proposed framework was capable of exploiting the MR image database with the prior model, regardless of the changes in MR imaging acquisition settings. Also, the reconstruction was achieved by a series of estimations those employed the maximum likelihood of posterior with the image prior, that is, applying the Bayesian estimation repeatedly. The reconstruction iterated over the data fidelity enforcement in k-space and the image refinement, using the Bayesian estimation. During the iteration, the projected sub-gradient algorithm was used to maximize the posterior. The method is theoretically described, which was adapted from the methodology proposed by others,¹³ and then demonstrated in different MRI acquisition scenarios, including parallel imaging, compressed sensing, and non-Cartesian reconstructions. The robustness and the reproducibility of the algorithm were also experimentally validated.

2 | THEORY

The proposed method applied a generative neural network, as a data-driven MRI prior, to an MRI reconstruction method. This section contained an MRI reconstruction method using Bayes' theorem and a generative neural network-based MRI prior model, a pixel-wise joint probability distribution for images, using the PixelCNN++.¹³

2.1 | MRI reconstruction using Bayes' theorem

With Bayes' theorem, one could write the posterior as a product of likelihood and prior:

$$f(\mathbf{x}|\mathbf{y}) = \frac{f(\mathbf{y}|\mathbf{x})g(\mathbf{x})}{f(\mathbf{y})} \propto f(\mathbf{y}|\mathbf{x})g(\mathbf{x}) \quad (1)$$

where $f(\mathbf{y}|\mathbf{x})$ is probability of the measured k-space data $\mathbf{y} \in \mathbb{C}^M$ for a given image $\mathbf{x} \in \mathbb{C}^N$, N is the number of pixels and M is the number of measured data points. $g(\mathbf{x})$ is the prior model that estimates the distribution of MR images. In order to avoid the confusion with the likelihood occurs in modelling the prior, $f(\mathbf{y}|\mathbf{x})$ is referred to as k-space likelihood model. The image reconstruction is achieved by exploring the posterior $f(\mathbf{x}|\mathbf{y})$ with an appropriate estimator. The maximum a posterior estimation (MAP) could provide the reconstructed image $\hat{\mathbf{x}}$ that is given by:

$$\hat{\mathbf{x}}_{\text{MAP}}(\mathbf{y}) = \arg \max_{\mathbf{x}} f(\mathbf{x}|\mathbf{y}) = \arg \max_{\mathbf{x}} f(\mathbf{y}|\mathbf{x})g(\mathbf{x}) \quad (2)$$

In this way, the reconstruction problem is recast as posterior probability calculations. Equation (2) indicates 2 models are required for the estimation: the prior model, $g(\mathbf{x})$, and k-space likelihood model, $f(\mathbf{y}|\mathbf{x})$.

2.2 | Expressing prior using model parameters

The proposed method estimated the distribution of MR image features concerning parameters of a mixture distribution model given by a generative network.¹³ The mixture distribution model (or the distribution of MR image features) was then used to compute the likelihood of images, serving as a prior model for reconstruction (as illustrated in Figure 1). The generative network was commonly used as a parameterized model, approximating the distribution of image features.¹³ There were 3 reasons for using such mixture distribution model and generative network approximation¹³: (a) hierarchical architectures allowed the network to capture complex structure in the image, (b) the mixture of conditional distribution characterized an intrinsic dependence

among pixels and factorized the probability density function for an image over its pixels, (c) they discretized/instanced logistic mixture likelihood for using pixelCNN++.¹³ For MRI reconstruction, the number of image channels was changed from 3 (ie, RGB channels for color image) to 2 (ie, real and imaginary parts for MR image). For each image pixel, the continuous variable ν denoted the real or the imaginary part of a pixel, $\nu = \text{Re}(x_{i,j})$ or $\text{Im}(x_{i,j})$. Like in the VAE and pixelCNN++,^{13,16} the distribution of ν was a mixture of the logistic distribution, given by

$$\nu \sim \sum_{k=1}^K \pi_k \text{logistic}(\mu_k, s_k). \quad (3)$$

Here, π_i was the mixture indicator, μ_i and s_i were the mean and scale of logistic distribution, K was the number of mixture distributions, respectively. It should be noted that the logistic distribution function was interchangeable with any distribution function for creating a mixture with sufficient representation capacity. However, for a tractable computation, the logistic distribution was recommended in the previous study.¹³ Then the probability on each observed ν of the pixel was computed as¹³

$$P(\nu; \pi, \mu, s) = \sum_{k=1}^K \pi_k [\sigma((\nu + d/2 - \mu_k)/s_k) - \sigma((\nu - d/2 - \mu_k)/s_k)], \quad (4)$$

where σ was the logistic sigmoid function, d was the smallest discretized interval for ν . Furthermore, in,^{13,14} each pixel was dependent on all previous pixels up and to the left in an image, as shown in Figure 1. The derivation from Equations (3) to (4) is provided in Appendix A. The conditional distribution of the subsequent pixel ($\text{Re}(x_{i,j})$, $\text{Im}(x_{i,j})$) at position (i, j) was given by¹³

$$\begin{aligned} p(x_{i,j}|C_{i,j}) &= p(\text{Re}(x_{i,j}), \text{Im}(x_{i,j})|C_{i,j}) \\ &= P(\text{Re}(x_{i,j}); \pi(C_{i,j}), \mu_{\text{Re}}(C_{i,j}), s_{\text{Re}}(C_{i,j})) \\ &\quad \times P(\text{Im}(x_{i,j}); \pi(C_{i,j}), \hat{\mu}_{\text{Im}}(C_{i,j}, \text{Re}(x_{i,j})), s_{\text{Im}}(C_{i,j})) \\ \hat{\mu}_{\text{Im}}(C_{i,j}, \text{Re}(x_{i,j})) &= \mu_{\text{Im}}(C_{i,j}) + \alpha(C_{i,j})\text{Re}(x_{i,j}), \end{aligned}$$

where the $C_{i,j} = \{x_{i-1,j}, x_{i-2,j}, \dots, x_{i,1}\}$ denoted the context information which was comprised of the mixture indicator and the previous pixels as showed in Figure 1, α was the coefficient related to mixture indicator and previous pixels, μ_{Re} was the μ of the real channel of $C_{i,j}$, and μ_{Im} , s_{Re} , s_{Im} were subscripted in the same way. $p(\text{Re}(x_{i,j}), \text{Im}(x_{i,j})|C_{i,j})$ was also a joint distribution for both real and imaginary channels. The real part of the first pixel, that is, $x^{(1)} = x_{i,1}$ in Figure 1, was predicted by a mixture of logistics as described in Equation (3). This definition assumed that the mean of mixture components of the imaginary channel was linearly dependent on the real channel. In this study, the number of mixture components was 10. In this model, mixture indicator was shared

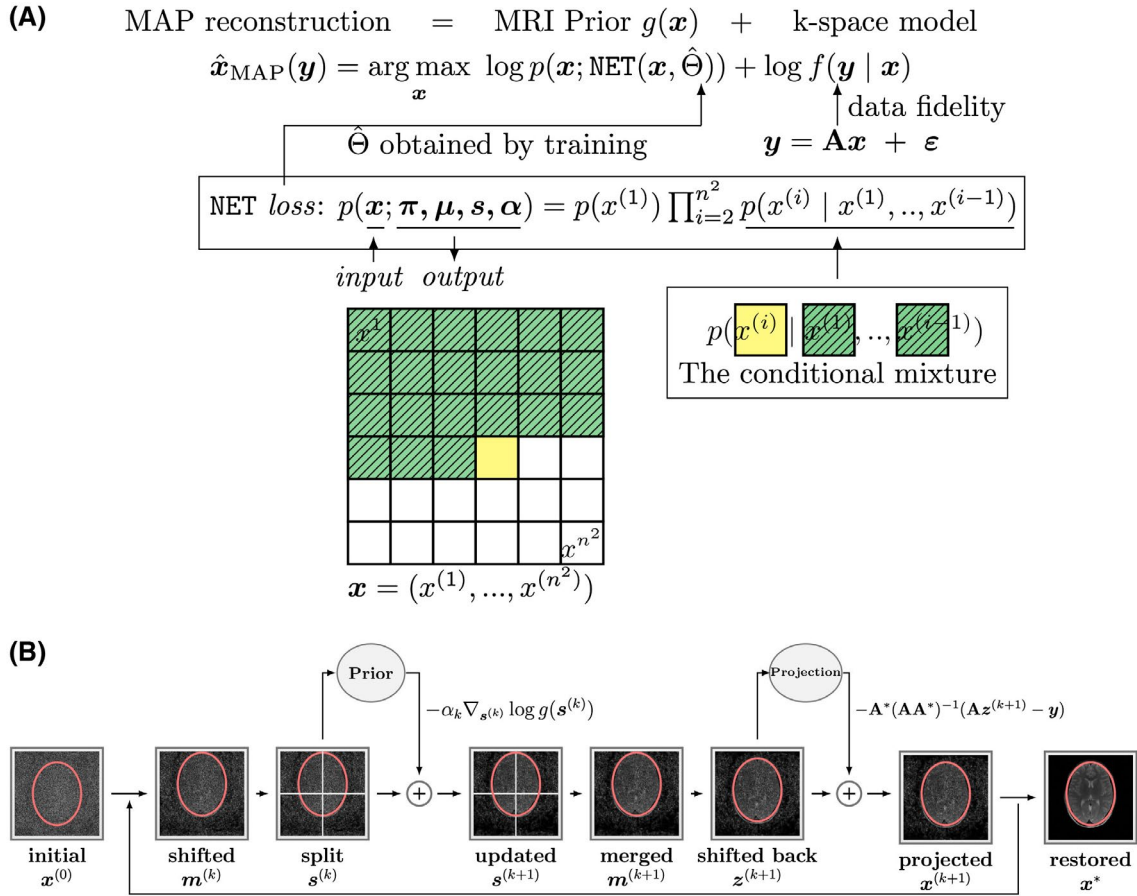


FIGURE 1 A, Overview of the MAP reconstruction. Conditional model in^{13,14} defined the probability of image pixel (yellow) x_{ij} dependent on all the pixels from its up and left side (green). B, In this method, we reconstructed images with 256×256 matrix size, using the prior model $g(\mathbf{x})$ that was trained with 128×128 images and illustrated in Supporting Information Figure S1. To reconcile this mismatch, we split one 256×256 image into four 128×128 patches for applying the prior model. After updating $\mathbf{s}^{(k+1)}$, 4 patches for 1 image were merged to form an image with the original size of 256×256 . Then the merged image was projected onto $\{\mathbf{y} = \mathbf{A}\mathbf{x} + \varepsilon\}$ in Equation (11). Furthermore, the random shift along phase encoding direction was applied to mitigate the stitching line in-between patches

between 2 channels. The $n \times n$ image could be considered as an vectorized image $\mathbf{x} = (x^{(1)}, \dots, x^{(n^2)})$ by stacking pixels from left to right and up to bottom of one another, that is, $x^{(1)} = x_{1,1}, x^{(2)} = x_{2,1}, \dots$, and $x^{(n^2)} = x_{n,n}$, as illustrated in Figure 1. The joint distribution of the image vector could be expressed as following¹³:

$$p(\mathbf{x}; \boldsymbol{\pi}, \boldsymbol{\mu}, \mathbf{s}, \boldsymbol{\alpha}) = p(x^{(1)}) \prod_{i=2}^{n^2} p(x^{(i)} | x^{(1)}, \dots, x^{(i-1)}). \quad (5)$$

$\boldsymbol{\pi}, \boldsymbol{\mu}, \mathbf{s}, \boldsymbol{\alpha}$ were the parameters of mixture distribution for each pixel as were defined in Equations (4)-(6). The generative network PixelCNN++ was expected to predict the joint probability distribution of all pixels in the input image.¹³ Therefore, the network $\text{NET}(\mathbf{x}, \Theta)$ was trained by maximizing the likelihood in Equation (5), as the training loss was given by

$$\hat{\Theta} = \arg \max_{\Theta} p(\mathbf{x}; \text{NET}(\mathbf{x}, \Theta)), \quad (6)$$

where Θ contained the trainable parameters within the network. After training, the network could be used as the image prior. Here, we defined the prior model $g(\mathbf{x})$ as

$$g(\mathbf{x}) = p(\mathbf{x}; \text{NET}(\mathbf{x}, \hat{\Theta})). \quad (7)$$

For example, if the complex images \mathbf{x} for training had a size of (128, 128), and the number of mixture distribution $K = 10$, then $\boldsymbol{\pi}, \boldsymbol{\mu}, \mathbf{s}, \boldsymbol{\alpha}$ would get dimensions of (128, 128, 10, 1), (128, 128, 10, 2), (128, 128, 10, 2), and (128, 128, 10, 1), respectively. To summarize, a prior model of \mathbf{x} was defined in Equations from (2) to (7) that could be considered as a data-driven model, utilizing the knowledge learned from an image database.

2.3 | Image reconstruction by MAP

The measured k-space data \mathbf{y} was given by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \varepsilon, \quad (8)$$

where \mathbf{A} was the encoding matrix, \mathbf{x} was MR image, and $\boldsymbol{\varepsilon}$ was the noise. The matrix $\mathbf{A} = \mathbf{GFS}$, where $S: \mathbb{C}^N \rightarrow \mathbb{C}^{N \times \gamma}$ coil sensitivity maps, γ the number of coils, $F: \mathbb{C}^{N \times \gamma} \rightarrow \mathbb{C}^{N \times \gamma}$ the Fourier operator, and $G: \mathbb{C}^{N \times \gamma} \rightarrow \mathbb{C}^{M \times \gamma}$ the k-space sampling operator/mask. In this study, the additive noise $\boldsymbol{\varepsilon}$ was assumed to follow a Gaussian distribution with zero mean. Substituting Equation (7) into the log-likelihood for Equation (2) yielded

$$\hat{\mathbf{x}}_{\text{MAP}}(\mathbf{y}) = \arg \max_{\mathbf{x}} \log f(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}; \text{NET}(\mathbf{x}, \hat{\Theta})). \quad (9)$$

From the data model, the log-likelihood term for $f(\mathbf{y}|\mathbf{x})$ had less uncertainty, considering the MR imaging principles, for a given image \mathbf{x} , the probability for k-space, \mathbf{y} , that is, $f(\mathbf{y}|\mathbf{x})$ when $\mathbf{y} = \mathbf{Ax} + \boldsymbol{\varepsilon}$, was close to a constant with the uncertainty from noise that was irrelevant and additive to \mathbf{x} . Hence, Equation (9) could be rewritten as

$$\hat{\mathbf{x}}_{\text{MAP}}(\mathbf{y}) = \arg \max_{\mathbf{x}} \log p(\mathbf{x}; \text{NET}(\mathbf{x}, \hat{\Theta})) \quad \text{s.t.} \quad \mathbf{y} = \mathbf{Ax} + \boldsymbol{\varepsilon} \quad (10)$$

The equality constraint for data consistency was the result of eliminating the first log-likelihood term in Equation (9). The projected subgradient method was used to solve the equality constrained problem.^{17,18} In¹⁸ authors proposed a stochastic backpropagation method for computing gradients through random variables for deep generative models. In PixelCNN++, the stochastic backpropagation provided the subgradient $\nabla_{\mathbf{x}} \log g(\mathbf{x})$, where $g(\mathbf{x}) = p(\mathbf{x}; \text{NET}(\mathbf{x}, \hat{\Theta}))$, for minimizing the log-likelihood in Equation (10). We empirically found that the dropout (which applied to gradient update) was necessary, when using the gradient to update \mathbf{x} in Equation (10).¹⁹ To summarize, the MAP-based MRI reconstruction had the following iterative steps:

Repeat

Get the descent direction $\nabla_{\mathbf{x}^{(k)}} \log g(\mathbf{x}^{(k)})$

Pick up a step size $\alpha_k = 1/k$ or use a fixed step size

Update $\mathbf{z}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla_{\mathbf{x}^{(k)}} \log g(\mathbf{x}^{(k)})$

Projection $\mathbf{x}^{(k+1)} = \arg \min_{\mathbf{x} \in X} \frac{1}{2} \|\mathbf{x} - \mathbf{z}^{(k+1)}\|_2^2$

Until $\|\mathbf{Az} - \mathbf{y}\|_2^2 < \epsilon$ or $k > \text{maxIter}$

The projection of \mathbf{z} onto $\{\mathbf{x} | \mathbf{y} = \mathbf{Ax} + \boldsymbol{\varepsilon}\}$ was given by

$$\mathcal{P}(\mathbf{z}) = \mathbf{z} - \mathbf{A}^*(\mathbf{AA}^*)^{-1}(\mathbf{Az} - \mathbf{y}). \quad (11)$$

Therefore, the generative network as a prior model was incorporated into the reconstruction of \mathbf{x} through the Bayesian inference based on MAP.

3 | METHODS

3.1 | MRI data and preprocessing

Both knee and brain MRI data were used to test the reconstruction performance of the proposed method. The knee MRI data (multi-channel k-space data, 973 scans) were downloaded from fastMRI reconstruction database.²⁰ As such, NYU fastMRI investigators provided data but did not participate in analysis or writing of this report. A listing of NYU fastMRI investigators, subject to updates, can be found at: fastmri.med.nyu.edu. The primary goal of fastMRI is to test whether machine learning can aid in the reconstruction of medical images. The knee data had 2 contrast weightings: proton-density with and without fat suppression (PDFS and PD). Scan parameters included 15-channel knee coil and 2D multislice turbo spin-echo (TSE) acquisition, and other settings which could be found in Ref. [20].

For brain MRI, we collected 2D multislice T_1 weighted, T_2 weighted, T_2 weighted fluid-attenuated inversion recovery (FLAIR), and T_2^* weighted brain images from 16 healthy volunteers examined with clinical standard-of-care protocols, approved by the Institutional Review Board of The University of Hong Kong/Hospital Authority Hong Kong West Cluster. The consent was obtained from all volunteers. All brain data were acquired using our 3T MR scanner (Philips, Achieva), and an 8-channel brain RF coil. T_1 weighted, T_2 weighted, and T_2 weighted FLAIR images were all acquired with TSE readout. Meanwhile, T_2^* -weighted images were obtained using a gradient-echo sequence. Brain MRI parameters for 4 contrast weightings were listed in Supporting Information Table S1.

Training images were reconstructed from multichannel k-space data without undersampling. Then, these image datasets after coil combination were scaled to a magnitude range of $[-1, 1]$ and resized to an image size of 256×256 . The training of PixelCNN++ model required a considerable computational capacity when a large image size was used. In this study, the 128×128 was the largest size that our 4-graphics processing unit (GPU) server could handle. Hence, the original 256×256 images were resized into 128×128 low-resolution images by cropping in k-space for knee MRI. For brain MRI, we split each raw 256×256 image into four 128×128 image patches, before fed into the network for training. Real and imaginary parts of all 2D images were separated into 2 channels when inputted into the neural network. For the data partitioning, we first separated all multi-slice volumes into training and testing groups. Then we split the volume into slices (ie, 2D images). For knee MRI, 1541 images were used as the training dataset, and 170 images were used for testing. For brain MRI, 1300 images were used as the training dataset, and 300 images were used for testing. For data analysis, root-mean-square error (RMSE in percent), peak

signal-to-noise ratio (PSNR, in dB) and structural similarity index (SSIM, in percent) were used to quantify the image accuracy.

3.2 | Deep neural network

As illustrated in Figure 1, when predicting the current pixel, that is, yellow square in Figure 1 or $x_{i,j}$, pixels to the left and above were used as the inputs for the estimation, that is, green squares or $C_{i,j} = \{x_{i-1,j}, x_{i-2,j}, \dots, x_{1,1}\}$. In,²¹ the convolution stream was split into 2 network stacks: one conditioned on the “current row so far,” that is, $\{x_{i-1,j}, x_{i-2,j} \dots, x_{1,j}\}$, and another conditioned on all rows above, that is, $\{x_{n,j-1}, x_{n-1,j-1} \dots, x_{1,1}\}$, creating upstream and downstream as shown in Supporting Information Figure S1. The upstream and downstream first pass through 2 blocks that had “padding” and “shifting” functions for removing blinding spots.²¹ Then paired residual blocks (Res-blocks) that contained 3 gated ResNets were applied. A gated ResNet had 3 convolutional layers with the middle layer as a gated layer.²¹ In the Res-block, each convolution layer had 100 filters with a kernel size of 3×3 . In between the first and second Res-blocks, as well as the second and third Res-blocks, the network contained subsampling operations, implemented using a 2×2 stride convolution. In between the fourth and fifth Res-blocks, as well as the fifth and sixth Res-blocks, the network had a transpose stride convolution, that is, 2×2 -upsampling. These subsampling and upsampling caused information loss; therefore, the network also employed short-cut connections in-between Res-blocks to recover information. The short-cut connections went from the layers in the first Res-block to the corresponding layers in the sixth Res-block, and similarly between Res-blocks 2 and 5, and Res-blocks 3 and 4.

The PixelCNN++ was modified from the code in <https://github.com/openai/pixel-cnn>. We implemented the reconstruction algorithm using Python, as explained in Equation (11) and Appendix. With the trained prior model, we implemented the iterative reconstruction algorithm for maximizing the posterior while enforcing the k-space data fidelity (as explained in Appendix and Figure 1). Only 2 deep learning models were trained and utilized, 1 for knee MRI with 2 contrast weightings, and another for brain MRI with 4 contrast weightings. These 2 models can support all experiments performed in this study with variable undersampling patterns, coil sensitivity maps, channel numbers, image sizes, and trajectory types. Our networks were trained in Tensorflow software, and on 4 NVIDIA RTX-2080Ti graphic cards. Other parameters were 500 epochs, batch size = 4, and Adam optimizer. It took about 4 days to train the network for knee dataset and 2 days for brain dataset under the abovementioned configuration.

3.3 | Parallel imaging and ℓ_1 or ℓ_2 regularization reconstruction

The generalized autocalibrating partial parallel acquisition (GRAPPA) reconstruction was performed with a block size of 4 and 20 central k-space lines as the autocalibration area.²² We simulated GRAPPA accelerations with undersampling factors from 2 to 4. The representative undersampling masks were shown in Supporting Information Figure S2. We chose l_1 -eigenvalue approach to autocalibrating parallel MRI (ESPIRiT),^{1,23,24} MODL,⁴ and VN⁷ as baseline methods for comparison. They were originated from analytical regularizations methods. The ℓ_1 -ESPIRiT exploited the sparsity of image, and the MODL was a deep learning method for compressed sensing reconstruction, trained via minimizing ℓ_2 or ℓ_1 reconstruction error. In the ℓ_1 -ESPIRiT reconstruction, we set the ℓ_1 regularization parameter to be 0.01. For the training of MODL, the setting followed Ref. [4] when training MODL to reconstruct the undersampled knee data. The only difference was the k-space mask in Ref. [4] was 2D undersampled, while in the current study, the 1D undersampling was applied. The central 20 k-space lines were sampled which account for 7% of the full k-space of one 256×256 image. The others in the outer region were picked randomly with certain undersampling rate for each single slice. VN method required fixing the sampling mask.⁷ When training VN, we repeated the process multiple times for using different undersampling masks in this study. The sensitivity maps feed into our method, MODL, and VN were estimated from the central 20×20 k-space region, using the ESPIRiT function in Berkeley advanced reconstruction toolbox (BART).²⁴

For the proposed method, MR images with 256×256 matrix size were reconstructed, using the prior model in Equation (7) that was trained by 128×128 images or image patches. During inference, the 256×256 image was split into four 128×128 patches for applying the prior model, as shown in Figure 1. After updating $s^{(k+1)}$, 4 patches for 1 image were concatenated to form an image with the original size of 256×256 , before it was projected onto $\{x|y = Ax + \varepsilon\}$ in Equation (11). The detailed algorithm was presented in the Appendix B. When using the patch-based approach, the random shift along the phase encoding direction was applied to eliminate stitching lines. The random shift would move the discontinuity from edge to the center of patches, which was then erased by the next gradient updates. In Supporting Information Video S1, the evolution of restored image was presented.

3.4 | Non-Cartesian k-space acquisition

In this experiment, spiral sampled k-space from the acquired T_2^* -weighted k-space data was simulated. For spiral k-space sampling, the ground truth came from Cartesian sampling. The method proposed in Ref. [25] was used to design the spiral

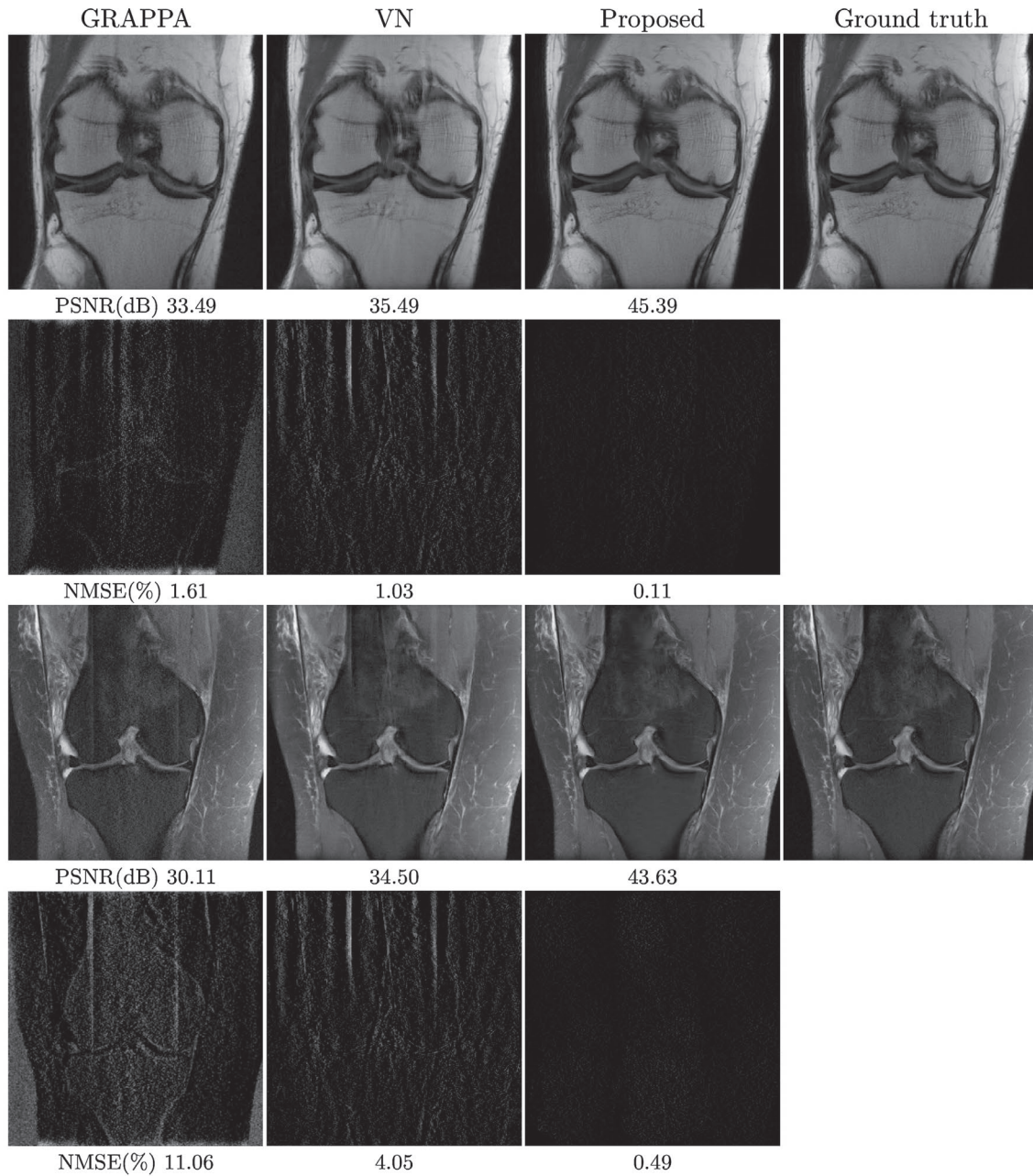


FIGURE 2 Comparisons on PD and PDFS contrasts using GRAPPA, VN, and the proposed reconstructions with $R = 3$ acceleration and 256×256 matrix size. The intensity of error maps was 5 times magnified. The proposed method effectively eliminated noise amplification and aliasing artifact in GRAPPA reconstruction

trajectory. The full k-space coverage required 24 spiral interleaves for the spatial resolution used in this study. Besides, the implementation of nonuniform fast Fourier transform was based on the method in Ref. [26]. For comparison, we used the conjugate gradient sensitivity encoding (CG SENSE), proposed in Ref. [27], as a baseline method. Our implementation exactly followed the original paper of CG SENSE, which applied sensitivity encoding spatially in an L2 minimization, that is, $\min \|Ax - y\|_2$, where A the sensitivity encoding and Fourier transform, x the image and y the undersampled k-space, CG was used as the solver. In the iterative CG reconstruction,

forward and backward Fourier transforms were performed with NUFFT. The NUFFT function was compiled using the code from Jeffrey A. Fessler's Lab (<https://web.eecs.umich.edu/fessler/>), which used min-max interpolation.

3.5 | Prospective experiment in vivo

To further validate the feasibility of proposed method, we implemented the prospective k-space undersampling for GRAPPA and compressed sensing in a rapid

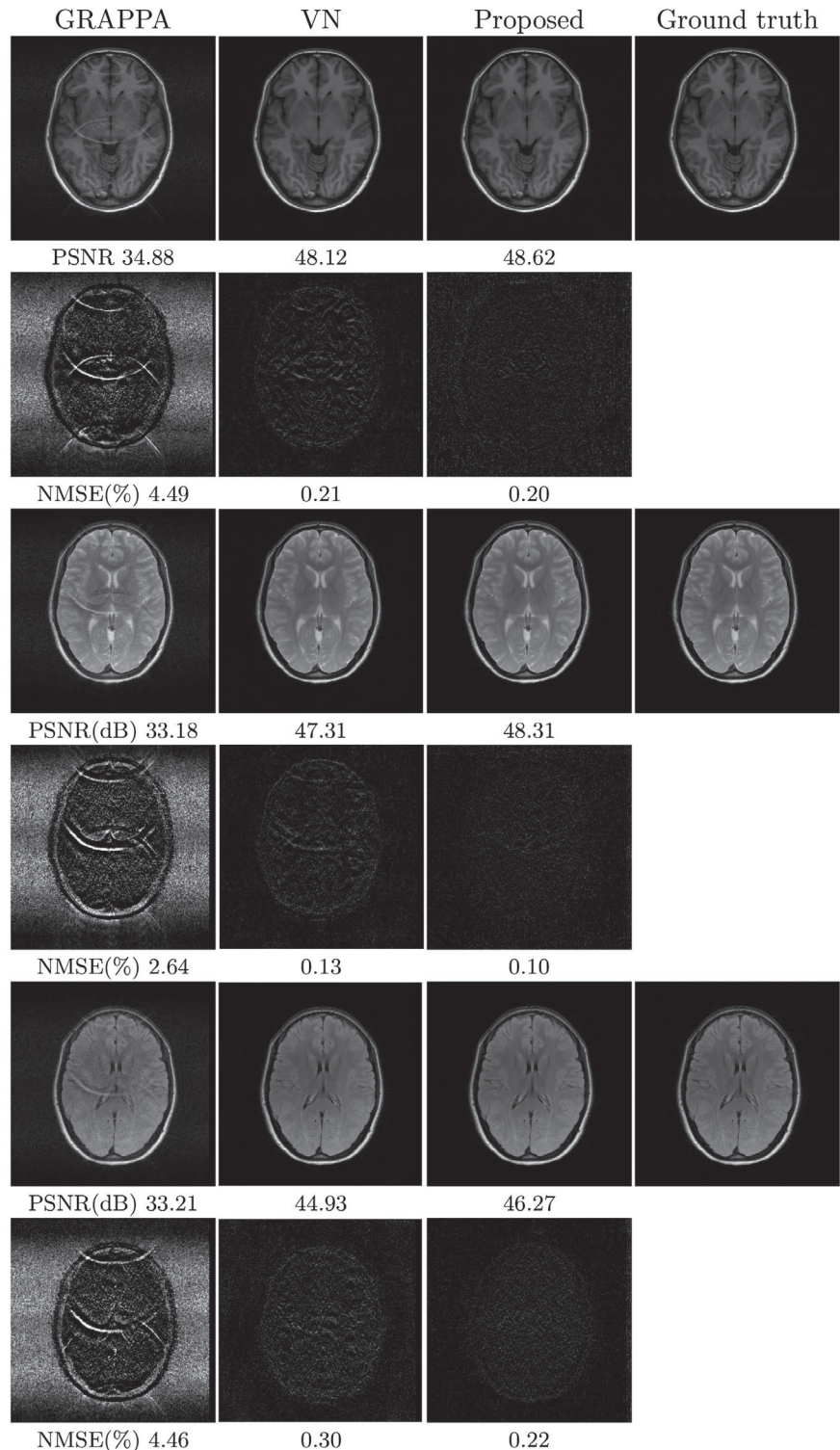
gradient-echo sequence with TE/TR = 16/770 ms. The image size was 256×256 , and the resolution was $0.9 \times 0.9 \text{ mm}^2$. For GRAPPA, the acquisition was accelerated by a factor of 3, and 20 center lines were sampled. For compressed sensing, only 15% phase-encoding lines were acquired, and the 20 center lines were kept. Then, acquired k-space data were normalized by the maximum magnitude of zero-filled and reconstructed images. Finally, the images reconstructed via GRAPPA, ℓ_1 -ESPIRiT, our method, and VN were compared.

4 | RESULTS

4.1 | Parallel imaging

Figures 2 and 3 show the comparison of knee and brain MRI reconstructed using GRAPPA, VN, and the proposed method. The proposed method had an improved performance in recovering brain and knee image details and reducing the aliasing artifacts, compared with GRAPPA and

FIGURE 3 Comparisons on T_1 , T_2 , and FLAIR- T_2 weighted image reconstruction, using parallel imaging, VN, and the proposed reconstruction with $R = 3$ acceleration and 256×256 matrix size. The intensity of error maps was 15 times magnified. The proposed method effectively eliminated the noise amplification in GRAPPA reconstruction



VN. As expected, parallel imaging amplified the noise in the low coil sensitivity regions and along the undersampled dimension. In contrast, error maps demonstrated in Figures 2 and 3 showed that both the proposed method and VN effectively eliminated the noise amplification and the aliasing artifacts. In the row (a) of Figure 4, the proposed method was demonstrated in prospective accelerations in parallel imaging and compressed sensing scenarios, confirmed the results in retrospective experiments. The performances of VN and the proposed methods were largely similar; however, the proposed method showed slight better preservation of boundaries between gray matter and white matter. More importantly, the proposed method uniquely supported different undersampling masks without the need for retraining the deep learning model. Table 1 presents the comparison of GRAPPA, VN, and the proposed method for knee ($N = 170$) and brain ($N = 300$) MRI testing images. With the increase of the undersampling factor, the PSNR of the proposed method decreased less, compared with that of GRAPPA. In addition, with acceleration factor $R = 2$ in brain MRI, the proposed method showed 8 dB more improvement in the PSNR than GRAPPA. In Supporting Information Video S2, the brain images of T_2^* weighting reconstructed with 3-fold prospective acceleration k-space data from a volunteer were presented.

4.2 | Compressed sensing reconstruction

In Figures 5 and 6, the ℓ_1 -ESPIRiT and VN had caused apparent blurring in the reconstructed images for both knee and brain MRI data. Both the ℓ_1 -ESPIRiT and MODL methods caused residual aliasing artifacts. Meanwhile, the proposed reconstruction recovered most anatomical structures and sharp boundaries in knee and brain MR images, compared with those from ℓ_1 -ESPIRiT, MODL and VN reconstructions, as shown on error maps in Figures 5 and 6. Besides, the row (b) of Figure 4 shows that the proposed method had slight better edge preservation and artifact suppression compared with VN. This observation was consistent with our prospective results. Table 1 summarized reconstruction results using ℓ_1 -ESPIRiT, MODL, VN and the proposed method. The proposed method generally showed more than 5 dB PSNR improvement compared with ℓ_1 -ESPIRiT, MODL and VN. In Supporting Information Video S3, the brain images of T_2^* weighting reconstructed with 22% prospective k-space data from a volunteer were presented.

4.3 | Preliminary result in non-Cartesian MRI reconstruction

In this study, we used a T_2^* weighted gradient-echo images to simulate the spiral k-space data with 4-fold acceleration. The

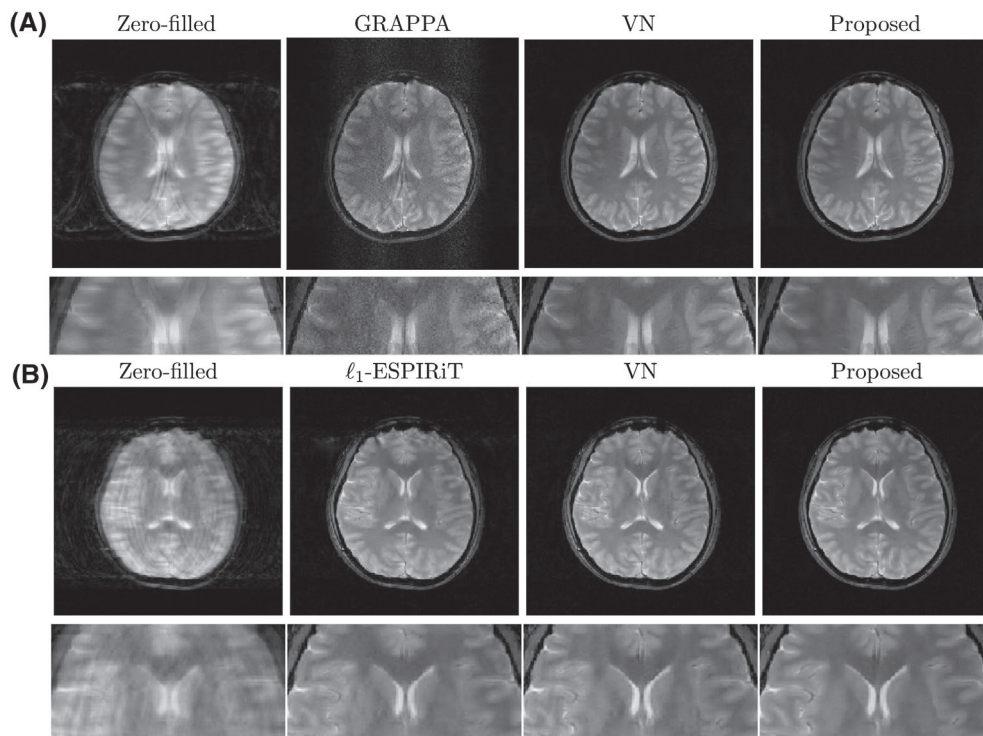


FIGURE 4 Comparison of different methods with prospective accelerations. The row (A) shows the images reconstructed from k-space data sampled in GRAPPA mask ($R = 3$). The row (B) shows the images reconstructed from k-space data sampled in CS mask ($R = 22\%$)

TABLE 1 PSNR/SSIM comparisons (in dB/%, mean \pm standard deviation) for parallel imaging, compressive sensing, and the proposed method on knee and brain MRI

<i>R</i> factor	Organ	GRAPPA	-	VN	Proposed
<i>Parallel imaging</i>					
PSNR					
<i>R</i> = 2	knee	35.24 \pm 4.53	-	38.23 \pm 2.35	57.31 \pm 3.84
<i>R</i> = 3	knee	31.48 \pm 3.21	-	35.53 \pm 1.75	45.64 \pm 2.20
<i>R</i> = 4	knee	26.39 \pm 1.67	-	33.52 \pm 1.37	39.51 \pm 2.76
<i>R</i> = 2	brain	38.05 \pm 4.73	-	46.19 \pm 3.03	51.18 \pm 2.77
<i>R</i> = 3	brain	31.60 \pm 3.33	-	43.67 \pm 2.74	45.43 \pm 2.52
<i>R</i> = 4	brain	28.27 \pm 2.84	-	40.51 \pm 2.39	43.58 \pm 2.66
SSIM					
<i>R</i> = 2	knee	87.85 \pm 8.01	-	93.99 \pm 3.31	99.65 \pm 0.32
<i>R</i> = 3	knee	76.82 \pm 10.56	-	90.31 \pm 4.18	97.95 \pm 1.18
<i>R</i> = 4	knee	62.34 \pm 6.47	-	86.39 \pm 4.33	93.41 \pm 3.03
<i>R</i> = 2	brain	83.48 \pm 3.00	-	97.90 \pm 1.37	98.10 \pm 0.19
<i>R</i> = 3	brain	69.24 \pm 5.53	-	96.74 \pm 2.03	96.85 \pm 4.36
<i>R</i> = 4	brain	58.32 \pm 6.56	-	95.31 \pm 2.45	95.54 \pm 5.10
Sampling rate	Organ	ℓ_1 -ESPIRiT	MODL	VN	Proposed
<i>Compressive sensing</i>					
PSNR					
15% + 7%	knee	30.81 \pm 2.01	26.78 \pm 3.19	31.87 \pm 0.92	35.34 \pm 3.13
20% + 7%	knee	31.81 \pm 2.23	31.30 \pm 2.93	30.72 \pm 0.62	37.14 \pm 3.23
15% + 7%	brain	32.73 \pm 3.46	29.06 \pm 3.24	33.95 \pm 2.25	39.80 \pm 2.73
20% + 7%	brain	34.51 \pm 3.94	30.70 \pm 3.25	34.73 \pm 2.26	41.18 \pm 2.70
SSIM					
15% + 7%	knee	78.97 \pm 8.38	72.40 \pm 9.29	82.50 \pm 4.55	88.12 \pm 5.66
20% + 7%	knee	81.74 \pm 8.30	86.50 \pm 3.77	81.66 \pm 3.84	91.27 \pm 4.23
15% + 7%	brain	87.58 \pm 6.26	78.85 \pm 7.23	89.86 \pm 3.47	91.52 \pm 6.26
20% + 7%	brain	90.08 \pm 6.03	82.67 \pm 7.43	91.24 \pm 3.54	92.48 \pm 5.66

reconstructed images from the CG SENSE and the proposed method were compared in Figure 7. The proposed method showed apparent improvement regarding the aliasing artifact reduction and the preservation of T_2^* contrast between gray matter and white matter. The proposed method also showed a slight denoising effect on the reconstructed image compared with the ground truth. Noted that the same deep learning model used in the previous Cartesian k-space reconstruction experiments in Figures 3 and 5 was applied to spiral reconstruction, without the need of re-training the deep learning model.

4.4 | Phase maps

Figure 8 shows the phase maps from the proposed accelerated reconstruction with 4-fold GRAPPA. Noted that the same deep learning model used in the previous brain experiments was applied to this experiment, with phase information preserved in all reconstructed images. The proposed

deep learning method also showed slight de-noising effect on phase maps, while still preserved the major phase contrast even with high acceleration.

5 | DISCUSSION

The proposed method can reliably and consistently recover the nearly aliased-free images with relatively high acceleration factors. Meanwhile, as expected, the increase of image smoothing with high acceleration factors was noticed, reflecting the loss of intrinsic resolution. The estimated image from the maximum of the posterior can not guarantee the full recovery of the image details, that is, PSNR > 40 dB/SSIM > 95% for a full recovery. However, at modest acceleration, the reconstruction from a maximum of posterior showed the successful reconstruction of the detailed anatomical structures, such as vessels, cartilage, and membranes in-between muscle bundles.

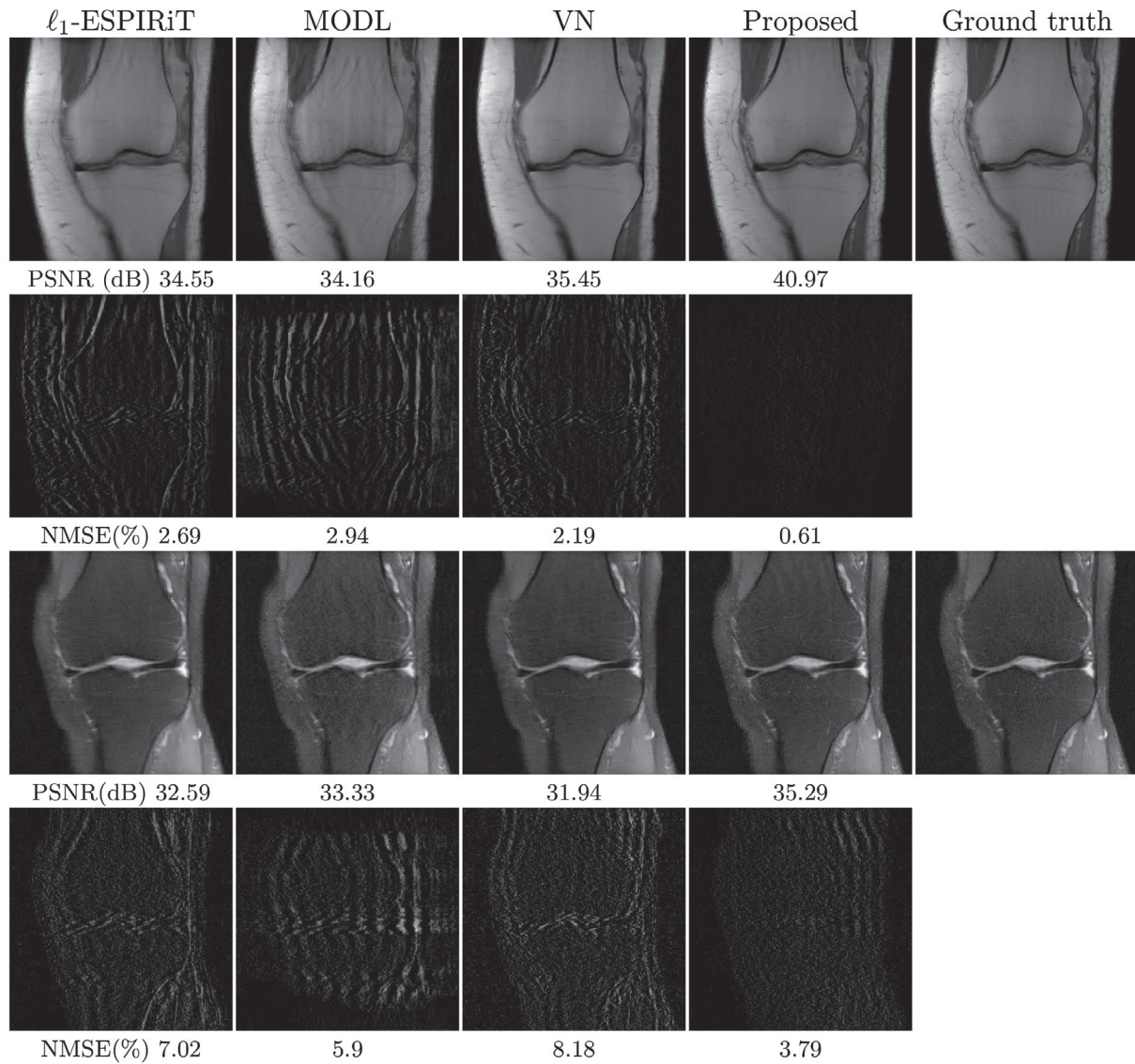


FIGURE 5 Comparison of different methods on PD and PDFS contrasts, using 27% 1D undersampled k-space and 256×256 matrix size. The intensity of error maps was 5 times magnified. The proposed method substantially reduced the aliasing artifact and preserved image details in compressed sensing reconstruction

In this study, the results demonstrated the successful reconstruction of high-resolution image (ie, 256×256 matrix) with low-resolution prior (ie, trained with 128×128 matrix), confirming the feasibility of reconstructing images of different sizes without the need for retraining the prior model. The prior model was trained by 128×128 images; it was still valid and applicable for the reconstruction of a high-resolution image. The proposed methods provided more than 8 dB improvement over the conventional GRAPPA reconstruction at the 4-fold acceleration in knee MRI. Besides, in contrast to other deep learning-based methods, which focused on the ℓ_2 loss, the likelihood that was conditioned by pixel-wise dependencies of the whole image showed an improved representation capacity, leading to a higher reconstruction accuracy. The applicability of the proposed method in the

patch-based reconstruction also suggested its high representation capacity and flexibility. Even when the inputs were image patches, the prior model could still recover the whole image.

The projected subgradient approach to solving Equation (10) was computationally inexpensive but converged slowly, as shown in Figure 9. For a random initialization, the algorithm needed about 500 iterations to converge with a fixed step size. Meanwhile, we noticed that if the zero-filled-reconstructed image was used for initialization, the number of iterations required could be reduced to 100. Besides, the decay of residual norm stopped earlier than that of the log-likelihood, that is, when the residual norm stopped decaying, the likelihood can still penalize the error. This evidence indicated that using the residual norm as the ℓ_2 fidelity alone was sub-optimal, and the deep learning-based statistical regularization

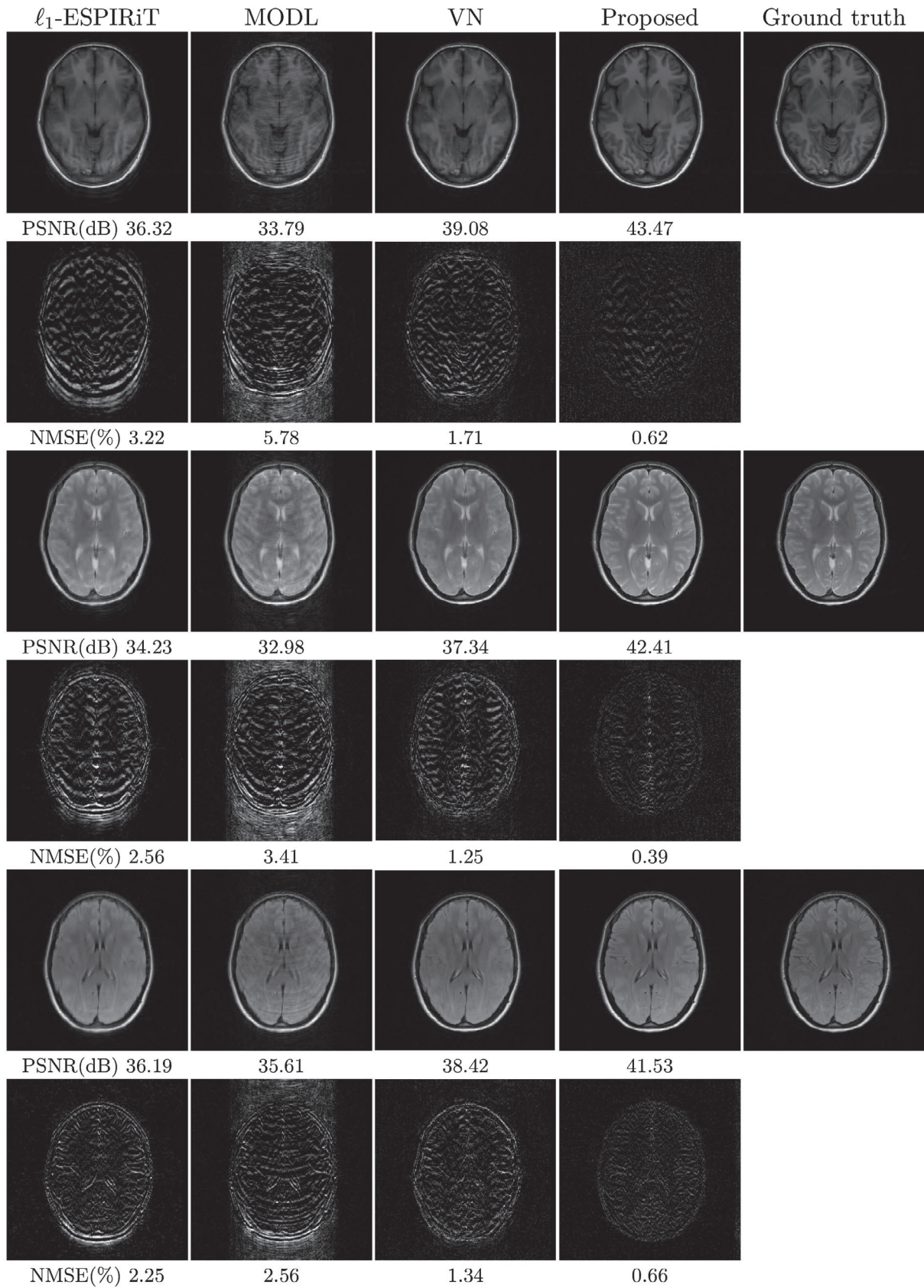


FIGURE 6 Comparison of compressed sensing and deep learning approaches for T₁, T₂, and FLAIR weighted image reconstructions, using 22% 1D undersampled k-space and 256 × 256 matrix size. The intensity of error maps was 10 times magnified. The proposed method substantially reduced the aliasing artifact and preserved image details in compressed sensing reconstruction

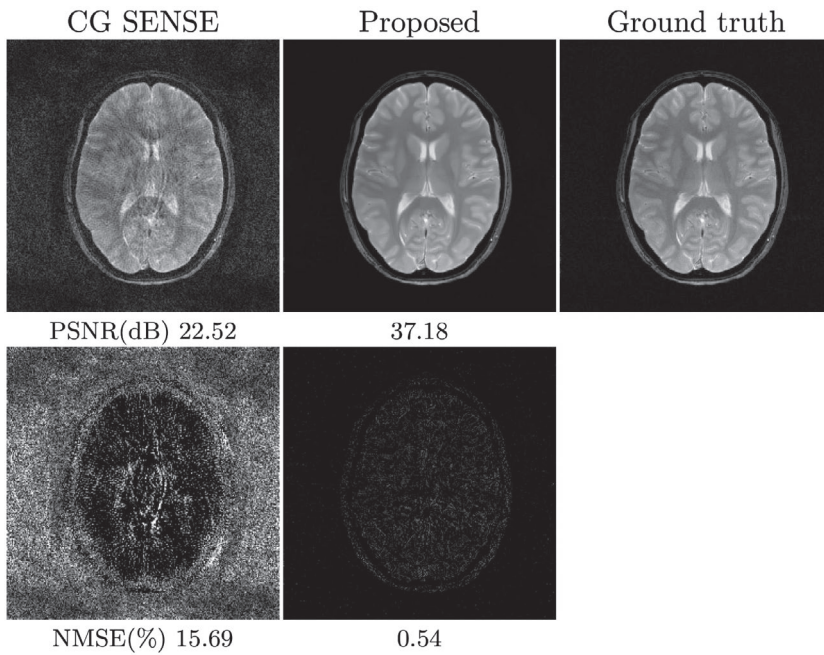


FIGURE 7 Comparison of the CG SENSE and proposed reconstruction for simulated spiral k-space with 4-fold acceleration (ie, 6 out of 24 spiral interleaves), acquired by T_2^* weighted gradient echo sequence. The intensity of error maps was 5 times magnified. The proposed method substantially reduced the aliasing artifact in spiral reconstruction. Noted that the same deep learning model used in the previous Cartesian k-space reconstruction was applied to spiral reconstruction, without the need of re-training the deep learning model

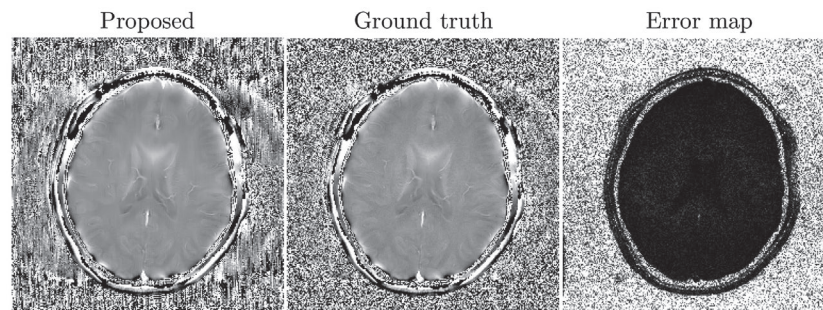


FIGURE 8 Phase maps from the ground truth and the proposed accelerated reconstruction with $R = 4$ and GRAPPA type of 1D undersampling. The raw images were acquired by T_2^* weighted gradient echo sequence. The proposed deep learning method preserved the major phase contrast even with high acceleration, $R = 4$. The error phase was 10 times magnified

can lead to a better reconstruction result compared with the ℓ_2 fidelity. Deep learning-based statistical regularization in the proposed method outperformed other conventional regularizations trained by image-level ℓ_2 loss. ℓ_2 loss did not give an explicit description of the relationship amid all pixels in the image, while the likelihood used in conjunction with the proposed image prior model was conditioned by the pixel-wise relationship and demonstrated superior performance compared with the conventional methods, under the current experimental setting.

Furthermore, the demonstrated image prior can be extended to a more elaborated form with clinical information, such as organs and contrast types, as the model inputs. For example, one could input the image prior with labels such as brain or knee. Then hypothetically, the image prior can be designed as a conditional probability for the given image label. In other words, the posterior would be dependent on both the k-space data and image labels. Moreover, the MR pulse sequence parameters could serve as image labels for the

prior, such as echo time and repetition time. In short, the prior model can be used to describe clinical information or acquisition parameters. This setting opens up a future direction on a more elaborated image prior, incorporating clinical information and MR sequence parameters, for more intelligent image representation and pattern detection.

In this study, the generative network solely served as an image prior model, in contrast to how neural network was used in other deep learning-based reconstructions.³⁻⁶ Specifically, in previous studies,³⁻⁶ embedding k-space fidelity term into the network made the algorithm inflexible because image prior and undersampling artifacts were mixed during the training. The proposed method used the standard analytical term for fidelity enforcement; therefore, its flexibility was comparable to the traditional optimization algorithm, such as ℓ_1 regularization. Due to unavoidable changes of the encoding scheme, for example, the image size and the RF coils during MRI experiment in practice, it was essentially needed to separate the learned component (the image

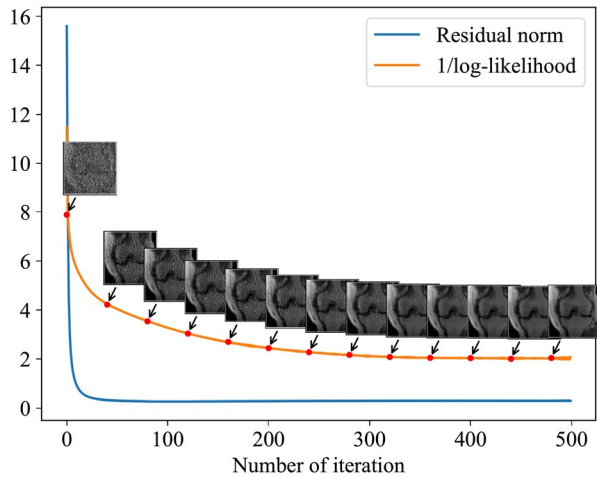


FIGURE 9 Convergence curves reflected stabilities of iterative steps: (a) maximizing the posterior, which effectively minimized the log-likelihood of MRI prior model and (b) k-space fidelity enforcement, which reduced the residual norm on k-space fidelity. The 22% sampling rate and 1D undersampling scheme were used in this simulation. The residual norm was written as $\|y - Ax\|_2^2$ in Equation (11), and the reciprocal of log-likelihood for MRI prior model was given in Equation (7).

prior) from the encoding matrix used in the fidelity term in reconstruction. Besides, the proposed method showed the feasibility of incorporating the coil sensitivity information in the fidelity term, which enabled the changeable encoding scheme without the need of retraining the model.^{15,27} In summary, the separation of the image prior and the encoding matrix embedded in the fidelity term made the proposed method more flexible and generalizable compared with conventional deep learning approaches.

6 | CONCLUSION

In summary, this study presented the application of Bayesian inference in MR imaging reconstruction with the deep learning-based prior model. We demonstrated that the deep MRI prior model was a computationally tractable and effective tool for MR image reconstruction. The Bayesian inference significantly improved the reconstruction performance over that of conventional ℓ_1 sparsity prior in compressed sensing. More importantly, the proposed reconstruction framework was generalizable for most reconstruction scenarios.

ACKNOWLEDGMENT

This work is supported by The University of Hong Kong, Seed Fund for Basic Research.

ORCID

Peng Cao  <http://orcid.org/0000-0002-1116-1171>

REFERENCES

- Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med*. 2007;58:1182–1195.
- Ravishankar R, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans Med Imaging*. 2010;30:1028–1041.
- Yang Y, Sun J, Li H, Xu Z. Deep adm-net for compressive sensing MRI. *Adv Neural Inform Process Syst*. 2016;10–18.
- Aggarwal HK, Mani MP, Jacob M. Modl: model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging*. 2018;38:394–405.
- Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging*. 2017;37:491–503.
- Mardani M, Gong E, Cheng JY, et al. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Trans Med Imaging*. 2018;38:167–179.
- Hammernik K, Klatzer T, Kobler E, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. 2018;79:3055–3071.
- Rezende DJ, Mohamed S, Wierstra D. Stochastic backpropagation and approximate inference in deep generative models. arXiv preprint arXiv:1401.4082; 2014.
- Arridge S, Maass P, Öktem O, Schönlieb CB. Solving inverse problems using data-driven models. *Acta Numerica*. 2019;28:1–174.
- Goodfellow Ian, Bengio Yoshua, Courville Aaron. *Deep Learning* (Chapter 16). Cambridge, MAMIT Press; 2017:558.
- Tezcan KC, Baumgartner CF, Luechinger R, Pruessmann KP, Konukoglu E. MR image reconstruction using deep density priors. *IEEE Trans Med Imaging*. 2018;38:1633–1642.
- Rosca M, Lakshminarayanan B, Mohamed S. Distribution matching in variational inference. arXiv preprint arXiv:1802.06847, 2018.
- Salimans T, Karpathy A, Chen X, Kingma DP. Pixelcnn++: improving the pixelcnn with discretized logistic mixture likelihood and other modifications. arXiv preprint arXiv:1701.05517, 2017.
- van den Oord A, Kalchbrenner N, Kavukcuoglu K. Pixel recurrent neural networks. arXiv preprint arXiv:1601.06759, 2016.
- Fessler JA. Model-based image reconstruction for MRI. *IEEE Signal Process Mag*. 2010;27:81–89.
- Kingma DP, Salimans T, Jozefowicz R, Chen X, Sutskever I, Welling M. Improving variational inference with inverse autoregressive flow. arXiv preprint arXiv:1606.04934, 2016.
- Gregor K, Danihelka I, Mnih A, Blundell C, Wierstra D. Deep autoregressive networks. arXiv preprint arXiv:1310.8499, 2013.
- Boyd S, Xiao L, Mutapcic A. Subgradient methods. *Lecture Notes of EE392o*. Stanford University, Santa Clara. Autumn Quarter. 2003.
- Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of The 33rd International Conference on Machine Learning, New York, NY, USA. JMLR: W&CP vol 48.
- Zbontar J, Knoll F, Sriram A, et al. FASTMRI: an open dataset and benchmarks for accelerated MRI. arXiv preprint arXiv:1811.08839, 2018.
- Van den Oord A, Kalchbrenner N, Espeholt L, et al. Conditional image generation with pixelcnn decoders. In *Advances in Neural Information Processing Systems*; 2016:4790–4798.

22. Griswold MA, Jakob PM, Heidemann RM, et al. Generalized autocalibrating partially parallel acquisitions (grappa). *Magn Reson Med.* 2002;47:1202–1210.
23. Uecker M, Lai P, Murphy MJ, et al. Esprit—an eigenvalue approach to autocalibrating parallel MRI: where sense meets grappa. *Magn Reson Med.* 2014;71:990–1001.
24. Uecker M, Ong F, Tamir JJ, et al. Berkeley advanced reconstruction toolbox. In Proceedings of the 23rd Annual Meeting of ISMRM, Toronto, Canada, 2015. p. 2486.
25. Lustig M, Kim S-J, Pauly JM. A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories. *IEEE Trans Med Imaging.* 2008;27:866–873.
26. Fessler JA, Sutton BP. Nonuniform fast fourier transforms using min-max interpolation. *IEEE Trans Signal Process.* 2003;51:560–574.
27. Pruessmann KP, Weiger M, Börner P, Boesiger P. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magn Reson Med.* 2001;46:638–651.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

VIDEO S1 The evolution of image during the process of reconstruction

VIDEO S2 Reconstruction by the proposed method of a healthy volunteer for T_2^* -weighted brain images with under-sampling factor $R = 3$ (GRAPPA prospective experiment)

VIDEO S3 Reconstruction by the proposed method of a healthy volunteer for T_2^* -weighted brain images with 22% k-space (Compressive sensing prospective experiment)

TABLE S1 The scan parameters of different weightings used in brain MRI experiments

FIGURE S1 The diagram shows the PixelCNN++ network in,¹³ which was the prior model used in this study, that is, $g(\mathbf{x})$ in Equation (7). Each ResNet block (gray) has a gated-resnet components. The last layer is fully connected. The input of network was \mathbf{x} , outputs of network were parameters of mixture distribution $(\boldsymbol{\pi}, \boldsymbol{\mu}, s, \boldsymbol{\alpha})$, which were fed into the conditional probability model in Equation (5)

FIGURE S2 k-space masks used in the compressed sensing, parallel imaging, and deep learning reconstructions. Bright lines indicate the sampled frequency encoding lines in the 2D k-space, that is, the 1D undersamplings were simulated

How to cite this article: Luo G, Zhao N, Jiang W, Hui ES, Cao P. MRI reconstruction using deep Bayesian estimation. *Magn Reson Med.* 2020;84:2246–2261. <https://doi.org/10.1002/mrm.28274>

APPENDIX A DERIVATION FOR THE PROBABILITY OF LOGISTIC DISTRIBUTION IN EQUATION 4

The probability density function of logistic distribution is as follow

$$\begin{aligned} f(x; \boldsymbol{\mu}, s) &= \frac{e^{-(x-\boldsymbol{\mu})/s}}{s(1+e^{-(x-\boldsymbol{\mu})/s})^2} \\ &= \frac{1}{s(e^{(x-\boldsymbol{\mu})/(2s)} + e^{-(x-\boldsymbol{\mu})/(2s)})^2} \end{aligned}$$

The cumulative distribution function of logistic distribution is as follow

$$F(x; \boldsymbol{\mu}, s) = \int_{-\infty}^x f(x; \boldsymbol{\mu}, s) = \frac{1}{1+e^{-(x-\boldsymbol{\mu})/s}}$$

Therefore if we assume the ν follows the logistic distribution, the probability of ν in range from $x - d/2$ to $x + d/2$ is

$$\begin{aligned} P(\nu; \boldsymbol{\mu}, s) &= \frac{1}{1+e^{-(\nu+d/2-\boldsymbol{\mu})/s}} - \frac{1}{1+e^{-(\nu-d/2-\boldsymbol{\mu})/s}} \\ &= \sigma((\nu+d/2-\boldsymbol{\mu})/s) - \sigma((\nu-d/2-\boldsymbol{\mu})/s). \end{aligned}$$

Similarly, if we assume the ν follows the mixture logistic distribution

$$\nu \sim \sum_{k=1}^K \pi_k \text{logistic}(\boldsymbol{\mu}_k, s_k),$$

then, we have

$$P(\nu; \boldsymbol{\pi}, \boldsymbol{\mu}, s) = \sum_{k=1}^K \pi_k [\sigma((\nu+d/2-\boldsymbol{\mu}_k)/s_k) - \sigma((\nu-d/2-\boldsymbol{\mu}_k)/s_k)].$$

APPENDIX B

RECONSTRUCTION FOR THE VARIED IMAGE SIZE WITH DEEP PRIOR MODEL

Algorithm 1 Reconstruction algorithm with deep prior model

-
- Input:**
 \mathbf{y} - k-space data
 \mathbf{A} - encoding matrix
 λ - maximum iteration
- Output:**
 \mathbf{x} - the restored image
- 1: Give a random initial point $\mathbf{x}^{(0)}$, $k = 1$ ▷ Initialization
 - 2: **while** $\|g^{(k)}\|_2^2 > \varepsilon$ and $k < \lambda$ **do** ▷ Iteration
 - 3: Generate a random shifting offset d
 - 4: Shift $\mathbf{x}^{(k)}$ d pixels away from the center circularly
 - 5: Split $\mathbf{x}^{(k)}$ into pieces $\mathbf{s}^{(k)}$ for feeding to network
 - 6: Get subgradient $\nabla_{\mathbf{x}} \log(g(\mathbf{x}))$ at $\mathbf{x}^{(k)}$
 - 7: Pick a step size $\alpha_k = 1/k$ or use a fixed step size
 - 8: Update $\mathbf{s}^{(k+1)} = \mathbf{s}^{(k)} - \alpha_k \nabla_{\mathbf{s}^{(k)}} \log g(\mathbf{s}^{(k)})$
 - 9: Merge pieces $\mathbf{s}^{(k+1)}$ into $\mathbf{z}^{(k+1)}$ for projection
 - 10: Shift $\mathbf{z}^{(k+1)}$ $-d$ pixels away from the center circularly
 - 11: Projection $\mathbf{x}^{(k+1)} = \arg \min_{\mathbf{x} \in X} \frac{1}{2} \|\mathbf{x} - \mathbf{z}^{(k+1)}\|_2^2$
 - 12: $k = k + 1$
 - 13: **return** $\mathbf{x}^{(k+1)}$
-