MRI Reconstruction Using Deep Bayesian Estimation

Guanxiong Luo¹, Na Zhao², Wenhao Jiang¹, Edward S. Hui¹, and Peng Cao, *¹

¹Department of Diagnostic Radiology, The University of Hong Kong, Hong Kong
²Department of Statistics and Actuarial Science, The University of Hong Kong

February 18, 2022

Abstract

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, ℓ₁-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.

Conclusion: The Bayesian estimation significantly improved the reconstruction performance, compared with the conventional ℓ₁-sparsity prior in compressed sensing reconstruction tasks. More importantly, the proposed reconstruction framework can be generalized for most MRI reconstruction scenarios.

Keywords – Generative network, Bayesian estimation, Deep learning reconstruction, Compressed sensing, Parallel Imaging

1 Introduction

In compressed sensing MRI reconstruction, the commonly used analytical regularization such as ℓ₁ 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 ℓ₁ 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 generative [6] deep neural networks. Especially, the method proposed in [3] 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 embed-
ded in an unrolled gradient descent scheme [7]. These deep learning methods may show improved performance
in some predetermined acquisition settings or pre-trained imaging tasks. However, they also lack flexibility
when used with changes in MRI under-sampling scheme, the number of radio-frequency coils, and matrix size
or spatial resolution. Such restriction is caused by the embedding 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 repre-
sentation 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 exp-
licit 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 cal-
culated 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 like-
lihood 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 writ-
ten 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 estimator [9] for the high dimensional image reconstruction
problem tackled in this study. To summarize, the Bayesian estimation for MRI reconstruction had two 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 cap-
able 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 max-
imum likelihood of posterior with the image prior, i.e., applying the Bayesian estimation repeatedly. The re-
construction 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:

\[
    f(x|y) = \frac{f(y | x)g(x)}{f(y)} \propto f(y | x)g(x)
\]  

(1)

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

\[
\hat{x}_{\text{MAP}}(y) = \arg \max_x f(x \mid y) = \arg \max_x f(y \mid x) g(x)
\]

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

**MAP reconstruction**

\[
\tilde{x}_{\text{MAP}}(y) = \arg \max_x \log p(x; \text{NET}(x, \hat{\Theta})) + \log f(y \mid x)
\]

\[y = Ax + \varepsilon \]

\[
\Theta \text{ obtained by training}
\]

\[
\text{NET loss: } p(x; \pi, \mu, s, \alpha) = p(x^{(1)}) \prod_{i=2}^{n^2} p(x^{(i)} \mid x^{(1)}, \ldots, x^{(i-1)})
\]

\[
\text{The conditional mixture}
\]

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

### 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 three reasons for using such mixture distribution model and generative network approximation [13]: 1) hierarchical architectures allowed the network...
to capture complex structure in the image, 2) the mixture of conditional distribution characterized an intrinsic dependence among pixels and factorized the probability density function for an image over its pixels, 3) they discretized/instanced logistic mixture likelihood for using pixelCNN++ [13]. For MRI reconstruction, the number of image channels was changed from three (i.e., RGB channels for color image) to two (i.e., real and imaginary parts for MR image). For each image pixel, the continuous variable \( \nu \) 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 \( \nu \) was a mixture of the logistic distribution, given by

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

Here, \( \pi_i \) was the mixture indicator, \( \mu_i \) 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 \( \nu \) of the pixel was computed as [13]

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

where \( \sigma \) was the logistic sigmoid function, \( d \) was the smallest discretized interval for \( \nu \). 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 Equation 3 to Equation 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 by [13]

\[
p(x_{i,j} | C_{i,j}) = p(\text{Re}(x_{i,j}), \text{Im}(x_{i,j}) | C_{i,j}) = p(\text{Re}(x_{i,j}); \pi(C_{i,j}), \mu_{\text{Re}}(C_{i,j}), \sigma_{\text{Re}}(C_{i,j})) \times p(\text{Im}(x_{i,j}); \pi(C_{i,j}), \mu_{\text{Im}}(C_{i,j}, \text{Re}(x_{i,j})), \sigma_{\text{Im}}(C_{i,j}))
\]

\[
\hat{\mu}_{\text{Re}}(C_{i,j}, \text{Re}(x_{i,j})) = \mu_{\text{Im}}(C_{i,j}) + \alpha(C_{i,j}) \text{Re}(x_{i,j}),
\]

where the \( C_{i,j} = \{x_{i-1,j}, x_{i-2,j}, ..., x_{1,j}\} \) denoted the context information which was comprised of the mixture indicator and the previous pixels as showed in Figure 1, \( \alpha \) was the coefficient related to mixture indicator and previous pixels, \( \mu_{\text{Re}} \) was the \( \mu \) of the real channel of \( C_{i,j} \), and \( \mu_{\text{Im}}, \sigma_{\text{Re}}, \sigma_{\text{Im}} \) 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, i.e., \( x_1^{(1)} = x_{1,1} \) 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 between two channels. The \( n \times n \) image could be considered as an vectorized image \( x = (x^{(1)}, ..., x^{(n^2)}) \) by stacking pixels from left to right and up to bottom of one another, i.e., \( x^{(1)} = x_{1,1}, x^{(2)} = x_{2,1}, ..., x^{(n^2)} = x_{n,n} \), as illustrated in Figure 1. The joint distribution of the image vector could be expressed as following [13]:

\[
p(x; \pi, \mu, s, \alpha) = p(x^{(1)}) \prod_{i=2}^{n^2} p(x^{(i)} | x^{(1)}, ..., x^{(i-1)}). \tag{5}
\]

\( \pi, \mu, s, \alpha \) 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 \( \text{NET}(x, \Theta) \) was trained by maximizing the likelihood in Equation 5, as the training loss was given by

\[
\hat{\Theta} = \arg \max_{\Theta} p(x; \text{NET}(x, \Theta)), \tag{6}
\]

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

\[
g(x) = p(x; \text{NET}(x, \hat{\Theta})). \tag{7}
\]

For example, if the complex images \( x \) for training had a size of \((128,128)\), and the number of mixture distribution \( K = 10 \), then \( \pi, \mu, s, \alpha \) 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 \( x \) was defined in Eqs. 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 $y$ was given by

$$ y = A x + \varepsilon, $$

(8)

where $A$ was the encoding matrix, $x$ was MR image, and $\varepsilon$ was the noise. The matrix $A = GFS$, where $S : \mathbb{C}^N \rightarrow \mathbb{C}^{N \times \gamma}$ coil sensitivity maps, $\gamma$ 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 $\varepsilon$ was assumed to follow a Gaussian distribution with zero mean. Substituting Equation 7 into the log-likelihood for Equation 2 yielded

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

(9)

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

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

(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]. In [18], authors proposed a stochastic backpropagation method for computing gradients through random variables for deep generative models. In PixelCNN++, the stochastic backpropagation provided the subgradient $\nabla_x \log g(x)$, where $g(x) = p(x; \text{NET}(x, \hat{\Theta}))$, 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 $x$ in Equation 10 [19]. To summarize, the MAP-based MRI reconstruction had the following iterative steps:

Repeat

- Get the descent direction $\nabla_{x^{(k)}} \log g(x^{(k)})$
- Pick up a step size $\alpha_k = 1/k$ or use a fixed step size
- Update $x^{(k+1)} = x^{(k)} - \alpha_k \nabla_{x^{(k)}} \log g(x^{(k)})$
- Projection $x^{(k+1)} = \arg\min_{x \in X} \frac{1}{2} \|x - x^{(k+1)}\|^2$

Until $\|A z - y\|^2 < \epsilon$ or $k > \text{maxIter}$

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

$$ P(z) = z - A^* (A A^*)^{-1} (Az - y). $$

(11)

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

3 Methods

3.1 MRI data and pre-processing

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 two contrast weightings: proton-density with and without fat suppression (PDWS and PD). Scan parameters included 15-channel knee coil and 2D multi-slice turbo spin-echo (TSE) acquisition, and other settings which could be found in Ref. [20].

For brain MRI, we collected 2D multi-slice T1 weighted, T2 weighted, T2 weighted fluid-attenuated inversion recovery (FLAIR), and T2* 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 eight-channel brain RF coil. T1 weighted, T2 weighted, and T2 weighted FLAIR images were all acquired with TSE readout. Meanwhile, T2*-weighted images were obtained using a gradient-echo sequence. Brain MRI parameters for four contrast weightings were listed in Supporting Information Table S1.

Training images were reconstructed from multi-channel 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 $256 \times 256$. The training of PixelCNN++ model required a considerable computational capacity when a large image size was used. In this study, the $128 \times 128$ was the largest size that our 4-graphics processing unit (GPU) server could handle. Hence, the original $256 \times 256$ images were resized into $128 \times 128$ low-resolution images by cropping in k-space for knee MRI. For brain MRI, we split each raw $256 \times 256$ image into four $128 \times 128$ image patches, before fed into the network for training. Real and imaginary parts of all 2D images were separated into two 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 (i.e., 2D images). For knee MRI, 15541 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 %), peak signal-to-noise ratio (PSNR, in dB) and structural similarity index (SSIM, in %) were used to quantify the image accuracy.

### 3.2 Deep neural network

As illustrated in Figure 1, when predicting the current pixel, i.e., yellow square in Figure 1 or $x_{i,j}$, pixels to the left and above were used as the inputs for the estimation, i.e., green squares or $C_{i,j} = \{x_{i-1,j}, x_{i-2,j}, \ldots, x_{1,j}\}$. In [21], the convolution stream was split into two network stacks: one conditioned on the “current row so far”, i.e., $\{x_{i-1,j}, x_{i-2,j}, \ldots, x_{1,j}\}$, and another conditioned on all rows above, i.e., $\{x_{n,j-1}, x_{n-1,j-1}, \ldots, x_{1,1}\}$, creating up-stream and downstream as shown in Supporting Information Figure S1. The upstream and downstream first pass through two blocks that had “padding” and “shifting” functions for removing blinding spots [21]. Then paired residual blocks (Res-blocks) that contained three gated ResNets were applied. A gated ResNet had three 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 $3 \times 3$. 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 $2 \times 2$ 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, i.e., $2 \times 2$-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 two and five, and Res-blocks three and four.

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 two deep learning models were trained and utilized, one for knee MRI with two contrast weightings, and another for brain MRI with four contrast weightings. These two 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 four NVIDIA RTX-2080Ti graphic cards. Other parameters were 500 epochs, batch size $= 4$, and Adam optimizer. It took about four days to train the network for knee dataset and two days for brain dataset under the above-mentioned configuration.

### 3.3 Parallel imaging and $\ell_1$ or $\ell_2$ 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 auto-calibration 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 VN [7] as baseline methods for comparison. They were originated from analytical regularizations methods. The $\ell_1$-ESPIRiT exploited the sparsity of image, and the MODL was a deep learning method for compressed sensing reconstruction, trained via minimizing $\ell_2$ or $\ell_1$ reconstruction error. In the $\ell_1$-ESPIRiT reconstruction, we set the $\ell_1$ 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 $\times$ 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 $\times$ 20 k-space region, using the ESPIRiT function in Berkeley advanced reconstruction toolbox (BART) [24].

For the proposed method, MR images with 256 $\times$ 256 matrix size were reconstructed, using the prior model in Equation 7 that was trained by 128 $\times$ 128 images or image patches. During inference, the 256 $\times$ 256 image was split into four 128 $\times$ 128 patches for applying the prior model, as shown in Figure 1. After updating $s^{(k+1)}$, four patches for one image were concatenated to form an image with the original size of 256 $\times$ 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 T2$^*$-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 trajectory. The full k-space coverage required 24 spiral interleaves for the spatial resolution used in this study. Besides, the implementation of non-uniform 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, i.e., Ref [26].

The full k-space coverage required 24 spiral interleaves for the spatial resolution used in this study. The sensitivity maps feed into our method, MODL, and VN were estimated from the central 20 $\times$ 20 k-space region, using the ESPIRiT function in Berkeley advanced reconstruction toolbox (BART) [24].

For the proposed method, MR images with 256 $\times$ 256 matrix size were reconstructed, using the prior model in Equation 7 that was trained by 128 $\times$ 128 images or image patches. During inference, the 256 $\times$ 256 image was split into four 128 $\times$ 128 patches for applying the prior model, as shown in Figure 1. After updating $s^{(k+1)}$, four patches for one image were concatenated to form an image with the original size of 256 $\times$ 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 T2$^*$-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 trajectory. The full k-space coverage required 24 spiral interleaves for the spatial resolution used in this study. Besides, the implementation of non-uniform 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, i.e., $\min||Ax - y||_2$, 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 $\times$ 256, and the resolution was 0.9 $\times$ 0.9 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, $\ell_1$-ESPIRiT, our method, and VN were compared.

4 Results

4.1 Parallel imaging

Figure 2 and Figure 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 VN. As expected, parallel imaging amplified the noise in the low coil sensitivity regions and along the undersampled dimension. On the other hand, error maps demonstrated in Figure 2 and Figure 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 T2\(^*\) weighting reconstructed with 3-fold prospective acceleration k-space data from a volunteer were presented.

![Image of PD and PDFS contrasts using GRAPPA, VN, and the proposed reconstructions.](image)

### Table 1: Comparisons on PD and PDFS contrasts using GRAPPA, VN, and the proposed reconstructions with R=3 acceleration and 256 × 256 matrix size.

| Method   | PSNR (dB) | NMSE (%) |
|----------|-----------|----------|
| GRAPPA   | 33.49     | 1.61     |
| VN       | 35.49     | 1.03     |
| Proposed | 45.39     | 0.11     |
| Ground truth | 45.39     | 0.11     |

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 five times magnified. The proposed method effectively eliminated noise amplification and aliasing artifact in GRAPPA reconstruction.

#### 4.2 Compressed sensing reconstruction

In Figure 5 and Figure 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 Figure 5 and Figure 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. Tables 1 summarized reconstruction results using $\ell_1$-ESPIRiT, MODL, VN and the proposed method. The proposed method generally showed more than 5 dB PSNR improvement compared with $\ell_1$-ESPIRiT, MODL and VN. In Supporting Information Video S3, the brain images of T2* weighting reconstructed with 22% prospective k-space data from a volunteer were presented.

| Method   | PSNR | NMSE(%) | PSNR | NMSE(%) | PSNR | NMSE(%) |
|----------|------|---------|------|---------|------|---------|
| GRAPPA   | 34.88| 4.49    | 48.62| 0.20    | 48.31| 0.10    |
| VN       | 48.12| 0.21    | 47.31| 0.13    | 46.93| 0.30    |
| Proposed | 48.62| 0.20    | 48.31| 0.10    | 46.27| 0.22    |
| Ground truth | 48.62| 0.20    | 48.31| 0.10    | 46.27| 0.22    |

Figure 3. Comparisons on T1, T2, and FLAIR-T2 weighted image reconstruction, using parallel imaging, VN, and the proposed reconstruction with R=3 acceleration and $256 \times 256$ matrix size. The intensity of error maps was 15 times magnified. The proposed method effectively eliminated the noise amplification in GRAPPA reconstruction.
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 ± standard deviation) for parallel imaging, compressive sensing and the proposed method on knee and brain MRI.

| Parallel imaging          | R factor | Organ | GRAPPA     | VN        | Proposed       |
|---------------------------|----------|-------|------------|-----------|----------------|
| **PSNR**                  |          |       |            |           |                |
| R=2 knee                  |          |       | 35.24±4.53 | -         | 38.23±2.35     |
| R=3 knee                  |          |       | 31.48±3.21 | -         | 35.53±1.75     |
| R=4 knee                  |          |       | 26.39±1.67 | -         | 33.52±1.37     |
| R=2 brain                 |          |       | 38.05±4.73 | -         | 46.19±3.03     |
| R=3 brain                 |          |       | 31.60±3.33 | -         | 43.67±2.74     |
| R=4 brain                 |          |       | 28.27±2.84 | -         | 40.51±2.39     |
| **SSIM**                  |          |       |            |           |                |
| R=2 knee                  |          |       | 87.85±8.01 | -         | 93.99±3.31     |
| R=3 knee                  |          |       | 76.82±10.56 | -       | 90.31±4.18     |
| R=4 knee                  |          |       | 62.34±6.47 | -         | 86.39±4.33     |
| R=2 brain                 |          |       | 83.48±3.00 | -         | 97.90±1.37     |
| R=3 brain                 |          |       | 69.24±5.53 | -         | 96.74±2.03     |
| R=4 brain                 |          |       | 58.32±6.56 | -         | 95.31±2.45     |

| Compressive sensing       | Sampling rate | Organ | ℓ₁-ESPIRiT | MODL | VN        | Proposed       |
|---------------------------|---------------|-------|------------|------|-----------|----------------|
| **PSNR**                  |               |       |            |      |           |                |
| 15% + 7% knee             |               |       | 30.81±2.01 | 26.78±3.19 | 31.87±0.92 | 35.34±3.13     |
| 20% + 7% knee             |               |       | 31.81±2.23 | 31.30±2.93 | 30.72±0.62 | 37.14±3.23     |
| 15% + 7% brain            |               |       | 32.73±3.46 | 29.06±3.24 | 33.95±2.25 | 39.80±2.73     |
| 20% + 7% brain            |               |       | 34.51±3.94 | 30.70±3.25 | 34.73±2.26 | 41.18±2.70     |
| **SSIM**                  |               |       |            |      |           |                |
| 15% + 7% knee             |               |       | 78.97±8.38 | 72.40±9.29 | 82.50±4.55 | 88.12±5.66     |
| 20% + 7% knee             |               |       | 81.74±8.30 | 86.50±3.77 | 81.66±3.84 | 91.27±4.23     |
| 15% + 7% brain            |               |       | 87.58±6.26 | 78.85±7.23 | 89.86±3.47 | 91.52±6.26     |
| 20% + 7% brain            |               |       | 90.08±6.03 | 82.67±7.43 | 91.24±3.54 | 92.48±5.66     |
4.3 Preliminary result in non-Cartesian MRI reconstruction

In this study, we used a T2* weighted gradient-echo images to simulate the spiral k-space data with 4-fold acceleration. The 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 T2* 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 Figure 3 and Figure 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.
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 $256 \times 256$ matrix size. The intensity of error maps was ten 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 (i.e., 6 out of 24 spiral interleaves), acquired by T2$^*$ weighted gradient echo sequence. The intensity of error maps was five 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 T2$^*$ 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.

5 Discussion

The proposed method can reliably and consistently recover the nearly alias-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, i.e., 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.

In this study, the results demonstrated the successful reconstruction of high-resolution image (i.e., 256 $\times$ 256 matrix) with low-resolution prior (i.e., trained with 128 $\times$ 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 x 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 $\ell_2$
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, i.e., when the residual norm stopped decaying, the likelihood can still penalize the error. This evidence indicated that using the residual norm as the $\ell_2$ fidelity alone was sub-optimal, and the deep learning-based statistical regularization can lead to a better reconstruction result compared with the $\ell_2$ fidelity. Deep learning-based statistical regularization in the proposed method outperformed other conventional regularizations trained by image-level $\ell_2$ loss. $\ell_2$ 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 $\ell_1$ regularization. Due to unavoidable changes of the encoding scheme, e.g., the image size and the RF coils during MRI
experiment in practice, it was essentially needed to separate the learned component (the image 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 $\ell_1$ sparsity prior in compressed sensing. More importantly, the proposed reconstruction framework was generalizable for most reconstruction scenarios.

7 Acknowledgment

None

References

[1] Michael Lustig, David Donoho, and John M Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 58(6):1182–1195, 2007.

[2] Saiprasad Ravishankar and Yoram Bresler. Mr image reconstruction from highly undersampled k-space data by dictionary learning. IEEE transactions on medical imaging, 30(5):1028–1041, 2010.

[3] Jian Sun, Huibin Li, Zongben Xu, et al. Deep admm-net for compressive sensing mri. In Advances in neural information processing systems, pages 10–18, 2016.

[4] Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. Modl: Model-based deep learning architecture for inverse problems. IEEE transactions on medical imaging, 38(2):394–405, 2018.

[5] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. IEEE transactions on Medical Imaging, 37(2):491–503, 2017.

[6] Morteza Mardani, Enhao Gong, Joseph Y Cheng, Shreyas S Vasaniwala, Greg Zaharchuk, Lei Xing, and John M Pauly. Deep generative adversarial neural networks for compressive sensing mri. IEEE transactions on medical imaging, 38(1):167–179, 2018.

[7] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated mri data. Magnetic Resonance in Medicine, 79(6):3055–3071, 2018.

[8] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. arXiv preprint arXiv:1401.4082, 2014.

[9] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. Acta Numerica, 28:1–174, 2019.

[10] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep learning, chapter 16, page 558. MIT press, 2017.

[11] Kerem C Tezcan, Christian F Baumgartner, Roger Luechinger, Klaas P Pruessmann, and Ender Konukoglu. Mr image reconstruction using deep density priors. IEEE transactions on medical imaging, 2018.

[12] Mihaela Rosca, Balaji Lakshminarayanan, and Shakir Mohamed. Distribution matching in variational inference. arXiv preprint arXiv:1802.06847, 2018.
[13] Tim Salimans, Andrej Karpathy, Xi Chen, and Diederik P Kingma. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. arXiv preprint arXiv:1701.05517, 2017.

[14] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. arXiv preprint arXiv:1601.06759, 2016.

[15] Jeffrey A Fessler. Model-based image reconstruction for mri. IEEE Signal Processing Magazine, 27(4):81–89, 2010.

[16] Diederik P Kingma, Tim Salimans, Rafał Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improving variational inference with inverse autoregressive flow. (nips), 2016. URL http://arxiv.org/abs/1606.04934, 2016.

[17] Karol Gregor, Ivo Danihelka, Andriy Mnih, Charles Blundell, and Daan Wierstra. Deep autoregressive networks. arXiv preprint arXiv:1310.8499, 2013.

[18] Stephen Boyd, Lin Xiao, and Almir Mutapcic. Subgradient methods. Lecture notes of EE392o, Stanford University, Autumn Quarter, 2004:2004–2005, 2003.

[19] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of The 33rd International Conference on Machine Learning, volume 48 of Proceedings of Machine Learning Research, pages 1050–1059, New York, New York, USA, 20–22 Jun 2016. PMLR.

[20] Jure Zbontar, Florian Knoll, Anuroop Sriram, Matthew J. Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, James Pinkerton, Duo Wang, Nafissa Yakobova, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastmri: An open dataset and benchmarks for accelerated MRI. CoRR, abs/1811.08839, 2018.

[21] Aaron Van den Oord, Nal Kalchbrenner, Lasse Espeholt, Oriol Vinyals, Alex Graves, et al. Conditional image generation with pixelcnn decoders. In Advances in neural information processing systems, pages 4790–4798, 2016.

[22] Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (grappa). Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 47(6):1202–1210, 2002.

[23] Martin Uecker, Peng Lai, Mark J Murphy, Patrick Virtue, Michael Elad, John M Pauly, Shreyas S Vasanawala, and Michael Lustig. Espirit—an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa. Magnetic resonance in medicine, 71(3):990–1001, 2014.

[24] Martin Uecker, Frank Ong, Jonathan I Tamir, Dara Bahri, Patrick Virtue, Joseph Y Cheng, Tao Zhang, and Michael Lustig. Berkeley advanced reconstruction toolbox. In Proc. Intl. Soc. Mag. Reson. Med, volume 23, page 2486, 2015.

[25] Michael Lustig, Seung-Jean Kim, and John M Pauly. A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories. IEEE transactions on medical imaging, 27(6):866–873, 2008.

[26] Jeffrey A Fessler and Bradley P Sutton. Nonuniform fast fourier transforms using min-max interpolation. IEEE transactions on signal processing, 51(2):560–574, 2003.

[27] Klaas P Pruessmann, Markus Weiger, Peter Börnert, and Peter Boesiger. Advances in sensitivity encoding with arbitrary k-space trajectories. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine, 46(4):638–651, 2001.
A Appendix A: derivation for the probability of logistic distribution in Equation 4

The probability density function of logistic distribution is as follow

\[ f(x; \mu, s) = \frac{e^{-(x-\mu)/s}}{s \left(1 + e^{-(x-\mu)/s}\right)^2} \]

and

\[ f(x; \mu, s) = \frac{1}{s \left(e^{(x-\mu)/(2s)} + e^{-(x-\mu)/(2s)}\right)^2} \]

The cumulative distribution function of logistic distribution is as follow

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

Therefore if we assume the \( \nu \) follows the logistic distribution, the probability of \( \nu \) in range from \( x - d/2 \) to \( x + d/2 \) is

\[ P(\nu; \mu, s) = \frac{1}{1 + e^{-(\nu+d/2-\mu)/s}} - \frac{1}{1 + e^{-(\nu-d/2-\mu)/s}} \]

\[ = \sigma((\nu + d/2 - \mu)/s) - \sigma((\nu - d/2 - \mu)/s). \]

Similarly, if we assume the \( \nu \) follows the mixture logistic distribution

\[ \nu \sim \sum_{k=1}^{K} \pi_k \text{logistic}(\mu_k, s_k), \]

then, we have

\[ P(\nu; \pi, \mu, s) = \sum_{k=1}^{K} \pi_k [\sigma((\nu + d/2 - \mu_k)/s_k) - \sigma((\nu - d/2 - \mu_k)/s_k)]. \]
B Appendix B: reconstruction for the varied image size with deep prior model

Algorithm 1 Reconstruction algorithm with deep prior model

Input:
\( y \) - k-space data
\( A \) - encoding matrix
\( \lambda \) - maximum iteration

Output:
\( x \) - the restored image

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

C List of supporting information

The following support information will be available on this webpage\(^1\),\(^2\):

- Supporting Information Table S1: The scan parameters of different weightings used in brain MRI experiments.
- Supporting Information Figure S1: The diagram shows the PixelCNN++ network in [13], which was the prior model used in this study, i.e., \( g(x) \) in Equation 7. Each ResNet block (gray) has a gated-resnet components. The last layer is fully connected. The input of network was \( x \), outputs of network were parameters of mixture distribution \((\pi, \mu, s, \alpha)\), which were fed into the conditional probability model in Equation 5.
- Supporting Information 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, i.e., the 1D undersamplings were simulated.
- Supporting Information Video S1: The evolution of image during the process of reconstruction.
- Supporting Information Video S2: Reconstruction by the proposed method of a healthy volunteer for T\(_2^*\)-weighted brain images with undersampling factor R=3 (GRAPPA prospective experiment).
- Supporting Information Video S3: Reconstruction by the proposed method of a healthy volunteer for T\(_2^*\)-weighted brain images with 22% k-space (Compressie sensing prospective experiment).

\(^1\)https://github.com/mrirecon/spreco
\(^2\)https://onlinelibrary.wiley.com/doi/full/10.1002/mrm.28274