

Learning-Assisted Fast Determination of Regularization Parameter in Constrained Image Reconstruction

Yue Guan ¹, Member, IEEE, Yudu Li ², Member, IEEE, Ziwen Ke ¹, Xi Peng ¹, Member, IEEE, Ruihao Liu ¹, Yao Li ¹, Senior Member, IEEE, Yiping P. Du ¹, Senior Member, IEEE, and Zhi-Pei Liang ¹, Fellow, IEEE

Abstract—Objective: To leverage machine learning (ML) for fast selection of optimal regularization parameter in constrained image reconstruction. **Methods:** Constrained image reconstruction is often formulated as a regularization problem and selecting a good regularization parameter value is an essential step. We solved this problem using an ML-based approach by leveraging the finding that for a specific constrained reconstruction problem defined for a fixed class of image functions, the optimal regularization parameter value is weakly subject-dependent and the dependence can be captured using few experimental data. The proposed method has four key steps: a) solution of a given constrained reconstruction problem for a few (say, 3) pre-selected regularization parameter values, b) extraction of multiple approximated quality metrics from the initial reconstructions, c) predicting the true quality metrics values from the approximated values using pre-trained neural networks, and d) determination of the optimal regularization parameter by fusing the predicted quality metrics. **Results:** The effectiveness of the proposed method was demonstrated in two constrained reconstruction problems. Compared with L-curve-based method, the

proposed method determined the regularization parameters much faster and produced substantially improved reconstructions. Our method also outperformed state-of-the-art learning-based methods when trained with limited experimental data. **Conclusion:** This paper demonstrates the feasibility and improved reconstruction quality by using machine learning to determine the regularization parameter in constrained reconstruction. **Significance:** The proposed method substantially reduces the computational burden of the traditional methods (e.g., L-curve) or relaxes the requirement of large training data by modern learning-based methods, thus enhancing the practical utility of constrained reconstruction.

Index Terms—Regularization parameter selection, constrained image reconstruction, machine learning, optimization.

I. INTRODUCTION

IMAGE reconstruction using a priori constraints, such as sparsity, low-rankness, and machine learning priors, is playing an increasingly important role in MRI applications [1], [2], [3]. A common approach to enforcing a priori constraints is to use the regularization framework [4], [5]. This framework consists of a weighted sum of two objective functions, one promoting data consistency and the other incorporating a priori information. The relative “weight” of the two objective functions is controlled by a regularization parameter whose value can have a significant impact on the quality of the resulting reconstructions.

A number of methods have been proposed for automatic selection of the regularization parameter, which include L-curve-based methods [6], [7], [8], [9], [10], generalized cross validation (GCV)-based methods [11], [12], [13], [14], [15], [16], [17] and Stein’s unbiased risk estimate (SURE)-based methods [14], [18], [19], [20], [21], [22]. These methods have been applied to various constrained reconstruction problems in MRI, e.g., parallel imaging [23], [18], chemical shift imaging [24], quantitative susceptibility mapping [25], [26] and functional MRI [27]. However, they often require long computation time due to repeated solution of the underlying optimization problem [24], [28] and/or lead to sub-optimal reconstruction (e.g., blurred images) [14], [19], [29], [30].

Deep learning (DL)-based methods have also been proposed for selection of the regularization parameter. These methods are computationally efficient after training. Currently, there are three

Manuscript received 30 June 2023; revised 16 November 2023 and 2 February 2024; accepted 15 February 2024. Date of publication 20 February 2024; date of current version 20 June 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62001293 Grant 81627901, in part by the National Key Research and Development Program of China under Grant 2016YFC0103905, in part by the Shanghai Pilot Program for Basic Research-Shanghai Jiao Tong University under Grant 21TQ1400203, and in part by China Postdoctoral Science Foundation under Grant 2022M720086. (Corresponding author: Zhi-Pei Liang.)

Yue Guan, Ziwen Ke, Yao Li, and Yiping P. Du are with the National Engineering Research Center of Advanced Magnetic Resonance Technologies for Diagnosis and Therapy, School of Biomedical Engineering, Shanghai Jiao Tong University, China.

Yudu Li is with the Beckman Institute for Advanced Science and Technology and National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign USA.

Zhi-Pei Liang is with the Department of Electrical and Computer Engineering and Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: z-liang@illinois.edu).

Xi Peng is with the Department of Radiology, University of Iowa, USA.

Ruihao Liu is with the School of Biomedical Engineering, Shanghai Jiao Tong University, China, and also with the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign USA.

This article has supplementary downloadable material available at <https://doi.org/10.1109/TBME.2024.3367762>, provided by the authors.

Digital Object Identifier 10.1109/TBME.2024.3367762

main approaches to DL-based selection of the regularization parameter: 1) learning a nonlinear mapping that predicts the optimal regularization parameter directly from the measured data [31], 2) simultaneously learning both the regularization parameter and regularization functional through an unrolled network [32], [33], [34], [35], [36], [37], [38], [39], and 3) using reinforcement learning to tune the regularization parameter value [40]. Although the existing DL-based approaches show improvement over traditional methods in some applications, they are limited by the requirement of large experimental data. When very limited experimental data are available as often the case in medical imaging applications, these approaches are prone to have poor generalization capability [41], [42]. Although a transfer-learning-based approach has been proposed to address this issue, it still requires more than tens of training images for network fine-tuning, which may be expensive for some applications [43].

In this work, we propose a novel learning-based method for fast determination of the regularization parameter, which requires very limited experimental data for training (e.g., 2 subjects). Our method is based on the observation that for a particular reconstruction problem and a given class of image functions (e.g., brain images acquired from different subjects with the same data acquisition sequence), the optimal regularization parameter value is weakly dependent on the subject scanned. In other words, for a given constrained image reconstruction problem (e.g., sparsity-constrained parallel imaging), if the optimal parameter values were learned from a few representative samples of an image class, the optimal value can be predicted for other members of the image class. The proposed method has four key steps: a) solution of a given constrained reconstruction problem for a few (say, 3 in this work) pre-selected values of the regularization parameter, b) extraction of multiple approximated image quality metrics from the initial reconstructions, c) predicting the true quality metrics values from the approximated values using pre-trained neural networks, and d) determination of the optimal value for the regularization parameter by fusing the predicted image quality metrics values. Details of the proposed method are given in the subsequent sections.

II. METHODS

A. Problem Formulation

The measured imaging data in many MRI applications can be expressed in matrix-vector form as:

$$\mathbf{d} = \mathbf{E}\boldsymbol{\rho} + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{E} \in \mathbb{C}^{P \times Q}$ denotes the encoding operator, $\mathbf{d} \in \mathbb{C}^{P \times 1}$ the measured data, $\boldsymbol{\rho} \in \mathbb{C}^{Q \times 1}$ the desired (ground truth) image, and $\boldsymbol{\epsilon} \in \mathbb{C}^{P \times 1}$ the measurement noise. Given the measured data \mathbf{d} , reconstruction of $\boldsymbol{\rho}$ is often solved using the regularization framework:

$$\boldsymbol{\rho}^*(\lambda) = \arg \min_{\boldsymbol{\rho}} \|\mathbf{d} - \mathbf{E}\boldsymbol{\rho}\|_2^2 + \lambda R(\boldsymbol{\rho}), \quad (2)$$

where $R(\cdot)$ is a regularization functional incorporating a priori information about $\boldsymbol{\rho}$ (e.g., the sparsity constraint through an L_1 -based regularization [44]); λ is the regularization parameter.

Here, we explicitly express the dependence of the final reconstruction on λ as $\boldsymbol{\rho}^*(\lambda)$.

This paper addresses the problem of selecting an ‘‘optimal’’ value for λ , denoted as λ^* . Mathematically, this problem can be formulated as the following optimization problem:

$$\lambda^* = \arg \min_{\lambda} \mathcal{L}(\boldsymbol{\rho}, \boldsymbol{\rho}^*(\lambda)), \quad (3)$$

where $\mathcal{L}(\cdot)$ is a cost function (or optimality criterion often called quality metric) quantifying the quality of the reconstruction $\boldsymbol{\rho}^*(\lambda)$. This work considers two optimality criteria: the mean squared error (MSE) and the structural similarity index measure (SSIM). The MSE measures the difference between the reconstruction $\boldsymbol{\rho}^*(\lambda)$ and the ground truth $\boldsymbol{\rho}$:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \|\boldsymbol{\rho} - \boldsymbol{\rho}^*(\lambda)\|_2^2. \quad (4)$$

The SSIM [45] is defined as:

$$\mathcal{L}_{\text{SSIM}} = -\frac{[(2\mu_{\rho^*}\mu_{\rho} + c_1)(2\sigma_{\rho^*\rho} + c_2)]}{[(\mu_{\rho^*}^2 + \mu_{\rho}^2 + c_1)(\sigma_{\rho^*}^2 + \sigma_{\rho}^2 + c_2)]}, \quad (5)$$

where μ_{ρ^*} , μ_{ρ} , σ_{ρ^*} , σ_{ρ} , $\sigma_{\rho^*\rho}$ are the voxel mean, variance and covariance of the reconstructed and ground truth images, c_1 and c_2 the constants used to stabilize the division, N the total number of image pixels. We denote \mathcal{L}_{MSE} or $\mathcal{L}_{\text{SSIM}}$ as $\mathcal{L}_{\text{Target}}$ for notation simplicity.

A major practical problem in utilizing $\mathcal{L}_{\text{Target}}$ to guide the selection of λ is that the optimization problem in (3) involves repetitively solving (2) for a number of candidate λ values, leading to long computational time (especially when (2) admits no analytical solutions and an iterative optimization algorithm is needed). In addition, calculating $\mathcal{L}_{\text{Target}}$ requires the knowledge of the ground truth image, $\boldsymbol{\rho}$, which is not available in practice. We solve the aforementioned problems using a novel machine learning-based method that requires a very few experimental images for training, which are available in most MRI applications.

B. Learning-Based Determination of Optimal Regularization Parameter

The proposed method uses a machine learning approach to directly predict the true quality metrics values, including $\mathcal{L}_{\text{Target}}$, as well as the data consistency, $\|\mathbf{d} - \mathbf{E}\boldsymbol{\rho}^*(\lambda)\|_2^2$ (denoted as \mathcal{L}_{DC}), and the regularization level, $R(\boldsymbol{\rho}^*(\lambda))$ (denoted as \mathcal{L}_{REG}), as calculated in the L-curve-based methods. A key feature of the proposed method is that it requires neither the knowledge of the ground truth image, $\boldsymbol{\rho}$, nor the solutions to the constrained reconstruction in (2) for a range of candidate λ values (i.e., the true reconstructions $\{\boldsymbol{\rho}^*(\lambda_m)\}_{m=1}^M$).

The approach is motivated by the finding that for a particular reconstruction problem and a given class of image functions (e.g., the set of images obtained from the same organ (e.g., the brain) using the same data acquisition set-up (e.g., sequence, imaging parameters, and type of hardware)), the variations of \mathcal{L}_{DC} , \mathcal{L}_{REG} and $\mathcal{L}_{\text{Target}}$ across different subjects showed very small variations for a given λ , which can be seen from the surface plots (as illustrated in Fig. 1 obtained using T1POST brain images from fastMRI database [46]) along the horizontal axis (labeled as ‘‘different subjects’’). The optimal regularization parameter λ^* for different subjects also showed small

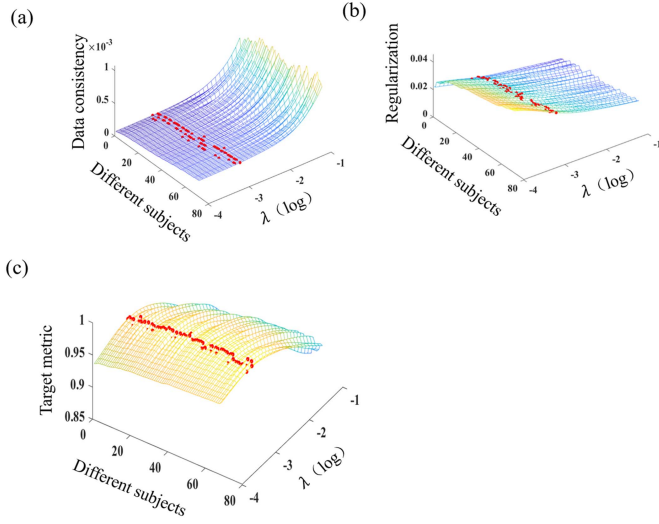


Fig. 1. Illustration of the weak subject dependency of the optimal regularization parameter in sparsity-constrained parallel imaging reconstruction of the brain images using fastMRI dataset: (a) small variation of the data consistency metric and optimal regularization parameter value with respect to different subjects, (b) and (c) are similar results for the regularization and the target metric (SSIM), respectively. The red points indicate the values of different metrics for the optimal regularization parameter values.

fluctuations, which corresponds to the surface points marked in red. Their dependence on the subjects can be learned using a very small set of experimental data (whose ground truth images are available) and then be used for prediction for a new data. More specifically, the proposed method first solves a given constrained reconstruction problem for a very few (say, 3 in this study) pre-selected values of λ ; based on these initial reconstructions, approximated image quality metrics, $\tilde{\mathcal{L}}_{\text{DC}}$, $\tilde{\mathcal{L}}_{\text{REG}}$ and $\tilde{\mathcal{L}}_{\text{Target}}$, are extracted and used to predict the true \mathcal{L}_{DC} , \mathcal{L}_{REG} and $\mathcal{L}_{\text{Target}}$, respectively, using pre-trained neural networks; the predicted image quality metrics are then fused together to determine the desired λ^* using another pre-trained neural network. An overview of the proposed method is illustrated in Fig. 2, and its details are provided below.

First, the proposed method solves (2) to perform initial reconstructions using 3 pre-selected regularization parameters, denoted as λ_1 , λ_2 and λ_3 . Here, λ_2 is set as the average of the optimal regularization parameters of the training data, representing a typical value for λ^* . λ_1 and λ_3 are set as the under-enforced and over-enforced regularizations, respectively, representing the lower and upper bounds for the search interval of λ^* . The specific choice for λ_1 and λ_3 are flexible as long as $\lambda^* \in [\lambda_1, \lambda_3]$. The resulting initial reconstructions (i.e., $\rho^*(\lambda_1)$, $\rho^*(\lambda_2)$ and $\rho^*(\lambda_3)$) are three sparse samples of the function $\rho^*(\lambda)$. Since determination of λ^* often requires dense sampling of $\rho^*(\lambda)$, to avoid repeatedly solving the image reconstruction problem, we use interpolation to efficiently generate samples between $\rho^*(\lambda_1)$ and $\rho^*(\lambda_3)$. More specifically, we use the following linear interpolation in our current implementation:

$$\tilde{\rho}(\lambda) = \begin{cases} c(\lambda; \lambda_1, \lambda_2) \cdot \rho^*(\lambda_1) \\ + (1 - c(\lambda; \lambda_1, \lambda_2)) \cdot \rho^*(\lambda_2), & \text{if } \lambda_1 \leq \lambda < \lambda_2 \\ c(\lambda; \lambda_2, \lambda_3) \cdot \rho^*(\lambda_2) \\ + (1 - c(\lambda; \lambda_2, \lambda_3)) \cdot \rho^*(\lambda_3), & \text{if } \lambda_2 \leq \lambda < \lambda_3 \end{cases}$$

$$\text{with } c(\lambda; \lambda_1, \lambda_2) = \frac{\log_{10}(\lambda) - \log_{10}(\lambda_1)}{\log_{10}(\lambda_2) - \log_{10}(\lambda_1)},$$

$$c(\lambda; \lambda_2, \lambda_3) = \frac{\log_{10}(\lambda) - \log_{10}(\lambda_2)}{\log_{10}(\lambda_3) - \log_{10}(\lambda_2)}. \quad (6)$$

The resulting interpolated images are denoted as $\{\tilde{\rho}(\lambda_m)\}_{m=1}^M$, corresponding to $\{\lambda_m\}_{m=1}^M$. Note that although the interpolated images $\{\tilde{\rho}(\lambda_m)\}_{m=1}^M$ were different from the true reconstructions $\{\rho^*(\lambda_m)\}_{m=1}^M$, the image quality metrics resulting from them were highly correlated (as demonstrated in the Results section). Therefore, $\{\tilde{\rho}(\lambda_m)\}_{m=1}^M$ can be viewed as a good surrogate of $\{\rho^*(\lambda_m)\}_{m=1}^M$.

With the interpolated images, $\{\tilde{\rho}(\lambda_m)\}_{m=1}^M$, instead of directly using them as the network input to predict \mathcal{L}_{DC} , \mathcal{L}_{REG} and $\mathcal{L}_{\text{Target}}$, we extract three approximated metrics from $\{\tilde{\rho}(\lambda_m)\}_{m=1}^M$ to reduce the learning complexity, which include the data consistency $\tilde{\mathcal{L}}_{\text{DC}}$ (i.e., $\|\mathbf{d} - \mathbf{E}\tilde{\rho}(\lambda_m)\|_2^2$), the regularization level $\tilde{\mathcal{L}}_{\text{REG}}$ (i.e., $R(\tilde{\rho}(\lambda_m))$), as well as the target image quality metric $\tilde{\mathcal{L}}_{\text{Target}}$ (i.e., MSE in (4) or SSIM in (5)) as the network input. While the first two approximated metrics (i.e., $\tilde{\mathcal{L}}_{\text{DC}}$ and $\tilde{\mathcal{L}}_{\text{REG}}$) are straightforward to calculate, the target image quality metric requires the knowledge of the desired image function ρ that is not available. To address this problem, we use a neural network to generate an approximate image ρ_{ref} as a surrogate for the ground truth image ρ . More specifically, we adopt the widely used conditional generative model, Pix2Pix GAN [47], which takes the initial reconstruction $\rho^*(\lambda_2)$ as input and is trained to predict ρ . Since the experimental training data are very limited, we train the network in a patch-to-patch manner following [48]. The loss function is the same as that used in [48], which is a combination of the L_1 distance loss and the adversarial loss [47]. Note that although the generated image ρ_{ref} is different from the ground truth ρ , it is sufficient for extracting approximated target image quality metric as demonstrated in the Results section.

After the approximated image quality metrics are extracted, we use curve-to-curve nonlinear mapping networks to predict the “true” image quality metrics (i.e., \mathcal{L}_{DC} , \mathcal{L}_{REG} and $\mathcal{L}_{\text{Target}}$ calculated using $\{\rho^*(\lambda_m)\}_{m=1}^M$) from the approximated ones (i.e., $\tilde{\mathcal{L}}_{\text{DC}}$, $\tilde{\mathcal{L}}_{\text{REG}}$ and $\tilde{\mathcal{L}}_{\text{Target}}$ calculated using $\{\tilde{\rho}(\lambda_m)\}_{m=1}^M$). More specifically, these nonlinear mapping networks are designed as three independent multi-layer perceptron (MLP) neural networks with each network performing one nonlinear mapping task for the data consistency, regularization, and the target quality metric, respectively. Taking the nonlinear mapping task for data consistency as an example, the MLP takes $\{\tilde{\mathcal{L}}_{\text{DC}}(\lambda_m)\}_{m=1}^M$ as input and is trained to predict $\{\mathcal{L}_{\text{DC}}(\lambda_m)\}_{m=1}^M$. Each nonlinear mapping network is trained independently by minimizing the MSE error between the predicted and the desired image quality metric values. The network architecture is shared among different networks, which includes 1 input layer, 3 hidden layers and 1 output layer; each hidden layer has 32 neurons followed by batch normalization and rectified linear unit (ReLU) activation; the output layer is activated by the sigmoid function.

Finally, we fuse the predicted image quality metrics to determine the optimal regularization parameter λ^* . This is achieved

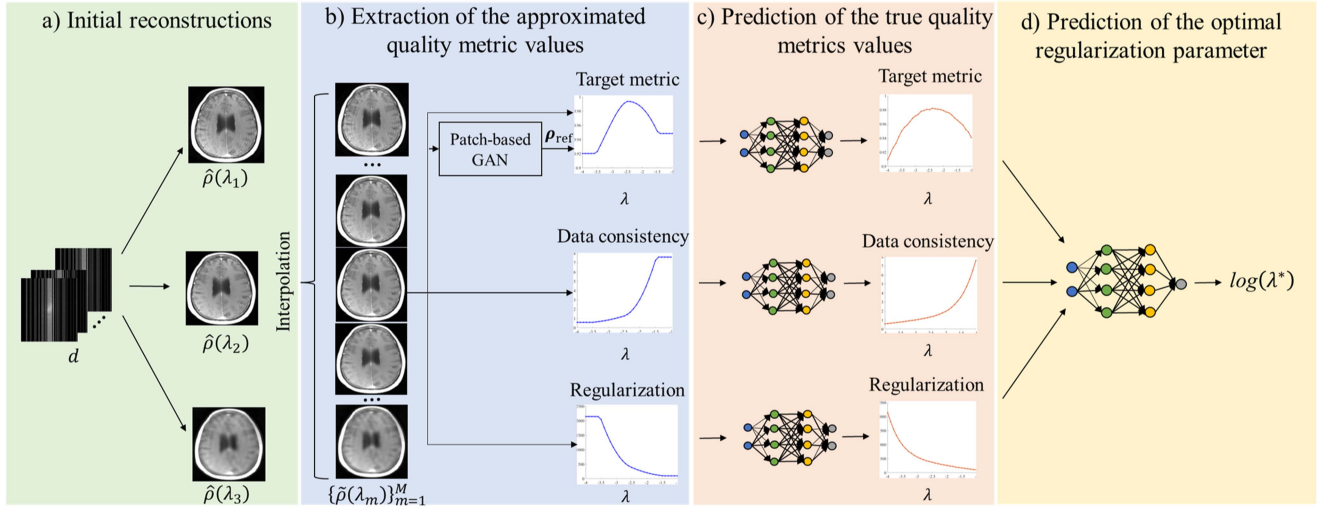


Fig. 2. Schematic overview of the proposed machine learning-based scheme for fast regularization parameter selection. Our method consists of four main steps: (a) initial reconstructions of the given constrained reconstruction problem using a few pre-selected regularization parameter values, (b) extraction of approximated quality metrics from the initial reconstructions, (c) prediction of the true quality metrics values from the approximated values using pre-trained neural networks, and (d) prediction of the optimal regularization parameter value based on fusion of the image quality metrics.

using another MLP network that takes all the predicted metrics as the input and outputs the predicted λ^* . More specifically, the designed network (see Fig. S1 in the supplementary material for an illustration) consists of two branches, one branch taking the predicted data consistency and regularization metrics values (concatenated as a vector) as the input and the other taking the target image quality metric values as the input. Each branch has 4 fully connected layers; the first 3 layers have 16 neurons and the last layer has one neuron. The last one neuron from each branch is then concatenated and passes to the final decision neuron. All neurons use ReLU as the activation function. The network is trained to minimize the MSE error between the predicted and true optimal regularization parameter value determined under the desired optimality criterion. Note this fusion step takes advantage of the complementary information that each metric contains for predicting λ^* . We demonstrate its benefit over using only a single metric in the Results section.

After all the networks are properly trained, determination of the regularization parameter can be done very efficiently at the inference stage using the following algorithm:

- 1) Solution of a given constrained reconstruction problem for 3 pre-selected values of the regularization parameter, denoted as $\rho^*(\lambda_1)$, $\rho^*(\lambda_2)$ and $\rho^*(\lambda_3)$.
- 2) Interpolating $\rho^*(\lambda_1)$, $\rho^*(\lambda_2)$ and $\rho^*(\lambda_3)$ based on (6) to generate $\{\tilde{\rho}(\lambda_m)\}_{m=1}^M$.
- 3) Generating the reference image ρ_{ref} for the target image quality metric extraction (i.e., MSE in (4) or SSIM in (5)) using the pre-trained patch-based pix2pix GAN with the initial reconstruction $\rho^*(\lambda_2)$ as input.
- 4) Extraction of the approximated image quality metrics values including the data consistency, the regularization as well as the target image quality metric value.
- 5) Mapping each approximated image quality metric value to the corresponding “true” image quality metric value

using pre-trained curve-to-curve nonlinear mapping network.

- 6) Determination of the final value of the optimal regularization parameter λ^* using the pre-trained fusion-based network with all the predicted metrics as input.

C. Implementation Details

For all the MLP networks, they were implemented in Tensorflow. Network training was done using the Adam algorithm [49] with a learning rate of 10^{-6} , batch size 32. Each network was trained independently using 1 NVIDIA TITAN Xp GPU. The training time for each network was about 1 hour. For patch-based Pix2Pix GAN, the original public code implemented in Pytorch was used with the patch size setting to 32×32 . The network was also trained using the Adam algorithm with a learning rate of 10^{-3} , batch size 256. The training time was about 2 hours using 4 NVIDIA TITAN Xp GPUs.

III. EXPERIMENTAL SETTING

A. Constrained Reconstruction Scenarios and Optimality Criterion for λ^*

We evaluated the performance of the proposed method in 2 reconstruction scenarios with different regularizers: a) parallel imaging reconstruction with sparsity-promoting regularization, and b) accelerated dynamic imaging with low-rank regularization. Details of each reconstruction scenario are provided below.

For parallel imaging reconstruction, the following compressed sensing (CS-SENSE) formulation was considered:

$$\rho^* = \arg \min_{\rho} \sum_{n=1}^N \|d_n - \Omega F S_n \rho\|_2^2 + \lambda \|\Psi \rho\|_1 \quad (7)$$

where N is the number of receiver coils; \mathbf{d}_n and \mathbf{S}_n are the measured data and sensitivity profile of the n^{th} coil; Ω , \mathbf{F} , and Ψ are the operators of k -space sampling, Fourier transform, and total variation, respectively. The optimization problem was solved using the alternating direction method of multipliers (ADMM) algorithm [50].

For dynamic imaging, the following low-rankness constrained formulation was adopted:

$$\boldsymbol{\rho}^* = \arg \min_{\boldsymbol{\rho}} \sum_{n=1}^N \|\mathbf{d}_n - \Omega \mathbf{F} \mathbf{S}_n \boldsymbol{\rho}\|_2^2 + \lambda \|\mathbf{R} \boldsymbol{\rho}\|_*, \quad (8)$$

where \mathbf{R} is a linear operator that regroups $\boldsymbol{\rho}$ to a Casorati matrix, and $\|\cdot\|_*$ the nuclear norm that promotes low-rankness. The optimization problem was solved using the iterative singular value thresholding (IST) algorithm [51].

To demonstrate the flexibility of the proposed method to optimize different target image quality metrics, we used SSIM as the optimality criterion for choosing the optimal regularization parameter value in parallel imaging and MSE in dynamic imaging. The selection interval for λ^* was set as $[10^{-4}, 10^{-1}]$ (i.e., $\lambda_1 = 10^{-4}$, $\lambda_3 = 10^{-1}$) which contains the optimal regularization parameter (with λ_m uniformly distributed in log scale, leading to a total number of 31 candidates).

B. Datasets and Preprocessing

For parallel imaging, multi-coil T1POST brain images from the NYU fastMRI dataset [46] were used. Undersampled measurements were retrospectively obtained using 1D random Cartesian sampling pattern with center k -space (24 lines) fully sampled and were scaled to have a maximum magnitude value of 1 prior to reconstruction. The total acceleration factor was 3.5. The sensitivity maps were estimated from the fully sampled center k -space using ESPIRT [52]. To test the generalization capability, images without lesions were used in training whereas lesion images were used for testing.

For dynamic imaging, cardiac cine images were collected from healthy volunteers on a 3T Siemens Trio scanner (Erlangen, Germany) with a 20-channel receiver coil array using the retrospective electrocardiogram (ECG)-gated segmented bSSFP sequence. The experiments were approved by the local Institutional Review Board, and informed consents were obtained from the volunteers. The following sequence parameters were used: FOV = 330×330 mm, acquisition matrix = 256×256 , slice thickness = 6 mm, and TR/TE = 3.0 ms/1.5 ms. Retrospective undersampling was performed using a variable density incoherent spatiotemporal acquisition (VISTA) mask [53] with a factor of 6 acceleration. The undersampled measurement was also scaled to have a maximum magnitude value of 1 prior to reconstruction. The sensitivity maps were estimated from the central fully sampled k -space region using ESPIRT [52].

For each reconstruction scenario, a total number of 10 slices from 2 subjects were selected as training experimental data and a total of 100 slices from other 20 subjects were used for testing. In addition, in the training stage, we performed data augmentation to mimic the effect of different experimental variations on λ^* . More specifically, a simulator was designed to generate varying noise levels and experimental settings that may be encountered

in practical situations. For data augmentation with noise, we retrospectively added white Gaussian noise to the measured data to produce a range of SNR levels (15dB-30dB). For data augmentation with different experimental settings, we rescaled the original images with variable sizes (75%-110%) to mimic the variations in FOV in practice. In our current implementation, we augmented each experimental training sample with 25 different SNR levels and 6 different FOVs, leading to 150 times expansion of the training experimental data.

C. Baseline Models

To evaluate the performance of the proposed method, we compared it with both traditional and advanced learning-based methods, including the L-curve method, deep neural network (DNN)-mapping [31] and End-to-End Variational Network End2End VN) [36]. The L-curve method selected the optimal regularization parameter by balancing between the data fidelity term, $\|\mathbf{d} - \mathbf{E} \hat{\boldsymbol{\rho}}(\lambda)\|_2^2$, and the regularization term, $R(\hat{\boldsymbol{\rho}}(\lambda))$, which corresponds to the maximal curvature of the L-shaped curve [7]. The search interval for λ^* was set as the same as the proposed method for fair comparison. The Triangle method [54] was used to locate the optimal regularization parameter.

DNN-mapping directly predicts the optimal regularization parameter from the measured data using neural networks [31]. Here, we used a convolutional neural network as in the original work with some minor modifications for our applications. Specifically, the network contains 3 convolutional blocks and 3 fully connected layers. Each convolutional block consisted of a 5×5 convolution layer followed by batch normalization, a ReLU and a 2×2 max pooling layer with a stride of 2 for down-sampling; each fully connected layer had one neuron with ReLU activation function. Note that, for dynamic imaging, the kernel sizes were set to $3 \times 3 \times 3$ and $2 \times 2 \times 2$ in the convolution layer and max pooling layer, respectively, to accommodate 2D+t dataset [39]. The network parameters were optimized by minimizing the MSE loss as in the original paper using the Adam algorithm with a learning rate of 10^{-5} .

End2End VN learns both the regularization parameter and regularization function from training data for CS-SENSE. The network parameters were optimized by minimizing the SSIM loss between the reconstruction and the ground truth, which was the same with the proposed method and follows the original paper. We trained the network using the open-source code provided by the authors (code address: https://github.com/facebookresearch/fastMRI/tree/main/fastmri_examples/varnet) without changing the network architecture. In addition, we also added more data augmentations by including random image flipping, rotating and shifting as commonly done in neural network training.

IV. RESULTS

A. Comparison to Other Methods

Fig. 3 compared a set of representative reconstruction results of parallel imaging obtained using different methods at various noise levels and with multiple FOV settings. As can be seen, the L-curve method produced very smooth

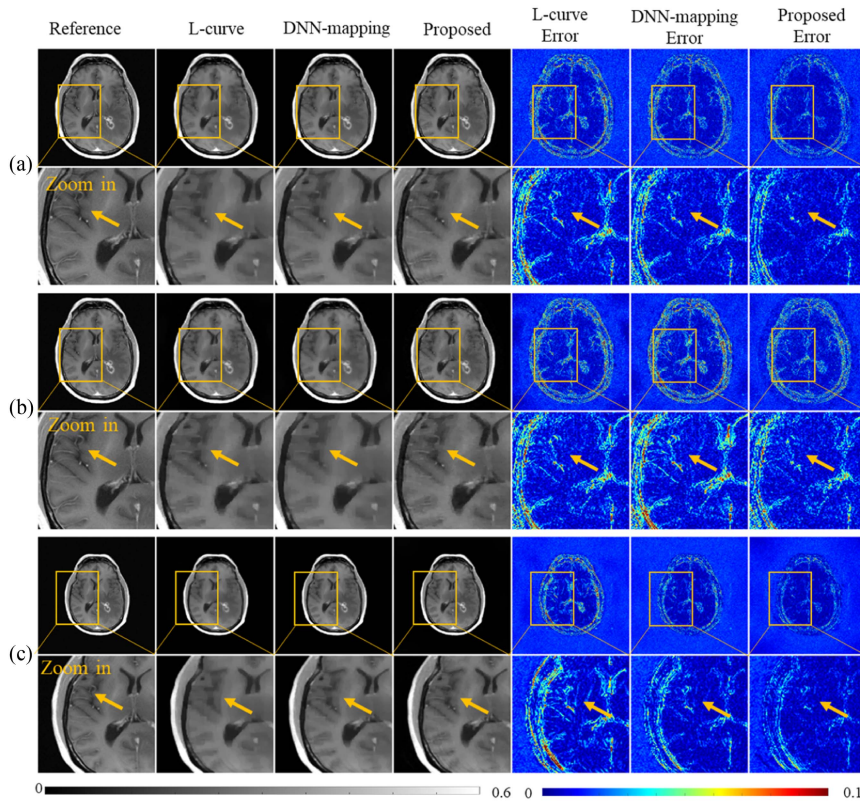


Fig. 3. Representative sparsity-constrained parallel imaging reconstruction results obtained from L-curve, DNN-mapping and proposed method under various scenarios, including (a) low noise level (SNR = 25 dB), (b) high noise level (SNR = 18 dB), and (c) different field of view compared to (a). The error bars were used to indicate the scale of the reconstruction error maps (same for Figs. 4, 6 and 9).

reconstructions (due to over-regularization). DNN-mapping relied on the high-dimensional image features to predict the regularization parameters, which produced reconstructions missing subtle image features (marked by the yellow arrow in the zoomed-in region). This may be due to poor generalization to handle the unseen lesion features. These problems were better handled with the proposed method.

Fig. 4 displayed the reconstruction results for dynamic imaging. The L-curve method produced suboptimal values for the regularization parameter, which led to reconstructions with noticeable aliasing artifacts (marked by the yellow arrow in the error map). The DNN-mapping method yielded reconstructions with less artifacts but at the expense of some subtle image structures (marked by the red arrow in the zoomed-in region). These problems were better handled by our proposed method.

Fig. 5 compared the optimality of the selected regularization parameters across different testing subjects for both parallel imaging and dynamic imaging. As can be seen, L-curve tended to select over- and under-regularized parameters for parallel imaging and dynamic imaging, respectively. DNN-mapping failed to yield optimal parameters consistently for all testing images. In contrast, the proposed method better handled these problems.

We have also quantitatively evaluated the reconstruction accuracy of our proposed method in comparison with the L-curve and DNN methods. As shown in Table I, the proposed method produced the most accurate image reconstructions in terms of

peak signal-to-noise ratio (PSNR), SSIM and MSE. In particular, for parallel imaging, the PSNR value of the proposed method was improved by 1.42 dB and 0.45 dB over L-curve and DNN-mapping, respectively. For dynamic imaging, the proposed method improved the PSNR by 0.28 dB and 0.64 dB compared to L-curve and DNN-mapping, respectively.

B. Comparison to Unrolled Deep Network

Fig. 6 shows results comparing the proposed method with the unrolled deep network End2End VN for parallel imaging in the presence of limited training data (i.e., 10 slices from 2 subjects). As can be seen, when testing data was included in the training set (Fig. 6(a)), the End2End VN outperformed the proposed method substantially, which is expected since End2End VN not only learns the regularization parameter but also the data adaptive regularization function. However, when the testing data with lesion was not included in the training set (Fig. 6(b)), the reconstruction quality of End2End VN degraded substantially due to the overfitting issue [55]. Although we performed data augmentation which artificially increases the training data number, it could not create real image features (e.g., lesions) reflecting the true physiological variations across population. In contrast, the proposed method was able to yield reconstructions consistently close to those obtained at optimal regularization parameters for all those cases.

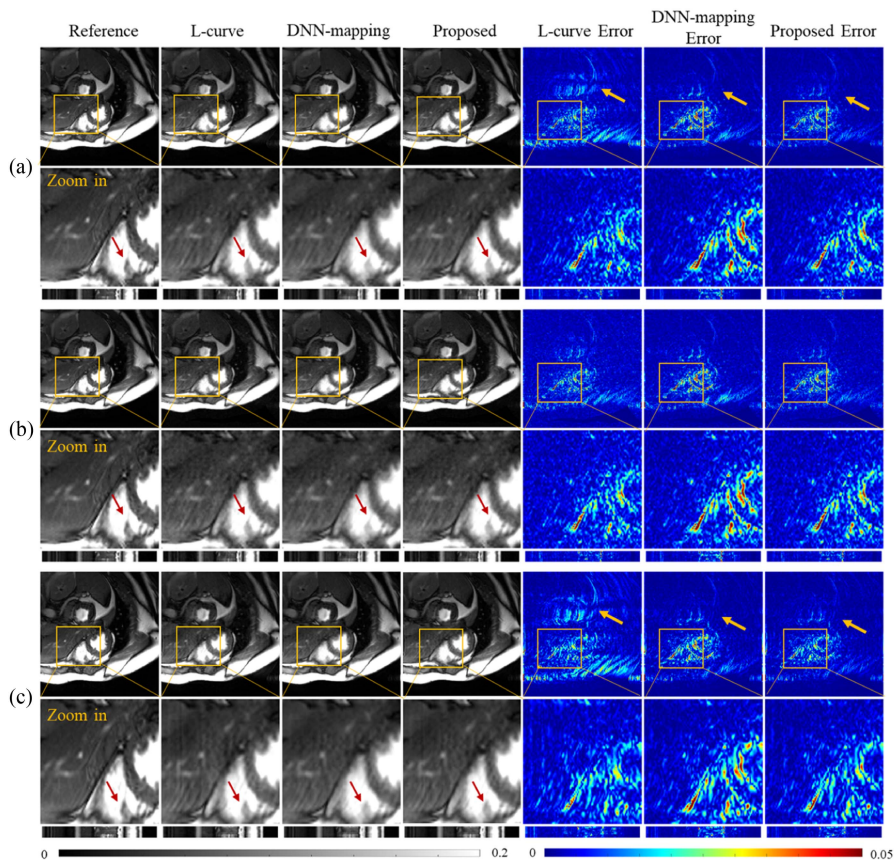


Fig. 4. Representative accelerated dynamic imaging reconstruction results with the same layout as in Fig. 3. In each scenario, the last row represents the y-t image and the corresponding error (extraction along the y and temporal dimensions).

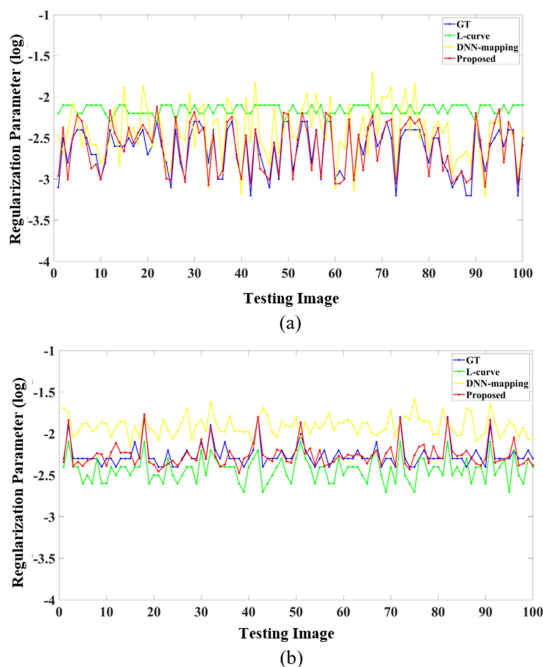


Fig. 5. Comparison of the regularization parameter value selected using the L-curve, DNN-mapping and the proposed method for (a) sparsity-constrained parallel imaging reconstruction and (b) accelerated dynamic imaging reconstruction.

We have also quantitatively compared the performance of End2End VN and the proposed method in terms of PSNR, SSIM and MSE at the training and testing stage, respectively.

As can be seen in Table II, the quantitative results were consistent with our observations in Fig. 6. Specifically, at the training stage, the PSNR, SSIM and MSE of End2End VN were all much better than those of the proposed method. However, at the testing stage, the performance of End2End VN was decreased by a large margin with the PSNR being reduced by 7.19 dB. In contrast, the proposed method nicely overcame this issue, which consistently yielded results close to the ones obtained using the optimal regularization parameters for CS-SENSE reconstructions at both the training and testing stages.

As a result, the proposed method outperformed the End2End VN substantially at the testing stage with the PSNR improved by 1.46 dB

C. Effectiveness of Linear Interpolation for Approximated Quality Metrics Extraction

To demonstrate the effectiveness of the linear interpolation, we compared the approximated quality metrics values extracted from the linear interpolated images, the quality metrics values predicted from the trained curve-to-curve nonlinear mapping networks (with the approximated quality metrics as input) and the true quality metrics values extracted from the true images

TABLE I

QUANTITATIVE COMPARISON OF RECONSTRUCTIONS OBTAINED BY DIFFERENT METHODS FOR PARALLEL IMAGING AND ACCELERATED DYNAMIC IMAGI

Applications	Methods	PSNR	MSE ($\times 10^{-4}$)	SSIM
Parallel imaging	L-curve	36.14 \pm 1.06	2.50 \pm 0.60	0.9462 \pm 0.0086
	DNN-mapping	37.11 \pm 1.88	2.13 \pm 0.95	0.9515 \pm 0.0150
	Proposed	37.56\pm1.61	1.87\pm0.65	0.9559\pm0.0111
Accelerated dynamic imaging	L-curve	34.48 \pm 2.08	3.99 \pm 1.88	0.9459 \pm 0.0127
	DNN-mapping	34.12 \pm 2.05	4.30 \pm 1.94	0.9499 \pm 0.0125
	Proposed	34.76\pm2.15	3.75\pm1.76	0.9510\pm0.0132

TABLE II

QUANTITATIVE COMPARISON OF RECONSTRUCTIONS OBTAINED BY END2END VN AND CS-SENSE WITH THE OPTIMAL AND THE LEARNED (BY PROPOSED METHOD) REGULARIZATION PARAMETER IN THE PRESENCE OF FEW EXPERIMENTAL DATA

Stage	Methods	PSNR	MSE ($\times 10^{-4}$)	SSIM
Training	End2End VN	43.29\pm2.24	0.53\pm0.27	0.9906\pm0.0044
	CS-SENSE (Optimal λ)	39.43 \pm 1.95	1.25 \pm 0.52	0.9730 \pm 0.0112
	CS-SENSE (Proposed λ)	38.79 \pm 2.37	1.51 \pm 0.76	0.9590 \pm 0.0220
	End2End VN	36.10 \pm 1.15	2.54 \pm 0.64	0.9453 \pm 0.0097
Testing	CS-SENSE (Optimal λ)	37.65\pm1.61	1.83\pm0.63	0.9566\pm0.0109
	CS-SENSE (Proposed λ)	37.56 \pm 1.61	1.87 \pm 0.65	0.9559 \pm 0.0111

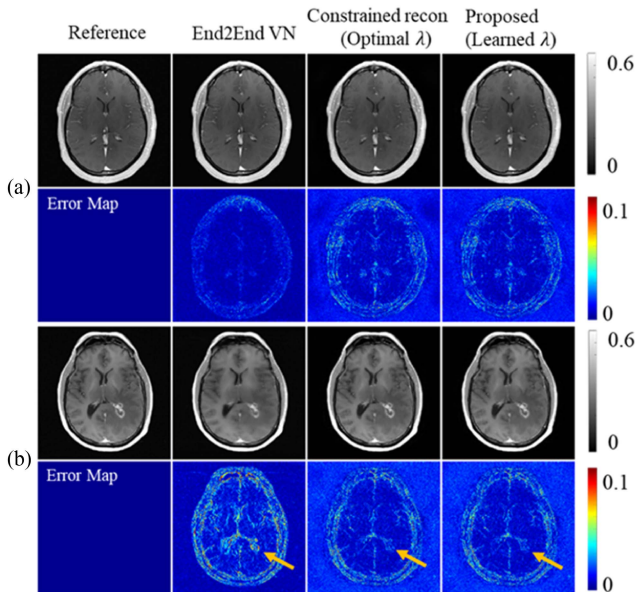


Fig. 6. Comparison of End2End VN and CS-SENSE with the optimal and learned regularization parameter for parallel imaging reconstruction in the presence of few experimental data. Here, we show the reconstruction results from different testing data, including (a) data included in training, (b) data with lesion not seen in the training stage. As can be seen, End2End VN produced reconstructions with much larger error for the lesion data as compared to CS-SENSE with the regularization parameter learned by the proposed method.

under different SNRs. As shown in Fig. 7, the approximated quality metrics values extracted from the linear interpolated images correlated well with the corresponding true metrics values under different SNRs and they can be accurately mapped to the true metrics values with an error less than 2%, indicating

the effectiveness of the linear interpolation for the approximated quality metrics extraction.

D. Effectiveness of the Generated Reference Image

As described in Section II-B, we adopted a patch-based network to “generate” a reference image, which allows us to incorporate the target image quality metric (e.g., SSIM) for improved regularization parameter selection (demonstrated in Fig. 9). Here, we demonstrated the effectiveness of the generated reference image in the context of parallel imaging under different SNRs. As can be seen in Fig. 8(a), given the limited experimental data, the generated reference image failed to preserve fine image details (marked as red arrow), thus not useful as the final reconstruction. However, it served as a rather effective surrogate for evaluating the effect of regularization, i.e., whether the reconstruction is too noisy or too blurring. Fig. 8(b) shows a similar trend of the approximated quality metrics obtained using the true and the generated reference image across a wide range of the regularization parameter values, confirming the effectiveness of the learned reference image as a surrogate for evaluating the regularization effect.

E. Effectiveness of Incorporating Two Kinds of the Quality Metrics

To demonstrate of the advantage of incorporating two kinds of the quality metrics in the proposed method, we compared the reconstructions evaluated at the learned regularization parameters obtained using two kinds of the quality metrics (i.e., the target image quality metric and the L-curve-based metrics) and only one of them alone. As can be seen in Fig. 9, using only L-curve-based metrics tended to produce over-regularized (i.e., overly smooth) reconstructions, while using only SSIM metric

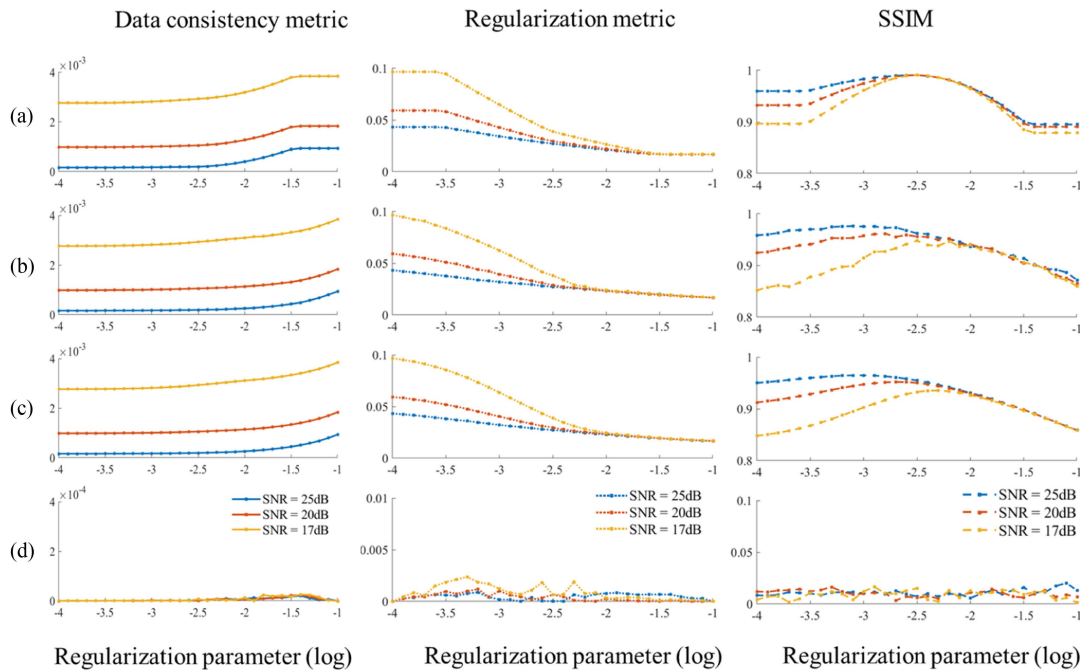


Fig. 7. Illustration of the effectiveness of the approximated quality metrics extracted from linear interpolated images under sparsity-constrained parallel imaging for different SNR levels: (a) extracted approximated image quality metrics, (b) predicted image quality metrics using pre-trained neural networks, (c) true image quality metrics, and (d) error between the predicted and true metrics. As can be seen, the approximated metrics under different SNRs correlated well with the true metrics and they were well mapped to the true metrics with the error less than 2%, indicating the effectiveness of the linear interpolation practice for the approximated quality metrics extraction.

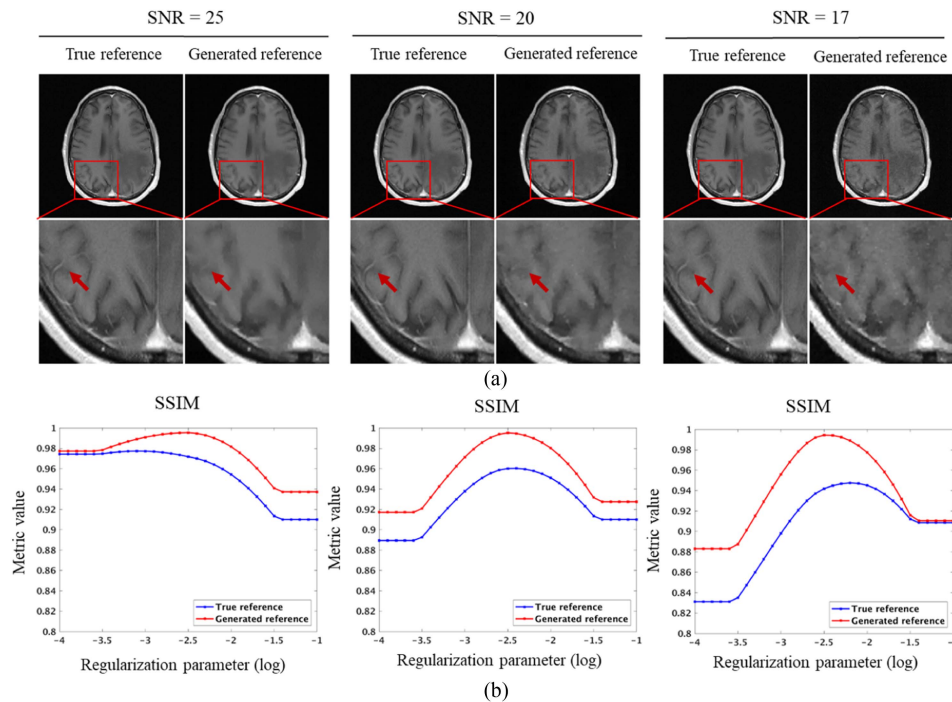


Fig. 8. Illustration of the effectiveness of the generated reference images for indicating the regularization effect under different SNR levels. Although the generated reference image suffers from the loss of detailed image features (red arrow marked in the zoomed-in region in (a)), the approximated quality metric (i.e., structural similarity index (SSIM)) obtained using the generated reference correlated well with that obtained using the true reference image (b).

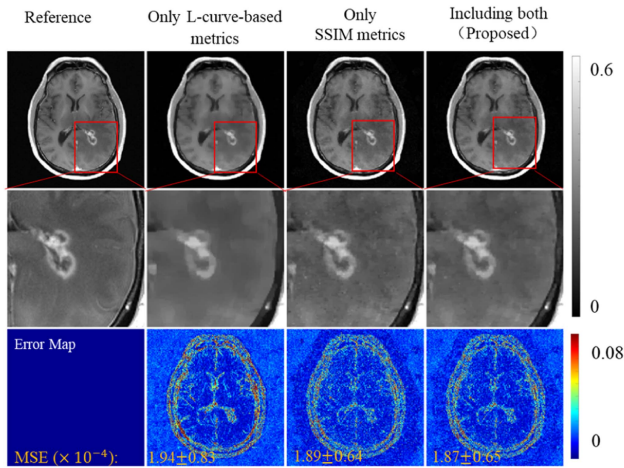


Fig. 9. Illustration of the advantage of including both the L-curve-based metrics (i.e., the data consistency and the regularization) and the SSIM of the proposed method in sparsity-constrained parallel imaging. The numbers represent the mean and standard deviation of the mean square error between the reconstructions and the references for the testing set.

TABLE III
COMPARISON OF COMPUTATIONAL TIME OF DIFFERENT ALGORITHMS AT THE TRAINING AND TESTING STAGE

Applications	Methods	Training Stage	Testing Stage
Parallel imaging	L-curve	-	10 min
	DNN-mapping	2 hours	<1 s
	End2End VN	10 hours	<1 s
	Proposed	78 hours	~1 min
Accelerated dynamic imaging	L-curve	-	40 min
	DNN-mapping	4 hours	<1 s
	Proposed	300 hours	~4 min

tended to produce under-regularized (i.e., noisy) reconstructions. The proposed method overcame the problems, yielding the best reconstruction.

F. Computational Time

We also compared the computational efficiency of the four methods (L-curve, DNN, End2End VN and the proposed) at both the training and testing stages under each reconstruction scenario. As shown in Table III, the proposed method needed much longer time at the training stage (including pre-processing all the training data) mainly because it required repetitively solving (2) under different regularization parameter values to obtain the true quality metrics values. However, once trained, the proposed method was much faster. The computational time for each slice was about 10 times shorter than that of the traditional L-curve method. Our method was slower than the reference deep learning-based methods mainly due to solving the initial reconstructions using the iterative algorithm (i.e., solving (2)). With more optimized implementation (e.g., using GPU and parallel computing [56], [57]), its computation time can be further reduced for practical applications.

V. DISCUSSION

This paper presents a learning-based method for fast selection of regularization parameter that requires only a very small set of experimental data for training. The proposed method has been evaluated in two different constrained image reconstruction scenarios, demonstrating improved computational efficiency and reconstruction quality.

Our method is based on a useful finding that for a particular reconstruction problem and a given class of image functions, the optimal regularization parameter value is weakly dependent on the subject scanned. This finding can be understood by considering the following: a) if the image model is appropriate, the data fidelity term $\|d - E\rho\|_2^2$, in the vicinity of λ^* , is mostly dependent on the noise level rather than subject-specific image features. In other words, the measurement noise dominates the residual term, $d - E\rho$, as λ approaches λ^* ; hence, this term has a very weak subject dependence; b) the regularization functional $R(\rho_{\text{true}})$ often measures some intrinsic image properties (such as sparsity, low-rankness, etc). These image properties (say, the rank of a dynamic cardiac sequence) can be reasonably assumed to be subject-invariant for the same image class; hence, the regularization term is also weakly subject dependent. As both the data fidelity and regularization terms have weak subject dependence, so does the optimal regularization parameter. One may argue that this property is being used implicitly whenever a pre-set regularization value is used for solving a (computationally expensive) constrained reconstruction problem as is often done in practice.

An advantage of the proposed method is that it provides a flexible framework, in which any other image quality metrics besides MES/SSIM and any other features useful for the selection of the optimal regularization parameter can be easily incorporated. The only modification is to add another branch in curve-to-curve nonlinear mapping and/or fusion-based optimal regularization parameter determination.

Our method is also compatible with more advanced regularized reconstruction formulation that incorporates deep learning-based image priors [3]. After a deep learning prior is obtained using those methods, the regularization parameter can be learned using the proposed method, which could further improve the performance of those deep learning-based image reconstruction methods.

In this study, we normalized the datasets to the range of [0, 1], as usually done in machine learning-based imaging applications. This strategy could significantly improve model accuracy by making intensity of different images on a similar scale, so that no single image dominates model training process just because it has large intensity; as a result, the trained model would have better robustness and generalization ability. Our method is also workable for other normalization schemes because the procedures to determine the optimal regularization weight (guided by the quality metric of reconstruction) and to determine the pre-selected regularization parameters (based on the under- and over-regularization) are compatible with any scaling schemes.

This study focuses on reconstruction problems with only one regularization parameter. The proposed method can be extended

to handle multiple regularization parameters in a relatively straightforward manner. For example, in the case of two regularization parameters, one can start with 5 initial reconstruction with different parameter combinations, such as $\{\lambda_1^1, \lambda_1^2\}$, $\{\lambda_1^1, \lambda_3^2\}$, $\{\lambda_3^1, \lambda_1^2\}$ and $\{\lambda_3^1, \lambda_3^2\}$ and $\{\lambda_2^1, \lambda_2^2\}$, followed by 2D linear interpolation to generate images, $\{\tilde{\rho}(\lambda_m^1, \lambda_n^2)\}_{m,n=1}^M$, over the grid of $\{\lambda_m^1, \lambda_n^2\}_{m,n=1}^M$. Then, a similar set of approximated quality metrics can be extracted but in the form of 2D curves. Then, 2D curve-to-curve nonlinear mapping networks can be used to predict the “true” metrics, from which one can determine the final optimal parameters. We will address this kind of problem in our follow-up work.

For practical imaging applications, our method could significantly reduce the computational time for constrained reconstruction and has the potential to enable inline image reconstruction on the scanner. Particularly, all the models or neural networks are trained for a specific image class of interest offline using available training data. Once the networks are properly trained, they can be used for each new patient inline for fast selection of the regularization parameter. Incorporating our method for on-the-fly image reconstruction will be explored in our future work.

The proposed method is designed to learn the optimal regularization specifically for fixed imaging operator(s) and class of feasible image functions. If the imaging operator or functional class changes, then our key assumption that “the variations of quality metrics and optimal regularization parameters across different subjects are small” is likely to be invalid. So, for any new data acquisition scheme (e.g., different sampling trajectories and acceleration rates) and/or different imaging operators (say, T1-weighted imaging vs Diffusion tensor imaging (DTI)) and/or new image functional classes (e.g., new image contrast and objects), a new network needs to be trained (but with only a few experimental training images).

The present study has several limitations. First, the current implementation determined the optimal regularization parameter based on target image quality metrics (i.e., MSE and SSIM) widespread use in practice, which are not necessarily the optimal choices for achieving the highest diagnostic value [58]. With the learning framework in place, future work will investigate other potential metrics which may better guide the image reconstruction process. Second, we extracted the approximated quality metrics from the linear interpolated images of initial reconstructions and demonstrated the effectiveness of the linear interpolation experimentally in the Results Section. Theoretical analysis and alternative approach to extract the approximated quality metrics need further investigation. Third, we retrospectively scaled the training images to mimic different FOVs and/or resolution settings in this study. Future study is needed to evaluate the proposed method using prospective experimental data.

VI. CONCLUSION

This paper presents a new learning-based method for fast and optimal determination of the regularization parameter in constrained image reconstruction. The proposed method requires

only a very small set of experimental data for learning. The performance of the proposed method has been evaluated using in vivo experimental data, producing substantially improved parameter determination accuracy and computational efficiency over the existing methods. The proposed method would enhance the utility of constrained image reconstruction in practical applications.

REFERENCES

- [1] Z. Meng et al., “Accelerating T2 mapping of the brain by integrating deep learning priors with low-rank and sparse modeling,” *Magn. Reson. Med.*, vol. 85, no. 3, pp. 1455–1467, Mar. 2021.
- [2] F. Lam et al., “Ultrafast magnetic resonance spectroscopic imaging using SPICE with learned subspaces,” *Magn. Reson. Med.*, vol. 83, no. 2, pp. 377–390, 2020.
- [3] S. Wang et al., “Accelerating magnetic resonance imaging via deep learning,” in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2016, pp. 514–517.
- [4] J. P. Haldar et al., “Anatomically constrained reconstruction from noisy data,” *Magn. Reson. Med.*, vol. 59, no. 4, pp. 810–818, 2008.
- [5] T. Zhang, J. M. Pauly, and I. R. Levesque, “Accelerating parameter mapping with a locally low rank constraint,” *Magn. Reson. Med.*, vol. 73, no. 2, pp. 655–661, 2015.
- [6] P. C. Hansen and D. P. O’Leary, “The use of the L-curve in the regularization of discrete ill-posed problems,” *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, 1993.
- [7] P. C. Hansen, “The L-curve and its use in the numerical treatment of inverse problems,” in *Computational Inverse Problems in Electrocardiology*, vol. 4, P. Johnston, Ed. Adv. Comput. Bioeng., Ashurst, Southampton U.K.: WIT Press, 2000, pp. 119–142.
- [8] P. J. McCarthy, “Direct analytic model of the L-curve for Tikhonov regularization parameter selection,” *Inverse Problems*, vol. 19, no. 3, pp. 643–663, Jun. 2003.
- [9] F. Lin et al., “Parallel imaging reconstruction using automatic regularization,” *Magn. Reson. Med.*, vol. 51, no. 3, pp. 559–567, 2004.
- [10] P. C. Hansen, “Analysis of discrete ill-posed problems by means of the L-curve,” *SIAM Rev.*, vol. 34, no. 4, pp. 561–580, 1992.
- [11] G. H. Golub, M. Heath, and G. Wahba, “Generalized cross-validation as a method for choosing a good ridge parameter,” *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.
- [12] J. D. Carew et al., “Optimal spline smoothing of fMRI time series by generalized cross-validation,” *Neuroimage*, vol. 18, no. 4, pp. 950–961, Apr. 2003.
- [13] G. H. Golub, M. Heath, and G. Wahba, “Generalized a method cross-validation for choosing parameter,” *Technometrics*, vol. 21, no. 2, pp. 215–223, 2008.
- [14] S. Ramani et al., “Regularization parameter selection for nonlinear iterative image restoration and mri reconstruction using GCV and SURE-based methods,” *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3659–3672, Aug. 2012.
- [15] A. M. Thompson et al., “A study of methods of choosing the smoothing parameter in image restoration by regularization,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 4, pp. 326–339, Apr. 1991.
- [16] N. P. Galatsanos and A. K. Katsaggelos, “Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation,” *IEEE Trans. Image Process.*, vol. 1, no. 3, pp. 322–336, Jul. 1992.
- [17] D. A. Girard, “The fast monte-carlo cross-validation and cl procedures—comments, new results and application to image recovery problems,” *Comput. Statist.*, vol. 10, no. 3, pp. 205–231, 1995.
- [18] D. S. Weller et al., “Monte Carlo SURE-based parameter selection for parallel magnetic resonance imaging reconstruction,” *Magn. Reson. Med.*, vol. 71, no. 5, pp. 1760–1770, 2014.
- [19] S. Ramani et al., “Non-cartesian MRI reconstruction with automatic regularization via Monte-Carlo SURE,” *IEEE Trans. Med. Imag.*, vol. 32, no. 8, pp. 1411–1422, Aug. 2013.
- [20] B. Y. C. M. Stein, “Estimation of the mean of a multivariate normal distribution,” *Statistics*, vol. 9, no. 6, pp. 1135–1151, 2011.
- [21] R. Giryas, M. Elad, and Y. C. Eldar, “The projected GSURE for automatic parameter tuning in iterative shrinkage methods,” *Appl. Comput. Harmon. Anal.*, vol. 30, no. 3, pp. 407–422, 2011.

- [22] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, Feb. 2009.
- [23] F. H. Lin et al., "Parallel imaging reconstruction using automatic regularization," *Magn. Reson. Med.*, vol. 51, no. 3, pp. 559–567, 2004.
- [24] E. Von Harbou et al., "Quantitative mapping of chemical compositions with MRI using compressed sensing," *J. Magn. Reson.*, vol. 261, pp. 27–37, 2015.
- [25] B. Kressler et al., "Nonlinear regularization for per voxel estimation of magnetic susceptibility distributions from MRI field maps," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 273–281, Feb. 2010.
- [26] C. Milovic et al., "Comparison of parameter optimization methods for quantitative susceptibility mapping," *Magn. Reson. Med.*, vol. 85, no. 1, pp. 480–494, Jan. 2021.
- [27] J. D. Carew et al., "Optimal spline smoothing of fMRI time series by generalized cross-validation," *Neuroimage*, vol. 18, no. 4, pp. 950–961, 2003.
- [28] C. Chen et al., "Sparsity adaptive reconstruction for highly accelerated cardiac MRI," *Magn. Reson. Med.*, vol. 81, no. 6, pp. 3875–3887, Jun. 2019.
- [29] P. C. Hansen, T. K. Jensen, and G. Rodriguez, "An adaptive pruning algorithm for the discrete L-curve criterion," *J. Comput. Appl. Math.*, vol. 198, no. 2, pp. 483–492, 2007.
- [30] G. Varela-Mattatall, C. A. Baron, and R. S. Menon, "Automatic determination of the regularization weighting for wavelet-based compressed sensing MRI reconstructions," *Magn. Reson. Med.*, vol. 86, no. 3, pp. 1403–1419, 2021.
- [31] B. M. Afkham, J. Chung, and M. Chung, "Learning regularization parameters of inverse problems via deep neural networks," *Inverse Problems*, vol. 37, no. 10, 2021, Art. no. 105017.
- [32] K. Hammernik et al., "Learning a variational network for reconstruction of accelerated MRI data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [33] Y. Yang et al., "Deep ADMM-net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 10–18.
- [34] J. Schlemper et al., "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 491–503, Feb. 2018.
- [35] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, Feb. 2019.
- [36] A. Sriram et al., "End-to-End variational networks for accelerated MRI reconstruction," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2020, pp. 64–73.
- [37] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1828–1837.
- [38] J. Cheng et al., "Model learning: Primal dual networks for fast MR imaging," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2019, pp. 21–29.
- [39] W. Huang et al., "Deep low-rank plus sparse network for dynamic MR imaging," *Med. Image Anal.*, vol. 73, 2021, Art. no. 102190.
- [40] C. Shen et al., "Intelligent parameter tuning in optimization-based iterative CT reconstruction via deep reinforcement learning," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1430–1439, Jun. 2018.
- [41] V. Antun et al., "On instabilities of deep learning in image reconstruction and the potential costs of AI," *Proc. Natl. Acad. Sci. USA*, vol. 117, no. 48, pp. 30088–30095, 2020.
- [42] F. Knoll et al., "Assessment of the generalization of learned image reconstruction and the potential for transfer learning," *Magn. Reson. Med.*, vol. 81, no. 1, pp. 116–128, Jan. 2019.
- [43] S. U. H. Dar et al., "A transfer-learning approach for accelerated MRI using deep neural networks," *Magn. Reson. Med.*, vol. 84, no. 2, pp. 663–685, 2020.
- [44] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [45] S. S. Channappayya et al., "SSIM-optimal linear image restoration," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2008, pp. 765–768.
- [46] F. Knoll et al., "fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning," *Radiol. Artif. Intell.*, vol. 2, no. 1, Jan. 2020, Art. no. e190007.
- [47] P. Isola et al., "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE 30th Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5967–5976.
- [48] D. Nie et al., "Medical image synthesis with deep convolutional adversarial networks," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2720–2730, Dec. 2018.
- [49] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015.
- [50] J. Yang and Y. Zhang, "Alternating direction algorithms for ℓ_1 -problems in compressive sensing," *SIAM J. Sci. Comput.*, vol. 33, no. 1, pp. 250–278, 2011.
- [51] S. Shalev-shwartz, N. Srebro, and T. Zhang, "A singular value thresholding algorithm for matrix completion," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 6, pp. 2807–2832, 2010.
- [52] M. Uecker et al., "ESPIRiT—An eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA," *Magn. Reson. Med.*, vol. 71, no. 3, pp. 990–1001, Mar. 2014.
- [53] R. Ahmad et al., "Variable density incoherent spatiotemporal acquisition (VISTA) for highly accelerated cardiac MRI," *Magn. Reson. Med.*, vol. 74, no. 5, pp. 1266–1278, 2015.
- [54] J. L. Castellanos, S. Gómez, and V. Guerra, "The triangle method for finding the corner of the L-curve," *Appl. Numer. Math.*, vol. 43, no. 4, pp. 359–373, 2002.
- [55] D. Liang et al., "Deep magnetic resonance image reconstruction: Inverse problems meet neural networks," *IEEE Signal Process. Mag.*, vol. 37, no. 1, pp. 141–151, Jan. 2020.
- [56] X.-L. Wu et al., "Advanced MRI reconstruction toolbox with accelerating on GPU," *Proc. SPIE*, vol. 78720, pp. 241–250, 2011.
- [57] S. S. Stone et al., "Accelerating advanced MRI reconstructions on GPUs," *J. Parallel Distrib. Comput.*, vol. 68, no. 10, pp. 1307–1318, Oct. 2008.
- [58] M. J. Muckley et al., "Results of the 2020 fastMRI challenge for machine learning MR image reconstruction," *IEEE Trans. Med. Imag.*, vol. 40, no. 9, pp. 2306–2317, Sep. 2021.