Supplemental information

Stabilizing deep tomographic

reconstruction: Part A. Hybrid

framework and experimental results

Weiwen Wu, Dianlin Hu, Wenxiang Cong, Hongming Shan, Shaoyu Wang, Chuang Niu, Pingkun Yan, Hengyong Yu, Varut Vardhanabhuti, and Ge Wang

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Table of Contents

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

I. Deep Networks & Datasets

I.A. Ell-50 Network

I.A.1. Narrative. Ell-50 is a special form of FBPConvNet, which is a classic neural network for CT imaging proposed in Ref. 1. The FBPConvNet with multiple-solution decomposition and residual learning[2] was proposed to remove sparse-data artifacts and preserve image features and structures. The reconstruction performance of the FBPConvNet was validated, outperforming the total variation-regularized iterative reconstruction using the realistic phantoms. Besides, it was very fast to reconstruct an image on GPUs. In this study, the training dataset mainly contains ellipses with different intensities, sizes and locations. The network is named Ell-50, indicating that the measurements were collected from 50 different views. This network was trained by the authors of Ref. 1, which can be freely downloaded (https://github.com/panakino/FBPConvNet).

I.A.2. Network Architecture. The Ell-50 network was trained to reconstruct f from measurements $p = Af$, where A represents a subsampling system matrix, with which only 50 uniformly spaced radial lines are collected. Because the FBPConvNet is an image post-processing network, it is trained from filtered backprojection (FBP) reconstruction images rather than directly learning a mapping from p to f. The network first employs FBP to convert p to $\hat{f} = A^+p$ where A^+ represents the FBP and is considered as the first layer of the neural network.

The FBPConvNet is a useful model based on U-Net[3], which is considered as an encoder-decoder pair. The main features of U-Net based FBPConvNet are summarized as the following three features: multilevel decomposition, multichannel filtering, and skip connections (including concatenation operator and residual learning). The network input is an image with 512 ×512 pixels, where it is first down-sampled 4 times for encoding, and then the resultant low-dimensional image features are up-sampled to 512 ×512 pixels. Besides, the skip concatenation operator is employed in this network. The Ell-50 network consists of convolutional and deconvolutional layers, and each convolutional and deconvolutional layer is followed by batch normalization (BN) and ReLU layers. The sizes of filters and stride in the Ell-50 network were set to 3×3 and 1×1, respectively. Moreover, the Ell-50 network details are shown in Fig. S1.

I.A.3. Network Training. The few-view and full-view FBP images are treated as the input and ground-truth of the Ell-50, respectively. In this study, the network was implemented using the MatConvNet[4] toolbox with a slight modification to train and evaluate the performance. To prevent the divergence of the cost function, the MatConvNet[4] toolbox was slightly modified by clipping the computed gradients to a fixed range[5]. In this study, we only used the pre-trained network weights of Ref. 1 that were publicly available at GitHub (https://github.com/panakino/FBPConvNet). Such a configuration is consistent with the literature [6]. The loss function plays an important role in controlling the image quality, and the mean square error (MSE) between the network output and the ground truth is considered in Ell-50. Since the employed network was performed on a TITAN Black GPU graphic processor (NVIDIA Corporation), the

total training time took about 15 hours with 101 epochs[6]. Regarding the learning rate, it was decreased logarithmically from 0.01 to 0.001. Besides, the batch size, momentum, and clipping value were set to 1, 0.99 and 10^{-2}, respectively. For the Ell-50 network, it was implemented in MATLAB with the MatConvNet platform based on Window 10 system with one NVIDIA TITAN XP graphics processing units (GPUs) installed on a PC (16 CPUs @3.70GHz, 32.0GB RAM and 8.0GB VRAM).

Fig. S1. Architecture for Ell-50.

I.A.4. Training Data. Regarding the training dataset, the number of training images is 475. The training images were reconstructed via FBP with sparse-view measurement. The dynamic range of the reconstructed images was controlled to the range of [−500, 500] HU. Since only ellipses with different sizes, locations and intensities were simulated, the projections were accurately and analytically computed[7]. The scanning geometry was set to produce parallel beams[8]. The number of full projections and the number of detector units were set to 1,000 and 729, respectively. Especially, the functions of *radon* and *iradon* in MATLAB were employed to realize the projection and backprojection operations. For sparse-data reconstruction, only 50 views were extracted from full projections, and then FBP reconstructed images were input to the selected networks in this paper. This case is a typical sparse-view reconstruction[9-11]. The ground truths are FBP images from full projections (i.e., 1000 views).

I.A.5. Test Data. To demonstrate the instability of neural network (i.e., Ell-50), the additional symbol " " and the text "CAN YOU SEE IT" were first embedded in the original image, which was provided by the authors of Ref. 6. These artificial features were to mimic the structure changes and further validate the instability of the neural network in this case (see Fig. S2). In this study, the image with the symbol " " and text "CAN YOU SEE IT" was also treated as case C1 to validate the instability of Ell-50 and the stability of our proposed ACID

2

method. Besides, a slightly complicated phantom with the inserted logo of a bird and text "A BIRD?" was provided by the authors of Ref. 12 and downloaded from Ref. 6, which is defined as case C2. The test image consists of 512×512 pixels, and it contains structural features without tiny perturbation. To generate adversarial attacks, the proposed method in [6] was employed to induce tiny perturbations. For case C3, an original image was randomly selected from the test datasets of https://github.com/panakino/FBPConvNet, which contains no perturbation. Here, the tiny perturbation is added to the original image with the same technique used in Ref. 6, and then we obtained the case C3. Regarding case C4, the same technique used in Ref. 6 was employed to generate the tiny perturbation and then embedded into case C1 to obtain C4. Furthermore, a Gaussian noise image with zero mean and standard deviation of 15 in HU over the pixel value range was superimposed to case 1 to obtain C5 image. To validate the ability of ACID against adversarial attacks, the adversarial samples (see Section III in this supplementary information for details) for the whole ACID were generated and added into the C3 and C1 images respectively to obtain the cases of C6 and C7. The searched adversarial attacks in the whole ACID flowchart are greater than those used in a single neural network (i.e., Ell-50) in terms of L_2-norm.
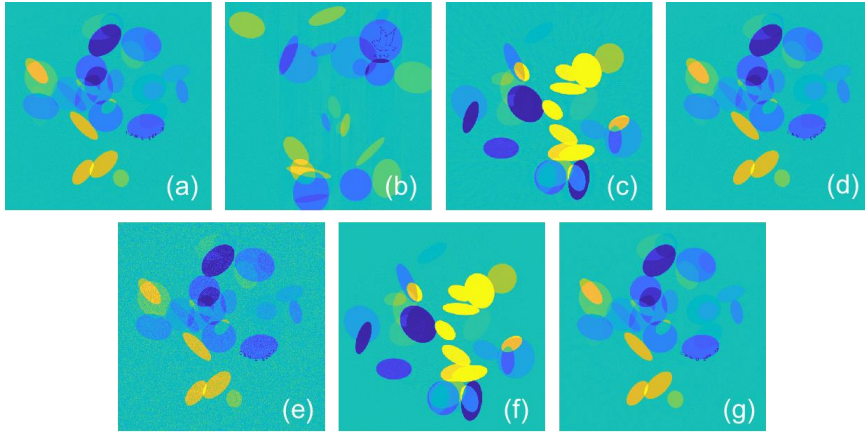
Fig. S2. Test images used to validate the effectiveness of ACID in stabilizing Ell-50 for CT study. (a)-(g) correspond to the C1-C7 cases, respectively. The display window is [-150, 150] HU.

I.B. DAGAN Network

I.B.1. Narrative. The DAGAN network is to reduce aliasing artifacts with the U-Net[3] based generator[13]. To enhance the ability of the reconstruction method in preserving image texture and edges, DAGAN incorporates an innovative content loss and adversarial loss. Besides, it also introduces frequency-domain features to encourage coherence in image and frequency domains. Compared with the traditional CS-based and some other deep learning methods[14-16], the DAGAN method achieved superior performance in retaining image details. Besides, as one of the post-processing methods, the speed of DAGAN reconstruction is very fast. In this study, the DAGAN network was tested on a single-coil MRI with 10% and 20% subsampling rates. The trained weights are not available online, however, the authors of Ref. 13 provide the implementation details of DAGAN. With this help, we retrained the DAGAN with different subsampling rates and masks. The architecture, training parameters, and test data are summarized in the following subsections.

I.B.2. Network Architecture. The DAGAN[13] network was proposed for fast MRI reconstruction from subsampled measurement data. In the case of DAGAN, the measurement data is $p = Af$, where A is the subsampled discrete Fourier transform. The aim of DAGAN is to recover f from the degraded image $\hat{f}$ that is reconstructed directly via inverse Fourier transform from the zero-filled measurement data.

To restore high-quality MR images from measurement, DAGAN adopted a conditional generative adversarial network (GAN)[17-19] model. It consists of two modules: generator and discriminator. The generator is to recover the image, and the discriminator is to distinguish the recovered image and the ground-truth. The goal is to make the discriminator fail, and hence improve the recovered image quality. The authors of Ref. [13] provided three variants of DAGAN, and we selected the full model version (Pixel-Frequency-Perceptual-GAN-Refinement) in our experiments. According to parameter settings in Ref. 13 and the codes provided by the authors of Ref. 6, we retrained the DAGAN.

The architecture of the generator is illustrated in Fig. S3. It adopted the basic U-Net type structure, which contains 8 convolutional layers and 8 deconvolutional layers. All of them are followed by batch normalization layers to accelerate training convergence and overcome overfitting. The leaky ReLU layers are adopted as the activation function with a slope equal to 0.2 when the input is less than 0. Additionally, skip connections are employed to concentrate on encoder and decoder features to gain reconstruction details and promote the information flow. The hyperbolic tangent function is applied as the activate function for the output of the last convolutional layer. Then a global skip connection, adding the input data and the output of the hyperbolic tangent function together, is then clipped by a ramp function to scale the output of the generator to the range [-1,1]. The global skip connection can accelerate the training convergence and improve the performance of the network. The DAGAN network architecture was shown in Fig. S3. For more detailed information on the DAGAN network, please refer to Ref. 13.
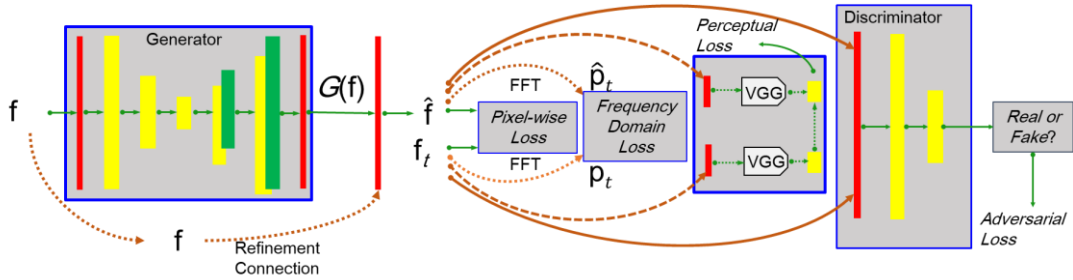

Fig. S3. The architecture of the DAGAN network.

I.B.3. Training Parameters. The loss function of DAGAN is formulated as follows:

$$L_{DAGAN} = \sigma_1 L_{Img} + \sigma_2 L_{frq} + \sigma_3 L_{VGG} + \sigma_4 L_D \quad s.t.\, \sigma_1, \sigma_2, \sigma_3, \sigma_4 > 0 , \quad \text{(S.1.1)}$$

where L_{Img} computes the Euclidean distance in the image domain between the generated image and ground truth, and L_{frq} accounts for the counterpart in the k-space. To constrain the similarity loss L_{VGG} in the feature space, the trained VGG-16 was used to optimize the L_2-distance between feature maps of the generated image and ground truth, which is the same as Ref. 20. In particular, the feature maps generated of the conv4 layer in VGG-16 were used to

calculate L_{VGG}. Last, L_D is the adversarial loss using a cross entropy to make the generated image more realistic. $\sigma_1, \sigma_2, \sigma_3, \sigma_4$ are the hyper-parameters to balance different constraint terms. According to Ref. 13, they were set to 15, 0.1, 0.0025 and 1, respectively. The generator and the discriminator were optimized using the Adam algorithm[21] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. Specifically, the learning rate was initially set to 0.0001, which was decreased half every 5 epochs, and the batch size is 25. To prevent overfitting, an early stopping strategy was adopted via measuring the loss L_{frq} on the validation set, and the stopping number was set to 10.

I.B.4. Training Data. The datasets for training the DAGAN network were provided by the MICCAI 2013 Grand Challenge and are publicly available in https://my.vanderbilt.edu/masi/workshops/. More details about the training datasets are given in https://github.com/tensorlayer/DAGAN. Specifically, to exclude the negative influence on the DAGAN network, all the images that have more T% background pixels were dropped. In our experiments, the threshold T was set to 90. After data preprocessing, there are 15,912 images for training and 4,977 images for testing. All the images are T1-weighted brain MR images. Again, the data augmentation methods were applied to eliminate overfitting, including image flipping, rotation, shifting, and so on[13].

In the experiments, the DAGAN is to recover images from 10% subsampling rate using a 2D Gaussian mask and the radial mask of a 20% subsampling rate, respectively. Two models of the DAGAN network were trained for these two subsampling masks. All the codes were implemented with TensorLayer and Tensorflow frameworks[13].

I.B.5. Testing Data. To test the robustness of DAGAN in terms of small structural changes, adversarial attacks and noise, the symbols "HELLO NATURE" and "CAN YOU SEE IT" were embedded in two different original images, which are denoted as Cases M1 and M2, respectively. Specifically, the image with the symbol "CAN YOU SEE IT" was provided by the authors of Ref. 6 (download in https://github.com/vegarant/Invfool). The original image with the symbol "HELLO NATURE" was produced (downloaded from https://github.com/tensorlayer/DAGAN). In cases M1 and M2, there are two test images used to demonstrate the instability of the DAGAN network with respect to small structural changes. Next, to explore the performance of the DAGAN network in terms of adversarial attacks and small structural changes, the tiny perturbations derived from Ref. 6 were added into cases M1 and M2 to generate cases M3 and M4. Last, to test the DAGAN network in terms of anti-noising, the noise was superimposed to cases M1 and M2 to obtain cases M5 and M6. In our ablation study of ACID, we randomly selected one original image as M7 from the DAGAN test dataset (https://github.com/tensorlayer/DAGAN). Furthermore, cases M8 and M9 were generated by applying the radial mask of a 20% subsampling rate on the M1 and M2 images, which were used to compare the performances between ACID and the classic Alternating Direction Method of Multipliers (ADMM)-net[22]. Regarding the stability of ACID, the tiny perturbations from ACID were added into M7, M1 and M2, and then the images with tiny perturbations were marked as M10, M11 and M12. The tiny perturbations from M11 and M12 are greater than the perturbations within M3 and M4 in terms of the L_2-norm. Except for M8 and M9, all the rest of the images were recovered from the k-space data collected at a 10% subsampling rate

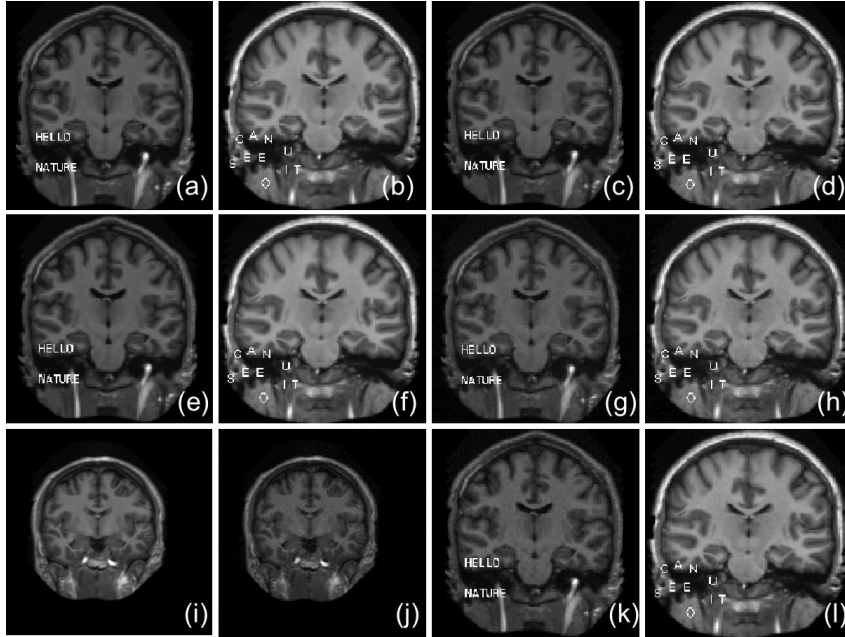using the Gaussian mask. All the images from the references of M1-M12 are shown in Fig. S4.

Fig. S4. Test images for showing the instability of neural networks. (a)-(l) correspond to the M1-M12 cases, respectively.

I.C. AUTOMAP Network

I.C.1. Narrative. Proposed as a framework for image reconstruction, the automated transform by manifold approximation (AUTOMAP) transfers sensor data to a high-quality image with a mapping function between the sensor and image domains[23]. The AUTOMAP demonstrated its advantages in various magnetic resonance imaging acquisition modes using the same architecture and hyperparameters. In this study, the AUTOMAP neural network was tested on the single-coil MRI with subsampled data. The trained ATUOMAP used in our experiments is provided by Ref. 6. The architecture, training details, and test data of AUTOMAP are in the following sub-sections.

I.C.2. Network Architecture. The AUTOMAP[23] presents a framework for image reconstruction by translating sensor-domain signals into the image domain directly via domain-transform manifold learning. For MRI reconstruction, four subsampling strategies were applied to access the performance of the AUTOMAP, which are Radon projection, spiral non-Cartesian Fourier, under-sampled Cartesian Fourier, and misaligned Fourier.

The AUTOMAP network takes a vectorized measurement data as input which is sub-sampled from the full-sampled k-space data. First, we can obtain the complex k-space data using the discrete Fourier transform on the MR images. Then, the subsampled k-space data are generated via a subsampling mask. Next, these measurement data are reshaped into vectors. Last, the vectorized measurement data are fed into the AUTOMAP network. In this paper, the images with the size of 128×128 and 60% subsampling rate are tested for MRI reconstruction. There are two fully connected layers in the AUTOMAP network, which have 25,000 and 128x128 nodes, respectively. The activate function of the first fully connected layer is the hyperbolic tangent function, and

the output of the second fully connected layer then subtracts the mean value of itself. Next, it is reshaped into a feature map with the same size as the reconstructed image. Furthermore, two convolutional layers are applied to extract essential features from their input data. Each of them contains 64 filters with a size of 5×5 and the stride of 1×1. The activation function of the first convolution layer is a hyperbolic tangent function and the other is rectified linear unit (ReLU). The last convolutional layer has one filter with the size of 7×7 and a stride of 1×1. The output of the network is the corresponding reconstruction image. The trained weights were provided by the authors of Ref. 6.

I.C.3. Training Parameters. The whole optimization problem of the AUTOMAP is defined as follow:

$$L_{AUTOMAP} = L_{rec} + \lambda_1 L_{fea} . \tag{S.1.2}$$

The loss function of AUTOMAP $L_{AUTOMAP}$ consists of two terms, i.e., L_{rec} and L_{fea}. L_{rec} is employed to evaluate the Euclidean distance between the predicted image provided by the AUTOMAP network and the ground-truth image. L_{fea} is ℓ_1-norm to constrain the feature maps produced by the activation function of the second convolutional layer. $\lambda_1 > 0$ is to balance the two terms. The total loss function is optimized by the RMSProp algorithm with momentum 0 and decay 0.9[23]. The learning rate is 0.00002 and the batch size is 100. The network was trained and stopped after 100 epochs.

I.C.4. Training Data. Selected in the MGH-USC HCP public dataset (http://www.humanconnectomeproject.org/data/), there are 50,000 images from 131 subjects in total. Specifically, the training images are 128×128 matrices, which were subsampled from the central part cropped from the original image. Meanwhile, all the training datasets were scaled to a given range. In the Fourier space, the subsampled measurement data were produced by a Poisson-disk mask of a 60% subsampling rate.

To improve the generalization ability of the AUTOMAP network, the data augmentation strategy was applied. 1.0% multiplicative noise was added to the input to promote manifold learning during the course of network training, and it is beneficial for the trained network learning robust representations from corrupted inputs. In fact, the specific additional noise distribution of the corruption process is not subject to the additive Gaussian noise during the process of evaluation. The corresponding training datasets with the size of 128×128 are cropped from original MR images by using four types of reflections. All the related codes were implemented in the TensorFlow framework[6].

I.C.5. Testing Data. To validate the instability of the AUTOMAP network, the symbol "♡" was first added to the original MR image, which was also provided by the author [6]. This simple symbol was used to simulate small structural changes in the patient and then test the instability of the AUTOMAP network reconstruction. All the test data were downloaded from Ref. 6. In addition, the "HELLO NATURE", "CAN U SEE IT" and "◇" were added to the original test image to generate A1 and A2 with the structural changes. The resultant tiny perturbations were added to A1 and A2 images to obtain A3 and A4 images (see Fig. S5).

I.C.6. Reconstruction Results. Here, to demonstrate the advantages of ACID, a typical reconstruction network, AUTOMAP, was selected as an example, and

the reconstruction results of Fig. S5 (a) are in Fig. S6. As shown in Fig. S6, ACID produced significantly better image quality than AUTOMAP. The PSNR was improved by ACID to 36.0 dB, well above 27.8 dB of AUTOMAP. Also, the SSIM of ACID reached 0.971, while the counterpart of AUTOMAP was 0.730. It further demonstrates that ACID achieves better image quality than AUTOMAP. The related reconstruction results of A1-A4 are in the main body of the paper.
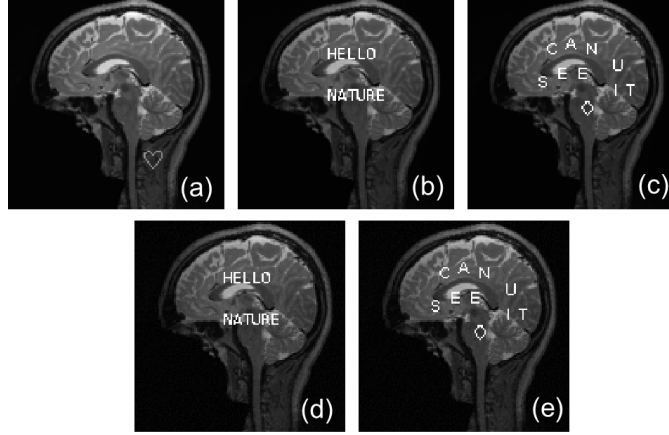

Fig. S5. Test images with a structural change and tiny perturbations for evaluation of the AUTOMAP stability. (a) was the test image provided in Ref. 6, and (b)-(e) represent the test images of A1-A4.
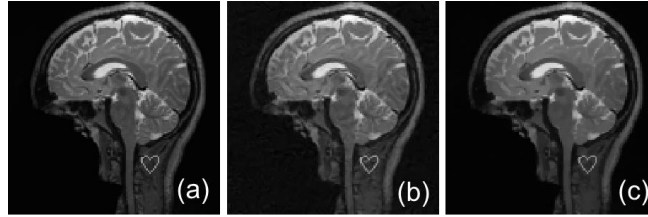
Fig. S6. ACID deep reconstruction with the embedded AUTOMAP network. (a) represents the original brain phantom, (b) and (c) represent the reconstructed results by AUTOMAP and ACID respectively.

I.D. ADMM-net
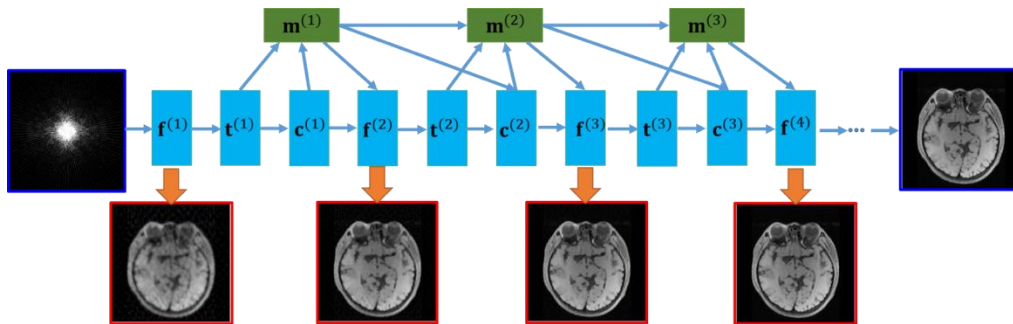
Fig. S7. The flowchart of the ADMM-net. $\boldsymbol{f}^{(k)}$, $\boldsymbol{c}^{(k)}$, $\boldsymbol{t}^{(k)}$ and $\boldsymbol{m}^{(k)}$ represent the construction layer, convolution layer, nonlinear transform layer, and multiplier update layer in the k-th stage.

I.D.1 Narrative. Inspired by the traditional Alternating Direction Method of Multipliers (ADMM) iterative optimization algorithm for CS-based MRI[24], the ADMM-net defined over a data flow graph was first proposed in Ref. 22. Regarding the training procedure, the network parameters (e.g. image shrinkage functions, transforms) are trained into an end-to-end architecture using the L-BFGS algorithm[25]. Regarding the testing step, it needs a similar

computational overhead with the ADMM. However, there is only one parameter to be chosen initially in the ADMM-net since others are automatically learned during the training step. The superior experiments on MRI image reconstruction demonstrate the advantages over fast MRI imaging and higher image quality. In this study, the ADMM-net is tested on the single-coil MRI with 20% subsampling and the trained weights are provided by the authors of Ref. 22. The architecture, training parameters and test data of ADMM-net are summarized in the following sections. The workflow of ADMM-net is given in Fig. S7.

I.D.2. Network Architecture. ADMM-net[22] is a classical unrolled iterative optimization algorithm for MRI reconstruction. Different from the traditional compressed sensing (CS) based methods[26] and data-driven based methods, ADMM-net can be trained end-to-end by incorporating a physic-guided model, and it achieves excellent performance in MR imaging with much less computational cost. The ADMM-net is derived from the ADMM algorithm via solving the sub-problem with deep learning networks. The CS-MRI model can be described as:

$$\underset{\boldsymbol{f}}{\mathrm{argmin}} \frac{1}{2} \left\| \boldsymbol{A}\boldsymbol{f} - \boldsymbol{p}^{(0)} \right\|_F^2 + \sum_{l=1}^{L} \lambda_l g(D_l(\boldsymbol{f})) \qquad (S.1.3)$$

where $\boldsymbol{f} \in C^N$ is the MR image to be reconstructed, $\boldsymbol{p}^{(0)} \in C^H$ denotes the under-sampled measurement data, $\boldsymbol{A}$ is the Fourier translation based system matrix with an under-sampled mask, D_l represents the transform operation, g is the regularization function, and λ_l is the regularization parameter. By introducing $t_l = D_l(\boldsymbol{f})$, $l = 1, \dots, L$, (S.1.3) is converted into the following constraint optimization problem:

$$\underset{\boldsymbol{f}, \{t_l\}_{l=1}^{L}}{\mathrm{argmin}} \frac{1}{2} \left\| \boldsymbol{A}\boldsymbol{f} - \boldsymbol{p}^{(0)} \right\|_F^2 + \sum_{l=1}^{L} \lambda_l g(t_l), \quad t_l = D_l(\boldsymbol{f}), l = 1, \dots L. \qquad (S.1.4)$$

(S.1.4) is a constraint programming procedure and it can be further converted into the following unconstraint problem

$$\underset{\boldsymbol{f}, \{t_l\}_{l=1}^{L}, \{\alpha_l\}_{l=1}^{L}}{\mathrm{argmin}} \frac{1}{2} \left\| \boldsymbol{A}\boldsymbol{f} - \boldsymbol{p}^{(0)} \right\|_F^2 + \sum_{l=1}^{L} \lambda_l g(t_l) - \sum_{l=1}^{L} \langle t_l - D_l(\boldsymbol{f}), \alpha_l \rangle + \frac{1}{2} \sum_{l=1}^{L} \gamma_l \| t_l -$$
$$D_l(\boldsymbol{f}) \|_F^2, \qquad (S.1.5)$$

where α_l $(l = 1, \dots, L)$ are Lagrange multipliers and γ_l $(l = 1, \dots, L)$ are the corresponding penalty parameters. (S.1.5) can be solved using the ADMM algorithm [27] as the following three sub-problems:

$$\boldsymbol{f}^{(k+1)} = \underset{\boldsymbol{f}}{\mathrm{argmin}} \frac{1}{2} \left\| \mathbf{A}\boldsymbol{f} - \boldsymbol{p}^{(0)} \right\|_F^2 - \sum_{l=1}^{L} \langle t_l^{(k)} - D_l(\mathbf{f}), \alpha_l^{(k)} \rangle + \frac{1}{2} \sum_{l=1}^{L} \gamma_l \left\| t_l^{(k)} - \right.$$
$$\left. D_l(\boldsymbol{f}) \right\|_F^2 , \qquad (S.1.6)$$

$$t_l^{(k+1)} = \underset{\{t_l\}_{l=1}^{L}}{\mathrm{argmin}} \lambda_l g(t_l) - \langle t_l - D_l(\boldsymbol{f}^{(k+1)}), \alpha_l \rangle + \frac{1}{2} \gamma_l \| t_l - D_l(\boldsymbol{f}^{(k+1)}) \|_F^2, l =$$
$$1, \dots, L, \qquad (S.1.7)$$

$$\alpha_l^{(k+1)} = \alpha_l^{(k)} + t_l^{(k+1)} - D_l(\boldsymbol{f}^{(k+1)}), l = 1, \dots, L. \qquad (S.1.8)$$

Finally, these three sub-problems can be updated iteratively using deep neural blocks. Regarding the ADMM-net, the above optimization with one separate variable update can be generalized as four type layers: reconstruction layer ($\boldsymbol{f}^{(k+1)}$), convolutional layer ($\{D_l(\boldsymbol{f}^{(k+1)})\}_{l=1}^{L}$), non-linear layer ($\{t_l^{(k+1)}\}_{l=1}^{L}$), and multiplier update layer ($\{\alpha_l^{(k+1)}\}_{l=1}^{L}$). More details related to the construction and organization of the layers in ADMM-net can be referred to Ref. 22. The

ADMM-net takes the sub-sampled k-space data as the input and finally generates the reconstructed image through the iterative process.

I.D.3. Training Parameters. The ADMM-net adopts the normalized mean square error (NMSE) as the loss function to optimize the neural network. The image reconstructed from fully-sampled k-space data was used as the reference image and the corresponding under-sampling data in the k-space was used as the input. The loss function is defined as

$$L_{NMSE} = \frac{1}{N_1} \sum_{n_1=1}^{N_1} \frac{\sqrt{\|\hat{f}_{n_1}(\Theta) - f_{n_1}\|_F^2}}{\sqrt{\|f_{n_1}\|_F^2}} \;, \tag{S.1.9}$$

where $\hat{f}_{n_1}$ and f_{n_1} are the generated image from ADMM-net and the reference image (as the label), respectively. N_1 is the number of training samples. Θ denotes parameters needed to be optimized in ADMM-net. The L-BFGS algorithm was used to minimize the loss function L_{NMSE}.

I.D.4. Training Data. The ADMM-net is trained with brain and chest MR image datasets (https://my.vanderbilt.edu/masi/workshops/). For each dataset, 100 images were randomly selected for training and 50 images for testing. In our experiments, all the under-sampled k-space data were generated with the radial mask of a 20% subsampling rate, as shown in Fig. S8. All the codes are in MATLAB with Intel core i7-4790k CPU, and the training and testing datasets were downloaded from https://github.com/yangyan92/Deep-ADMM-Net.
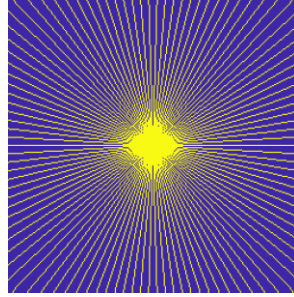

Fig. S8. Radial sampling mask of a 20% subsampling rate.

I.D.5. Testing Data. To validate the performance of ADMM-net about small structural changes, the same images as those in Fig. 2 in the main text of the paper with the symbols "CAN YOU SEE IT" and "HELLO NATURE" were employed with the radial mask of a 20% subsampling rate.

II. CS Based Reconstruction Methods

II.A. CS-inspired Reconstruction

II.A.1. Narrative. To demonstrate the advantages of our ACID in terms of stability against the benchmark compressed sensing (CS)-based methods[28-31], the related experiments are performed, and the reconstruction results are provided using the established methods[32-34]. Since the total variation minimization (individual or combination) is popular in the image reconstruction field with consideration of sparsity prior, it is respectively chosen for CT and MRI image reconstruction in this study. The specific details are given as follows.

II.A.2. X-ray CT Reconstruction. The re-weighting technique[35] combining both shearlets[36] and TV[37] was proposed to validate the stability in Ref. 6. In this study,

it served as a state-of-the-art CS-based comparison method for CT. The details can be found in Refs. 37 and 6. Here we only provide a brief summary as follows.

The mathematical model for this method is formulated as

$$\underset{f}{argmin}\frac{\omega}{2}\|Af - p\|_F^2 + \sum_{j=1}^{J}\vartheta_j\|W_j\psi_j f\|_1 + TGV_\varrho(f), \tag{S.2.1}$$

where ϑ_j represents the j-th balance factor, W_j is a diagonal matrix, and ψ_j represents the j-th subband from the corresponding shearlet transformation. The $TGV_\varrho(f)$ stands for the total generalized variation with the parameter ϱ. $TGV_\varrho(f)$ is combined with the components from the first and second orders of the total variation of the reconstructed image. Furthermore, the parameters ϱ is introduced to balance these two terms. $\omega > 0$ is to balance data fidelity and regularization of sparsity prior.

To solve the optimization problem (S.2.1), $d = \psi'f$ is introduced to represent the matrix format of $\sum_{j=1}^{J}\psi_j f$ and (S.2.1) is split into three sub-problems:

$$\{f^{(k+1)}, d^{(k+1)}\} = \underset{f,d}{argmin}\frac{\omega}{2}\|Af - y^{(k)}\|_F^2 + \|Wd\|_1 + \frac{\omega_1}{2}\|d - \psi'f - b^{(k)}\|_F^2 +$$
$$TGV_\varrho(f), \tag{S.2.2}$$
$$b^{(k+1)} = b^{(k)} + \psi'f^{(k+1)} - d^{(k+1)}, \tag{S.2.3}$$
$$y^{(k+1)} = y^{(k)} + p - Af^{(k+1)}. \tag{S.2.4}$$

where W is the matrix format of $\vartheta_j W_j$. In (S.2.2)-(S.2.4), the four variables are updated iteratively. First, the minimization problem in (S.2.2) is optimized utilizing the multiple non-linear block Gauss-Seidel iterations[38]. Compared with the original re-weighting strategy[35], the weights in W are not only updated after convergence to the solution of (S.2.2), but also are put into the following split process. This unique weight updating strategy is further described in Ref. 6. In this work, the same strategy and configuration[6] were used (including the parameters, the number of iterations, etc.). Note that the number of iterations and the regularization parameters can be further optimized.

II.A.3. MRI Reconstruction. By extending the iteratively regularized Gauss–Newton method (IRGN) with variational penalties[39,40], the total generalized variation (TGV) based IRGN (IRGN-TGV) was proposed[41], and better reconstruction quality was achieved by combining estimation of image and coil sensitivities with TGV regularization. Indeed, the IRGN-TGV had superior noise suppression because of the TGV regularization. In addition, the IRGN-TGV approach can remove sampling artifacts arising from pseudorandom and radial sampling patterns. In this study, it was employed as a state-of-the-art to perform CS-based MRI experiments. Here we also give a brief summary of this method.

Mathematically, MRI is a typical inverse problem with the sampling operator A and the correspondingly k-space data p from the receivers. Besides, the spin density is given as h, and c represents the unknown set of coil sensitivities. For the current iteration index k with the given $f^{(k)} := (h^{(k)}, c^{(k)})$, the solution $\Delta f := (\Delta h, \Delta c)$ is sought to minimize the following objective function:

$$\underset{\Delta f}{argmin}\frac{1}{2}\|A'(f^{(k)})\Delta f + A(f^{(k)}) - y\|_F^2 + \frac{\alpha_k}{2}\|W_1(c^{(k)} + \Delta c)\|_1 + \rho_k R(h^{(k)} + \Delta h).$$
$$\tag{S.2.5}$$

Given $\alpha_k > 0$, $\rho_k > 0$, we have $f^{(k+1)} := f^{(k)} + \Delta f$, $\alpha_{k+1} = q_a\alpha_k$ and $\rho_{k+1} =$

$q_b \rho_k$ and $0 < q_a, q_b < 1$. $\boldsymbol{A}'\left(\boldsymbol{f}^{(k)}\right)$ represents the derivative of $\boldsymbol{A}\left(\boldsymbol{f}^{(k)}\right)$ with respect to $\boldsymbol{f}^{(k)}$. The term $W_1\left(\boldsymbol{c}^{(k)} + \Delta \boldsymbol{c}\right)$ represents the penalty on the Fourier coefficients, and R is a regularization term. In the original IRGN method, the conventional L_2 was considered. Since the TV regularization can introduce stair-casing artifacts and reduce the image quality if the penalty parameter is too large, the authors of Ref. 41 considered the second-order TGV (total generalization variation, TGV), which is a generalized TV. Compared with the conventional TV, the TGV avoids stair-casing in regions of smooth signal changes and improves the image quality[30,42]. Therefore, the authors of Ref. 41 employed TGV in IRGN and then generated IRGN-TGV for MRI. More details are in Ref. 30 and the corresponding code can be downloaded from https://www.tugraz.at/fileadmin/user_upload/Institute/IMT/files/misc/irgntv.zip. The parameters can be further tuned, depending on experimental designs[30].

II.B. Dictionary Learning-Based Reconstruction

II.B.1. Narrative. As a successful example, dictionary learning-based methods were developed for tomographic reconstruction, including MRI[43-45], Optical Coherence Tomography[46-49] and CT[33,50-52]. Dictionary learning based reconstruction methods explored the intrinsic properties using the trained dictionary with initial reconstruction results. The reconstruction process is usually divided into two steps: dictionary learning and image reconstruction. Without loss of generality, we compare the dictionary learning-based reconstruction method with our proposed ACID for CT and MRI.

II.B.2. Dictionary Learning Model. A number of image patches $\boldsymbol{f}_{i_d} \in \mathcal{R}^{s \times s}$, $i_d = 1, \ldots, I_d$, are extracted from the training datasets $\boldsymbol{f}$, and s represents the size of image patches. The set of $\boldsymbol{f}_{i_d}$, $i_d = 1, \ldots, I_d$ is employed to train the global dictionary $\boldsymbol{D}_{ic} \in \mathcal{R}^{S \times T_d}$, where $S = s \times s$ and T_d is the number of atoms. The aim of dictionary learning is to search representation coefficients with sparse-level space constrained by $\boldsymbol{q} \in \mathcal{R}^{T_d \times I_d}$ based on the dictionary $\boldsymbol{D}_{ic}$. It can be explained by solving the following optimization expression:

$$\{\boldsymbol{D}_{ic}^*, \boldsymbol{q}^*\} = argmin \ \frac{1}{2}\sum_{i_d=1}^{I_d}\left\| \boldsymbol{f}_{i_d} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d} \right\|_2^2 \quad \text{s.t.} \ \left\| \boldsymbol{q}_{i_d} \right\|_0 \le L_{dl}, \qquad \text{(S.2.6)}$$

where L_{dl} is the sparsity level of dictionary learning, $\|\cdot\|_0$ represents the quasi-l_0 norm, $\boldsymbol{q}_{i_d} \in \mathcal{R}^{T_d \times 1}$ represents sparse representation coefficients for the i_d-th image patch. (S.2.6) is a constrained problem, and it is equivalent to the following unconstrained problem under a certain condition:

$$\{\boldsymbol{D}_{ic}^*, \boldsymbol{q}^*\} = argmin \left(\Sigma_{i_d=1}^{I_d} \left(\frac{1}{2} \left\| \boldsymbol{f}_{i_d} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d} \right\|_2^2 + h_{i_d} \left\| \boldsymbol{q}_{i_d} \right\|_0 \right) \right), \quad \text{(S.2.7)}$$

where h_{i_d} represents a Lagrange multiplier, which needs to be optimized. Furthermore, (S.2.7) can be solved by an alternating minimization scheme. First, we need to update $\boldsymbol{q}_{i_d}$ with a fixed dictionary $\boldsymbol{D}_{ic}$,

$$\boldsymbol{q}^* = argmin \ \Sigma_{i_d=1}^{I_d} \left(\frac{1}{2} \left\| \boldsymbol{f}_{i_d} - \boldsymbol{D}_{ic}\boldsymbol{q}_{i_d} \right\|_2^2 + h_{i_d} \left\| \boldsymbol{q}_{i_d} \right\|_0 \right). \qquad \text{(S.2.8)}$$

(S.2.8) can be solved using the matching pursuit (MP)[53] or orthogonal matching pursuit (OMP) algorithm[54]. Then, we can update the dictionary with a fixed sparse representation coefficients $\boldsymbol{q}$. Many methods can be employed to train the dictionary $\boldsymbol{D}_{ic}$, such as K-SVD[55], discriminate K-SVD[56], coupled dictionary training[57], online learning technique[58] and online robust learning[59]. In this study, the K-SVD was employed.

II.B.3. Dictionary Learning-Based CT Reconstruction. The conventional dictionary learning was first employed to MR reconstruction from under-sampled k-space data[43]. Then, the dictionary learning was utilized to low-dose CT imaging in our previous work[33], few-view CT reconstruction[28] and material decomposition[60]. In this study, we only consider the dictionary learning-based sparse data CT reconstruction. The mathematical model of dictionary learning-based CT reconstruction can be written as follows:

$$\underset{f,q}{\operatorname{argmin}} \frac{1}{2}\left\|p^{(0)} - Af\right\|_2^2 + \varsigma \sum_{i_d=1}^{I_d} \left(\frac{1}{2}\left\|\wp_{i_d}f - D_{ic}q_{i_d}\right\|_2^2 + h_{i_d}\left\|q_{i_d}\right\|_0\right), \quad \text{(S.2.9)}$$

where $\varsigma > 0$ represents the regularization penalty parameter. $\wp_{i_d}$ is an operator to extract i_d–th image patch from f. Regarding the optimization of (S.2.9), there are many strategies to reach such a goal. Here, the split-Bregman method is used to obtain its solution. First, we introduce a new variable b to replace f and (S.2.9) can be converted into the following constraint programming problem

$$\underset{f,q}{\operatorname{argmin}} \frac{1}{2}\left\|p^{(0)} - Af\right\|_2^2 + \varsigma \sum_{i_d=1}^{I_d} \left(\frac{1}{2}\left\|\wp_{i_d}b - D_{ic}q_{i_d}\right\|_2^2 + h_{i_d}\left\|q_{i_d}\right\|_0\right), s.t., f = b.$$

$$\text{(S.2.10)}$$

To optimize (S.2.10), it can be further converted into

$$\underset{f,b,\chi,q}{\operatorname{argmin}} \frac{1}{2}\left\|p^{(0)} - Af\right\|_2^2 + \varsigma \sum_{i_d=1}^{I_d} \left(\frac{1}{2}\left\|\wp_{i_d}b - D_{ic}q_{i_d}\right\|_2^2 + h_{i_d}\left\|q_{i_d}\right\|_0\right) +$$

$$\frac{\varsigma_1}{2}\|f - b - \chi\|_2^2, \quad \text{(S.2.11)}$$

where $\varsigma_1 > 0$ represents the coupling factor, and χ is the error feedback. In (S.2.11), there are four variables f, b, q and χ. It can be split into the following three sub-problems:

$$\underset{f}{\operatorname{argmin}} \frac{1}{2}\left\|p^{(0)} - Af\right\|_2^2 + \frac{\varsigma_1}{2}\left\|f - b^{(k)} - \chi^{(k)}\right\|_2^2, \quad \text{(S.2.12)}$$

$$\underset{b,q}{\operatorname{argmin}} \sum_{i_d=1}^{I_d} \left(\frac{1}{2}\left\|\wp_{i_d}b - D_{ic}q_{i_d}\right\|_2^2 + h_{i_d}\left\|q_{i_d}\right\|_0\right) + \frac{\varsigma_1}{2}\left\|f^{(k+1)} - b - \chi^{(k)}\right\|_2^2,$$

$$\text{(S.2.13)}$$

$$\chi^{(k+1)} = \chi^{(k)} - \tau_d\left(f^{(k+1)} - b^{(k+1)}\right), \quad \text{(S.2.14)}$$

where $\tau_d > 0$ represents the step length and it was set to 1 in this study. Regarding (S.2.12), it is solved by using the separable surrogate method[61]

$$f_{j_1j_2}^{(k+1)} = f_{j_1j_2}^{(k)} - \frac{\left[A^T\left(Af^{(k)} - p^{(0)}\right)\right]_{j_1j_2} + \varsigma_1\left[f^{(k)} - b^{(k)} - \chi^{(k)}\right]_{j_1j_2}}{\left[A^TA + \varsigma_1\right]_{j_1j_2}}, \quad \text{(S.2.15)}$$

where $[.]_{j_1j_2}$ represents the $(j_1, j_2)^{th}$ pixel in the matrix. In practice, (S.2.15) was performed using two steps:

$$f_{j_1j_2}^{\left(k+\frac{1}{2}\right)} = f_{j_1j_2}^{(k)} - \frac{\left[A^T\left(Af^{(k)} - p^{(0)}\right)\right]_{j_1j_2}}{\left[A^TA + \varsigma_1\right]_{j_1j_2}}, \quad \text{(S.2.16)}$$

and

$$f_{j_1j_2}^{(k+1)} = f_{j_1j_2}^{\left(k+\frac{1}{2}\right)} - \frac{\varsigma_1\left[f^{(k)} - b^{(k)} - \chi^{(k)}\right]_{j_1j_2}}{\left[A^TA + \varsigma_1\right]_{j_1j_2}}. \quad \text{(S.2.17)}$$

In fact, the number of iterations for $f_{j_1j_2}^{\left(k+\frac{1}{2}\right)}$ in (S.2.16) needs to be set to a good number (it was set to 10 in this study), and then $f_{j_1j_2}^{(k+1)}$ is updated. Since ς_1 is specific to scanning geometry, it is normalized into a new parameter γ_1 so that $\varsigma_1 = \gamma_1\|A^TA\|$; that is, we only need to select a geometrically-invariant γ_1. Regarding the optimization of (S.2.13), it is a typical dictionary learning-based

signal recovery problem, and there are a large number of algorithms to solve this problem[55,57]. To control the image recovery via dictionary learning, the parameters of sparsity level L_{dl} and the precision level ζ should be chosen; for more details, see our previous studies[28,33,62].

II.B.4. Dictionary Learning-Based MRI Reconstruction. The conventional dictionary learning methods are common for MR reconstruction[43,63-65]. In this study, the dictionary learning-based MRI (DLMRI)[43] was employed to highlight the advantages of the ACID with built-in DAGAN and TV. Regarding the reconstruction process of DLMRI, it is similar to the process for CT reconstruction. It is also divided into two steps: dictionary learning and image updating. Regarding the dictionary learning step, both MRI and CT are the same except that training images are different. Again, the dictionary used in CT reconstruction was trained from FBP results or updated results within the iteration process. In contrast, the dictionary utilized in MRI was trained from the inverse Fourier transform results. Regarding the image updating step, it is not necessary to update the image based on the fast Fourier transform. More details can be found in Ref. 43.

II.B.5. Experimental Results. To validate the outperformance of ACID in comparison with the dictionary learning-based CT reconstruction method (DLCT), we repeated the experiments design for the cases C1 and C2. Here, we adopt the FBP method to reconstruct images. Then, the FBP results were employed to train the dictionaries. In this study, only 1.0×10^4 image patches were extracted from FBP images to train the dictionary by the K-SVD algorithm. The size of extracted image patches was set to 6×6. The dictionary $\boldsymbol{D}_{ic}$ is overcomplete, and it can benefit the sparsity level enforcement. The number of atoms was set to 512. The sparsity level L_{dl} in the dictionary training can be set empirically, and it was chosen as 6. The number of iterations for the training dictionary was set to 100.

Note that the total variation is still treated as the compressed sensing-based sparsity for the built-in component in the ACID. Here, the parameters of γ_1, L_{dl} and ζ in DLCT were set to 0.22, 8 and 0.06, respectively. The number of outer iterations was set to 200. The implementation environment for training and reconstruction is the same as Ell-50. Specifically, the computational costs of dictionary training and reconstruction consume 139 and 561 seconds. However, the whole ACID with the built-in Ell-50 consumes about 70.5 seconds. In other words, the ACID is faster than the DLCT method.

The reconstruction results from DLCT and ACID with C1 and C2 are in Fig. S9. It is observed that DLCT provides higher image quality than that obtained by the CS method. However, it is still worse than those obtained by the ACID. Besides, the proposed ACID method obtains better image edges and avoids blurred artifacts compared with the DLCT method. Especially, the insert texts in DLCT results (i.e., "CAN U SEE IT" and "A BIRD?") are very blurry, and they failed to be discriminated against. These texts are clearly observed in ACID results. Regarding small features (*i.e.*, the symbol " "), they are totally missing in the DLCT result. However, they were still recovered by the ACID. In terms of quantitative assessment, our proposed ACID obtained the best results remarkably. More details for codes and test data are at https://zenodo.org/record/5497811.

To show the advantages of ACID with the built-in DAGAN, the reconstruction results in the M2 case from ACID and DLMRI are given in Fig. S10. The DLMRI obtained higher image quality than that obtained by the conventional CS method in the main text. However, it is still worse than those obtained by the ACID. Besides, the proposed ACID method obtained better image edges and avoided blurry artifacts compared with the DLMRI method. Especially, the inserted texts in DLCT results (i.e., "CAN U SEE IT") are very blurry, and they could not be discriminated. These texts are clearly observed in the ACID results. The small symbol was totally blurred in DLMRI result, which was still recovered by ACID. In terms of quantitative assessment, our proposed ACID obtained better results than those achieved by DLMRI method. The MATLAB code of the MRIDL method can be downloaded from http://www.ifp.illinois.edu/~yoram/DLMRI-Lab/DLMRI.html. The reconstruction parameters within DLMRI were optimized. Regarding the computational cost, under the same computing environment, DLMRI took 606.3 seconds, which is higher than that of CS reconstruction methods in the main text (i.e., 127.8 seconds). Our proposed ACID only took 9.2 seconds.

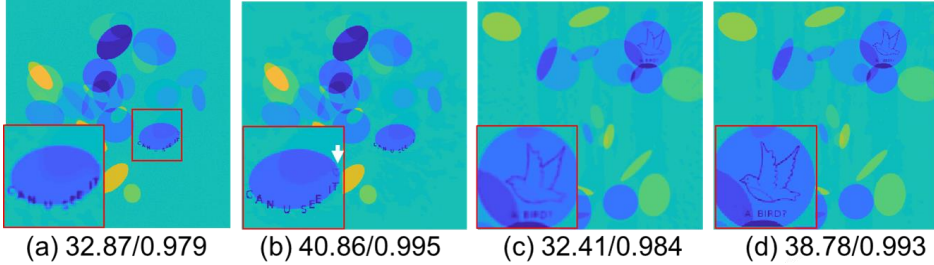(a) 32.87/0.979   (b) 40.86/0.995   (c) 32.41/0.984   (d) 38.78/0.993

Fig. S9. Comparison study on the DLCT and ACID methods. (a) and (c), (b) and (d) are reconstructed results from DLCT and ACID, respectively. The numbers represent the quantitative results in terms of PSNR and SSIM, and the display window is [-150, 150] HU.
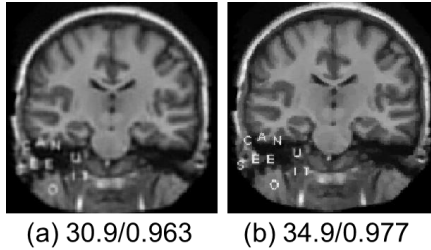

(a) 30.9/0.963   (b) 34.9/0.977

Fig. S10. Comparison study on the DLMRI and ACID methods. (a) and (b) are the reconstructed results from DLCT and ACID respectively. The numbers represent the quantitative results in terms of PSNR and SSIM.

III. ACID Implementation & More Results

III.A. ACID Implementation

For an implementation of the whole ACID system, it is considered as an iterative framework and listed in Algorithm 1. In the whole ACID framework, we need to have input data $p^{(0)}$, a neural network Φ and a system matrix A. Then, we should specify the stopping condition; i.e., the maximum number of iterations K. Finally, the parameters λ and ε should be given to control the iterative process and the regularization strength, all of which can be empirically picked up. When k=1, we need to compute $\Phi(p^{(0)})$ and then normalize $\Phi(p^{(0)})$. The goal of the normalization operator is to facilitate the adjustment of the regularization parameters for different applications. Then, we obtain the updated $b^{(1)}$ using

the second formula in (1) in the main body of this paper. Next, $\boldsymbol{f}^{(1)}$ is updated by de-normalizing $\boldsymbol{b}^{(1)}$. When 1<k<K+1, we need to compute the residual data by $\boldsymbol{p}^{(k+1)} = \frac{\lambda(\boldsymbol{p}^{(0)} - A\boldsymbol{f}^{(k)})}{1+\lambda}$. Since the residual data are not in the dynamic range of the original data, the residual data should be normalized into the original range to make sure the efficiency of the neural network (Line #9 in Algorithm 1). After the neural network predicts a residual image, the de-normalization operator should be applied on the prediction to ensure the consistency of the reconstruction results. Then, $\boldsymbol{f}^{(k)} + \frac{1}{\lambda}\Phi(\boldsymbol{p}^{(k+1)})$ is normalized and fed to the compressed sensing-based regularization module to encourage image sparsity. Finally, we obtain the updated image $\boldsymbol{f}^{(k+1)}$ after the de-normalization. More details on our codes and other materials are available at https://zenodo.org/record/5497811.

Algorithm 1. Pseudocode of the ACID workflow.

Input: Data $\boldsymbol{p}^{(0)}$, neural network Φ, system matrix $\boldsymbol{A}$, maximum number of iterations K, auxiliary parameters λ, ε, and k=1;
1. If k<K+1 do
2.   if k=1 do
3.   Computing $\Phi(\boldsymbol{p}^{(0)})$;
4.   Normalizing $\Phi(\boldsymbol{p}^{(0)})$;
5.   Updating $\boldsymbol{b}^{(1)}$ where the normalized $\Phi(\boldsymbol{p}^{(0)})$ is treated as the input;
6.   Updating $\boldsymbol{f}^{(1)}$ by de-normalizing $\boldsymbol{b}^{(1)}$;
7.   else do
8.   Computing residual data using $\boldsymbol{p}^{(k+1)} = \frac{\lambda(\boldsymbol{p}^{(0)} - A\boldsymbol{f}^{(k)})}{1+\lambda}$;
9.   Normalizing the residual data $\boldsymbol{p}^{(k+1)}$ into the input range of neural network to obtain $\overline{\boldsymbol{p}}^{(k+1)}$;
10.   Inputting $\overline{\boldsymbol{p}}^{(k+1)}$ into the neural network Φ and obtaining $\Phi(\overline{\boldsymbol{p}}^{(k+1)})$;
11.   De-normalizing $\Phi(\overline{\boldsymbol{p}}^{(k+1)})$ to obtain $\Phi(\boldsymbol{p}^{(k+1)})$;
12.   Normalizing $\boldsymbol{f}^{(k)} + \frac{1}{\lambda}\Phi(\boldsymbol{p}^{(k+1)})$
13.   Updating $\boldsymbol{b}^{(k+1)}$;
14.   Updating $\boldsymbol{f}^{(k+1)}$ by de-normalizing $\boldsymbol{b}^{(k+1)}$;
15.   end
16. end
17. return $\boldsymbol{f}^{(K)}$
Output: Reconstructed image $\boldsymbol{f}^{(K)}$

III.B. Difference Images for Figures 2-4

Here we provide the difference images for figures 2-4 in the main text. Fig. S11 includes the difference images of Figure 2 in the main text. It can be seen that the results provided by ACID is closer to the ground truth. In addition, the differences of the text symbols in the competing approaches are more obvious than that obtained by our ACID. It further demonstrates our ACID can effective stabilize the deep tomographic network against structure changes.

Fig. S12 includes the difference images of Figure 3 in the main text. It can be seen that the results provided by ACID is also closer to the ground truth. In addition, the differences of inserted text symbols in the competing algorithms are more obvious than that obtained by our ACID. Fig. S12 further demonstrates our ACID can effectively stabilize the deep tomographic network against adversarial attacks. The difference images of ACID against noise are also provided in Fig. S13. It can be observed that the difference images with our ACID are the smallest among all the algorithms.
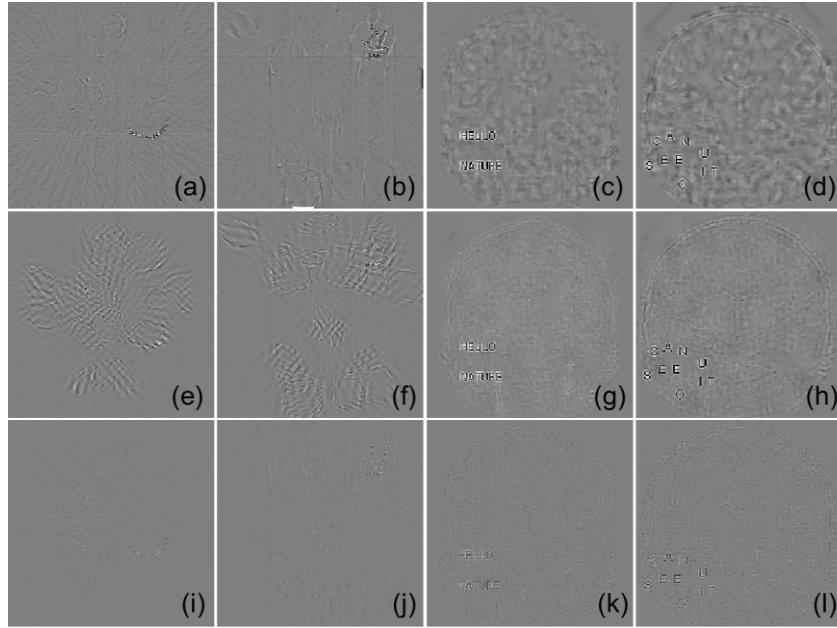
Fig. S11. **Difference images of ACID with small structural changes in the CT and MRI cases.** The 1st- 3rd rows are the difference images of Ell-50, CS-inspired and ACID results respectively with respect to the ground truth. The 1st-4th columns correspond to CT cases C1 and C2, and MRI cases M1 and M2, respectively. The sub-sampling rate of MRI is 10%. The display windows for CT and MRI are [-70 70] HU and [-0.5 0.5], respectively.
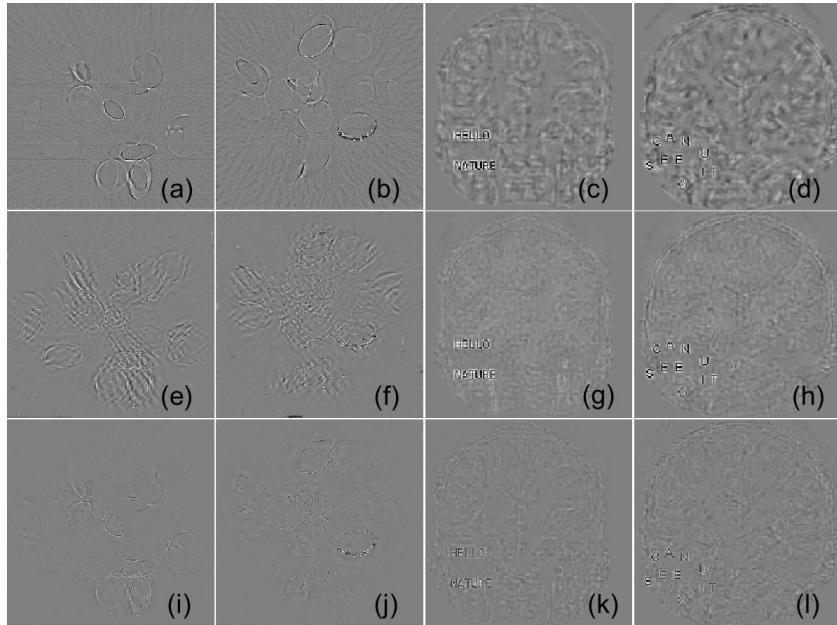

Fig. S12. **Difference images of ACID with adversarial attack in the CT and MRI cases.** The 1st -3rd rows are the difference images of Ell-50, CS-inspired and ACID results respectively with respect to the ground truth. The 1st-4th columns correspond to CT cases C3 and C4, and MRI cases M3 and M4, respectively. Each CT dataset contains 50 projections, and the sub-sampling rate of MRI is 10%. The display windows for CT and MRI are [-70 70] HU and [-0.5 0.5], respectively.

III.C. ACID Performance on Real CT Dataset

To validate the proposed ACID on real datasets, here we retrained the

FBPConvNet on mayo clinical datasets and the reconstruction results are
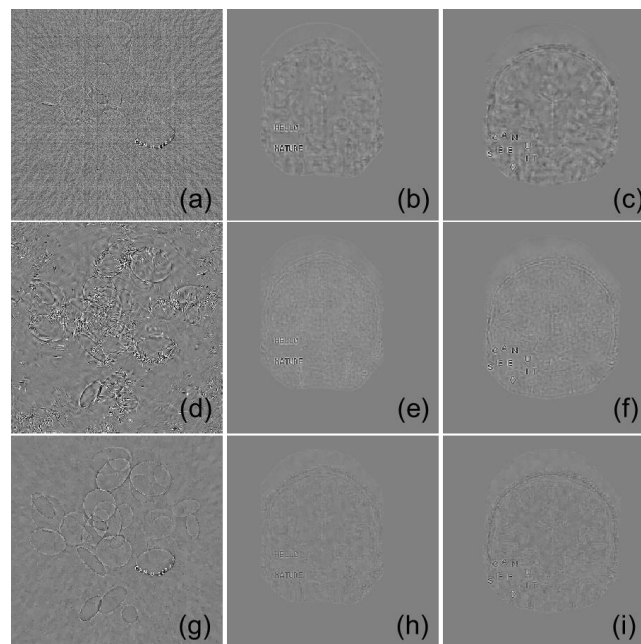

Fig. S13. **Difference images in the C5, M5 and M6 cases against noise.** In the first column, (a), (d) and (g) present the difference images between ground truth and that of Ell-50, CS and ACID results on C5. The second and third columns are the counterparts of M5 and M6 showing the difference images between the ground truth and that of DAGAN, CS and ACID results, respectively.
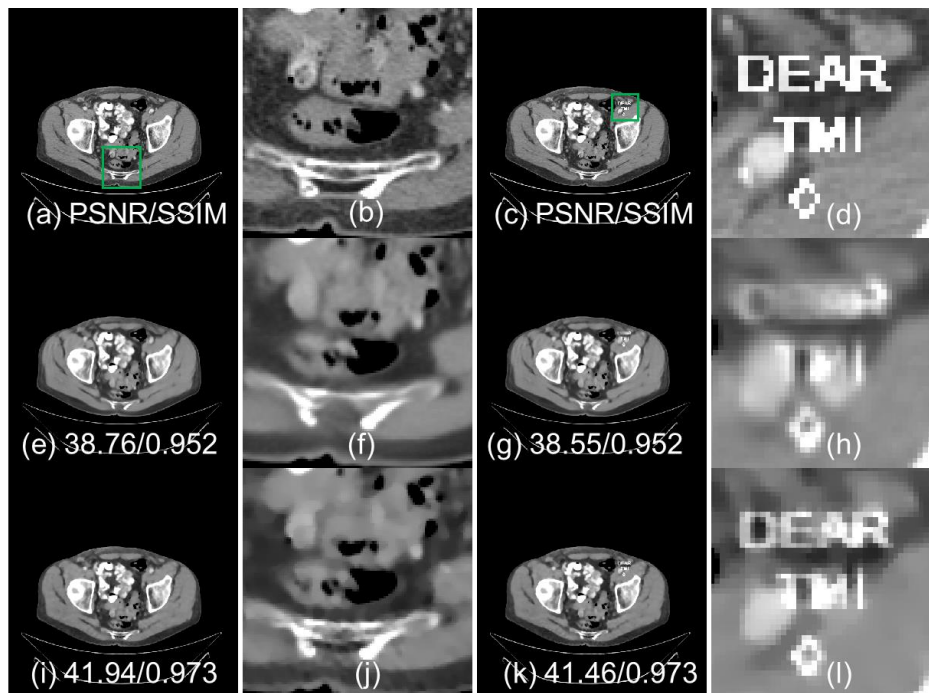

Fig. S14. **Reconstruction performance of ACID on Mayo clinical datasets.** In the first column, (a), (e) and (i) are the ground truth, FBPConvNet and ACID results. The second column are the magnified ROI in the first column. The third and fourth columns are the counterparts of the first and second columns with structural changes. The numbers indicate the PSNR and SSIM values, and the display window is [-160, 240] HU.

further given in Fig. S14. There are two cases in Fig. S14, where one has no

structural changes and the other has structural changes. For the case without structural changes, our ACID built with FBPConvNet can provide higher reconstructed image quality than FBPConvNet itself. For the case with structural changes, our ACID obviously provides clearer insert symbols than FBPConvNet. Furthermore, ACID has higher quantitative PSNR and SSIM results than the FBPConvNet.

III.D. ACID Against Distributional Robustness

As for robustness, distributional robustness is very important for image reconstruction. To further demonstrate the advantages of the generalization ability from our ACID, here we first use the test dataset of DAGAN to test AUTOMAP. Then, we also use the test datasets from AUTOMAP to test DAGAN. The results of Fig. S15 demonstrate the DAGAN has a relative weakness distributional robustness since its results contain several structural artifacts. However, these artifacts can be removed by our ACID framework with built-in DAGAN network. In addition, our ACID also provides higher PSNR as well as SSIM. The results of Fig. S16 demonstrate the AUTOMAP has good distributional robustness, but it also contains structure and other artifacts. However, these artifacts can be removed by our ACID framework with built-in AUTOMAP network. In addition, our ACID also provides higher PSNR as well as SSIM.
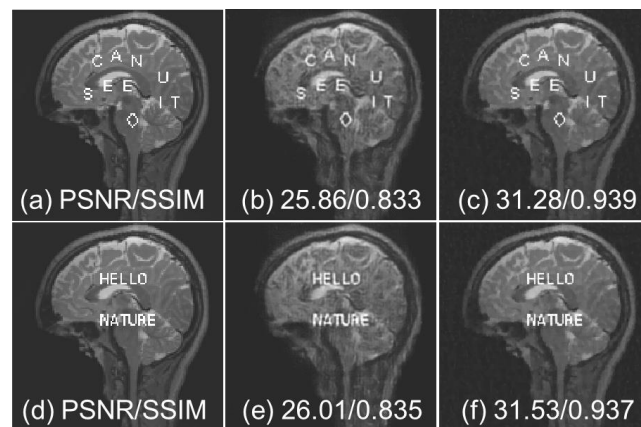

Fig. S15. Distributional robustness of DAGAN against AUTOMAP test datasets. (a)-(c) are the ground truth, DAGAN result and ACID result with built-in DAGAN network. (d)-(f) are the counterparts of (a)-(c) for another case. The numbers indicate PSNR and SSIM values.
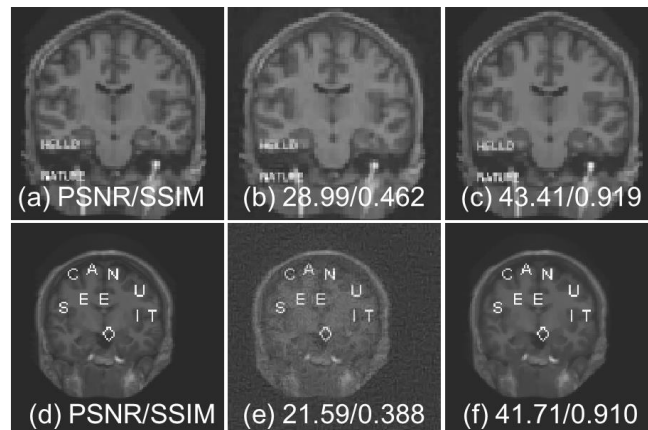

Fig. S16. Distributional robustness of AUTOMAP against DAGAN test datasets. (a)-(c) are ground truth, AUTOMAP result and ACID result with built-in AUTOMAP network. (d)-(f) are

counterparts of (a)-(c) for another case. The numbers indicate PSNR and SSIM values.

References

1   Jin, K. H., McCann, M. T., Froustey, E., and Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing,* 26, 4509-4522. 10.1109/TIP.2017.2713099.

2   He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, p.*770-778. 10.1109/CVPR.2016.90.

3   Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. in *International Conference on Medical Image Computing and Computer-assisted Intervention, LNCS, Vol. 9531, p.*234-241.

4   Vedaldi, A., and Lenc, K. (2015). Matconvnet: Convolutional neural networks for matlab. in *Proceedings of the 23rd ACM international conference on Multimedia,* p.689-692. 10.1145/2733373.2807412.

5   Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. in *International conference on machine learning*, PMLR, 28(3), 1310-1318.

6   Antun, V., Renna, F., Poon, C., Adcock, B., and Hansen, A. C. (2020). On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences,* 117, 30088-30095. 10.1073/pnas.1907377117.

7   Yu, H.Y., Zhao, S.Y., and Wang, G. (2005). A differentiable Shepp–Logan phantom and its applications in exact cone-beam CT. *Physics in Medicine & Biology,* 50, 5583-5595. 10.1088/0031-9155/50/23/012.

8   Averbuch, A., Sedelnikov, I., and Shkolnisky, Y. (2011). CT reconstruction from parallel and fan-beam projections by a 2-D discrete Radon transform. *IEEE Transactions on Image Processing,* 21, 733-741. 10.1109/TIP.2011.2164416.

9   Zhang, Z., Liang, X., Dong, X., Xie, Y., and Cao, G. (2018). A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE Transactions on Medical Imaging,* 37, 1407-1417. 10.1109/TMI.2018.2823338.

10   Zang, G., Aly, M., Idoughi, R., Wonka, P., and Heidrich, W. (2018). Super-resolution and sparse view CT reconstruction. in *Proceedings of the European Conference on Computer Vision (ECCV),* p.137-153.

11   Li, H.Y., Zhao, T.H., Wei, M.L., Ruan, H.X., Shuang, Y., Cui, T.J., Del Hougne, P., and Li, L.L (2020). Intelligent electromagnetic sensing with learnable data acquisition and processing. *Patterns* 1, Article ID: 100006. 10.1016/j.patter.2020.100006.

12   Gottschling, N. M., Antun, V., Adcock, B., and Hansen, A. C. (2020). The troublesome kernel: why deep learning for inverse problems is typically unstable. *arXiv preprint, arXiv:2001.01258*.

13   Yang, G., Yu, S.M., Dong, H., Slabaugh, G., Dragottyi, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., *et al.* (2017). DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Transactions on Medical Imaging,* 37, 1310-1321. 10.1109/TMI.2017.2785879.

14   Hwang, H., Rehman, H.Z.U., and Lee, S. (2019). 3D U-Net for skull stripping in brain MRI. *Applied Sciences,* 9, Article ID: 569, 515 pages. 10.3390/app9030569.

15   Dolz, J., Desrosiers, C., and Ayed, I.B. (2018). IVD-Net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet. in *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging,* p.130-143, Springer.

16   Kermi, A., Mahmoudi, I., and Khadir, M.T. (2018). Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes. in *International MICCAI Brainlesion Workshop,* p.37-48, Springer.

17   Zhao, J., Mathieu, M., and LeCun, Y. (2016). Energy-based generative adversarial network. *arXiv preprint, arXiv:1609.03126*.

18   You, C.Y., Li, G., Zhang, Y., Zhang, X., Shan, H.M., Li, M.Z., Ju, S.H., Zhao, Z., Zhang, Z.Y., Cong, W.X., *et al.* (2019). CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). *IEEE Transactions on Medical Imaging,* 39, 188-203. 10.1109/TMI.2019.2922960.

19   Wang, G., Ye, J. C., Mueller, K., and Fessler, J.A. (2018). Image reconstruction is a new frontier of machine learning. *IEEE Transactions on Medical Imaging,* 37, 1289-1296. 10.1109/TMI.2018.2833635.

20   Yang, Q., Yan, P.K., Zhang, Y.B., Yu, H.Y., Shi, Y.Y., Mou, X.Q., Kalra, M.K., Zhang, Y., Sun., L., and Wang, G. (2018). Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging,* 37, 1348-1357. 10.1109/TMI.2018.2827462.

21   Kingma, D.P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint, arXiv:1412.6980*.

22   Yang, Y., Sun, J., Li, H., and Xu, Z (2016). Deep ADMM-Net for compressive sensing MRI. in *Advances in neural information processing systems,* p.10-18.

23   Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., and Rosen, M.S. (2018). Image reconstruction by domain-transform manifold learning. *Nature,* 555, 487-492. 10.1038/nature25988.

24   Liu, J., Lefebvre, A., and Nadar, M. (2013). Alternating direction of multipliers method for parallel MRI reconstruction, Google Patents, Patent # 8,879,811.

25   Lustig, M., Donoho, D.L., Santos, J.M. and Pauly, J.M. (2008). Compressed sensing MRI. *IEEE Signal Processing Magazine,* 25, 72-82. 10.1109/MSP.2007.914728.

26   Haldar, J. P., Hernando, D., and Liang, Z.P. (2011). Compressed-sensing MRI with random encoding. *IEEE Transactions on Medical Imaging,* 30, 893-903.

10.1109/TMI.2010.2085084.

27   Shi, W., Ling, Q., Yuan, K., Wu, G., and Yin, W. (2014). On the linear convergence of the ADMM in decentralized consensus optimization. *IEEE Transactions on Signal Processing,* 62, 1750-1761. 10.1109/TSP.2014.2304432.

28   Xu, M., Hu, D.L., Luo, F.L., Liu, F.L., Wang, S.Y., and Wu, W.W. (2021). Limited angle X ray CT Reconstruction using Image Gradient $\ell0$ norm with Dictionary Learning. *IEEE Transactions on Radiation and Plasma Medical Sciences,* 5, 78-87. 10.1109/TRPMS.2020.2991887.

29   Ma, S., Yin, W., Zhang, Y., and Chakraborty, A. (2008). An efficient algorithm for compressed MR imaging using total variation and wavelets. in *2008 IEEE Conference on Computer Vision and Pattern Recognition,* p.1-8.

30   Knoll, F., Bredies, K., Pock, T. , and Stollberger, R. (2011). Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine,* 65, 480-491. 10.1002/mrm.22595.

31   Knoll, F., Holler, M., Koesters, T., Otazo, R., Bredies, K., and Sodickson, D.K. (2017). Joint MR-PET reconstruction using a multi-channel image regularizer. *IEEE transactions on medical imaging,* 36, 1-16. 10.1109/TMI.2016.2564989.

32   Yu, H.Y., Wang, G., Hsieh, J., Entrikin, D.W., Ellis, S., Liu, B.D., and Carr, J.J. (2011). Compressive Sensing–Based Interior Tomography: Preliminary Clinical Application. *Journal of Aomputer Assisted Tomography,* 35, 762-764. 10.1097/RCT.0b013e318231c578.

33   Xu, Q., Yu. H.Y., Mou, X.Q., Zhang, L., Hsieh, J. and Wang, G. (2012). Low-dose X-ray CT reconstruction via dictionary learning. *IEEE Transactions on Medical Imaging,* 31, 1682-1697. 10.1109/TMI.2012.2195669.

34   Wu, W.W., Zhang, Y.B., Wang, Q., Liu, F.L., Chen, P.J., and Yu, H.Y. (2018). Low-dose spectral CT reconstruction using image gradient $\ell0$–norm and tensor dictionary. *Applied Mathematical Modelling,* 63, 538-557. 10.1016/j.apm.2018.07.006.

35   Candes, E.J., Wakin, M.B., and Boyd, S.P. (2008). Enhancing sparsity by reweighted $\ell1$ minimization. *Journal of Fourier Analysis and Applications,* 14, 877-905. 10.1007/s00041-008-9045-x.

36   Ma, J., März, M., Funk, S., Schulz-Menger, J., Kutyniok, G., Schaeffter, T., and Kolbitsch, C. (2018). Shearlet-based compressed sensing for fast 3D cardiac MR imaging using iterative reweighting. *Physics in Medicine & Biology,* 63, Article ID: 235004. 10.1088/1361-6560/aaea04.

37   Ma, J., and März, M. (2016). A multilevel based reweighting algorithm with joint regularizers for sparse recovery. *arXiv preprint, arXiv:1604.06941*.

38   Yoon, S., and Jameson, A. (1988). Lower-upper symmetric-Gauss-Seidel method for the Euler and Navier-Stokes equations. *AIAA Journal,* 26, 1025-1026 (1988). 10.2514/3.10007.

39   Block, K.T., Uecker, M., and Frahm, J. (2007). Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine,* 57, 1086-1098. 10.1002/mrm.21236.

40   Ehrhardt, M.J., and Betcke, M.M. (2016). Multicontrast MRI reconstruction with structure-guided total variation. *SIAM Journal on Imaging Sciences,* 9, 1084-1106. 10.1137/15M1047325.

41   Knoll, F., Clason, C., Bredies, K., Uecker, M., and Stollberger, R. (2012). Parallel imaging with nonlinear reconstruction using variational penalties. *Magnetic Resonance in Medicine,* 67, 34-41. 10.1002/mrm.22964.

42   Bredies, K., Kunisch, K., and Pock, T. (2010). Total generalized variation. *SIAM Journal on Imaging Sciences,* 3, 492-526. 10.1137/090769521.

43   Ravishankar, S., and Bresler, Y. (2011). MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging,* 30, 1028-1041. 10.1109/TMI.2010.2090538.

44   Caballero, J., Price, A.N., Rueckert, D., and Hajnal, J.V. (2014). Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Transactions on Medical Imaging,* 33, 979-994. 10.1109/TMI.2014.2301271.

45   Zhan, Z.F., Cai, J.F., Guo, D., Liu, Y., Chen, Z., and Qu, X. (2015). Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction. *IEEE Transactions on Biomedical Engineering,* 63, 1850-1861. 10.1109/TBME.2015.2503756.

46   Fang, L., Li, S., McNabb, R.P., Nie, Q., Kuo, A.N., Toth, C.A., Izatt, J.A., and Farsiu, S. (2013). Fast acquisition and reconstruction of optical coherence tomography images via sparse representation. *IEEE Transactions on Medical Imaging,* 32, 2034-2049. 10.1109/TMI.2013.2271904.

47   Esmaeili, M., Dehnavi, A. M., Rabbani, H., and Hajizadeh, F. (2017). Speckle noise reduction in optical coherence tomography using two-dimensional curvelet-based dictionary learning. *Journal of Medical Signals and Sensors,* 7, 86-91. 10.1364/BOE.377021.

48   Albarrak, A., Coenen, F., and Zheng, Y. (2017). Volumetric image classification using homogeneous decomposition and dictionary learning: a study using retinal optical coherence tomography for detecting age-related macular degeneration. *Computerized Medical Imaging and Graphics,* 55, 113-123. 10.1016/j.compmedimag.2016.07.007.

49   Fang, L., Li, S., Cunefare, D., and Farsiu, S. (2016). Segmentation based sparse reconstruction of optical coherence tomography images. *IEEE Transactions on Medical Imaging,* 36, 407-421. 10.1109/TMI.2016.2611503.

50   Chen, Y., Shi, L., Feng, Q.J., Yang, J., Shu, H.Z., Luo, L.M., Coatrieux, J.L., and Chen, W.F. (2014). Artifact suppressed dictionary learning for low-dose CT image processing. *IEEE Transactions on Medical Imaging,* 33, 2271-2292. 10.1109/TMI.2014.2336860.

51   Fang, R., Chen, T., and Sanelli, P. C. (2013). Towards robust deconvolution of low-dose perfusion CT: Sparse perfusion deconvolution using online dictionary learning. *Medical image analysis,* 17, 417-428. 10.1016/j.media.2013.02.005.

52   Tan, S.Q., Zhang, Y.B., Wang, G., Mou, X.Q., Cao, G.H., Wu. Z.F., and YU, H.Y. (2015). Tensor-based dictionary learning for dynamic tomographic reconstruction. *Physics in Medicine & Biology,* 60, 2803-2818. 10.1088/0031-

9155/60/7/2803.

53   Mallat, S.G., and Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing,* 41, 3397-3415. 10.1109/78.258082.

54   Chen, S., Billings, S A., and Luo, W. (1989). Orthogonal least squares methods and their application to non-linear system identification. *International Journal of Control,* 50, 1873-1896. 10.1080/00207178908953472.

55   Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing,* 54, 4311-4322. 10.1109/TSP.2006.881199.

56   Zhang, Q., and Li, B. (2010). Discriminative K-SVD for dictionary learning in face recognition. in *2010 IEEE computer society conference on computer vision and pattern recognition,* p.2691-2698. 10.1109/CVPR.2010.5539989.

57   Yang, J., Wang, Z., Lin, Z., Cohen, S., and Huang, T. (2012). Coupled dictionary training for image super-resolution. *IEEE Transactions on Image Processing,* 21, 3467-3478. 10.1109/TIP.2012.2192127.

58   Mairal, J., Bach, F., Ponce, J. and Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research,* 11, 19-60. 10.1145/1756006.1756008.

59   Lu, C., Shi, J., and Jia, J. (2013). Online robust dictionary learning. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* p.415-422. 10.1109/CVPR.2013.60.

60   Wu, W.W., Yu, H.J., Chen, P.J., Luo, F.L., Wang, Q., Zhu, Y.N., Zhang, Y.B., Feng, J., and Yu, H.Y. (2020). Dictionary learning based image-domain material decomposition for spectral CT. *Physics in Medicine & Biology,* 65, Article ID: 245006. 10.1088/1361-6560/aba7ce.

61   Elbakri, I. A. and Fessler, J.A. (2002). Statistical image reconstruction for polyenergetic X-ray computed tomography. *IEEE Transactions on Medical Imaging,* 21, 89-99. 10.1109/42.993128.

62   Zhang, Y.B., Mou, X.Q., Wang, G. and Yu, H.Y. (2017). Tensor-based dictionary learning for spectral CT reconstruction. *IEEE Transactions on Medical Imaging* 36(1), 142-154. 10.1109/TMI.2016.2600249.

63   Merlet, S., Caruyer, E., and Deriche, R. (2012). Parametric dictionary learning for modeling EAP and ODF in diffusion MRI. in *International Conference on Medical Image Computing and Computer-Assisted Intervention,* p.10-17, Springer. 10.1007/978-3-642-33454-2_2.

64   Lei, Y., Shu, H.K., Tian, S.B., Jeong, J.J., Liu, T., Shim, H., Mao, H., Wang, T.H., Jani, A.B., Curran, W.J., et al. (2018). Magnetic resonance imaging-based pseudo computed tomography using anatomic signature and joint dictionary learning. *Journal of Medical Imaging,* 5, Article ID: 034001. 10.1117/1.JMI.5.3.034001.

65   Lingala, S.G., and Jacob, M. (2013). Blind compressive sensing dynamic MRI. *IEEE Transactions on Medical Imaging,* 32, 1132-1145. 10.1109/TMI.2013.2255133.