

# DEEP LEARNING-BASED SOLVABILITY OF UNDERDETERMINED INVERSE PROBLEMS IN MEDICAL IMAGING \*

CHANG MIN HYUN<sup>†</sup>, SEONG HYEON BAEK<sup>†</sup>, MINGYU LEE<sup>†</sup>, SUNG MIN LEE<sup>†</sup>, AND  
JIN KEUN SEO<sup>†‡</sup>

**Abstract.** Recently, with the significant developments in deep learning techniques, solving underdetermined inverse problems has become one of the major concerns in the medical imaging domain, where underdetermined problems are motivated by the willingness to provide high resolution medical images with as little data as possible, by optimizing data collection in terms of minimal acquisition time, cost-effectiveness, and low invasiveness. Typical examples include undersampled magnetic resonance imaging(MRI), interior tomography, and sparse-view computed tomography(CT), where deep learning techniques have achieved excellent performances. However, there is a lack of mathematical analysis of why the deep learning method is performing well. This study aims to explain about learning the causal relationship regarding the structure of the training data suitable for deep learning, to solve highly underdetermined problems. We present a particular low-dimensional solution model to highlight the advantage of deep learning methods over conventional methods, where two approaches use the prior information of the solution in a completely different way. We also analyze whether deep learning methods can learn the desired reconstruction map from training data in the three models (undersampled MRI, sparse-view CT, interior tomography). This paper also discusses the nonlinearity structure of underdetermined linear systems and conditions of learning (called  $\mathcal{M}$ -RIP condition).

**Key words.** underdetermined linear inverse problem, deep learning, medical imaging, magnetic resonance imaging, computed tomography

**AMS subject classifications.** 15A29, 65F22, 68T05, 68Q32

**1. Introduction.** In medical imaging, we want to improve our visual ability to provide meaningful expression and useful description of diagnosis and treatment, while optimizing data collection in terms of minimal acquisition time, cost-effectiveness, and low invasiveness. The goal is to find a function  $f$  that maps from inputs (what we measure) to outputs (reconstructing useful medical images):

$$(1.1) \quad f(\text{input data}) = \text{useful output}.$$

The output could be the two- or three-dimensional visual representation of the interior of a body, such as computerized tomography (CT) [30, 50] and magnetic resonance imaging (MRI) [43, 23]. To achieve the output of reasonable resolution and accuracy, we have used a suitable means of measurement (input), which allows to reconstruct the output.

Conventional CT and MRI data collections are designed to obtain a well-posed reconstruction method in the sense that the corresponding forward models are well-posed. The forward model can be expressed as the well-posed linear system as follows:

$$(1.2) \quad \mathbf{A}_{\text{full}}\mathbf{y} = \mathbf{b}_{\text{full}}$$

where  $\mathbf{b}_{\text{full}}$  denotes the “fully sampled” data (e.g, sinogram in CT and k-space data in MRI),  $\mathbf{y}$  denotes the CT or MRI image, and  $\mathbf{A}_{\text{full}}$  denotes the invertible matrix representing discrete Radon transform for CT and discrete Fourier transform for MRI. More precisely, the forward model of CT is based on the assumption that the measured

\*This work was supported by Samsung Science & Technology Foundation (No. SRFC-IT1902-09).

<sup>†</sup>Department of Computational Science and Engineering, Yonsei University, Seoul, Korea

<sup>‡</sup>To whom the corresponding author (seoj@yonsei.ac.kr)

X-ray projection data is the Radon transform of the output image. To achieve the given resolution of the output and invert the corresponding discrete Radon transform, a sufficient number of projection angles are required so that the number of equations (measured data) becomes greater than the number of unknowns (the number of pixels in the image). Hence, given the resolution of the output, the traditional approaches require a certain amount of data acquisition in order to make the reconstruction problem well-posed in the sense of Hadamard[24].

Because of the great needs to reduce the radiation dose in CT and data acquisition time in MRI, considerable attention has been given to solve underdetermined problems (or ill-posed inverse problems) that violate the Nyquist criteria [52] in the sense that the number of equations is much smaller than the number of unknowns. The demand for undersampled MRI is because of the long scan time with the human body trapped in an inconvenient narrow bore; shortening the MRI scan time can increase the satisfaction of patient, reduce the artifacts caused by patient movement, and reduce the medical costs [41, 45, 8, 15, 63]. The need for low-dose CT arises from the cancer risks associated with the exposure of patients to ionizing radiation [33, 42, 6]. A highly underdetermined problem (far less equations than unknowns) corresponding to (1.2) can be expressed as follows:

$$(1.3) \quad \underbrace{\mathbf{A}}_{\mathcal{S}_{\text{sub}}(\mathbf{A}_{\text{full}})} \mathbf{y} = \underbrace{\mathbf{b}}_{\mathcal{S}_{\text{sub}}(\mathbf{b}_{\text{full}})}$$

where  $\mathcal{S}_{\text{sub}}$  denotes a subsampling operator. For example, in undersampled MRI,  $\mathbf{b} = \mathcal{S}_{\text{sub}}(\mathbf{b}_{\text{full}})$  denotes an undersampled k-space data violating the Nyquist sampling criterion, and  $\mathbf{y}$  denotes MRI image reconstructed using a fully sampled k-space data. Because  $\mathbf{A}$  is not an invertible matrix, there exist infinitely many solutions.

Solving the underdetermined problem (1.3) depends on the appropriate use of *a priori* information about medical CT or MRI images as solutions. However, the conventional approaches using prior knowledge, such as regularization and compressed sensing(CS) approaches, may not be appropriate for medical images in which small anomalous details are more important than the overall feature [35]. Currently, deep learning techniques have exhibited excellent achievement in various underdetermined problems such as undersampled MRI, interior tomography, and sparse view CT. They seem to overcome the limitations of the existing mathematical methods in handling various ill-posed problems [36, 25, 31]. It is highly expected that deep learning methodologies will improve their performance, as training data and experience are accumulated over time. However, there is a tremendous lack of a rigorous mathematical foundation that would allow us to understand the reasons for the remarkable performance of deep learning methods [38].

This study aims to provide a systematic basis for learning the causal relationship regarding the structure of the training data suitable for deep learning to solve highly underdetermined problems. The goal of the undersampled problem (1.3) is to find a reconstruction map  $f_b : \mathbf{b} \rightarrow \mathbf{y}$  that maps from the highly undersampled data  $\mathbf{b}$  to the image  $\mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}}$  in (1.2). Here, for ease of explanation, we ignore the noise in  $\mathbf{b}$  and abuse the notation of  $\mathbf{A}_{\text{full}}^{-1}$ , which should be understood as representing the filtered backprojection (FBP) in CT and the inverse Fourier transform in MRI [61]. Without using a constraint on  $\mathbf{y}$ , one cannot find the reconstruction map  $f_b$ . Hence, we must use the prior knowledge on the data distribution of all possible images to be reconstructed. To extract prior knowledge on solutions, deep learning-based techniques use training data  $\{(\mathbf{b}^{(k)}, \mathbf{y}^{(k)})\}_{k=1}^{\text{n}_{\text{data}}}$ , where  $\mathbf{y}^{(k)} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}}^{(k)}$  and  $\mathbf{b}^{(k)} = \mathcal{S}_{\text{sub}} \mathbf{b}_{\text{full}}^{(k)}$ .

Learning  $f_b : \mathbf{b} \rightarrow \mathbf{y}$  can be achieved by learning the following map:

$$(1.4) \quad f : \mathbf{x} = \mathbf{A}^\sharp \mathbf{b} \mapsto \mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}}.$$

where  $\mathbf{A}^\sharp = \mathbf{A}_{\text{full}}^{-1} \mathcal{S}_{\text{sub}}^*$  and  $\mathcal{S}_{\text{sub}}^*$  is the dual of  $\mathcal{S}_{\text{sub}}$ , which can be understood as the zero padding operator corresponding to the subsampling  $\mathcal{S}_{\text{sub}}$ . Using the transformed training data  $\{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})\}_{k=1}^{\text{n}_{\text{data}}}$ , where  $\mathbf{x}^{(k)} = \mathbf{A}^\sharp \mathbf{b}^{(k)}$ , we consider the learning objective as follows:

$$(1.5) \quad f = \underset{f \in \mathbb{NN}}{\operatorname{argmin}} \sum_{k=1}^{\text{n}_{\text{data}}} \|f(\mathbf{x}^{(k)}) - \mathbf{y}^{(k)}\|_{\ell_2}$$

where  $\mathbb{NN}$  denotes a set of functions described in a special form of neural network and  $\|\cdot\|_{\ell_p}$  is the  $\ell_p$  norm of the vector. Notably, this  $f$  in (1.5) is designed to work well only on a low-dimensional solution manifold obtained by regressing the training data  $\{\mathbf{y}^{(k)}\}_{k=1}^{\text{n}_{\text{data}}}$ , not on the entire image domain.

This paper aims to provide some mathematical grounds for the learnability of  $f$  by using various performance experiments. In Section 2.2, we present a particular low-dimensional solution model to highlight the advantage of a deep learning method over conventional methods using PCA, wavelets, and total variation regularization. In Section 3, we discuss the nonlinearity structure of underdetermined linear systems and conditions of learning (called  $\mathcal{M}$ -RIP condition). We observe that highly underdetermined linear systems in medical imaging are highly non-linear. We also examine whether a desired reconstruction map  $f : \mathbf{x} \rightarrow \mathbf{y}$  can be learnable from the training data. The learning ability depends on the subsampling strategy  $\mathcal{S}_{\text{sub}}$ , and quality and quantity of training data. Section 3.1 investigates the learnability of undersampled MRI. It depends on the sampling pattern as follows: (i) If  $\mathcal{S}_{\text{sub}}$  is a uniform subsampling,  $f$  is not learnable because there exist two different realistic images, namely,  $\mathbf{y}$  and  $\mathbf{y}'$ , such that  $\mathbf{A}^\sharp \mathbf{A}(\mathbf{y} - \mathbf{y}') = 0$ . (ii) If  $\mathcal{S}_{\text{sub}}$  denotes a uniform sampling with one additional phase encoding line, then  $f$  is learnable. We also deal with the learnability of  $f$  in interior tomography (see Section 3.2) and sparse-view CT (see Section 3.3). In interior tomography,  $f$  is learnable because  $f(\mathbf{x}) - \mathbf{x}$  is directionally analytic; therefore, it is determined by the very local information of it. In sparse-view CT,  $f$  is somehow learnable because  $f(\mathbf{x}) - \mathbf{x}$  has common repetitive local patterns that are very different from realistic images. Finally, in Section 4, we discuss some issues related to deep learning-based solvability for underdetermined problems.

**2. Analysis on Underdetermined Inverse Problem.** This section considers the underdetermined problem (1.3), where  $\mathbf{b}$  represents the undersampled data (e.g.,  $k$ -space data in undersampled MRI and sinogram data in sparse-view CT and interior tomography). We denote the dimensions of row and column vectors by  $n$  and  $m$ , respectively. Note that  $n$  is the same as the dimension of image. The relation between the undersampled data  $\mathbf{b}$  and the corresponding fully sampled data  $\mathbf{b}_{\text{full}}$  can be expressed by

$$(2.1) \quad \mathbf{b} = \mathcal{S}_{\text{sub}}(\mathbf{b}_{\text{full}})$$

where  $\mathcal{S}_{\text{sub}}$  denotes the subsampling operator. Since  $\mathbf{A}$  is  $m \times n$  matrix with  $m \ll n$ , the underdetermined problem (1.3) has infinitely many solutions, which constitute the  $n - m$  dimensional subplane given by the followings:

$$(2.2) \quad \mathcal{N}_b(\mathbf{A}) := \{\mathbf{y} \in \mathbb{R}^n (\text{or } \mathbb{C}^n) : \mathbf{A}\mathbf{y} = \mathbf{b}\}$$

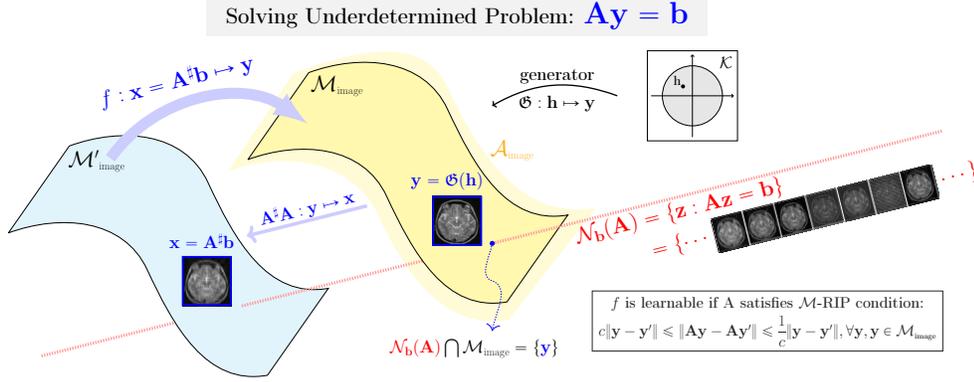


FIG. 1. Description of solvability of underdetermined inverse problem  $\mathbf{A}\mathbf{y} = \mathbf{b}$ . Solving  $\mathbf{A}\mathbf{y} = \mathbf{b}$  can be achieved by learning  $f: \mathbf{x} = \mathbf{A}^\sharp \mathbf{b} \mapsto \mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}}$  in (1.4) with probing the solution manifold  $\mathcal{M}_{\text{image}}$ . If  $\mathbf{A}$  satisfies the  $\mathcal{M}$ -RIP condition, then  $\mathbf{A}^\sharp \mathbf{A}: \mathcal{M}_{\text{image}} \mapsto \mathcal{M}'_{\text{image}}$  is one-to-one, i.e.,  $\mathcal{N}_{\mathbf{b}}(\mathbf{A}) \cap \mathcal{M}_{\text{image}} = \{\mathbf{y}\}$  is unique. In general,  $f$  is nonlinear and the degree of non-linearity depends on the sampling strategy of  $\mathbf{b}$  and the degree of bending the solution manifold.

To find  $f$  in (1.4), it is necessary to find a way to convert the distorted image  $\mathbf{x} = \mathbf{A}^\sharp \mathbf{b}$  to the desired image  $\mathbf{y}$ , which is selected from the set  $\mathcal{N}_{\mathbf{b}}(\mathbf{A})$ . In order to determine the unique solution among  $\mathcal{N}_{\mathbf{b}}(\mathbf{A})$ , we have to restrict the solution by invoking the prior knowledge of expected solutions.

**2.1. Constrained reconstruction problem.** Assume that  $\mathcal{A}_{\text{image}}$  is a set of all realistic images that include the set of all  $\mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}} \in \mathbb{R}^n$  (or  $\mathbb{C}^n$ ), where  $\mathbf{b}_{\text{full}}$  denotes the fully sampled medical data. We consider the following constraint problem:

$$(2.3) \quad \begin{cases} \text{Solve } \mathbf{A}\mathbf{y} = \mathbf{b} \\ \text{subject to the constraint } \mathbf{y} \in \mathcal{A}_{\text{image}}. \end{cases}$$

Ideally, we hope that  $\mathcal{N}_{\mathbf{b}}(\mathbf{A}) \cap \mathcal{A}_{\text{image}} = \{\mathbf{y} \in \mathcal{A}_{\text{image}} : \mathbf{A}\mathbf{y} = \mathbf{b}\} \neq \emptyset$ , and that all the images in the set  $\mathcal{N}_{\mathbf{b}}(\mathbf{A}) \cap \mathcal{A}_{\text{image}}$  are visually same for radiologists. Hence, it seems to be necessary to describe a similarity measure between two images,  $\mathbf{y}, \mathbf{y}' \in \mathcal{A}_{\text{image}}$ , by defining the distance  $\text{dist}_{\text{radiologist}}(\mathbf{y}, \mathbf{y}')$ ; e.g.,  $\text{dist}_{\text{radiologist}}(\mathbf{y}, \mathbf{y}') = 0$  means that both images are visually the same for radiologists. Currently, it seems to be considerably difficult to develop a concept of  $\text{dist}_{\text{radiologist}}(\mathbf{y}, \mathbf{y}')$  that agrees with the perspective of medical radiologists. To simplify the problem (2.3) along with avoiding complex similarity issues in terms of radiologists, let us assume the following:

- [H1] Any image in  $\mathcal{A}_{\text{image}}$  lies on or near a low-dimensional manifold, which is denoted by  $\mathcal{M}_{\text{image}}$ , whose Hausdorff dimension, which is denoted by  $\mathfrak{d}_{\text{mfd}}$ , is smaller than  $m$  (i.e., the dimension of sampled vector  $\mathbf{b}$ ).
- [H2] There exists a generator  $\mathfrak{G}: \mathbf{h} \in \mathcal{K} \rightarrow \mathbf{y} \in \mathbb{R}^n$  such that the following hold:

$$(2.4) \quad \mathcal{M}_{\text{image}} = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} = \mathfrak{G}(\mathbf{h}) \text{ and } \mathbf{h} \in \mathcal{K}\}$$

where  $\mathcal{K}$  denotes a subset of  $\mathbb{R}^{\mathfrak{d}_{\text{mfd}}}$ . Moreover, there exists a constant  $c \in (0, 1]$  such that the following hold: For all  $\mathbf{h}, \mathbf{h}' \in \mathcal{K}$ ,

$$(2.5) \quad c\|\mathbf{h} - \mathbf{h}'\| \leq \|\mathfrak{G}(\mathbf{h}) - \mathfrak{G}(\mathbf{h}')\| \leq \frac{1}{c}\|\mathbf{h} - \mathbf{h}'\|$$

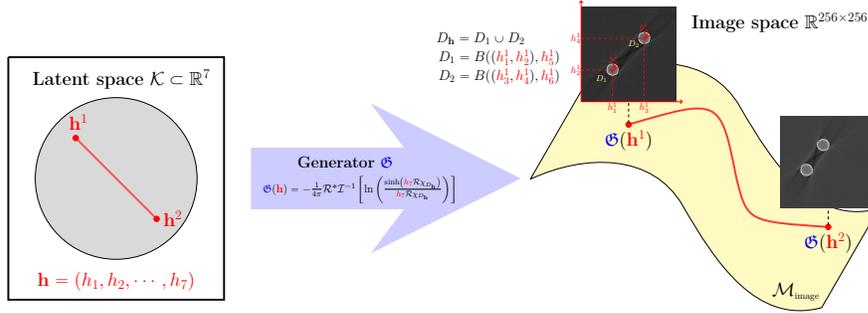


FIG. 2. Special solution manifold  $\mathcal{M}_{\text{image}} \subset \mathbb{R}^{256 \times 256}$  to highlight the advantage of deep learning method over conventional methods. It is generated by the generator function  $\mathfrak{G}$  in (2.10). The manifold  $\mathcal{M}_{\text{image}}$  is seven dimensional. Images in  $\mathcal{M}_{\text{image}}$  consist of two disks and a special streaking feature joining the two disks.

[H3] There exists a normalization map  $\mathfrak{N} : \mathcal{A}_{\text{image}} \mapsto \mathcal{M}_{\text{image}}$  such that if two images  $\mathbf{y}, \mathbf{y}' \in \mathcal{A}_{\text{image}}$  are visually the same for radiologists, then  $\mathfrak{N}(\mathbf{y}) = \mathfrak{N}(\mathbf{y}')$ .

The manifold  $\mathcal{M}_{\text{image}}$  can be viewed as a set of all  $256 \times 256$  human head-MR images in undersampled MRI problems, or as a set of all  $512 \times 512$  CT images in underdetermined CT problems. In the normalization map  $\mathfrak{N}$ , the difference  $\mathbf{y} - \mathfrak{N}(\mathbf{y})$  can be a noise that does not contain any diagnostic feature.

If we have both the generator  $\mathfrak{G}$  and the normalizer  $\mathfrak{N}$  in the above assumptions [H1]–[H3], the underdetermined problem (2.3) becomes a somewhat well-posed problem as follows:

$$(2.6) \quad \text{Given } \mathbf{x} = \mathbf{A}^\# \mathbf{b}, \text{ solve } \mathbf{A} \mathfrak{G}(\mathbf{h}) = \mathbf{A} \mathbf{x} \text{ for } \mathbf{h}$$

where the number of unknowns are smaller than the number of equations. A necessary condition for the solvability of (2.6) is  $\mathfrak{d}_{\text{mfd}} \leq n - m$ . Moreover, with the aid of the generator  $\mathfrak{G}$ , the very ambiguous distance  $\text{dist}_{\text{radiologist}}(\mathbf{y}, \mathbf{y}')$  from the viewpoint of radiologist can be clearly defined as  $\|\mathbf{h} - \mathbf{h}'\|$ , where  $\mathfrak{G}(\mathbf{h}) = \mathfrak{N}(\mathbf{y})$  and  $\mathfrak{G}(\mathbf{h}') = \mathfrak{N}(\mathbf{y}')$ . However, finding both the generator  $\mathfrak{G}$  and the normalizer  $\mathfrak{N}$  may be very difficult task, which is expected to be achieved via deep learning techniques using a training dataset  $\{\mathbf{y}^{(k)}\}_{k=1}^{\text{n}_{\text{data}}}$  in the near future.

The reconstruction map  $f : \mathbf{x} \rightarrow \mathbf{y}$  in (1.4) can be expressed as

$$(2.7) \quad f(\mathbf{x}) := \underset{\mathbf{y} \in \mathcal{M}_{\text{image}}}{\text{argmin}} \|\mathbf{A}^\# \mathbf{A} \mathbf{y} - \mathbf{x}\|_{\ell^2}^2,$$

by assuming that there exists a unique minimizer and that  $\mathcal{M}_{\text{image}}$  is known. Since it is very difficult to know the manifold  $\mathcal{M}_{\text{image}}$ , one can achieve the reconstruction map  $f$  as follows:

$$(2.8) \quad f := \underset{f \in \mathbb{NN}}{\text{argmin}} \sum_{k=1}^{\text{n}_{\text{data}}} \|f(\mathbf{x}^{(k)}) - \mathbf{y}^{(k)}\|_{\ell^2}^2,$$

where  $\mathbb{NN}$  denotes a set of functions described in a special form of neural network.

An important question is “*what is the minimum ratio of undersampling to provide guarantee of accurate reconstruction  $f$  in (2.7)?*”. It is closely related to the dimension of the manifold  $\mathcal{M}_{\text{image}}$  and the capability of finding the generator  $\mathfrak{G}$  in [H2]. Currently,

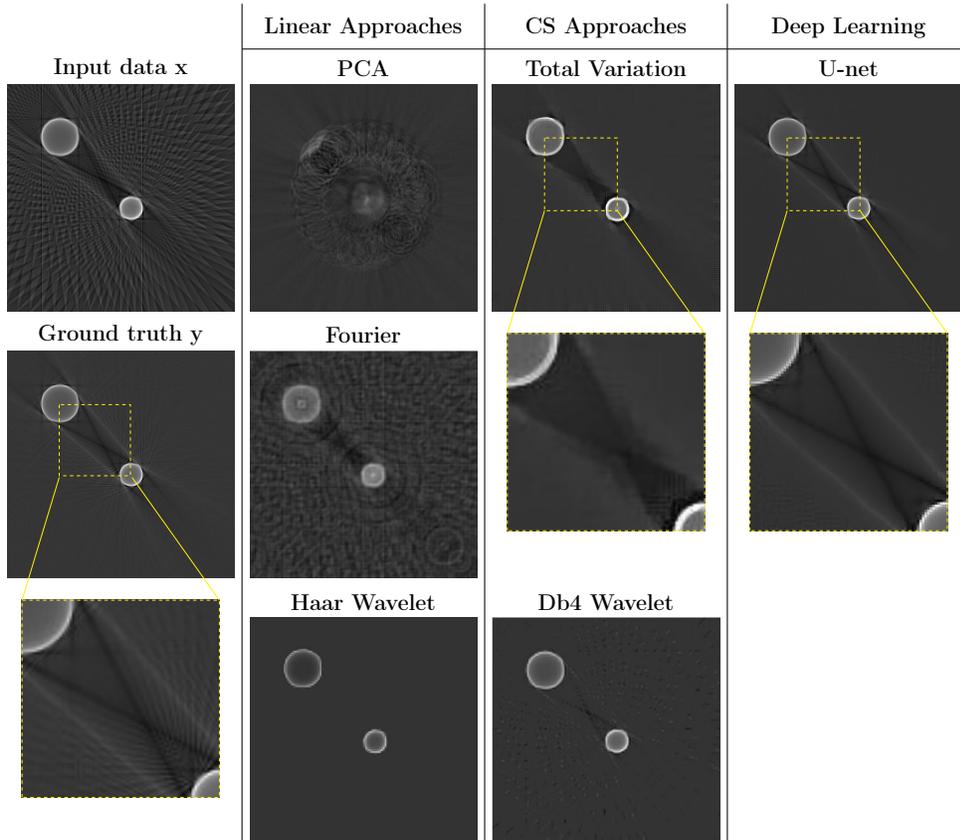


FIG. 3. Empirical results using various reconstruction approaches for the sparse-view CT problem with the special solution manifold in (2.9). For the linear projection approach, three different linear representations (PCA, Fourier, and Haar wavelet) of the input data were truncated at the 800th term after arranging the terms in the descending order according to the absolute value of their coefficients. For implementing the CS approach, we applied the  $\ell_1$  convex relaxation method using two different transforms (total variation(TV) and Daubechies four tap(Db4) wavelet). An optimal regularization parameter was empirically selected between 0.01 and 1. Lastly, as the deep learning technique, U-net, trained by 800 training data pairs, is used.

our explicit knowledge on the solution prior (i.e.,  $\mathcal{M}_{\text{image}}$ ) is very limited and hardly built.

To clarify a concept of manifold prior, we try to solve and analyze underdetermined problems subjected to the model manifold, which is well-understood in a mathematical framework.

**2.2. A special solution manifold: Comparison of conventional methods with deep learning method.** This section provides a novel example of a low-dimensional manifold  $\mathcal{M}_{\text{image}}$  to explain [H1]–[H3]. Using this manifold, we examine the capability of solving the sparse-view CT model using various exiting methods such as linear approaches (e.g. PCA, truncated Fourier and wavelet transform), sparse sensing (e.g. TV and Dictionary learning), and deep learning (e.g. U-net). This special solution manifold  $\mathcal{M}_{\text{image}}$  highlights the advantage of deep learning method

over conventional methods, where two approaches use prior information of the solution in a completely different way.

Our example of the manifold  $\mathcal{M}_{\text{image}}$  in [H1] is seven dimensional and given by

$$(2.9) \quad \mathcal{M}_{\text{image}} := \{ \mathfrak{G}(\mathbf{h}) \in \mathbb{R}^n : \mathbf{h} \in \mathcal{K} \}$$

where  $n = 256^2$ ,  $\mathcal{K}$  is a compact subset of  $\mathbb{R}^7$ , and the continuous version of  $\mathfrak{G}(\mathbf{h})$  is given by

$$(2.10) \quad \mathfrak{G}(\mathbf{h}) = -\frac{1}{4\pi} \mathcal{R}^* \mathcal{I}^{-1} \left[ \ln \left( \frac{\sinh(h_7 \mathcal{R} \chi_{D_{\mathbf{h}}})}{h_7 \mathcal{R} \chi_{D_{\mathbf{h}}}} \right) \right]$$

where the notations are the following:

- $\mathcal{R}^*$  is the dual of the Radon transform  $\mathcal{R}$ . (See Section 3.2 for details.)
- $\mathcal{I}^{-1}$  is the Riesz potential of degree -1.
- $\mathbf{h} = (h_1, h_2, \dots, h_7)$ .
- $D_{\mathbf{h}}$  is a union of two disks with centers  $(h_1, h_2)$ ,  $(h_2, h_3)$  and radii  $h_5, h_6$ .
- $\chi_D$  is the characteristic function of  $D$ .

This example originates from the paper [55], where the image of  $\mathfrak{G}(\mathbf{h})$  represents metal artifacts of CT in the presence of metallic objects occupied in the region  $D_{\mathbf{h}}$ . Fig. 2 shows images on the manifold  $\mathcal{M}_{\text{image}}$ .

Assuming that  $\mathfrak{G}$  is known, consider the highly underdetermined problem (1.3) to find the following reconstruction map:

$$(2.11) \quad f : \mathbf{x} \in \mathcal{M}'_{\text{image}} \mapsto \mathbf{y} \in \mathcal{M}_{\text{image}} \text{ satisfying } \mathbf{A}^\# \mathbf{A} \mathbf{y} = \mathbf{x}$$

where

$$(2.12) \quad \mathcal{M}'_{\text{image}} := \{ \mathbf{A}^\# \mathbf{A} \mathfrak{G}(\mathbf{h}) : \mathbf{h} \in \mathcal{K} \}$$

If  $m$  (the number of equations) is greater than seven, it is possible to find  $f$  and this  $f$  can be obtained as follows:

$$(2.13) \quad f(\mathbf{x}) = \mathfrak{G}(\mathbf{h}), \quad \mathbf{h} = \underset{\mathbf{h} \in \mathcal{K}}{\operatorname{argmin}} \|\mathbf{A}^\# \mathbf{A} \mathfrak{G}(\mathbf{h}) - \mathbf{x}\|_{\ell^2}^2$$

However, we do not know  $\mathfrak{G}$  in practice.

In the remaining part of this section, we examine the capability of various methods for finding a reconstruction map  $f : \mathbf{x} \mapsto \mathbf{y}$  in (2.11) using the sparse-view CT model described in Section 3.3. Let  $\{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})\}_{k=1}^{n_{\text{data}}}$  denote a training data set.

**2.2.1. Linear projection approach.** This subsection explains that there may not exist an appropriate low-dimensional linear projection that captures the variations in  $\mathfrak{G}(\mathbf{h})$ . Principal component analysis (PCA) is widely used for the dimensionality reduction in which the unknown manifold  $\mathcal{M}_{\text{image}}$  is approximated by a linear subspace spanned by the set of principal components  $\{\mathbf{d}_k\}_{k=1}^{n_{\text{PCA-basis}}}$ . To be precise, the first principal component,  $\mathbf{d}_1$ , is obtained as follows:

$$(2.14) \quad \mathbf{d}_1 = \underset{|\mathbf{d}|=1}{\operatorname{argmax}} \mathbf{d}^T \mathbf{Y}^T \mathbf{Y} \mathbf{d}$$

where  $\mathbf{Y} := (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{n_{\text{data}}})^T$ . Similarly, the second principal component,  $\mathbf{d}_2$ , is obtained by computing the first principal component of matrix  $\mathbf{Y}_1 := \mathbf{Y} - \mathbf{d}_1 \mathbf{d}_1^T$ .

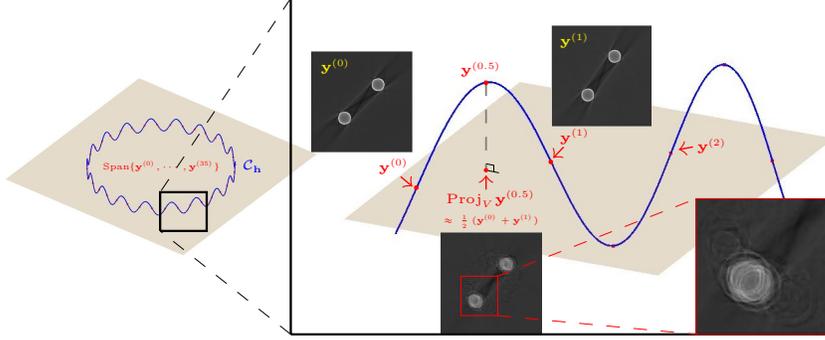


FIG. 4. Linear dimensionality reduction methods such as PCA may provide a poor approximation of the highly curved manifold in (2.16). Let  $V = \text{span}\{y^{(k)}\}_{k=0}^{35}$  where  $y^{(k)}$  is the  $k\pi/18$  degree rotated image of the image  $y^{(0)}$ . Let  $y^{(0.5)}$  be the  $\pi/36$  degree rotated image of the image  $y^{(0)}$ . The projection of  $y^{(0.5)}$  onto  $V$  is approximately equal to  $\frac{1}{2}(y^{(0)} + y^{(1)})$  that destroys the main characteristics of  $y^{(0.5)}$ .

Continuing this process, we obtain the orthogonal basis  $\{\mathbf{d}_k\}_{k=1}^{\text{nPCA-basis}}$ . Subsequently, the reconstruction map  $f$  is given by

$$(2.15) \quad f(\mathbf{x}) = \mathbf{D}\mathbf{h}, \quad \mathbf{h} = \underset{\mathbf{h}}{\text{argmin}} \|\mathbf{A}^\sharp \mathbf{A}\mathbf{D}\mathbf{h} - \mathbf{x}\|_{\ell_2}^2$$

where  $\mathbf{D}$  denotes the matrix whose columns are  $\{\mathbf{d}_k\}_{k=1}^{\text{nPCA-basis}}$ .

Fig. 4 depicts that PCA fails to provide satisfactory approximations of images in the unknown manifold  $\mathcal{M}_{\text{image}}$ , because the low dimensional subspace spanned by the principal components cannot sufficiently cover the nonlinearity of the solution manifold. In Fig. 4,  $\mathcal{C}_h$  represents the following one-dimensional curve lying on the manifold  $\mathcal{M}_{\text{image}}$ :

$$(2.16) \quad \mathcal{C}_h := \{\mathfrak{G}(\mathcal{T}_\theta \mathbf{h}) : 0 < \theta \leq 2\pi\}, \mathcal{T}_\theta = \begin{pmatrix} \mathcal{R}_\theta & 0 & 0 \\ 0 & \mathcal{R}_\theta & 0 \\ 0 & 0 & \mathbf{I} \end{pmatrix}$$

where  $\mathcal{R}_\theta$  is the rotation matrix with angle  $\theta$ ,  $\mathbf{I}$  is  $3 \times 3$  identity matrix, and 0 here denotes the corresponding zero matrix. The plane in Fig. 4 represents the 36-dimensional space spanned by the sampled images  $\{y^{(k)}\}_{k=0}^{35}$ , which are sampled at  $\theta = \frac{k}{36}2\pi$ ,  $k = 0, \dots, 35$ , on the curve  $\mathcal{C}_h$ . Although  $\mathcal{C}_h$  is the map of the simple circle  $\{\mathcal{T}_\theta \mathbf{h} : 0 < \theta \leq 2\pi\}$  (in the latent space) through the generator function  $\mathfrak{G}$ , it is highly curved and complex due to the severe nonlinearity of function  $\mathfrak{G}$ . Therefore, with the limited expressivity of PCA [56], one cannot adequately approximate the curve  $\mathcal{C}_h$  by using the plane spanned by  $\{y^{(k)}\}_{k=0}^{35}$ .

Fig. 3 depict that linear approaches, including PCA, in solving the underdetermined problem (1.3) result in the significant loss of information from the original image. The inability of the linear projection approach to provide the global approximation of the highly curved image manifold is the reason for such poor reconstruction results.

**2.2.2. Compressed sensing approach.** Compressed sensing(CS) is based on the assumption that  $\mathbf{y} \in \mathcal{M}_{\text{image}}$  has sparse representation under a basis  $\{\mathbf{d}_k\}_{k=1}^{\text{nCS-basis}}$ ,

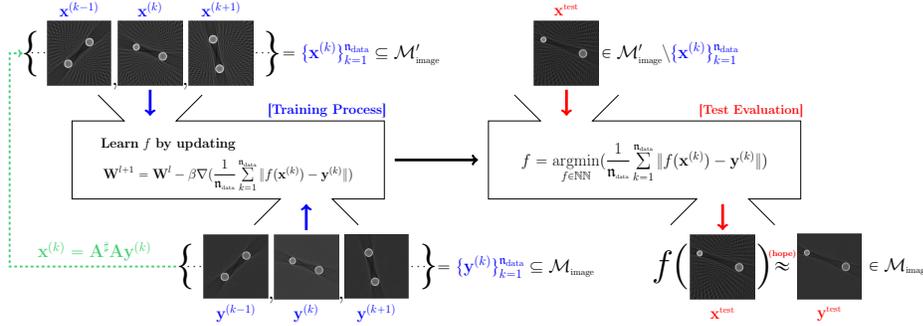


FIG. 5. Deep learning framework for estimating the reconstruction map  $f : \mathbf{x} \mapsto \mathbf{y}$ . For a given deep learning network and training dataset  $\{\mathbf{x}^{(k)}, \mathbf{y}^{(k)}\}_{k=1}^{n_{\text{data}}}$ , the reconstruction map  $f$  is obtained by minimizing the discrepancy between the network outputs  $\{f(\mathbf{x}^{(k)})\}_{k=1}^{n_{\text{data}}}$  and the corresponding labels  $\{\mathbf{y}^{(k)}\}_{k=1}^{n_{\text{data}}}$ . Despite the finite number of training data, we hope the function  $f$  to provide an accurate approximation of the test label  $\mathbf{y}^{\text{test}}$  for any unobserved test data  $\mathbf{x}^{\text{test}} \in \mathcal{M}'_{\text{image}} \setminus \{\mathbf{x}^{(k)}\}_{k=1}^{n_{\text{data}}}$ .

i.e.,

$$(2.17) \quad \mathbf{y} = \mathbf{D}\mathbf{h} \quad \text{s.t.} \quad \|\mathbf{h}\|_{\ell_0} \ll n_{\text{CS-basis}}$$

where  $\mathbf{D}$  is a matrix whose  $k$ -th column corresponds to  $\mathbf{d}_k$  and  $\|\mathbf{h}\|_{\ell_0}$  is the number of non-zero entries of  $\mathbf{h}$ . In CS,  $\ell^1$  convex relaxation methods are widely used to make the problem computationally feasible. A sparse approximation to the solution of the underdetermined problem (1.3) is obtained as follows:

$$(2.18) \quad f(\mathbf{x}) = \mathbf{D}\mathbf{h}, \quad \mathbf{h} = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathbf{A}^\# \mathbf{A}\mathbf{D}\mathbf{h} - \mathbf{x}\|_{\ell_2}^2 + \lambda \|\mathbf{h}\|_{\ell_1}$$

where  $\lambda$  is a regularization parameter that controls the trade-off between data fidelity and the regularity enforcing the sparsity of  $\mathbf{h}$ . Kindly refer to [16, 15, 9, 11, ?] for additional details. We implement the CS technique by using several wavelet bases, which are efficient in CS applications for natural images [14, 47]. However, the reconstruction results from Fig. 3 show that some details are not preserved in the CS process.

The total variation(TV)-based CS method imposes a sparsity of the image gradient, where  $f(\mathbf{x})$  can be obtained as follows:

$$(2.19) \quad f(\mathbf{x}) = \underset{\mathbf{y}}{\operatorname{argmin}} \|\mathbf{A}^\# \mathbf{A}\mathbf{y} - \mathbf{x}\|_{\ell_2}^2 + \lambda \|\nabla \mathbf{y}\|_{\ell_1}$$

Fig. 3 shows that TV-based CS method also eliminates some of the details. TV method does not selectively preserve the streaking feature lying between two disks, while removing the other artifacts.

Dictionary learning [54, 2] utilizes the given training data to find a (redundant and data-driven) basis  $\{\mathbf{d}_k\}_{k=1}^{n_{\text{dic}}}$  that can represent every  $\mathbf{y} \in \mathcal{M}_{\text{image}}$  as a sparse vector. A learned dictionary can handle a specific problem considerably better than analysis-driven dictionaries (e.g. wavelet and framelet) [17, 1, 46, 68, 59]. Dictionary learning approaches have a drawback in dealing with high-dimensional data due to the huge computational complexity; hence, the patch-based approach (e.g. image patch of size  $8 \times 8$  pixels) has been adopted in most image processing applications. However, this approach might not be fit for the tasks for which the global information should be sufficiently incorporated.

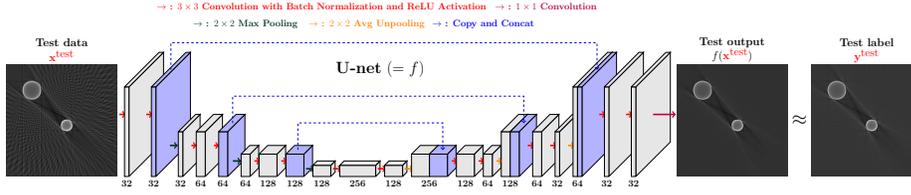


FIG. 6. Deep learning approach using U-net for solving the sparse-view CT model with the special solution manifold. By training U-net with 800 data pairs, the reconstruction function  $f$  is obtained. In Tensorflow environment, the minimization process was performed by using Batch normalization and Adam optimizer with learning rate 0.001, mini-batch size 16, and 1000 epochs. For a given test data  $\mathbf{x}^{\text{test}}$ , the reconstruction function  $f$  provides the test output  $f(\mathbf{x}^{\text{test}})$  which approximates the test label  $\mathbf{y}^{\text{test}}$ .

**2.2.3. Deep learning approach.** Deep learning techniques expand our ability to solve underdetermined problems via sophisticated learning process by using group-data fidelity of the training data; furthermore, they appear to effectively deal with the limitations of the existing mathematical methods in handling various ill-posed problems. In CS,  $f(\mathbf{x})$  in (2.19) can be viewed as a solution of the nonlinear Euler-Lagrange equation associated with the trade-off between two separate competitive objectives of maximizing the “single data fidelity” and minimizing TV (as a sparse prior of natural images). However, this sparse prior may not be appropriate for preserving small features that contain clinically useful information. In contrast, the deep learning approach (1.5) utilizes “group-data fidelity” to estimate the reconstruction map  $f : \mathbf{x} \mapsto \mathbf{y}$  by seemingly probing the relationship between unknown manifolds  $\mathcal{M}_{\text{image}}$  and  $\mathcal{M}'_{\text{image}}$ . The reconstruction  $f$  is obtained by minimizing the group-data discrepancy  $\sum_{k=1}^{\text{n}_{\text{data}}} \|f(\mathbf{x}^{(k)}) - \mathbf{y}^{(k)}\|$  (i.e. maximizing the group-data fidelity) on a finite number of training pairs  $\{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})\}_{k=1}^{\text{n}_{\text{data}}}$  lying on  $\mathcal{M}'_{\text{image}} \times \mathcal{M}_{\text{image}}$ , as shown in Fig. 5.

In particular, U-net [58] has achieved enormous success in finding the map for various underdetermined medical imaging problems [31, 36, 25]. In U-net, the network architecture of  $f$  comprises a contraction path  $\Phi : \mathbf{x} \mapsto \mathbf{h}$  and an expansion path  $\Psi : \mathbf{h} \mapsto \mathbf{y}$ ;  $f(\mathbf{x}) = \Psi \circ \Phi(\mathbf{x})$ . To be precise, the simplest form of the contraction path  $\Phi$  is expressed by

$$(2.20) \quad \mathbf{h} = \Phi(\mathbf{x}) = \sigma(\mathbf{w}_3 \otimes \mathcal{P}(\sigma(\mathbf{w}_2 \otimes \sigma(\mathbf{w}_1 \otimes \mathbf{x} + \mathbf{c}_1) + \mathbf{c}_2)) + \mathbf{c}_3)$$

and the corresponding expansive path  $\Psi$  is represented as

$$(2.21) \quad \Psi(\mathbf{h}) = \mathbf{w}_6 \odot (\sigma(\mathbf{w}_5 \otimes (\mathcal{C}_{\text{cat}}(\mathcal{U}_{\text{pool}}(\sigma(\mathbf{w}_4 \otimes \mathbf{h} + \mathbf{c}_4)), \mathbf{z})) + \mathbf{c}_5))$$

where  $\mathbf{z} = \sigma(\mathbf{w}_2 \otimes \sigma(\mathbf{w}_1 \otimes \mathbf{x} + \mathbf{c}_1) + \mathbf{c}_2)$ . Here,  $\sigma(z) = \text{ReLU}(z)$ ,  $\mathcal{P}$  is a pooling operator,  $\mathcal{U}_{\text{pool}}$  is an unpooling operator, and  $\mathcal{C}_{\text{cat}}$  is a concatenation operator. The work in [58] can be referred for a more detailed description. The overall structure of U-net is shown in Fig. 6.

The map  $f : \mathbf{x} \mapsto \mathbf{y}$ , as a function of parameters  $\mathbf{W} = \{\mathbf{w}_1, \mathbf{c}_1, \mathbf{w}_2, \mathbf{c}_2, \dots\}$ , is determined as follows:

$$(2.22) \quad f = \underset{f \in \text{NN}}{\text{argmin}} \frac{1}{\text{n}_{\text{data}}} \sum_{k=1}^{\text{n}_{\text{data}}} \|f(\mathbf{x}^{(k)}) - \mathbf{y}^{(k)}\|_{\ell^2}^2$$

where  $\text{NN}$  denotes a set of all the functions of the form  $f = \Psi \circ \Phi$  that vary with  $\mathbf{W}$ . Fig. 3 shows remarkable performance of U-net.

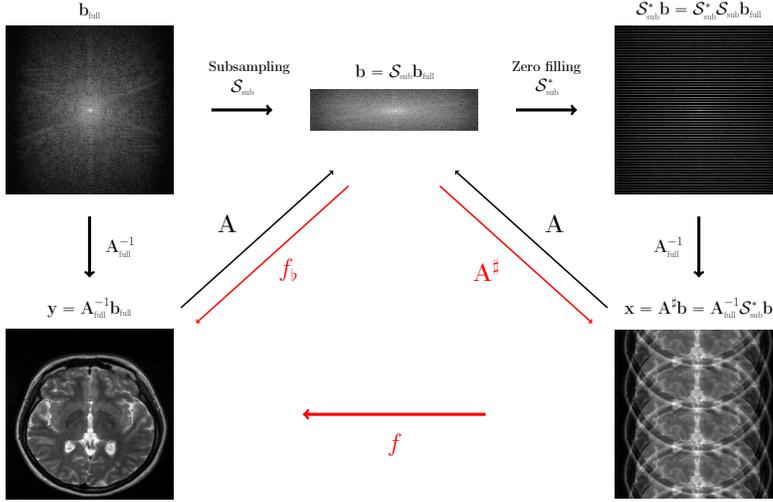


FIG. 7. *Undersampled MRI problem is to recover an image  $\mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}}$  from undersampled data  $\mathbf{b} = \mathcal{S}_{\text{sub}} \mathbf{b}_{\text{full}}$ , where  $\mathcal{S}_{\text{sub}}$  is a subsampling operator and  $\mathbf{b}_{\text{full}}$  is a fully-sampled data in the sense of Nyquist sampling. All images are displayed by taking their absolute values. Using the deep learning technique, we attempt to find a reconstruction function  $f$  that maps from  $\mathbf{x} = \mathbf{A}^{\#} \mathbf{b}$  to  $\mathbf{y}$ . Since the structure of  $\mathbf{x}$  is determined by the subsampling operator  $\mathcal{S}_{\text{sub}}$ , the learning  $f$  can be affected by the subsampling strategy.*

**3. Solvability of Underdetermined Linear System.** In undersampled problems, the subsampling strategy  $\mathcal{S}_{\text{sub}}$  inside  $\mathbf{A} = \mathcal{S}_{\text{sub}} \mathbf{A}_{\text{full}}$  is important for the uniqueness of solution  $\mathbf{y}$  on the manifold  $\mathcal{M}_{\text{image}}$  among all the possible solutions in  $\mathcal{N}_{\mathbf{b}}(\mathbf{A})$ . Precisely, a proper subsampling strategy  $\mathcal{S}_{\text{sub}}$  is related to the following manifold restricted isometry property (RIP) condition. The matrix  $\mathbf{A}$  associated with  $\mathcal{S}_{\text{sub}}$  is said to satisfy the  $\mathcal{M}$ -RIP condition if there exists a constant  $c \in (0, 1]$  such that

$$(3.1) \quad c \|\mathbf{y} - \mathbf{y}'\| \leq \|\mathbf{A}\mathbf{y} - \mathbf{A}\mathbf{y}'\| \leq \frac{1}{c} \|\mathbf{y} - \mathbf{y}'\| \text{ for all } \mathbf{y}, \mathbf{y}' \in \mathcal{M}_{\text{image}}.$$

The following two observations explain the necessary condition for constructing a suitable subsampling strategy:

OBSERVATION 3.1. *If  $\mathbf{A}$  satisfies the  $\mathcal{M}$ -RIP condition in (3.1), then*

$$(3.2) \quad \mathbf{A}^{\#} \mathbf{A} : \mathcal{M}_{\text{image}} \mapsto \mathcal{M}'_{\text{image}} \text{ is one-to-one.}$$

*Proof.* Suppose that there are two different  $\mathbf{y}$  and  $\mathbf{y}'$  such that  $\mathbf{A}^{\#} \mathbf{A}\mathbf{y} = \mathbf{A}^{\#} \mathbf{A}\mathbf{y}'$ . Since  $\mathbf{A}^{\#} = \mathbf{A}_{\text{full}}^{-1} \mathcal{S}_{\text{sub}}^*$ ,

$$\begin{aligned} 0 &= \|\mathbf{A}_{\text{full}}(\mathbf{A}^{\#} \mathbf{A}\mathbf{y} - \mathbf{A}^{\#} \mathbf{A}\mathbf{y}')\| = \|\mathcal{S}_{\text{sub}}^*(\mathbf{A}\mathbf{y} - \mathbf{A}\mathbf{y}')\| \\ &= \|\mathbf{A}\mathbf{y} - \mathbf{A}\mathbf{y}'\| \geq c \|\mathbf{y} - \mathbf{y}'\| \end{aligned}$$

where the last inequality follows from (3.1). Hence,  $\mathbf{y} - \mathbf{y}' = 0$ , which contradicts the assumption.  $\square$

OBSERVATION 3.2. *The reconstruction map  $f : \mathbf{x} \in \mathcal{M}'_{\text{image}} \mapsto \mathbf{y} \in \mathcal{M}_{\text{image}}$  is learnable if  $\mathbf{A}$  satisfies the  $\mathcal{M}$ -RIP condition.*

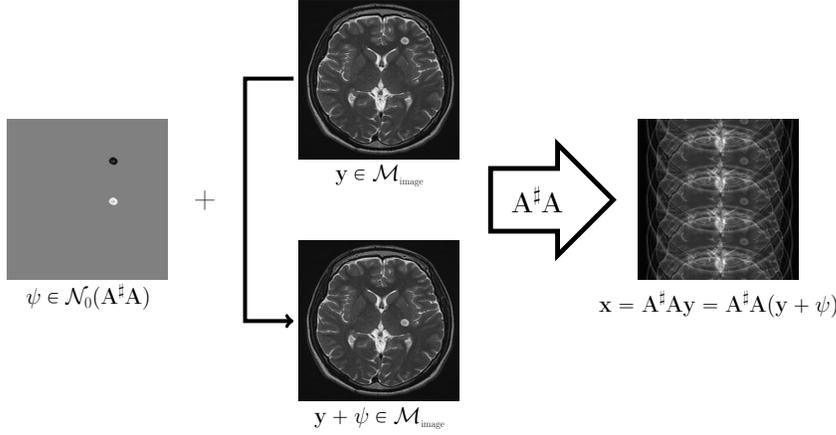


FIG. 8. Location uncertainty on the solution manifold when using a uniform subsampling with factor 4; let us consider two different MR images,  $\mathbf{y}$  and  $\mathbf{y} + \psi$ , where the location of a small anomaly is only different. When taking  $\mathbf{A}^\sharp \mathbf{A}$  to the images, we obtain the same output  $\mathbf{x} = \mathbf{A}^\sharp \mathbf{A} \mathbf{y} = \mathbf{A}^\sharp \mathbf{A} (\mathbf{y} + \psi)$ , where  $\mathbf{A}$  is given by  $\mathbf{A} = \mathcal{S}_{\text{sub}} \mathbf{A}_{\text{full}}$  and  $\mathcal{S}_{\text{sub}}$  denotes a uniform subsampling with factor 4.

If the corresponding matrix  $\mathbf{A}$  does not satisfy the  $\mathcal{M}$ -RIP condition, there exist  $\mathbf{y}_1 \neq \mathbf{y}_2$  such that  $\mathbf{x} = \mathbf{A}^\sharp \mathbf{A} \mathbf{y}_1 = \mathbf{A}^\sharp \mathbf{A} \mathbf{y}_2$ ; therefore, it is impossible to learn such  $f$  due to indistinguishability. The issue of learnability associated with  $\mathbf{A}$  that does not satisfy the  $\mathcal{M}$ -RIP condition will be addressed in Section 3.1 with a concrete example.

Given a highly undersampling operator  $\mathcal{S}_{\text{sub}}$ , the map  $f$  can be viewed as an image restoration function with filling-in missing data or unfolding image data; therefore,  $f(\mathbf{x})$  depends on the image structure. The nonlinearity of  $f$  is affected by  $\mathcal{S}_{\text{sub}}$  and the degree of bending of the manifold  $\mathcal{M}_{\text{image}}$ . The following observation explains that most problems of solving underdetermined linear systems in medical imaging are highly non-linear.

**OBSERVATION 3.3.** Suppose that  $\mathbf{A}$  satisfies the  $\mathcal{M}$ -RIP condition. Let  $\mathbf{V}_{\mathcal{M}_{\text{image}}}$  be the span of the set  $\{\frac{\partial}{\partial h_j} \mathfrak{G}(\mathbf{h}) : \mathbf{h} \in \mathcal{K}, j = 1, \dots, \mathfrak{d}_{\text{mfd}}\}$ . If  $\dim \mathbf{V}_{\mathcal{M}_{\text{image}}} > m$ , then the reconstruction map  $f : \mathbf{x} \in \mathcal{M}'_{\text{image}} \mapsto \mathbf{y} \in \mathcal{M}_{\text{image}}$  is non-linear.

*Proof.* Note that  $f$  satisfies  $f(\mathbf{x}) = \mathbf{y}$  with  $\mathbf{x} = \mathbf{A}^\sharp \mathbf{A} \mathbf{y}$  for all  $\mathbf{y} \in \mathcal{M}_{\text{image}}$ . Since  $\mathcal{M}_{\text{image}}$  is generated by  $\mathfrak{G}$ , we obtain

$$(3.3) \quad f(\mathbf{A}^\sharp \mathbf{A} \mathfrak{G}(\mathbf{h})) = \mathfrak{G}(\mathbf{h}), \quad \forall \mathbf{h} \in \mathcal{K}$$

Taking gradient with respect to  $\mathbf{h}$  on both sides, then

$$(3.4) \quad \nabla_{\mathbf{x}} f(\mathbf{A}^\sharp \mathbf{A} \mathfrak{G}(\mathbf{h})) \mathbf{A}^\sharp \mathbf{A} \nabla_{\mathbf{h}} \mathfrak{G}(\mathbf{h}) = \nabla_{\mathbf{h}} \mathfrak{G}(\mathbf{h}), \quad \forall \mathbf{h} \in \mathcal{K}$$

To derive a contradiction, suppose  $f$  is linear; i.e., there exists a fixed matrix  $\mathbf{B} \in \mathbb{R}^{n \times n}$  such that  $\nabla_{\mathbf{x}} f(\mathbf{x}) = \mathbf{B}$  for all  $\mathbf{x} \in \mathbf{A}^\sharp \mathbf{A} \mathcal{M}_{\text{image}}$ . Subsequently, (3.4) becomes

$$(3.5) \quad \mathbf{B} \mathbf{A}^\sharp \mathbf{A} \nabla \mathfrak{G}(\mathbf{h}) = \nabla \mathfrak{G}(\mathbf{h}), \quad \forall \mathbf{h} \in \mathcal{K}$$

Hence, denoting the eigenspace of  $\mathbf{B} \mathbf{A}^\sharp \mathbf{A}$  corresponding to the eigenvalue  $\lambda$  by  $E_\lambda(\mathbf{B} \mathbf{A}^\sharp \mathbf{A})$ , we have

$$(3.6) \quad E_1(\mathbf{B} \mathbf{A}^\sharp \mathbf{A}) \supseteq \mathbf{V}_{\mathcal{M}_{\text{image}}}$$

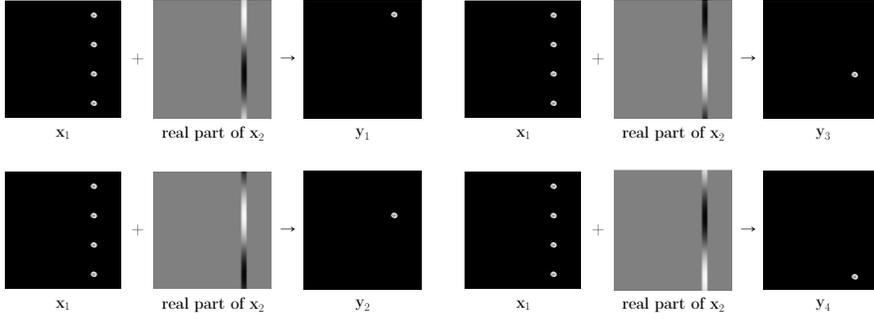


FIG. 9. Empirical observation regarding how to eliminate the location uncertainty by adding one phase encoding line in the  $k$ -space. Four images ( $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3$ , and  $\mathbf{y}_4$ ) containing one small anomaly in four different locations generate the same  $\mathbf{x}_1$ ; however, one additional phase encoding line information  $\mathbf{x}_2$  can deal with location uncertainty in  $\mathbf{x}_1$ .

and from the assumption on the dimension of  $\mathbf{V}_{\mathcal{M}_{\text{image}}}$ ,

$$(3.7) \quad \dim E_1(\mathbf{BA}^\sharp \mathbf{A}) > m.$$

Since  $\text{Rank}(\mathbf{BA}^\sharp \mathbf{A}) \leq \text{Rank}(\mathbf{A}) = m$ ,

$$(3.8) \quad \dim E_0(\mathbf{BA}^\sharp \mathbf{A}) \geq n - m.$$

This is a contradiction because

$$(3.9) \quad \dim E_0(\mathbf{BA}^\sharp \mathbf{A}) + \dim E_1(\mathbf{BA}^\sharp \mathbf{A}) > n. \quad \square$$

**3.1. Undersampled MRI.** In MRI, we apply an oscillating magnetic field to the imaging object in an MR scanner (being confined in a strong magnetic field) to acquire the  $k$ -space data ( $\mathbf{b}$ ), which is used to produce a cross-sectional MR image  $\mathbf{y}$ . In fully sampled MRI, the relation between a 2D MR image  $\mathbf{y}$  and the corresponding fully sampled  $k$ -space data  $\mathbf{b}_{\text{full}}$  can be expressed in the following form [62]:

$$(3.10) \quad \underbrace{\sum_{a,b=1,\dots,\sqrt{n}} e^{-2\pi i(ak_1\Delta k+bk_2\Delta k)} \mathbf{y}(a,b)}_{\mathbf{A}_{\text{full}} \mathbf{y}} = \mathbf{b}_{\text{full}}(k_1, k_2)$$

where  $\Delta k$  denotes the Nyquist sampling distance, which is chosen in such a way that

$$(3.11) \quad \mathbf{y}(a, b) = \sum_{k_1, k_2=1,\dots,\sqrt{n}} e^{2\pi i(ak_1\Delta k+bk_2\Delta k)} \mathbf{b}_{\text{full}}(k_1, k_2).$$

In other words, the Nyquist sampling make the problem  $\mathbf{A}_{\text{full}} \mathbf{y} = \mathbf{b}_{\text{full}}$  well-posed so that the standard reconstruction  $\mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}}$  can be obtained by 2D discrete inverse Fourier transform.

Assume that the frequency-encoding is along the  $k_1$ -axis and that the phase-encoding is along the  $k_2$ -axis in the  $k$ -space. Noting that the MRI scan time is roughly proportional to the number of time consuming phase-encoding steps in  $k$ -space, there have been numerous attempts to shorten the MRI scan time by skipping the phase-encoding lines in the  $k$ -space [63, 23]. In the undersampled MRI, we attempt to find

the optimal reconstruction function that maps the highly undersampled  $k$ -space data ( $\mathbf{b}$  that violates Nyquist sampling criterion) to an image ( $\mathbf{y}$ ) close to the MR image corresponding to the fully sampled data ( $\mathbf{b}_{\text{full}}$  that satisfies the Nyquist sampling criterion).

With undersampled data  $\mathbf{b}$ , the corresponding problem is

$$(3.12) \quad \underbrace{\mathcal{S}_{\text{sub}} \mathbf{A}_{\text{full}}}_{\mathbf{A}} \mathbf{u} = \underbrace{\mathcal{S}_{\text{sub}}(\mathbf{b}_{\text{full}})}_{\mathbf{b}}$$

where  $\mathcal{S}_{\text{sub}}$  denotes a subsampling operator and  $\mathbf{b} = \mathcal{S}_{\text{sub}}(\mathbf{b}_{\text{full}})$ . The image  $\mathbf{x} = \mathbf{A}^{\sharp} \mathbf{b}$  is one of the solutions of (3.12), because  $\mathbf{A}^{\sharp}$  is the pseudo-inverse of  $\mathbf{A}$  in this case. The undersampled MRI problem aims to find an image restoration map  $f : \mathbf{x} = \mathbf{A}_{\text{full}}^{-1} \mathcal{S}_{\text{sub}}^* \mathbf{b} \mapsto \mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}}$ . See Fig. 7.

**3.1.1. Uniform subsampling.** According to the Poisson summation formula, the discrete Fourier transform of the uniformly subsampled data with factor 4 produces the following four-folded image [61]:

$$(3.13) \quad \mathbf{x}(a, b) = \mathbf{A}^{\sharp} \mathbf{A} \mathbf{y} = \frac{1}{4} \sum_{b' \equiv b \pmod{\frac{\sqrt{n}}{4}}} \mathbf{y}(a, b')$$

where  $b' \equiv b \pmod{\frac{\sqrt{n}}{4}}$  means that both  $b$  and  $b'$  leave the same remainder when divided by  $\frac{\sqrt{n}}{4}$ . Unfortunately, there exists an uncertainty that makes it impossible to reconstruct  $\mathbf{y}$  from  $\mathbf{x}$ , and therefore  $f$  is not learnable. To see the reason, we consider the following:

$$\Psi_{\text{ufm}} := \mathcal{N}_0(\mathbf{A}^{\sharp} \mathbf{A}) = \text{Span}\{\psi_{a_*, b_*}^{0, \beta} : a_*, b_* \in \mathbb{Z}_{\sqrt{n}}, \beta = 1, 2, 3\}$$

where  $\mathbb{Z}_n := \{1, \dots, n\}$  for any positive integer  $n$  and  $\psi_{a_*, b_*}^{0, \beta}$  is given by

$$(3.14) \quad \psi_{a_*, b_*}^{0, \beta}(a, b) = \begin{cases} 1 & \text{if } (a, b) = (a_*, b_*) \\ -1 & \text{if } (a, b) = (a_*, b_*) + (0, \frac{\sqrt{n}}{4}\beta) \\ 0 & \text{otherwise} \end{cases}$$

Here,  $b_* + \frac{\sqrt{n}}{4}\beta$  should be understood as modulo  $\sqrt{n}$ .

**OBSERVATION 3.4.** *There exists a non-zero  $\psi \in \Psi_{\text{ufm}}$  and  $\mathbf{y} \in \mathcal{M}_{\text{image}}$  such that  $\mathbf{y} + \psi \in \mathcal{M}_{\text{image}}$ .*

The observation implies that the  $\mathcal{M}$ -RIP condition does not hold, as  $f$  requires the following contradictory two conditions  $f(\mathbf{x}) = \mathbf{y}$  and  $f(\mathbf{x}) = \mathbf{y} + \psi$ , where  $\mathbf{x} = \mathbf{A}^{\sharp} \mathbf{A} \mathbf{y} = \mathbf{A}^{\sharp} \mathbf{A}(\mathbf{y} + \psi)$ . The location of a small anomaly cannot be determined, and, therefore, there are many location uncertainties under the uniform subsampling, as shown in Fig. 8. This is the main reason why  $f$  is not learnable under the uniform subsampling.

**3.1.2. Uniform sampling with adding one phase encoding line.** This section provides a way to improve the separability by adding only one phase encoding line to a uniform subsampling. Let  $\mathcal{S}_{\text{sub}}$  be the uniform subsampling of factor 4 upon adding one phase encoding line. Then,  $\mathbf{x} = \mathbf{A}^{\sharp} \mathbf{b}$  can be decomposed into two parts:

$$(3.15) \quad \mathbf{x}(a, b) = \mathbf{x}_1(a, b) + \mathbf{x}_2(a, b)$$

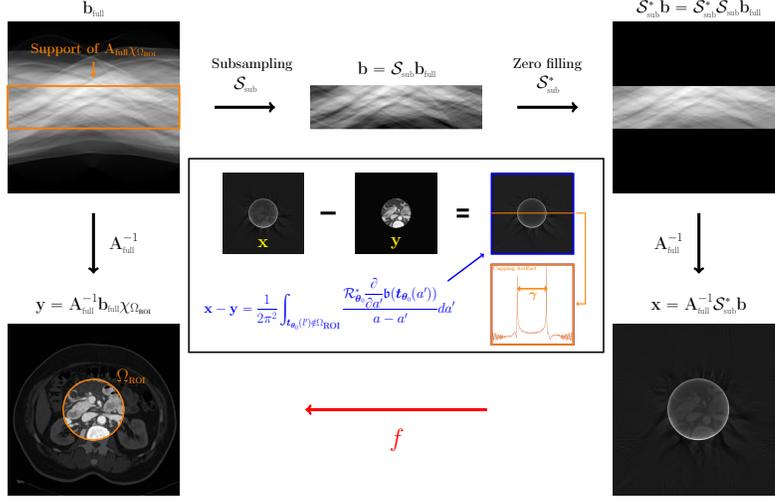


FIG. 10. Interior tomography problem is to recover an image  $\mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}} \chi_{\Omega_{\text{ROI}}}$  in our region of interest (ROI)  $\Omega_{\text{ROI}}$  by using the truncated data  $\mathbf{b} = \mathcal{S}_{\text{sub}} \mathbf{b}_{\text{full}}$ , where  $\mathcal{S}_{\text{sub}}$  is a subsampling operator and  $\chi$  is a characteristic function. Applying the deep learning method, the reconstruction map  $f: \mathbf{x} \mapsto \mathbf{y}$  is learnable because of the analyticity of residual  $\mathbf{x} - \mathbf{y}$ .

where  $\mathbf{x}_1$  is the uniform sampling part given by

$$(3.16) \quad \mathbf{x}_1(a, b) := \frac{1}{4} \sum_{b' \equiv b \pmod{\frac{\sqrt{\pi}}{4}}} \mathbf{y}(a, b')$$

and  $\mathbf{x}_2$  is the single phase encoding part given by

$$(3.17) \quad \mathbf{x}_2(a, b) := \sum_{b' \in \mathbb{Z}_{\sqrt{\pi}}} \mathbf{y}(a, b') e^{2\pi i(b-b')\Delta k}$$

Adding the additional low frequency line in the k-space (compared to the previous uniform sampling) provides the additional information of  $\mathbf{x}_2$ . Subsequently, the situation is dramatically changed to counter the anomaly-location uncertainty in uniform sampling. Fig. 9 shows why the  $\mathbf{x}_2$  information can effectively handle the location uncertainty in Observation 3.4.

**3.2. Interior tomography.** This section explains the underdetermined system for the interior tomography problem. For simplicity, let us consider a 2-D parallel beam system and assume that the projection data for the entire field of view(FOV) is given by

$$(3.18) \quad \mathbf{b}_{\text{full}}(\varphi, s) = \mathcal{R}\mathbf{u}(\varphi, s) := \int_{\mathbb{R}^2} \mathbf{u}(\mathbf{t}) \delta(\boldsymbol{\theta} \cdot \mathbf{t} - s) d\mathbf{t}$$

where  $\mathbf{u}$  represents an attenuation distribution on 2D-slice,  $\mathbf{t} = (t_1, t_2)$ ,  $\boldsymbol{\theta} = (\cos \varphi, \sin \varphi)$ , and  $\delta(\cdot)$  is the Dirac delta function. The discrete version of (3.18) can be expressed by the following linear system

$$(3.19) \quad \mathbf{A}_{\text{full}} \mathbf{u} = \mathbf{b}_{\text{full}}$$

where the Nyquist criterion must be considered in terms of the expected resolution of the CT image( $\mathbf{u}$ ) and sampling( $\mathbf{b}$ ). The standard reconstruction  $\mathbf{u} = \mathbf{A}_{\text{full}}^{-1}\mathbf{b}_{\text{full}}$  is based on the FBP algorithm, which is based on the following identity:

$$(3.20) \quad \mathbf{u}(\mathbf{t}) = \int_0^\pi \int_{\mathbb{R}} |\omega| \mathcal{F} \mathbf{b}_{\text{full}}(\varphi, \omega) e^{2\pi i \omega \mathbf{t} \cdot \boldsymbol{\theta}} d\omega d\varphi$$

where  $\mathcal{F}$  is 1 dimensional Fourier transform associated with the variable  $\omega$  and  $\mathbf{u}$  and  $\mathbf{b}$  are the continuous forms of  $\mathbf{u}$  and  $\mathbf{b}_{\text{full}}$ , respectively, in (3.19) [50].

Now, we are ready to explain the interior tomography problem. Let  $\Omega_{\text{ROI}} \subset \mathbb{R}^2$  denote the local region of interests(ROI), whose size in the interior tomography is smaller than that of a patient's body to be scanned, as depicted in Fig. 10. In the interior tomography, we attempt to reconstruct  $\mathbf{u}\chi_{\Omega_{\text{ROI}}}$  by using the truncated data  $\mathbf{b}_{\text{full}}\chi_D$ , where  $\chi_{\Omega_{\text{ROI}}}$  is the characteristic function of  $\Omega_{\text{ROI}}$  and  $D$  is the support of  $\mathcal{R}\chi_{\Omega_{\text{ROI}}}$ .

The discrete form of this interior tomography can be expressed as follows:

$$(3.21) \quad \text{Reconstruct } \mathbf{y} = \mathbf{u}\chi_{\Omega_{\text{ROI}}} \text{ satisfying } \mathbf{A}\mathbf{u} = \mathbf{b}$$

where  $\mathbf{b} = \mathcal{S}_{\text{sub}}(\mathbf{b}_{\text{full}})$ , with the subsampling  $\mathcal{S}_{\text{sub}}$  defined as  $\mathcal{S}_{\text{sub}}(\mathbf{b}_{\text{full}}) := \mathbf{b}_{\text{full}}\chi_D$ , and  $\mathbf{A} := \mathcal{S}_{\text{sub}}\mathbf{A}_{\text{full}}$ . The dual operator of  $\mathcal{S}_{\text{sub}}$ , which is denoted by  $\mathcal{S}_{\text{sub}}^*$ , can be interpreted as a zero-filling process in the unmeasured parts of  $\mathbf{b}$  in terms of  $\mathbf{b}_{\text{full}}$ , as shown in Fig 10. The standard FBP algorithm for the zero-filled data  $\mathcal{S}_{\text{sub}}^*\mathbf{b}$  provides the image  $\mathbf{x} = \mathbf{A}_{\text{full}}^{-1}\mathcal{S}_{\text{sub}}^*\mathbf{b}$  with cupping artifacts [18, 66]. Then, our reconstruction problem is the following:

$$(3.22) \quad \begin{aligned} &\text{Find a function } f : \mathbf{x} \mapsto \mathbf{y} \\ &\text{satisfying } \mathbf{A}^\# \mathbf{A}\mathbf{u} = \mathbf{x} \text{ and } \mathbf{y} = \mathbf{u}\chi_{\Omega_{\text{ROI}}}. \end{aligned}$$

To explain the learnability of  $f : \mathbf{x} \mapsto \mathbf{y}$ , we consider the continuous version. The Hilbert transform of  $\mathbf{u}$  with  $\boldsymbol{\theta}$  is defined by

$$(3.23) \quad \mathcal{H}_{\boldsymbol{\theta}}\mathbf{u}(\mathbf{t}) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{\mathbf{u}(\mathbf{t}_{\boldsymbol{\theta}}(a))}{a - \mathbf{t} \cdot \boldsymbol{\theta}} da$$

where  $\mathbf{t}_{\boldsymbol{\theta}}(a)$  is the point given by

$$(3.24) \quad \mathbf{t}_{\boldsymbol{\theta}}(a) = a\boldsymbol{\theta} + (\mathbf{t} \cdot \boldsymbol{\theta}^\perp)\boldsymbol{\theta}^\perp, \quad (\boldsymbol{\theta}^\perp = (-\sin \varphi, \cos \varphi)).$$

Note that  $\{\mathbf{t}_{\boldsymbol{\theta}}(a) : a \in \mathbb{R}\}$  is the  $\boldsymbol{\theta}$  directional line passing through  $\mathbf{t}$ . Let us define

$$(3.25) \quad \mathcal{R}_{\boldsymbol{\theta}_0}^* h(\mathbf{t}) = \int_{\varphi_0}^{\varphi_0 + \pi} h(\mathbf{t} \cdot \boldsymbol{\theta}, \varphi) d\varphi$$

Most interior tomography algorithms are based on the following identity [51, 66] :

$$(3.26) \quad \mathbf{u}(\mathbf{t}_{\boldsymbol{\theta}_0}(a)) = \frac{1}{2} \mathcal{R}_{\boldsymbol{\theta}_0}^* \mathcal{H}_{\boldsymbol{\theta}_0} \frac{\partial}{\partial a} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(a))$$

Applying the Hilbert transform to both sides of the above identity,

$$(3.27) \quad \mathcal{H}_{\boldsymbol{\theta}_0} \mathbf{u}(\mathbf{t}_{\boldsymbol{\theta}_0}(a)) = -\frac{1}{2\pi} \mathcal{R}_{\boldsymbol{\theta}_0}^* \frac{\partial}{\partial a} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(a))$$

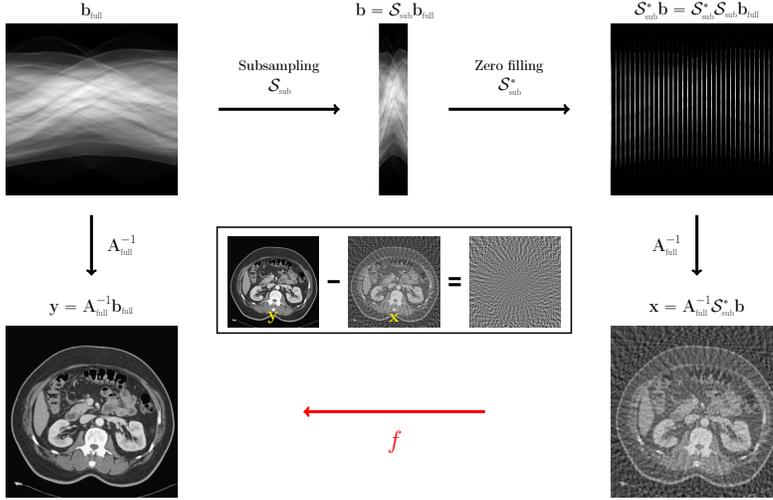


FIG. 11. Sparse-view CT aims to reconstruct an image  $\mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}}$  from uniformly under-sampled data  $\mathbf{b} = \mathcal{S}_{\text{sub}} \mathbf{b}_{\text{full}}$ , where  $\mathcal{S}_{\text{sub}}$  is the subsampling operator. Applying the deep learning method, we attempt to learn  $f$  that produces  $\mathbf{y}$  from the input  $\mathbf{x} = \mathbf{A}_{\text{full}}^{-1} \mathcal{S}_{\text{sub}}^* \mathbf{b}$ . The map  $f$  is learnable due to a simple structure of the residual  $\mathbf{x} - \mathbf{y}$ .

Given  $\mathbf{t} \in \Omega_{\text{ROI}}$  and  $\boldsymbol{\theta}_0$ , we have the following identity: For  $\mathbf{t}_{\boldsymbol{\theta}_0}(a) \in \Omega_{\text{ROI}}$ ,

$$(3.28) \quad \mathbf{u}(\mathbf{t}_{\boldsymbol{\theta}_0}(a)) = \Psi_{\boldsymbol{\theta}_0}^{\text{in}} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(a)) + \Psi_{\boldsymbol{\theta}_0}^{\text{out}} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(a))$$

where

$$(3.29) \quad \Psi_{\boldsymbol{\theta}_0}^{\text{in}} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(a)) = \frac{1}{2\pi^2} \int_{\mathbf{t}_{\boldsymbol{\theta}_0}(a') \in \Omega_{\text{ROI}}} \frac{\mathcal{R}_{\boldsymbol{\theta}_0}^* \frac{\partial}{\partial a'} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(a'))}{a' - a} da'$$

and

$$(3.30) \quad \Psi_{\boldsymbol{\theta}_0}^{\text{out}} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(a)) = \frac{1}{2\pi^2} \int_{\mathbf{t}_{\boldsymbol{\theta}_0}(a') \notin \Omega_{\text{ROI}}} \frac{\mathcal{R}_{\boldsymbol{\theta}_0}^* \frac{\partial}{\partial a'} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(a'))}{a' - a} da'.$$

The main point is that  $\Psi_{\boldsymbol{\theta}_0}^{\text{out}} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(a))$  is analytic in the line segment  $\gamma = \{a \in \mathbb{R} : \mathbf{t}_{\boldsymbol{\theta}_0}(a) \in \Omega_{\text{ROI}}\}$  [66]. This means that  $\Psi_{\boldsymbol{\theta}_0}^{\text{out}} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(\gamma))$  is completely determined by its knowledge in any open subset of  $\gamma$ . Consequently,  $\mathbf{u}(\mathbf{t}_{\boldsymbol{\theta}_0}(\gamma))$  can be recovered from  $\Psi_{\boldsymbol{\theta}_0}^{\text{in}} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(\gamma))$  and the information of  $\Psi_{\boldsymbol{\theta}_0}^{\text{out}} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(\gamma))$  in the small open subset of  $\gamma$ .

Now, we revisit the original discrete problem (3.22). Finding the function  $f : \mathbf{x} \mapsto \mathbf{y}$  is equivalent to finding the correction of the residual  $\mathbf{x} - \mathbf{y}$ . The analytic property of  $\Psi_{\boldsymbol{\theta}_0}^{\text{out}} \mathbf{b}(\mathbf{t}_{\boldsymbol{\theta}_0}(\gamma))$  explains the structure of the residual  $\mathbf{x} - \mathbf{y}$  in  $\Omega_{\text{ROI}}$ . Owing to the analytic structure of  $\mathbf{x} - \mathbf{y}$  along the line segments,  $\mathbf{x} - \mathbf{y}$  in  $\Omega_{\text{ROI}}$  has very different image structure from that of  $\mathbf{y}$  (medical image); therefore,  $\mathbf{x}$  can be decomposed into  $\mathbf{y}$  and  $\mathbf{x} - \mathbf{y}$ . Hence, the function  $f : \mathbf{x} \mapsto \mathbf{y}$  is learnable.

**3.3. Sparse-view CT.** The sparse-view CT problem aims to find a reconstruction function that maps from a sparse-view sinogram  $\mathbf{b}$  to an image whose quality is as high as that of a regular CT image reconstructed by full-view sinogram  $\mathbf{b}_{\text{full}}$ . Throughout this section, we will denote the sub-sampling operator by  $\mathcal{S}_{\text{sub}}$ , so  $\mathbf{b} = \mathcal{S}_{\text{sub}} \mathbf{b}_{\text{full}}$ .

Assuming that the subsampled data  $\mathbf{b} = \mathcal{S}_{\text{sub}}(\mathbf{b}_{\text{full}})$  violates the Nyquist's rule, the standard FBP algorithm using  $\mathbf{b}$  produces a streaking artifacted CT image, which can be expressed as

$$(3.31) \quad \mathbf{x} = \mathbf{A}_{\text{full}}^{-1} \mathcal{S}_{\text{sub}}^* \mathbf{b}$$

where  $\mathcal{S}_{\text{sub}}^* \mathbf{b}$  is a zero-filled data of  $\mathbf{b}$  and  $\mathcal{S}_{\text{sub}}^*$  is the dual operator of  $\mathcal{S}_{\text{sub}}$ .

The corresponding high quality image reconstructed from  $\mathbf{b}_{\text{full}}$  (satisfying the Nyquist's rule) is given by

$$(3.32) \quad \mathbf{y} = \mathbf{A}_{\text{full}}^{-1} \mathbf{b}_{\text{full}}$$

The goal is to learn the function  $f : \mathbf{x} \mapsto \mathbf{y}$  using  $\{(\mathbf{x}^{(k)}, \mathbf{y}^{(k)})\}_{k=1}^{\text{n}_{\text{data}}}$ .

The image structures of  $\mathbf{x} - \mathbf{y}$  and  $\mathbf{y}$  are very different from each other, as shown in Figure 11. Numerous researches have been conducted on image enhancement methods by suppressing noise  $\mathbf{x} - \mathbf{y}$ . The following CS technique is widely used to alleviate the noise  $\mathbf{x} - \mathbf{y}$  :

$$(3.33) \quad \mathbf{y} = \underset{\mathbf{y}}{\operatorname{argmin}} \|\mathbf{A}^\# \mathbf{A} \mathbf{y} - \mathbf{x}\|_{\ell^2}^2 + \lambda \|\nabla \mathbf{y}\|_{\ell^1}$$

where  $\|\nabla \mathbf{y}\|_{\ell^1}$  is used to penalize the undesired feature  $\mathbf{x} - \mathbf{y}$ . According to [10, 9], the convex minimization problem (3.33) is somehow close to the following problem:

$$(3.34) \quad \mathbf{y} = \underset{\mathbf{y} \in W_\alpha}{\operatorname{argmin}} \|\mathbf{A}^\# \mathbf{A} \mathbf{y} - \mathbf{x}\|_{\ell^2}^2$$

where  $W_\alpha := \{\mathbf{y} : \|\nabla \mathbf{y}\|_0 \leq \alpha\}$ ,  $\alpha$  is a positive integer, and  $\|\nabla \mathbf{y}\|_0$  indicates the number of non-zero entries of  $\nabla \mathbf{y}$ . Since  $W_\alpha$  is a finite union of the  $\alpha$ -dimensional space, the constraint  $\mathbf{y} \in W_\alpha$  shrinks the domain of solutions by enforcing sparsity. If  $\alpha$  is sufficiently smaller than  $m$  (the number of equations), then  $\mathbf{A}$  satisfies  $\alpha$ -RIP condition [9] within the sparse set  $W_\alpha$ . Namely,  $\|\mathbf{A} \mathbf{y} - \mathbf{A} \mathbf{y}'\| \neq 0$  for any different images  $\mathbf{y} \neq \mathbf{y}'$  in  $W_\alpha$ , so that the uniqueness of the problem (3.34) can be guaranteed within the sparse set  $W_\alpha$ .

Assuming that  $\mathbf{A}$  satisfies the  $\alpha$ -RIP condition and that  $\mathcal{M}_{\text{image}}$  is given by  $W_\alpha$ ,  $\mathbf{A}$  satisfies the  $\mathcal{M}$ -RIP condition in (3.1), and, thus, the reconstruction  $f$  is learnable from Observation 3.2. In the sparse-view CT problem, as several CS methods exhibit fairly successful reconstruction results [73, 42], the forward operator  $\mathbf{A}$  seems to possess some property somewhat closely related with RIP condition on the sparse set  $W_\alpha$ . However, the handmade set  $W_\alpha$  as prior knowledge can be viewed as a very rough approximation to the manifold  $\mathcal{M}_{\text{image}}$ , and hence it is limited in its ability to preserve small details with important medical information. Owing to the highly curved structure of  $\mathcal{M}_{\text{image}}$  as observed in Section 2.2, deep learning approaches [36, 26, 67] have been proposed to facilitate the machine-learned intrinsic regularizer using training data.

**REMARK 3.5.** *Let us give a brief comment on  $\mathbf{A}_{\text{full}}^{-1}$ . In general, the solution of  $\mathbf{A}_{\text{full}} \mathbf{y} = \mathbf{b}$  can be expressed as  $\mathbf{y} = (\mathbf{A}_{\text{full}}^* \mathbf{A}_{\text{full}})^{-1} \mathbf{A}_{\text{full}} \mathbf{b}$  and therefore  $\mathbf{A}_{\text{full}}^{-1}$  should be understood as  $(\mathbf{A}_{\text{full}}^* \mathbf{A}_{\text{full}})^{-1} \mathbf{A}_{\text{full}}$ . In CT,  $\mathbf{A}_{\text{full}}^{-1}$  represents the operator corresponding to the FBP algorithm as a practically feasible reconstruction.*

**4. Discussion.** In this section, we discuss several interesting issues related to deep learning-based solvability for underdetermined problems in medical imaging. An important question is ‘‘what is the minimum ratio of undersampling to provide guarantee of accurate reconstruction?’’. It is closely related to the dimension of the manifold.

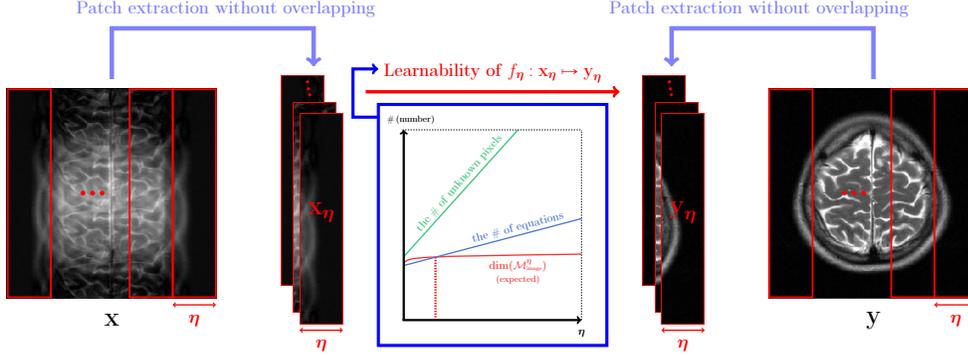


FIG. 12. Learnability of  $f_\eta: \mathbf{x}_\eta \mapsto \mathbf{y}_\eta$ . Let  $\mathbf{y}_\eta$  denote an image patch of size  $256 \times \eta$  extracted from a  $256 \times 256$  MR image  $\mathbf{y}$  and  $\mathbf{x}_\eta$  be an aliased image obtained by  $\mathbf{x}_\eta = \mathbf{A}^\# \mathbf{A} \mathbf{y}_\eta$ . The learnability of reconstruction map  $f_\eta$  seems to be related with the dimension of solution set  $\mathcal{M}_{image}^\eta$ . Here,  $\mathcal{M}_{image}^\eta$  is the set of all extracted patches from  $256 \times 256$  MR images.

**4.1. Solvability Issue.** This subsection discuss an interesting characteristic of the learning problem in the underdetermined MRI described in Section 3.1, where the uniform subsampling of factor 4 with additional phase encoding lines is used as the subsampling strategy. For the ease of explanation, we assume  $n = 256 \times 256$  (i.e.  $\mathbf{y}$  represents a  $256 \times 256$  MR image) and uniform subsampling of factor 4 adding 12 supplementary phase encoding lines.

For a given integer  $\eta \geq 1$ , let  $\{\mathbf{y}_\eta^{(j)} \in \mathbb{R}^{256 \times \eta}\}_{j=1}^{n_{\text{patch}}}$  be a set of the image patches extracted from an image  $\mathbf{y}$  and let  $\{\mathbf{x}_\eta^{(j)}\}_{j=1}^{n_{\text{patch}}}$  be the corresponding set of the aliased images, given by  $\mathbf{x}_\eta^{(j)} = \mathbf{A}^\# \mathbf{A} \mathbf{y}_\eta^{(j)}$ . We assume that there is no overlap between all the patches. (See Fig. 12.)

This section aims to investigate whether the factor of  $\eta$  is important in learning  $f_\eta: \mathbf{x}_\eta \mapsto \mathbf{y}_\eta$ . To observe the effects of  $\eta$ , we train the U-net by varying  $\eta$ , using the following training dataset:

$$(4.1) \quad \{(\mathbf{x}_\eta^{(j,k)}, \mathbf{y}_\eta^{(j,k)}) \mid j = 1, \dots, n_{\text{patch}} \text{ and } k = 1, \dots, n_{\text{data}}\}$$

where  $\mathbf{y}_\eta^{(j,k)}$  represents  $j$ -th image patch extracted from the  $n$ -th label MR image  $\mathbf{y}^{(k)}$ .

The reconstruction map  $f_\eta$  aims to solve the linear system that has  $76 \times \eta$  number of equations with  $256 \times \eta$  number of unknowns. As  $\eta$  increases, the number of unknowns increases more rapidly than the number of equations. See the middle box in Fig. 12. However, our experimental results, described in Fig. 13, demonstrate that the learning ability is gradually improved as  $\eta$  increases.

The experimental results in Fig. 13 can be explained by means of the dimensionality of the manifold given by

$$(4.2) \quad \mathcal{M}_{image}^\eta := \{\mathbf{y}_\eta \mid \mathbf{y}_\eta \text{ is a } 256 \times \eta \text{ patch extracted from } \mathbf{y}\}$$

The dimension of  $\mathcal{M}_{image}^\eta$ , denoted by  $g_{\mathcal{M}}(\eta)$ , can be viewed as a function of  $\eta$  variable. As shown in the middle box in Fig. 12,  $g_{\mathcal{M}}(\eta)$  seems to grow very slowly; therefore,  $g_{\mathcal{M}}(\eta)$  might intersect with the linear function  $g_{\# \text{equations}}(\eta) = 76\eta$  (i.e. the number of equations). Assuming  $\eta^*$  is the intersection point (i.e.  $g_{\mathcal{M}}(\eta^*) = g_{\# \text{equations}}(\eta^*)$ ),  $f_\eta$  can be regarded as learnable, provided  $\eta \geq \eta^*$ . This interpretation can be supported by the error estimations in Fig. 13.

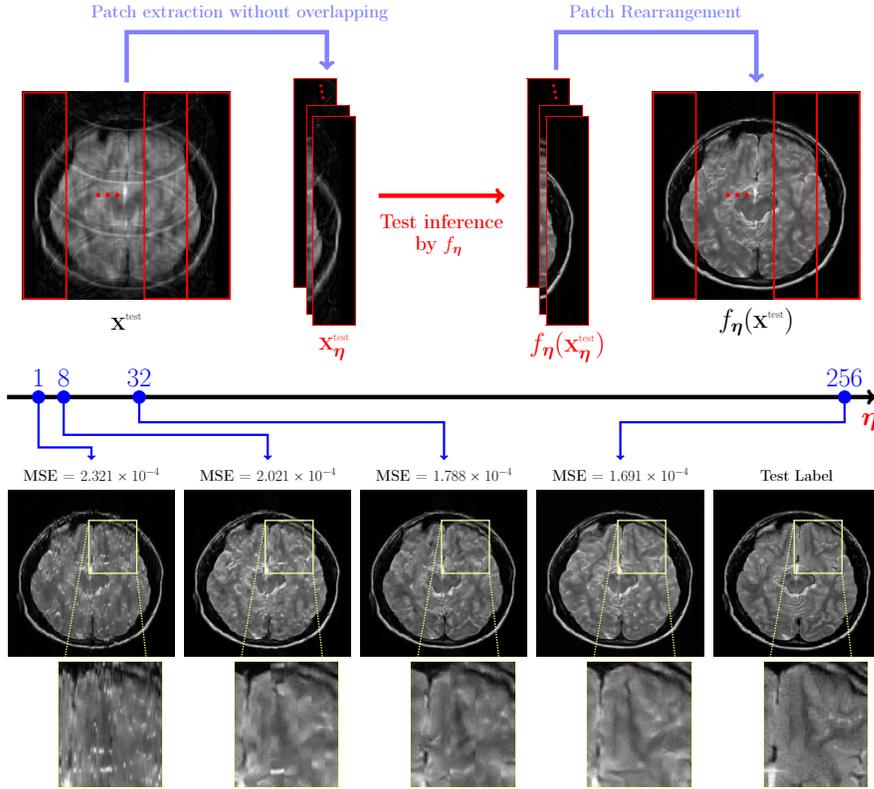


FIG. 13. Test performance evaluation of  $f_\eta$  with various  $\eta = 1, 8, 32,$  and  $256$ . For each  $\eta$ , the function  $f_\eta$  is obtained by training the U-net with the corresponding image patches extracted from 1500 MR images  $\{\mathbf{y}^{(k)}\}_{k=1}^{1500}$ . To compare quantitative performances with one another, a given test image  $\mathbf{x}^{test} \notin \{\mathbf{x}^{(k)}\}_{k=1}^{1500}$  is divided into image patches, where each patch is reconstructed through the trained U-net  $f_\eta$ , and the reconstruction outputs are rearranged into one image  $f_\eta(\mathbf{x}^{test})$ . We also qualitatively evaluate the test result qualitatively by computing the mean squared error (MSE) between the inference output  $f_\eta(\mathbf{x}^{test})$  and label  $\mathbf{y}^{test}$ .

Let us explain the reasons for expecting  $g_{\mathcal{M}}(\eta)$  to grow significantly slowly as  $\eta$  increases. Assume that  $\mathcal{M}_{\text{image}}$  is the set of all the human head MR images. Then, all the images in  $\mathcal{M}_{\text{image}}$  possess a similar anatomical structure that consists of skull, gray matter, white matter, cerebellum, among others. In addition, every skull and tissue in the image have distinct features that can be represented nonlinearly by a relatively small number of latent variables, and so does for the entire image. Notably, the skull and tissues of the image are spatially interconnected, and even if a part of the image is missing, the missing part can be recovered with the help of the surrounding image information. This is the reason that image inpainting techniques [5] have been successful in image restoration for filling-in the missing areas in images. These observations seem to indicate that  $g_{\mathcal{M}}(\eta)$  does not change much with  $\eta$  near  $\eta = 256$ , where  $g_{\mathcal{M}}(256)$  corresponds to the dimension of the entire image. Therefore, we expect that  $g_{\mathcal{M}}(256) \ll g_{\# \text{equations}}(256)$ , so that there exists  $\eta^*$  (the turning point for learnability) such that  $g_{\mathcal{M}}(\eta^*) = g_{\# \text{equations}}(\eta^*)$ . If the curse-of-dimensionality does not matter, it may be better to learn 3D images in total rather than dividing 3D images into multiple pieces. A rigorous mathematical analysis of this issue is the

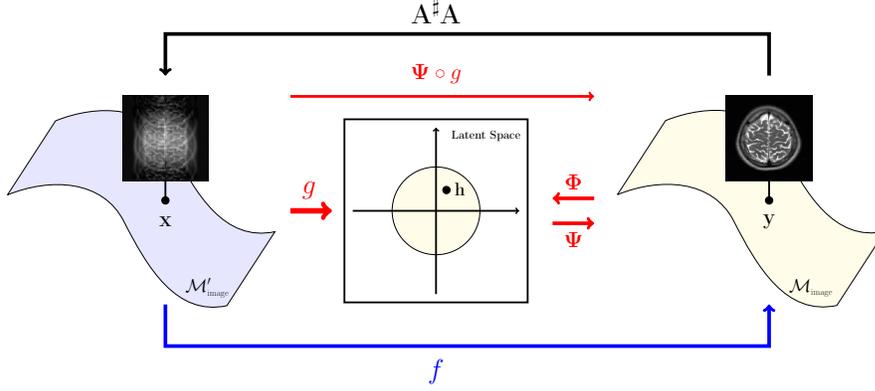


FIG. 14. What happens if a low dimensional latent representation is possible? Imagine that we have a low dimensional latent generator  $\Psi : \mathbf{h} \mapsto \mathbf{y}$  and an encoder  $\Phi : \mathbf{y} \mapsto \mathbf{h}$  such that  $\Psi \circ \Phi(\mathbf{y}) \approx \mathbf{y}$  for all  $\mathbf{y} \in \mathcal{M}_{\text{image}}$ . If we have a map  $g : \mathbf{x} \mapsto \mathbf{h} = \Psi(\mathbf{y})$ , then the reconstruction map  $f : \mathbf{x} \mapsto \mathbf{y}$  is given by  $f = \Psi \circ g$ .

subject of our future research topic.

**4.2. Some issues on learning low dimensional representation.** In our undersampled problems, the dimension of  $\mathbf{y}$  (i.e., the total number of pixels in the image) is considerably bigger than the dimension of  $\mathbf{b}$  (i.e., the number of independent components in the measurement data). Since there exist infinitely many images that solve the mathematical model  $\mathbf{A}\mathbf{y} = \mathbf{b}$ , we need to reflect prior information about the unknown solution manifold, either implicitly or explicitly, in the image restoration process.

Over several decades, various regularization approaches have been used with pre-defined convex regularization functionals in order to incorporate a-priori information on  $\mathbf{y}$ . CS methods involving  $\ell^1$ -norm regularization minimization have been powerful for noise removal, whereas they suffer from limitations in preserving small features. In medical imaging, there are a variety of small features, such that the difference in the data fidelity is very small as compared with that in normalization, whether or not those small features are present. Hence, finding a more sophisticated normalization to keep small features remains a challenging problem.

Unlike in CS (predetermined convex-norm-based approach), deep learning can be viewed as a black box model approach where training data is used for probing the solution manifold. To ensure the possibility of solving undersampled problems through deep learning, it would be desirable to investigate the performance of low dimensional representation learning for the unknown manifold from training data.

Autoencoder(AE) techniques (as the natural evolution of PCA) are widely used to find a low dimensional representation for the unknown  $\mathcal{M}_{\text{image}}$  from the known training data  $\{\mathbf{y}^{(k)}\}_{k=1}^{\text{n}_{\text{data}}}$  [29, 40]. The AE consists of a encoder  $\Phi : \mathbf{y} \rightarrow \mathbf{h}$  for a compressed latent representation and a decoder  $\Psi : \mathbf{h} \rightarrow \mathbf{y}$  for providing  $\Psi \circ \Phi(\mathbf{y}) \approx \mathbf{y}$  (i.e., an output image is similar to the original input image). Assuming that  $\Phi$  provides a satisfactory approximation of the solution manifold, the underdetermined problem (1.3) can be solved as follows:

$$(4.3) \quad f(\mathbf{x}) = \Psi(\mathbf{h}), \quad \mathbf{h} = \underset{\mathbf{h}}{\operatorname{argmin}} \|\mathbf{A}\Psi(\mathbf{h}) - \mathbf{A}\mathbf{x}\|$$

A deep learning technique can be used to solve the problem (4.3), as in the minimiza-

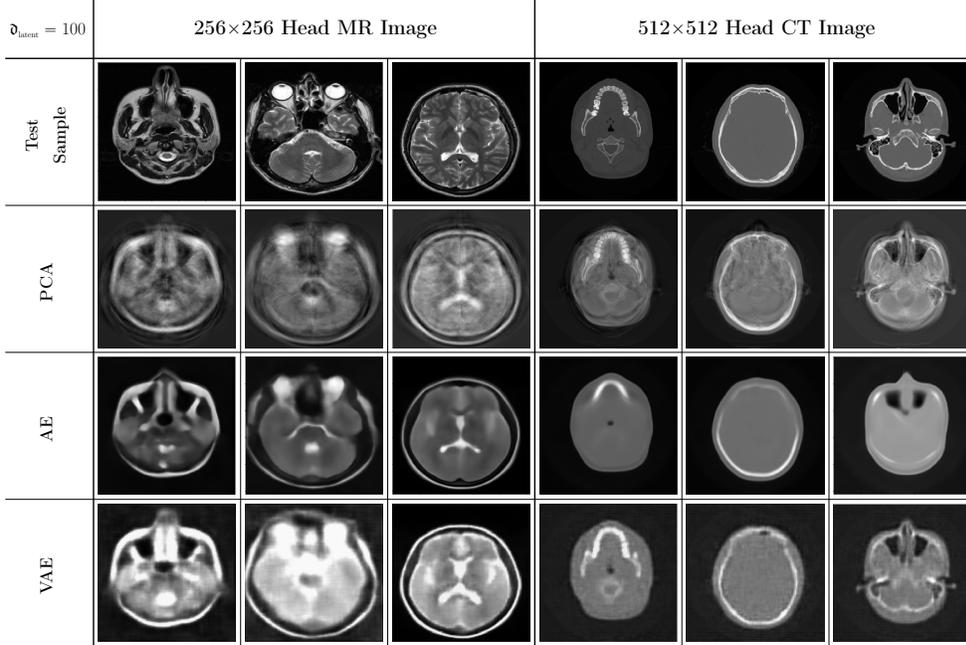


FIG. 15. Learning a low dimensional representation for MR and CT images with three different dimension reduction techniques (principal component analysis(PCA), auto-encoder(AE), and variational AE(VAE)). By using each technique, an encoder  $\Phi$  and decoder  $\Psi$  are trained so that MR or CT images are projected into a 100 dimensional space (i.e.  $d_{\text{latent}} = 100$ ) by using 1800 MR images of pixel dimension  $256 \times 256$  or using 3400 CT images of pixel dimension  $512 \times 512$ . PCA was performed by a built-in function `pca` in MATLAB. For AE and VAE, we used convolutional AE and VAE structures, respectively, consisting of modified residual blocks from ResNET [27, 28]. The networks were implemented in the Tensorflow environment. After the training, we tested three different samples for each MR and CT case, displayed in the first row, and the corresponding test results are displayed from the second to last row.

tion problem (4.3) it might be difficult to use the standard gradient descent method because of the complex deep learning structure of the decoder  $\Psi$ . If the dimension of the latent space is reasonably small, the reconstruction map is achieved by

$$(4.4) \quad f(\mathbf{x}) = \Psi \circ g(\mathbf{x}), \quad g := \operatorname{argmin}_{g \in \mathbb{N}\mathbb{N}} \sum_{k=1}^{n_{\text{data}}} \|g(\mathbf{x}^{(k)}) - \mathbf{h}^{(k)}\|$$

where  $\mathbf{h}^{(k)} = \Phi(\mathbf{y}^{(k)})$ . One can refer to Fig. 14 for the schematic understanding of the method. Recent papers have reported that the AE-based approaches show remarkable performances in several applications [12, 60, 34, 64]. However, for high dimensional data, AEs seem to suffer from the blurring and loss of small details, as depicted in Fig. 15. Improving performance of AEs in high dimensional medical image applications is still a challenging issue.

sGenerative adversarial networks (GANs) have been utilized to generate realistic images via interactions between learning and synthesis [20, 57, 71]. Typically, the architecture of GANs comprises two main parts; generator  $\Psi$  and discriminator  $\Gamma$ . The GAN network aims to find a  $\Psi$  that maps from a random noise vector  $\mathbf{h}$  in the latent space to an image in a real data distribution associated with the training

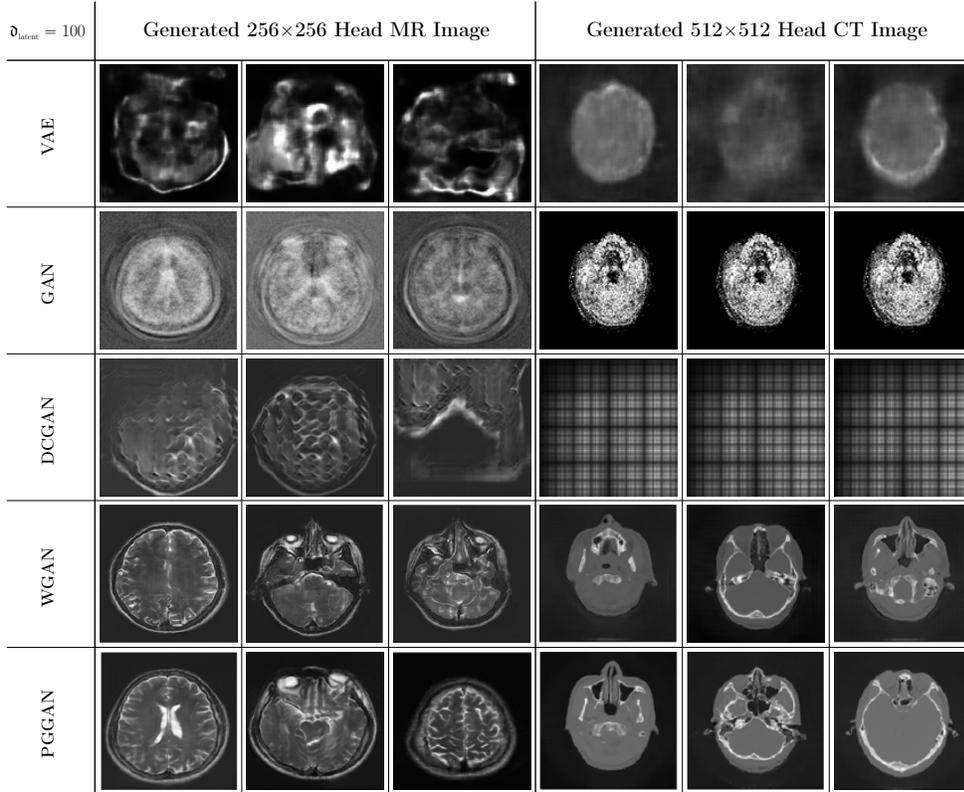


FIG. 16. Learning low dimensional representation for MR and CT images using five different generative models, variational auto-encoder (VAE), generative adversarial network (GAN), deep convolutional GAN (DGGAN), Wasserstein GAN (WGAN), and progressive generative GAN (PGGAN). Using these generative models, we learn a generator  $\Psi$  that synthesizes MR or CT images from points sampled in 100 dimensional Gaussian distribution (i.e.  $d_{\text{latent}} = 100$ ). In our experiments, we used a typical network structure for each generative model, as described in [40, 20, 57, 22, 37]. We replaced the cross-entropy of GAN into Hinge loss, and the Wasserstein loss of PGGAN into Hinge loss with spectral normalization in order to obtain the stability on learning [72]. All the networks were implemented in the Tensorflow environment with 1800 MR images and 3400 CT images. In each row, three different synthesized images from each trained generative models are displayed for each CT and MR case.

data  $\{\mathbf{y}^{(k)}\}_{k=1}^{n_{\text{data}}}$ . The generator  $\Psi$  is trained with the assistance of the discriminator  $\Gamma$  in such a way that  $\Gamma$  misclassifies  $\Psi(\mathbf{h})$  as a real image. The training procedure of GAN can be viewed as a performance competition between the generator and the discriminator. Although GANs have achieved remarkable success in generating various realistic images, there exist some limitations in synthesizing high resolution medical data. The GAN's approach makes it difficult to deal with high-dimensional data because the generated image can be easily distinguished from the training data, which can lead to collapse or instability during the training process [53]. Several variations of GAN, such as Wasserstein GAN (WGAN) [4, 22] and progressive growing GAN (PGGAN) [37], have been developed to deal with the training instability. WGAN uses Wasserstein distance, which may improve the loss sensitivity with respect to change of parameters, compared to the Jensen-Shannon distance used in the original

GAN. Fig. 16 shows that in our WGAN experiment, nearly plausible synthesis results are generated for high dimensional medical image, whereas the synthesized images still suffer from somewhat lack of reality. PGGAN facilitates synthesis of high dimensional data via hierarchical multi-scale learning fashion from low resolution to the desired high resolution; therefore, the network can focus on the overall structures at the beginning of the process, before shifting attention gradually to finer scale details via later connections as the training advances. Fig. 16 depicts our PGGAN experiment with the spectral weight normalization [48], that provides high quality synthesis.

Unfortunately, unlike bidirectional AE, which provides an explicit prior, GANs learn only the unidirectional mapping  $\Psi : \mathbf{h} \mapsto \mathbf{y}$  and, therefore, it is difficult to achieve (4.4). Recently, new strategies have been developed to use the implicit prior obtained using GANs as a solution prior for solving ill-posed inverse problems such as the undersampled MRI [49], image denoising [65], and inpainting [69]. Also, a scattering generator [3] with the advantages of both GAN and AE has been proposed. However, applications in high dimensional medical imaging problems are still far from satisfactory.

**5. Conclusion.** This work concerns with the solvability of undersampling problems with the use of training data. The undersampled MRI, sparse view CT, and interior tomography are typical examples of the underdetermined problem, which involve much fewer equations (measured data) than unknowns (pixels of the image). To compensate for the uncertainty of the huge number of free parameters (difference between the number of equations and the number of unknowns), we need to limit the solution manifold using prior information of the expected images. Regularization techniques have been widely used to impose very specific prior distributions on the expected images, such as penalizing a special norm of images (or promoting sparsity in expressions). However, norm-based regularization might not actually be able to provide a clinically useful image properly in advance. Well-known CS methods, which employ random sampling and are based on regularization methods, are effective in alleviating highly oscillatory noise while maintaining the overall structure; however, sparse sensing techniques tend to eliminate small anomalies, as shown in Section 2.2.2.

DL techniques appear to deal with various underdetermined inverse problems by effectively probing the unknown nonlinear manifold  $\mathcal{M}_{\text{image}}$  through training data. It seems to handle the uncertainty of solutions to the highly ill-posed problems.

According to Hadamard [24], the linear problem  $\mathbf{A}\mathbf{y} = \mathbf{b}$  is well-posed if the following two conditions hold (while ignoring the existence issue): first, for each  $\mathbf{b}$ , it has a unique solution, and second, the solution is stable under the perturbation of  $\mathbf{b}$ . However, we note that whether the problem is well-posed depends on the choice of the solution space. Many ill-posed problems can be well-posed within the constrained solution spaces (e.g. sparse solution spaces). For a simple example, we consider the Poisson's equation  $\nabla \cdot \nabla u = \mathbf{b}$  in the 2-D domain  $\Omega = \{(r \cos \theta, r \sin \theta) \mid 0 < r < 1, 0 < \theta < \frac{3\pi}{2}\}$  with the homogeneous Dirichlet boundary condition  $u|_{\partial\Omega} = 0$ . If  $u_*$  is a solution, so are  $u = u_* + (r^{\frac{2}{3}n} - r^{-\frac{2}{3}n}) \sin(\frac{2}{3}\theta)$  for  $n = 0, 1, 2, \dots$ . Hence, the problem is ill-posed without the constraint of the Sobolev space  $H^1(\Omega) = \{u \mid \int_{\Omega} |u|^2 + |\nabla u|^2 < \infty\}$ . Similarly, the underdetermined problem  $\mathbf{A}\mathbf{y} = \mathbf{b}$  can be well-posed under a suitable solution manifold  $\mathcal{M}_{\text{image}}$ . DL methods seem to possess ambiguous capability of learning data representation.

Recently, several experiments regarding adversarial classifications [19, 13] (e.g., false positive output of cancer) have shown that deep neural networks obtained via gradient descent-based error minimization procedure are vulnerable to various noisy-

like perturbations, resulting in incorrect output (that can be critical in medical environments). These experiments show that a well-trained function  $f$  in (1.5) works only in the immediate vicinity of a manifold, whereas producing incorrect results if the input deviates even slightly from the training data manifold. In practice, the measured data is exposed to various noise sources such as machine dependent noise; therefore, the developed algorithm must be stable against the perturbations due to noise sources. Hence, normalization of the input data is essential for improving robustness and generalizability of the deep learning network against adversarial attacks [21, 70, 19].

For input data normalization, we attempt to project the input  $\mathbf{x}$  to its normalized form  $\mathfrak{N}(\mathbf{x})$  in the way that two images  $\mathbf{x}$  and  $\mathfrak{N}(\mathbf{x})$  are almost the same from the viewpoint of radiologists. It is quite complicated to define the distance  $\text{dist}_{\text{radiologist}}(\mathbf{x}, \mathbf{x}')$  in terms of radiologist's view. If we have a good generator  $\mathfrak{G}$ , it can be defined as

$$(5.1) \quad \text{dist}_{\text{radiologist}}(\mathbf{x}, \mathbf{x}') = \|\mathbf{h} - \mathbf{h}'\|$$

where  $\mathfrak{G}(\mathbf{h}) = \mathfrak{N}(\mathbf{x})$  and  $\mathfrak{G}(\mathbf{h}') = \mathfrak{N}(\mathbf{x}')$ . The Euclidian distance  $\|\mathbf{h} - \mathbf{h}'\|$  can be somewhat equivalent to the geodesic distance between  $\mathfrak{N}(\mathbf{x})$  and  $\mathfrak{N}(\mathbf{x}')$  on the manifold. For example, given a noisy input  $\mathbf{x}$ , its normalization  $\mathfrak{N}(\mathbf{x})$  can be a denoised image while preserving the salient features of  $\mathbf{x}$ . The issues of finding the normalization  $\mathfrak{N}$  and the generator  $\mathfrak{G}$  would be very challenging tasks. Although there are still several challenging issues in deep learning associated with solving ill-posed inverse problems in medical imaging area, recent remarkable developments indicate that it has enormous potential to provide a useful means of overcoming the limitations of the traditional methods.

#### REFERENCES

- [1] M. Aharon and M. Elad, "Sparse and redundant modeling of image content using an image-signature-dictionary," *SIAM Journal on Imaging Sciences*, vol. 1, 2008, pp. 228–247.
- [2] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, 2006, pp. 4311–4322.
- [3] T. Angles and S. Mallat, "Generative networks as inverse problems with scattering transforms," *ICLR*, 2018.
- [4] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv:1701.07875*, 2017.
- [5] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 2000, pp. 417–424.
- [6] D. J. Brenner and E. J. Hall, "Computed tomography scan increasing source of radiation exposure," *New England Journal of Medicine*, vol. 357, 2007, pp. 2277–2284.
- [7] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, 51 (2008), p. 3481.
- [8] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trnas. on Information Theory*, vol. 52, 2006, pp. 489–509.
- [9] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, 2006, pp. 1207–1223.
- [10] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, 2005, pp. 4203–4215.
- [11] E. J. Candès and T. Tao, "Reflections on compressed sensing," *IEEE Information Theory Society Newsletter*, vol. 58, 2008, pp. 20–23.
- [12] J. H. R. Chang, C. Li, B. Poczos, B. V. K. V. Kumar, and A. C. Sankaranarayanan, "One network to solve them all solving linear inverse problems using deep projection models," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 5889–5898.

- [13] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, et al., "Opportunities and obstacles for deep learning in biology and medicine," *Journal of The Royal Society Interface*, vol. 15, 2018, p. 20170387.
- [14] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, 2004, pp. 1413–1457.
- [15] D. Donoho, "Compressed sensing," *IEEE Trans. on Information Theory*, vol. 52, 2006, pp. 1288–1306.
- [16] D. L. Donoho and M. Elad, "optimally sparse representation in general (non-orthogonal) dictionaries via  $\ell_1$  minimization," *Proc. Natl Acad. Sci. USA*, vol. 100, 2003, pp. 2197–2202.
- [17] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image processing*, vol. 15, 2006, pp. 3736–3745.
- [18] A. Faridani, E. L. Ritman, and K. T. Smith, "Local tomography," *SIAM Journal on Applied Mathematics*, vol. 52, 1992, pp. 459–484.
- [19] S. G. Finlayson, J. D. Bowers, J. Ito, J. L. Zittrain, A. L. Beam, and I. S. Kohane, "Adversarial attacks on medical machine learning," *Science*, vol. 363, 2019, pp. 1287–1289.
- [20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 27, 2014, pp. 2672–2680.
- [21] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *International Conference on Learning Representations*, 2015.
- [22] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein GANs," arXiv:1704.00028, 2017.
- [23] E. Haacke, R. Brown, M. Thompson, and R. Venkatesan, "Magnetic resonance imaging physical principles and sequence design," New York: Wiley, 1999.
- [24] J. Hadamard, "Sur les problèmes aux dérivées partielles et leur signification physique," *Bull. Univ. Princeton*, vol. 13, 1902, pp. 49–52.
- [25] Y. Han, J. Gu, and J. C. Ye, "Deep learning interior tomography for region-of-interest reconstruction," arXiv, 2017.
- [26] Y. Han and J. C. Ye, "Framing u-net via deep convolutional framelets: Application to sparse-view CT," *IEEE transactions on medical imaging*, vol. 37, 2018, pp. 1418–1429.
- [27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [28] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, Springer, 2016, pp. 630–645.
- [29] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, 2006, pp. 504–507.
- [30] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. description of system," *British Journal of Radiology*, vol. 46, 1973, pp. 1016–1022.
- [31] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled MRI reconstruction," *Physics in Medicine and Biology*, vol. 63, 2018, p. 135007.
- [32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.
- [33] M. K. Islam, T. G. Purdie, B. D. Norrlinger, H. Alasti, D. J. Moseley, M. B. Sharpe, J. H. Siewerdsen, and D. A. Jaffray, "Patient dose from kilovoltage cone beam computed tomography imaging in radiation therapy," *Medical physics*, vol. 33, 2006, pp. 1573–1582.
- [34] S. Jalali and X. Yuan, "Using auto-encoders for solving ill-posed linear inverse problems," arXiv:1901.05045, 2019.
- [35] O. N. Jaspán, R. Fleysher, and M. L. Lipton, "Compressed sensing MRI: a review of the clinical literature," *The British journal of radiology*, vol. 88, 2015.
- [36] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process*, vol. 26, 2017, pp. 4509–4522.
- [37] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," *ICLR*, 2018.
- [38] K. Kawaguchi, L. P. Kaelbling, and Y. Bengio, "Generalization in deep learning," In *Mathematics of Deep Learning*, Cambridge University Press, 2017.
- [39] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [40] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," arXiv:1312.6114, 2013.
- [41] D. A. Koff and H. Shulman, "An overview of digital compression of medical images: can we

- use lossy image compression in radiology?," *Canadian Association of Radiologists Journal*, vol. 57, 2006, pp. 211–217.
- [42] H. Kudo, T. Suzuki, and E. A. Rashed, "Image reconstruction for sparse-view CT and interior CT - introduction to compressed sensing and differentiated backprojection," *Quantitative imaging in medicine and surgery*, vol. 3, 2013, p. 147.
- [43] P. Lauterbur, "Image formation by induced local interactions: Examples of employing nuclear magnetic resonance," *Nature*, vol. 242, 1973, pp. 190–191.
- [44] C. P. Loizou, V. Murray, M. S. Pattichis, I. Seimenis, M. Pantziaris, and C. S. Pattichis, "Multiscale amplitude-modulation frequency-modulation (am-fm) texture analysis of multiple sclerosis in brain MRI images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 15, 2010, pp. 119–129.
- [45] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, 2007, pp. 1182–1195.
- [46] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, 2008, pp. 53–69.
- [47] S. G. Mallat, "A wavelet tour of signal processing," Academic Press, 2009.
- [48] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," *ICLR*, 2018.
- [49] D. Narnhofer, K. Hammernik, F. Knoll, and T. Pock, "Inverse GANs for accelerated MRI reconstruction," *Proc. SPIE : Wavelets and Sparsity XVIII*, vol. 11138, 2019, pp. 381–392.
- [50] F. Natterer, "The mathematics of computerized tomography," John Wiley and Sons, 1986.
- [51] F. Noo, R. Clackdoyle, and J. D. Pack, "A two-step hilbert transform method for 2D image reconstruction," *Physics in Medicine and Biology*, vol. 49, 2004, pp. 3903–3923.
- [52] H. Nyquist, "Certain topics in telegraph transmission theory," *Trans. AIEE*, vol. 47, 1928, pp. 617 – 644.
- [53] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," *ICML*, 2017.
- [54] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, 1996, p. 607.
- [55] H. S. Park, J. K. Choi, and J. K. Seo, "Characterization of metal artifacts in X-ray computed tomography," *Communications on Pure and Applied Mathematics*, vol. 70, 2017, pp. 2191–2217.
- [56] B. Poole, S. Lahiri, M. Raghu, J. Sohl-Dickstein, and S. Ganguli, "Exponential expressivity in deep neural networks through transient chaos," 2016, pp. 3360–3368.
- [57] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *ICLR*, 2016.
- [58] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, 2015.
- [59] R. Rubinstein, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proceedings of the IEEE*, vol. 98, 2010, pp. 1045–1057.
- [60] J. K. Seo, K. C. Kim, A. Jargal, K. Lee, and B. Harrach, "A learning-based method for solving ill-posed nonlinear inverse problems: A simulation study of lung EIT," *SIAM Journal on Imaging Sciences*, vol. 12, 2019, p. 12751295.
- [61] J. K. Seo and E. J. Woo, "Nonlinear inverse problems in imaging," John Wiley and Sons, 2013.
- [62] J. K. Seo, E. J. Woo, U. Katscher, and Y. Wang, "Electro-magnetic tissue properties MRI," Imperial College Press, 2014.
- [63] D. K. Sodickson and W. Manning, "Simultaneous acquisition of spatial harmonics (smash): fast imaging with radiofrequency coil arrays," *Magn. Reson. Med.*, vol. 38, 1997, pp. 591–603.
- [64] K. C. Tezcan, C. F. Baumgartner, R. Luechinger, K. P. Pruessmann, and E. Konukoglu, "MR image reconstruction using deep density priors," *IEEE transactions on medical imaging*, 2018.
- [65] S. Tripathi, Z. C. Lipton, and T. Q. Nguyen, "Correction by projection: Denoising images with generative adversarial networks," *arXiv*, 2018.
- [66] G. Wang and H. Yu, "Meaning of interior tomography," *Physics in medicine and biology*, vol. 58, 2013, pp. R161–R186.
- [67] S. Xie, X. Zheng, Y. Chen, L. Xie, J. Liu, Y. Zhang, J. Yan, H. Zhu, and Y. Hu, "Artifact removal using improved googlenet for sparse-view CT reconstruction," *Scientific reports*, vol. 8, 2018, p. 6700.
- [68] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, 2010, pp. 2861–2873.
- [69] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do,

- “Semantic image inpainting with deep generative models,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5485 – 5493.
- [70] X. Yuan, P. He, Q. Zhu, and X. Li, “Adversarial examples: Attacks and defenses for deep learning,” *IEEE transactions on neural networks and learning systems*, 2019.
- [71] Xin Yi, Ekta Walia, and Paul Babyn, “Generative adversarial network in medical imaging: A review,” *Medical Image Analysis*, 2019, p. 101552.
- [72] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, “Self-attention generative adversarial networks,” *arXiv:1805.08318*, 2018.
- [73] Z. Zhu, K. Wahid, P. Babyn, D. Cooper, I. Pratt, and Y. Carter, “Improved compressed sensing-based algorithm for sparse-view CT image reconstruction”, *Computational and mathematical methods in medicine*, 2013.

**Supplementary Material.** This appendix provides deep learning results for undersampled MRI, interior tomography, and sparse-view CT problem. To learn a reconstruction map  $f : \mathbf{x} \mapsto \mathbf{y}$ , we adopt the modified U-net architecture, described in Fig. 6, and set feature depths of the network to be multiples of 64.

To train U-net, we generate a training dataset  $\{\mathbf{x}^{(k)}, \mathbf{y}^{(k)}\}_{k=1}^{\text{n}_{\text{data}}}$  in the following sense: Let  $\{\mathbf{y}^{(k)}\}_{k=1}^{\text{n}_{\text{data}}}$  be a given set of medical images. Here,  $\mathbf{y}^{(k)}$  represents  $256 \times 256$  head MR image in undersampled MRI problem and  $256 \times 256$  abdominal CT image in interior tomography and sparse-view CT problem. The corresponding input data  $\mathbf{x}^{(k)}$  is generated by computing  $\mathbf{x}^{(k)} = \mathbf{A}_{\text{full}}^{-1} \mathbf{S}_{\text{sub}}^* \mathbf{A} \mathbf{y}^{(k)}$ . In our real implementation, 1500 MR images [44] are used for undersampled MRI problem and 1800 CT images are used for interior tomography and sparse-view CT problem.

With the generated training dataset, U-net is trained by minimizing  $\ell^2$  loss, as mentioned in (2.22). the minimization process was performed using Adam optimizer with learning rate 0.001, mini-batch size 16, and 3000 epochs. In addition, Batch normalization [39] is used for mitigating the overfitting issue.

	Test input	Reference image	U-net reconstruction	Error
Undersampled MRI <small>(only uniform subsampling)</small>				
Undersampled MRI <small>(uniform subsampling with additional one phase encoding line)</small>				
Sparse-view CT				
Interior tomography				

FIG. 17. Deep learning results using U-net for undersampled MRI, sparse-view CT, and interior tomography problem. The images in first, second, third, and fourth column represent test input images, reference images, U-net reconstruction results, and errors, respectively.