PINER: Prior-informed Implicit Neural Representation Learning for Test-time Adaptation in Sparse-view CT Reconstruction

Bowen Song
Stanford University
bowens18@stanford.edu

Liyue Shen*
University of Michigan
liyues@umich.edu

Lei Xing*
Stanford University
lei@stanford.edu

Abstract

Recently, deep learning has been introduced to solve important medical image reconstruction problems such as sparse-view CT reconstruction. However, the developed deep reconstruction models are generally limited in generalization when applied to out-of-distribution samples in unseen domains. Furthermore, privacy concerns may impede the availability of source-domain training data to retrain or adapt the model to the target-domain testing data, which are quite common in real-world medical applications. To address these issues, we introduce a source-free black-box test-time adaptation method for sparse-view CT reconstruction with unknown noise levels based on prior-informed implicit neural representation learning (PINER). By leveraging implicit neural representation learning to generate the image representations at various noise levels, the proposed method is able to construct the adapted input representations at test time based on the inference of black-box model and output analysis. We performed experiments of source-free test-time adaptation for sparse-view CT reconstruction with unknown noise levels on multiple anatomical sites with different black-box deep reconstruction models, where our method outperforms the state-of-the-art algorithms. Code: https://github.com/efzero/PINER

1. Introduction

Medical imaging aims at reconstructing the computational images from the measurements data acquired by physical sensors to visualize the internal structure of the living subjects [37, 29]. For example, projection data are measured for computed tomography imaging (CT) while the frequency space (k-space) data are sampled for magnetic resonance imaging (MRI). In practice, it is often desirable to reduce the number of measurements required for reconstructing high-quality medical images. Specifically, sparse-view CT enables lower radiation dose exposing to patients while under-sampling k-space data speeds up the MRI scanning procedure. However, due to the information loss from the sparse-sampling measurement process, the sparse-sampling image reconstruction becomes an ill-posed inverse problems, which requires additional prior knowledge to reconstruct artifacts-free images. Before the era of deep learning, previous methods have focused on adding prior information as the regularization term in the optimization objective such as compressed sensing[5, 23, 50, 8]. With the advance of deep learning techniques, many learning-based methods are developed to reconstruct images by learning the mapping function from the measurements to image domain, with the prior information driven from large-scale training data[3, 14, 49, 33, 34, 36, 54]. Especially, some works focus on incorporating the physical constraints of the measurement process into the deep learning-based reconstruction models [33].

However, deep learning models always suffer from performance drop with data distributional shift [2], which limits the generalization of the trained model to new testing domains. For example, in clinics, the CT scanners and parameter settings vary a lot among different hospitals. Moreover, the radiation dose delivered to different patients undergoing different treatments may be also different, which causes the various noise level of the measurements data. All these changing factors in practical applications pose a significant challenge for the robustness and generalization of the deep-learning-based reconstruction algorithms.

Besides, another big challenge in medical applications is the data privacy and scarcity. Generally, the patient data cannot be easily accessed or transferred across different medical centers for model development or deployment [47]. The specific model structures may also be private and protected by the intelligent properties. Thus, it is needed to develop algorithms that can adapt a black-box deep reconstruction model to an unknown testing environment without accessing the training data in many medical applications including sparse-view CT reconstruction. Generally speaking, this study of black-box test-time adaptation could serve as the fundamental for deploying the developed deep learn-

*Joint Corresponding Authors

ing models in practical clinical applications in the future.

In this paper, to address the aforementioned challenges, we propose a novel two-stage source-free black-box test-time adaptation algorithm for sparse-view CT reconstruction with unknown noise through prior-informed implicit neural representation learning (PINER). The first stage is called "input adaptation", which aims for constructing a good input for the black-box model. The second stage is called "physical consistency optimization", which aims for correcting the physical bias from the output of the black-box model. Compared to the relevant previous methods including "plug-and-play" (PnP) framework [1, 46] and the "prior-image-compressed-sensing" (PICCS) framework [7, 52], our method does not need to access ground truth testing images or training data to set the hyperparameters. We use only one test-time sample for our algorithm.

In the first stage, We propose a novel method that uses implicit neural representation learning (INR) to construct adapted inputs for the black-box model inference. To our best knowledge, we are the first to use INR for test-time adaptation for CT reconstruction. Implicit neural representation learning (INR) is an emerging methodology that represents an image or scene as a continuous function parameterized by a neural network. INR is able to learn to represent an image or object at various resolutions [35, 21, 16]. Instead of only using one final trained representation of an image, we propose to consider all the representations learned by the implicit neural representation model during the training process, and automatically detect the best representation that is the closest to the training data distribution based on the inference of black-box model. To sum up, we propose a two-step method to achieve this goal.

In the first step, we try to identify whether the noise from a test-time sample is very different from the training-time noise. In the second step, we use the original test-time sample input or construct an adapted input accordingly. We assume that the black-box model should produce good-quality reconstructions for inputs from the training distribution. Our hypothesis is that if the black-box model does not generalize well on the noise that is unseen in the training set, its output will change significantly when adding an out-of-distribution noise on an input from the training data distribution. Hence, we can utilize this rate of change to identify the change point that the model stops generalizing and then select a good adapted input by feeding representation images into the black-box model. Some other works adopt a similar ideas to detect out-of-distribution samples, but they are limited to classification tasks [20, 22].

In the second stage, we propose a method to further optimize the model output by leveraging the physical consistency based on the black-box model output. After we obtain the black-box model output corresponding to the adapted input, we can embed the model output into an implicit neural representation network as the prior image and then further optimize and refine the final reconstruction based on physical consistency. In contrast to the previous work with prior embedding [32], we use the model output as the prior image instead of previous scan. Especially, we use an early stopping strategy to obtain the best reconstruction during optimization. We conduct experiments on two CT image datasets with different anatomical sites and different black-box deep models, and demonstrate significant improvements over baselines. Our contributions can be summarized as below:

- We propose a novel two-stage framework for black-box test-time adaptation for sparse-view CT reconstruction with unknown noise. Our method does not require the access to the source data, ground truth testing images, and model parameters (only using model API), so that it could largely facilitate the generalization and robustness of deploying deep learning models in real clinical applications.

- Our algorithm exploits the training trajectory of implicit neural representation (INR) to construct adapted inputs to the black-box model. To our best knowledge, we are the first to use INR for test-time adaptation for sparse-view CT reconstruction.

- We perform experiments on different datasets and different pretrained models with continually changing noise. Our method produces better results in image reconstruction quality than existing approaches [52, 12, 14] by a significant margin.

2. Related Work

2.1. CT Reconstruction by Deep Learning

Since early 1970s, the Computed Tomography (CT) has been used as a non-invasive tool for inspecting objects' internal structures with various applications in medical diagnosis, material science, geoscience and industrial inspection [4]. Especially, the sparse-view CT reconstruction task aims at reducing the number of measurements to enable lower radiation dose exposing to patients. With the rise of deep learning, many methods have been proposed for sparse-view CT reconstruction using either supervised methods [13, 15, 33, 45, 54], or unsupervised methods [3, 14, 36]. For supervised methods, the common approaches are either applying backprojection on the measurements (sinogram) to obtain a raw image and then map that raw image to ground truth image [13, 15, 54], or solving a sinogram inpainting problem and then use backprojection [18], or combining both [45, 33, 34]. For unsupervised methods, typically the image prior is learned by a generative model and then a physical-consistent optimization [3, 14] or sam-

pling [36] is performed. However, these deep reconstruction approaches may suffer from performance degradation with out-of-distribution testing samples [2, 51].

2.2. Implicit Neural Representation Learning

Implicit neural representation learning (INR) represents an image as a continuous function that takes coordinates as the input and intensity values as the output. In natural image processing, INR has been demonstrated to be very effective for many vision tasks including novel view synthesis [24], 3D surface reconstruction [26], and data compression [11].

Moreover, several recent works have also demonstrated the successful applications of INR for medical image reconstruction [38, 51, 31, 10, 42, 32]. Specifically, some works [51, 31, 10] propose to use INR to learn the intensity-based image representation for sparse-view CT reconstruction with measurement data. Kim *et al.* [16] propose to use INR for zero-shot blind denoising. Vasconcelos *et al.* [42] propose to sample from a posterior distribution of INR network weights for sparse-view CT reconstruction. Shen *et al.* [32] proposes to enhance the performance of sparse-view CT reconstruction with prior image embedding on INR. However, few works address the model adaptation problem for CT reconstruction using INR.

2.3. Test-time Unsupervised Adaptation

Unsupervised domain adaptation (UDA) aims to improve the model's performance in the target domain without the ground truth labels, where the testing target domain and training source domain are different with a distribution shift [41, 28, 40]. Test-time unsupervised adaptation is a special form of UDA that further restricts the access to the training data from the source domain because of privacy. The adaptation is performed entirely based on test-time data without any ground truth information [43, 19, 48, 17]. In a more challenging scenario, the test-time data distribution may be continually changing, which makes the previous test-time adaptation methods not suitable since the assumption that test-time data comes from a consistent distribution is violated [44]. Most of these existing works of test-time unsupervised adaption are limited to classification, object detection and image segmentation tasks.

There are some works about test-time adaptation for medical image reconstruction. Gilton *et al.* [12] proposed a "RnR" (Reuse and Regularize) algorithm for model adaptation for MRI image reconstruction. Zhang *et al.* [52] propose a prior-constrained compressed sensing approach that uses the pretrained model output as a prior image. However, these models rely on ground truth images to tune the hyperparameters and assume a fixed noise level, which is not desirable in a continually changing testing environment without ground truth information.

Algorithm 1 Black-box Continual Test-time Adaptation to Unknown Noise for Sparse-view CT Reconstruction

Require FBP function B, image coordinates c, y_t, h, A , hyperparameters: k, α, learning rates: λ_1, λ_2, number of iterations: N

Randomly initialize the two INRs M_θ, M_ϕ
for i from 1 to N **do**
  $\theta \leftarrow \theta - \lambda_1 \nabla_\theta ||M_\theta - B(y_t)||$
  $R_i \leftarrow M_\theta(c)$
  $d_{i-k} \leftarrow \frac{||h(R_i) - h(R_{i-k})||}{||R_i - R_{i-k}||}$ if $i - k > 0$
Run change point (increasing) detection on d, and construct adapted input $I_{adapted}$ accordingly
$x_{adapted} \leftarrow h(I_{adapted})$
Embed M_ϕ by the loss function $||M_\phi - x_{adapted}||$
Estimate the noise level of test-time measurement as $\hat{\sigma^2}$
while *VAR*$(A(M_\phi) - y_t) > \hat{\sigma}^2$ **do**
  $\phi \leftarrow \phi - \lambda_2(\nabla_\phi ||AM_\phi - y_t||_2^2 + \alpha ||M_\phi - x_{adapted}||_1)$
Output the final reconstruction by $M_\phi(c)$

3. Our Method - PINER

Fig. 1 illustrates the framework of the proposed method (PINER). PINER mainly contains two stages to obtain the final reconstruction image. In the first stage, we either construct an adapted input or use the original input for the black-box model. This is achieved by exploiting the training trajectory of implicit neural representation learning. Then, an output is obtained from the black-box model inference corresponding to the input. Next, in the second stage, the black-box model output is embedded as the initialization of another implicit neural representation network. A physical consistency optimization is further performed to fine tune the network weights and refine the reconstruction. Finally, the reconstructed image can be obtained by querying the tuned network. Our proposed algorithm is summarized in Algorithm 1.

The reason that we propose this two-stage approach is that we want to improve the reconstruction quality both from a data-driven perspective and a physical consistency perspective. If we can provide better input data to the black-box model while enforcing physical consistency by modifying the output, we should ideally reduce the performance degradation of the black-box model.

3.1. Problem Definition

The target problem is defined as the source-free black-box model test-time adaptation. Specifically, in the context of deep learning-based sparse-view CT image reconstruction, we aim at adapting a trained model to a new testing domain with unknown noise without the access to the training data in the source domain and the model parameters (i.e.

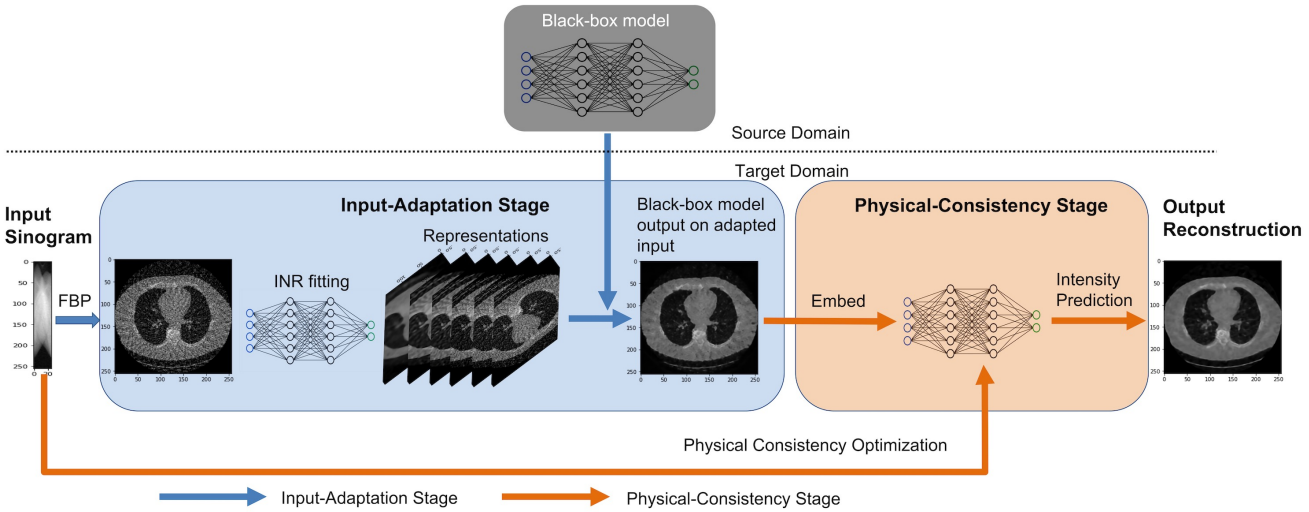

Figure 1. **Framework of our test-time adaptation approach (PINER)**: In the beginning, we obtain a raw image from test sample through filter-backprojection (FBP). Then we train an INR to fit the raw image. Representations exploited from the training trajectory of INR are analyzed to construct an adapted input to the black-box model. The adapted output is then embedded into another INR and further optimized by physical consistency to obtain the final reconstruction.

we can only access the API of that model). Furthermore, we assume that the noise level at test time is unknown and changing continuously. Therefore, the test-time input with unknown noise is given as the only input for running our adaptation algorithm.

Specifically, suppose x is the ground truth image that is given by the full-view noiseless CT reconstruction, A is the forward operator (radon transform) which includes the X-ray ray-marching operation along the defined trajectories at multiple view angles from x [4], and y is the noisy measurements given by $Ax + \epsilon$, where ϵ is the unknown noise. Solving the CT reconstruction problem amounts to recovering the ground truth image x from noisy measurements y. In deep-learning-based reconstruction methods, the neural network is trained to learn the mapping from y to x by leveraging a training dataset of paired images and measurements. In this work, we focus on a specific series of supervised learning methods that maps a raw image obtained from the filtered-backprojection of the measurements to the ground truth images. We denote x_s, y_s as the ground truth images and measurements from the training data (source domain). x_t, y_t are denoted as the ground truth images and measurement at test time, where y_t has continually changing noise levels not seen in the source domain. The function (filtered-backprojection) that transforms measurements to a raw image is denoted by B. The interface (API) of the black-box model is denoted by h. Thus, the output from the black-box model $h(B(y_t))$ at test-time often suffers from performance degradation due to changing noise in the target domain [51]. Hence, the goal of test-time unsupervised adaptation task here is to obtain a reconstruction image that could recover x_t by accessing only the test-time input and the API of the black-box model.

3.2. Adapt the Black-box Model to Test-time Inputs

It is challenging to adapt the black-box model given that we only have one test-time image as the input. The key idea we propose to delve the challenge is to generate inputs through the training trajectory of INR (implicit neural representation) fitting and construct the best input to the black-box model by analyzing the output change rate.

Generate Various Representations. We propose a novel method to exploit the training trajectory of the implicit neural representation learning to generated various representations. Specifically, we train a neural network M_θ that takes image coordinates as input and output the corresponding intensity such that $M_\theta(\mathbf{c}) = I_c$, where $\mathbf{c}$ is a coordinates, θ are the parameters of the neural network, and I_c is the corresponding intensity of the input coordinate. During training, θ is consistently changing to optimize the network parameters to fit into the test-time image so we can take a snapshot of the network prediction of pixel intensities for each θ. We call one such snapshot of the network parameter θ as a "neural representation" of the input image. We choose SIREN network [35] as the backbone of our implicit neural representation network (INR), which has been demonstrated to be able to preserve high-frequency details of the image [35, 16]. During the training process, The INR network fits the the low-frequent and noiseless signal faster than the noisy signal or information in the image [16, 21]. By leveraging this property, we are able to get various representations of the test-time input with different resolutions and noise levels.

Construct Adapted Input. After generating multiple

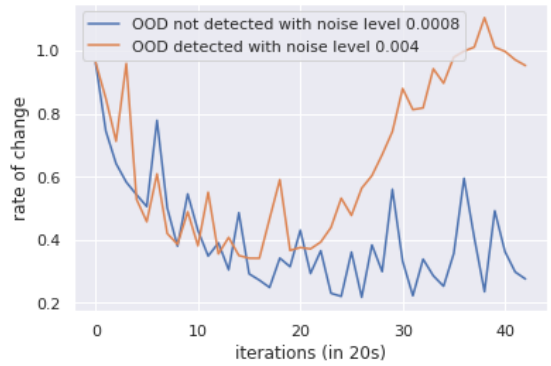

Figure 2. An example of using our proposed metric to detect test samples that contatin out-of-distribution noise.

representations from INR, we still need to decide which representation to use at test time. Ideally, if a test-time input is sampled from the training distribution, we should use the original input; otherwise we need to construct an adapted input. If a test-time sample has a noise that is very different from the training-time noise, then the black-box model may not be able to produce good-quality reconstructions. During INR fitting, we fit the clean signal first, noise afterwards. Hence, it is likely that there is a certain period during INR fitting that the representation images can serve as good adapted inputs to the black-box model since their noise may be close to the training-time noise.

To identify those representations, we propose to analyze the rate of change of the outputs by feeding different representations into the black-box model. The idea is that when the noise becomes out-of-distribution, the black-box model output may be significantly different from that when the noise is still in-distribution. Hence, if we observe the model output change rate starts to increase, then we may be starting to encounter out-of-distribution noise.

Based on this idea, we introduce a metric d_i to find out the rate of change of model output by sliding windows of neural representations of the original test-time input.

$$d_i = \frac{||h(R_{i+k}) - h(R_i)||}{||R_{i+k} - R_i||} \tag{1}$$

where h is the black-box model, R_i is the inferred image from the learned neural representations of $B(y_t)$ in different training iterations, and k is the size of a sliding window for comparing two representations.

According to the above hypothesis, when d_i starts to increase, R_{i+k} may contain out-of-distribution noise while R_i may still have in-distribution noise. Hence, we propose to select the representation that gives the minimum d_i before the first increasing change point. Let ch be the first increasing change point on d_i, we propose to select R_j such that $j = \arg\min_{i \leq ch} d_i$ as the adapted input. If there is no such change point, we can just use the original input.

3.3. Second-stage Optimization for Physical Consistency

We propose a novel second-stage physical-consistent optimization to further improve the reconstructed image quality. Specifically, we propose to embed the black-box model output as a prior image into another implicit neural representation network and then optimize through the physical consistency loss. Following the method proposed in [32] that embeds the patient's previous scan as a prior image, we use the output image inferred from the population-based black-box model as the prior image.

Let M_ϕ be a neural representation network where ϕ is the network parameters. We first pretrain the network by learning to represent the adapted test-time image, $x_{adapted}$ be the adapted output from the previous stage. such that

$$\phi_a = \arg\min_{\phi} ||M_\phi - x_{adapted}||_2^2 \tag{2}$$

We propose to conduct the prior-embedded optimization based on physical consistency, where we let the output image from the black-box model be the prior image. Since the output image corresponding to the adapted input actually obtains the initial reconstruction, we assume that the residual image between the prior image and the ground truth to be sparse. Our second-stage optimization objective is:

$$\tilde{\phi} = \arg\min_{\phi} ||AM_\phi - y||_2^2 + \alpha||(M_\phi - x_{adapted})||_1 \tag{3}$$

which consists of a physical-consistency loss term $||AM_\phi - y||_2^2$ and a prior-consistency loss term $\alpha||(M_\phi - x_{adapted})||$ where network parameters ϕ is initialized by ϕ_a. After obtaining $\tilde{\phi}$, we feed coordinates across the spatial grid into $M_{\tilde{\phi}}$ to obtain the pixel-based intensities of the final reconstructed image.

Note that we optimize the physical consistency loss in the function space instead of the pixel-based intensity space. In this way, since INR learns the underlying image representations with implicit regularization embedded in the parameterized continuous function, optimization in the function space could provide a better reconstruction.

α is a regularization hyperparmeter in the optimization objective to balance the physical consistency and prior consistency. Generally, we want α only slightly increases the physical-consistency loss within the same number of iterations compared to without this term. We will demonstrate that the performance is insensitive to change in α.

Following the idea in [16], the optimization or training process is stopped when the physical-consistency loss is lower than the estimated noise level of the test sample. Through early stopping strategy, we get rid of the extra regularization term such as the smoothness term which may blur out the details of the image.

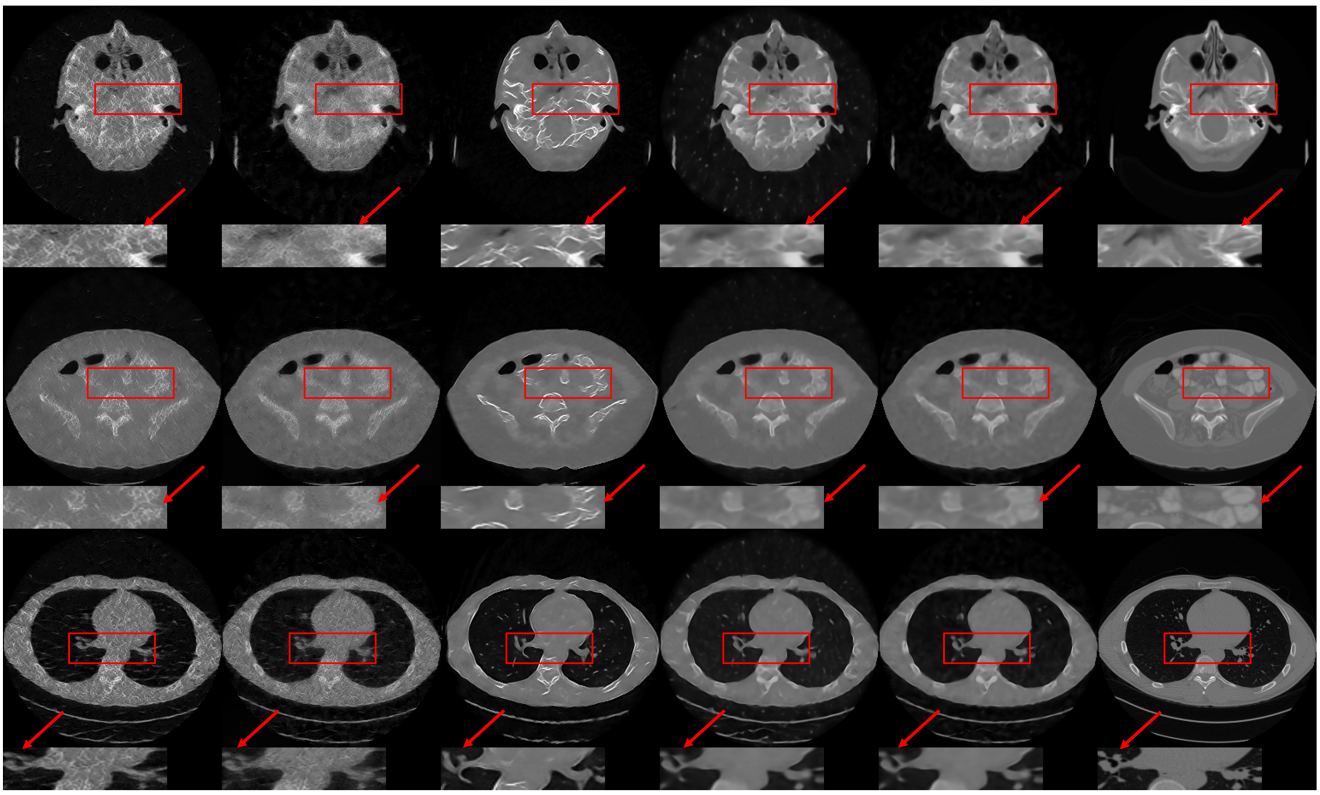

Figure 3. Examples of the reconstruction images from different adaptation methods. The critical image structures are annotated in red box and zoomed in, where the key differences are pointed out by red arrows.

Table 1. Performance (average PSNR/SSIM) of adaptation methods on the LDCT dataset with different pretrained models

Model	Method	Gaussian Noise			Poisson-Gaussian Noise		
		Abdominal	Head	Chest	Abdominal	Head	Chest
UNet [15]	None	28.67 / 0.799	26.58 / 0.707	26.68 / 0.771	27.21 / 0.749	25.94 / 0.683	25.44 / 0.724
	BP [14]	26.25 / 0.762	25.94 / 0.741	24.92 / 0.731	22.63 / 0.652	23.45 / 0.678	21.92 / 0.630
	RnR [12]	30.96 / 0.923	31.67 / 0.930	28.49 / 0.870	30.73 / 0.918	31.48 / 0.928	28.36 / 0.867
	PICCS [52]	30.79 / 0.870	31.07 / 0.885	28.26 / 0.812	30.05 / 0.869	30.52 / 0.880	27.89 / 0.821
	PINER (Ours)	**33.06 / 0.936**	**33.10 / 0.931**	**30.11 / 0.892**	**32.89 / 0.939**	**32.94 / 0.934**	**29.93 / 0.893**
DnCNN [53]	None	29.36 / 0.809	29.58 / 0.832	28.55 / 0.831	28.49 / 0.808	28.01 / 0.793	26.81 / 0.786
	BP [14]	25.77 / 0.723	25.45 / 0.727	24.80 / 0.712	22.16 / 0.626	23.03 / 0.661	21.50 / 0.607
	RnR [12]	27.84 / 0.901	28.58 / 0.919	26.33 / 0.865	27.78 / 0.901	28.15 / 0.914	26.11 / 0.861
	PICCS [52]	31.26 / 0.879	32.03 / 0.913	29.71 / 0.866	30.54 / 0.882	30.84 / 0.899	28.38 / 0.837
	PINER (Ours)	**34.22 / 0.948**	**34.40 / 0.947**	**31.30 / 0.912**	**34.10 / 0.950**	**33.20 / 0.943**	**30.89 / 0.908**

4. Experiments

We conduct experiments of black-box test-time adaptation for sparse-view CT reconstruction with different noise types and different noise levels with different deep network architectures for the black-box pretrained backbone models, respectively. In addition to different network architectures, we also investigate the impact of different training strategies for the pretrained backbone model on the adaptation performance. In the following, we first compare the performance of the proposed method (PINER) with the previous state-of-the-art algorithms. Then we discuss several desirable test-time properties of PINER, followed by the ablation study of several important submodules of PINER model that bring the performance gains.

4.1. Experimental Setup

Datasets and Metrics: We consider two datasets for CT experiments. The first is the Lung Image Database Consortium (LIDC) image collection dataset [9] where we randomly sample 3200 slices from 40 cases for training the

black-box models, and then sample 100 slices from 10 additional cases for testing. The second is the Low Dose CT (LDCT) Image and Projection dataset [25] that contains CT scans of multiple anatomic sites, including head, chest, and abdomen. We sample 3480 slices from all three anatomical sites of 40 patients for training black-box models, and then sample 300 slices with 100 slices for each anatomical sites from the remaining 10 patients for evaluation. Note that the training and testing data are sampled from different patient cases independently without overlap.All images have a size of (256, 256), and all pixel-wise intensities from the slices are normalized to a range of [0,1]. The algorithm performance for evaluating the final reconstructed images is measured by peak signal-to-noise ratio (PSNR) and structural similarity (SSIM).

CT Measurements Simulation and Pretrained Models: We pretrain a UNet model and a DnCNN model with the same training data. We also pretrain a UNet model with a robust training strategy by augmenting the training data with additive Gaussian noise. We refer to this base model as "UNet+". We simulate CT measurements (sinograms) with a parallel-beam geometry using 25 projection angles equally distributed across 180 degrees using the "torch-radon" package [30].

An independent Gaussian noise with a standard deviation of 0.0001 is added to the acquired measurements of the LDCT training dataset for UNet and DnCNN. For LIDC dataset, we use a Gaussian noise with a standard deviation of 0.0005 for the training data of UNet and DnCNN. For UNet+, we add another Gaussian noise with standard deviation sampled uniformly from $U[0, 0.0025]$ to the measurements for training. All models are trained to map the filtered back-projected image of the CT measurements to the ground truth images.

For testing, we assume that we can only access the API of the pretrained black-box models (i.e. a pretrained model is regarded as a black-box model here). We add a Poisson-Gaussian (signal dependent) noise with continually changing unseen noise levels for the LDCT testing dataset. We also add Gaussian noise with continually changing unseen noise levels for both the LDCT and LIDC testing dataset. For pure Gaussian noise on the LDCT dataset, the noise level is uniformly sampled from $U[0, 0.005]$ and $U[0, 0.006]$ respectively for UNet and DnCNN. For the Poisson-Gaussian noise, let P be the Poisson distribution, N be the Gaussian distribution, we use $\frac{a}{n}P(ny_t/a) + aN(0, 0.09)$, where a is sampled uniformly from $U[0.005, 0.013]$, $n = 256$. For LIDC dataset, we add Gaussian noise with noise levels uniformly drawn from $U[0, 0.007]$. We then apply adaptation methods on different pretrained models and compare the performances under different noise conditions.

Baselines: We compare our method with multiple test-time adaptation baselines for CT reconstruction task (BP [14], PICCS[52] and RnR[12]). These baselines either treat the black-box model output as a prior image (BP, PICCS), or use a "plug-and-play" approach to find a stationary point in the model output (RnR). The hyperparameters of PICCS and RnR are tuned based on a random test-time sample together with the corresponding ground truth image for each dataset and for each pretrained model. The hyperparameters are fixed for each dataset and for each pretrained model. Note that, our method does not need to access the ground truth image for parameter tuning.

Table 2. Performance (average PSNR/SSIM) of adaptation algorithms on the LIDC dataset with different pretrained models

Method	UNet		DnCNN	
	PSNR	SSIM	PSNR	SSIM
None	26.13	0.724	26.67	0.792
RnR [12]	28.69	0.885	27.26	0.876
PICCS [52]	27.79	0.816	28.72	0.845
PINER (input only)	27.42	0.798	28.14	0.825
PINER (physics only)	28.41	0.838	28.50	0.839
PINER (no-reg)	29.92	0.900	30.86	0.914
PINER (full)	**29.96**	**0.901**	**30.97**	**0.916**

Implementation Details: We use Pytorch [27] for all implementations. For the network architectures and hyperparameters of INRs, we follow the setting in [32] for CT reconstruction task. The maximum number of training iterations for both INRs is set to be 1000. The learning rate for the first-stage input adaptation is set to be 5e-5 for the LDCT dataset, and 3e-5 for the LIDC dataset. The learning rate is set based on the idea that we want to slowly increase the image fitting PSNR to obtain a more granular collection of input images. Due to privacy concerns, we only collect representations for each 20 epochs. The size of the sliding window is set to be 7 for all scenarios so that there is observable difference between R_{i+k} and R_i for most sliding windows. The penalty term of the changing point detection is set to be the BIC (Bayesian Information Criterion) of the d_i curve for balancing the precision and recall rate of the change point detection, which is given by $4\sigma^2 \log(n)$, where σ is the standard deviation of d_i, n is the number of data points in d_i. The learning rate of second-stage optimization is set to be 1e-5 for all scenarios following the setting in [32].α is set to 1.5e-4. We will demonstrate that the performance is insensitive to both α and k in Appendix. The noise level estimation and change point detection utilize the packages in [6, 39].

4.2. Results and Analysis

Our method outperforms all baselines on all anatomical sites with different pretrained models by a significant margin in both PSNR and SSIM as demonstrated in Ta-

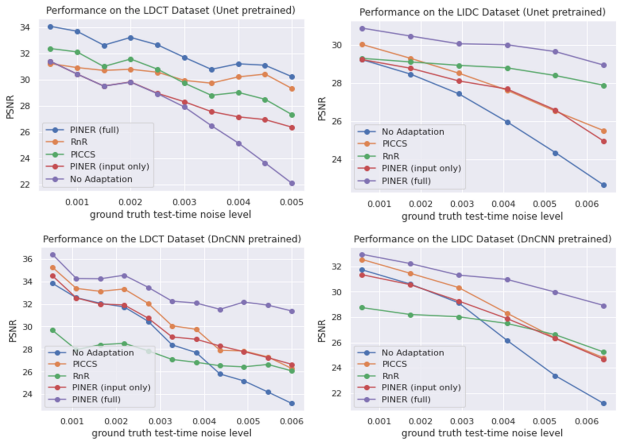

Figure 4. Performance of different adaptation algorithms with various test-time Gaussian noise levels

Table 3. Performance (average PSNR) of adaptation algorithms with different pretraining strategies for UNet on two datasets (Poisson-Gaussian noise for LDCT and Gaussian noise for LIDC)

Method	LDCT		LIDC	
	UNet	UNet+	UNet	UNet+
None	26.20	30.05	26.13	27.51
RnR [12]	30.19	30.80	28.69	29.18
PICCS [52]	29.49	31.31	27.79	29.04
PINER (input only)	28.28	30.27	27.42	27.89
PINER (full)	**31.92**	**32.73**	**29.96**	**30.17**

ble 1 and Table. 2. Visually, we observe that PINER is able to recover the fine details that are missing in the original black-box model output. PINER is also able to reconstruct higher-quality images with sharp organ boundaries, accurate bony details, and reduced noise and artifacts. RnR can produce smooth and sharp images by reusing the black-box model, but it can also generate inaccurate image structures due to the unpredictability of the black-box model output after reusing.

Fig. 4 shows the PSNR performance of each adaptation methods on each dataset with different pretrained models. We find that PINER is robust to different noise types and levels, different pretrained models, and different datasets. The performance of "plug-and-play" algorithms heavily depends on the black-box model structure since it finds a stationary point of the black-box model. For example from Fig. 4, we see that the RnR algorithm performs much better on UNet than on DnCNN for every noise level. On the contrary, the only assumption PINER makes is that the black-box model should perform well on the training data distribution. We observe that PINER's performance is stable and robust in all scenarios.

Ablation Studies. To investigate the contribution of each module of PINER to the performance gain, we perform an ablation study on the LIDC and LDCT datasets

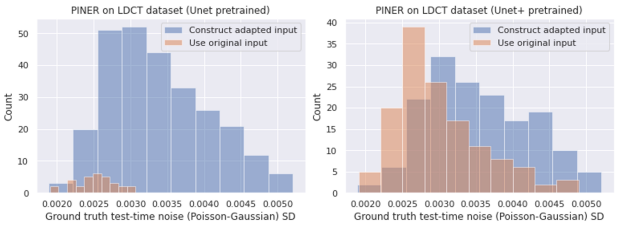

Figure 5. Histograms of ground truth noise levels of test samples

demonstrated in Table 2 and Appendix. We denote PINER only with the first-stage module as "PINER (input only)", PINER only with the second-stage module as "PINER (physics only)", PINER without the prior consistency loss as "PINER (no-reg)", and the complete PINER algorithm as "PINER (full)". We found that both PINER (input only) and PINER (physics only) have a significantly gain in performance compared to the black-box model. Nevertheless, PINER (full) and PINER (no-reg) perform significantly better than PINER (input only) and PINER (physics only), while PINER (no-reg)'s performance is very close to that of PINER (full). This observation implies that both the input adaptation module and the physical consistency module contribute significantly to the performance gain, while the impact of prior consistency regularization is only marginal.

We also study the behavior of the input adaptation module for a backbone model trained with noise-augmented inputs (UNet+) as demonstrated in Table 3 and Fig. 5. We found that the input adaptation module decides to construct adapted inputs much less frequently, especially for test samples that have a noise level close to the augmented training set, while the performance gain of PINER (input only) significantly decreases. This observation implies that PINER is able to detect samples that contain out-of-distribution noise

5. Conclusion

In this work, we introduced PINER, a novel source-free black-box test-time adaptation approach for sparse-view CT reconstruction with unknown measurement noise levels. We show that PINER outperforms existing approaches on black-box CT reconstruction model adaptation. PINER does not need to access training data or ground truth testing images for hyperparameters tuning, in contrast to other previous methods. According to experiments, PINER generalizes well across different pretrained models, noise, and datasets, which we believe are desirable properties of a test-time adaptation algorithm. However, this is only a proof-of-concept work which does not use real clinical CT measurement data. Nevertheless, we believe that since our method places minimum assumption on the measurement noise, we will experiment our method on real clinical measurement data in future work. We will also extend this work to different forward functions in future work.

References

[1] Rizwan Ahmad, Charles A Bouman, Gregery T Buzzard, Stanley H Chan, Edward T Reehorst, and Philip Schniter. Plug and play methods for magnetic resonance imaging. 2019.

[2] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C Hansen. On instabilities of deep learning in image reconstruction and the potential costs of ai. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, 2020.

[3] Ashish Bora, Ajil Jalal, Eric Price, and Alexandros G Dimakis. Compressed sensing using generative models. In *International Conference on Machine Learning*, pages 537–546. PMLR, 2017.

[4] Thorsten M Buzug. Computed tomography. In *Springer handbook of medical technology*, pages 311–342. Springer, 2011.

[5] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[6] Guangyong Chen, Fengyuan Zhu, and Pheng Ann Heng. An efficient statistical method for image noise level estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 477–485, 2015.

[7] Guang-Hong Chen, Jie Tang, and Shuai Leng. Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets. *Medical physics*, 35(2):660–663, 2008.

[8] Kihwan Choi, Jing Wang, Lei Zhu, Tae-Suk Suh, Stephen Boyd, and Lei Xing. Compressed sensing based cone-beam computed tomography reconstruction with a first-order method a. *Medical physics*, 37(9):5113–5125, 2010.

[9] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26(6):1045–1057, 2013.

[10] Abril Corona-Figueroa, Jonathan Frawley, Sam Bond-Taylor, Sarath Bethapudi, Hubert PH Shum, and Chris G Willcocks. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. *arXiv preprint arXiv:2202.01020*, 2022.

[11] Emilien Dupont, Adam Goliński, Milad Alizadeh, Yee Whye Teh, and Arnaud Doucet. Coin: Compression with implicit neural representations. *arXiv preprint arXiv:2103.03123*, 2021.

[12] Davis Gilton, Gregory Ongie, and Rebecca Willett. Model adaptation for inverse problems in imaging. *IEEE Transactions on Computational Imaging*, 7:661–674, 2021.

[13] Yoseob Han and Jong Chul Ye. Framing u-net via deep convolutional framelets: Application to sparse-view ct. *IEEE transactions on medical imaging*, 37(6):1418–1429, 2018.

[14] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Image-adaptive gan based reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3121–3129, 2020.

[15] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

[16] Chaewon Kim, Jaeho Lee, and Jinwoo Shin. Zero-shot blind image denoising via implicit neural representations. *arXiv preprint arXiv:2204.02405*, 2022.

[17] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9641–9650, 2020.

[18] Ziheng Li, Wenkun Zhang, Linyuan Wang, Ailong Cai, Ningning Liang, Bin Yan, and Lei Li. A sinogram inpainting method based on generative adversarial network for limited-angle computed tomography. In *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, volume 11072, pages 345–349. SPIE, 2019.

[19] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, pages 6028–6039. PMLR, 2020.

[20] Shiyu Liang, Yixuan Li, and R Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. In *6th International Conference on Learning Representations, ICLR 2018*, 2018.

[21] David B Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. Bacon: Band-limited coordinate networks for multiscale scene representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16252–16262, 2022.

[22] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in Neural Information Processing Systems*, 33:21464–21475, 2020.

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

[24] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020.

[25] Taylor R Moen, Baiyu Chen, David R Holmes III, Xinhui Duan, Zhicong Yu, Lifeng Yu, Shuai Leng, Joel G Fletcher, and Cynthia H McCollough. Low-dose ct image and projection dataset. *Medical physics*, 48(2):902–911, 2021.

[26] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.

[27] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.

[28] Vishal M Patel, Raghuraman Gopalan, Ruonan Li, and Rama Chellappa. Visual domain adaptation: A survey of recent advances. *IEEE signal processing magazine*, 32(3):53–69, 2015.

[29] Saiprasad Ravishankar, Jong Chul Ye, and Jeffrey A Fessler. Image reconstruction: From sparsity to data-adaptive methods and machine learning. *Proceedings of the IEEE*, 108(1):86–109, 2019.

[30] Matteo Ronchetti. Torchradon: Fast differentiable routines for computed tomography. *arXiv preprint arXiv:2009.14788*, 2020.

[31] Darius Rückert, Yuanhao Wang, Rui Li, Ramzi Idoughi, and Wolfgang Heidrich. Neat: Neural adaptive tomography. *arXiv preprint arXiv:2202.02171*, 2022.

[32] Liyue Shen, John Pauly, and Lei Xing. Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[33] Liyue Shen, Wei Zhao, Dante Capaldi, John Pauly, and Lei Xing. A geometry-informed deep learning framework for ultra-sparse 3d tomographic image reconstruction. *Computers in Biology and Medicine*, page 105710, 2022.

[34] Liyue Shen, Wei Zhao, and Lei Xing. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nature biomedical engineering*, 3(11):880–888, 2019.

[35] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020.

[36] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon. Solving inverse problems in medical imaging with score-based generative models. *arXiv preprint arXiv:2111.08005*, 2021.

[37] Paul Suetens. *Fundamentals of medical imaging*. Cambridge university press, 2017.

[38] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov. Coil: Coordinate-based internal learning for imaging inverse problems. *arXiv preprint arXiv:2102.05181*, 2021.

[39] Charles Truong, Laurent Oudre, and Nicolas Vayatis. ruptures: change point detection in python. *arXiv preprint arXiv:1801.00826*, 2018.

[40] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7472–7481, 2018.

[41] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014.

[42] Francisca Vasconcelos, Bobby He, Nalini Singh, and Yee Whye Teh. Uncertainr: Uncertainty quantification of end-to-end implicit neural representations for computed tomography. *arXiv preprint arXiv:2202.10847*, 2022.

[43] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020.

[44] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7201–7211, 2022.

[45] Haoyu Wei, Florian Schiffers, Tobias Würfl, Daming Shen, Daniel Kim, Aggelos K Katsaggelos, and Oliver Cossairt. 2-step sparse-view ct reconstruction with a domain-specific perceptual network. *arXiv preprint arXiv:2012.04743*, 2020.

[46] Kaixuan Wei, Angelica Aviles-Rivero, Jingwei Liang, Ying Fu, Carola-Bibiane Schönlieb, and Hua Huang. Tuning-free plug-and-play proximal algorithm for inverse imaging problems. In *International Conference on Machine Learning*, pages 10158–10169. PMLR, 2020.

[47] Rui Yan, Liangqiong Qu, Qingyue Wei, Shih-Cheng Huang, Liyue Shen, Daniel Rubin, Lei Xing, and Yuyin Zhou. Label-efficient self-supervised federated learning for tackling data heterogeneity in medical imaging. *arXiv preprint arXiv:2205.08576*, 2022.

[48] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Generalized source-free domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8978–8987, 2021.

[49] Xingde Ying, Heng Guo, Kai Ma, Jian Wu, Zhengxin Weng, and Yefeng Zheng. X2ct-gan: reconstructing ct from biplanar x-rays with generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10619–10628, 2019.

[50] Hengyong Yu and Ge Wang. Compressed sensing based interior tomography. *Physics in medicine & biology*, 54(9):2791, 2009.

[51] Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1960–1970, 2021.

[52] Chengzhu Zhang, Yinsheng Li, and Guang-Hong Chen. Accurate and robust sparse-view angle ct image reconstruction using deep learning and prior image constrained compressed sensing (dl-piccs). *Medical Physics*, 48(10):5765–5781, 2021.

[53] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.

[54] Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018.