Iterative Reconstruction for Low-Dose CT Using Deep Gradient Priors of Generative Model

Zhuonan He, Yikun Zhang, Yu Guan, Bing Guan, Shanzhou Niu, Yi Zhang, Senior Member, IEEE, Yang Chen, Senior Member, IEEE, and Qiegen Liu, Senior Member, IEEE

Abstract—Dose reduction in computed tomography (CT) is essential for decreasing radiation risk in clinical applications. Iterative reconstruction is one of the most promising ways to compensate for the increased noise due to the reduction of photon flux. Rather than most existing prior-driven algorithms that benefit from manually designed prior functions or supervised learning schemes, in this work, we integrate the data consistency as a conditional term into the iterative generative model for low-dose CT. At the stage of prior learning, the gradient of data density is directly learned from normal-dose CT images as a prior. Then, at the iterative reconstruction stage, the stochastic gradient descent is employed to update the trained prior with annealed and conditional schemes. The distance between the reconstructed image and the manifold is minimized along with data fidelity during reconstruction. Experimental comparisons demonstrated the noise reduction and detail preservation abilities of the proposed method.

Index Terms—Computed tomography (CT), gradient prior, iterative reconstruction, score-based generative network.

I. INTRODUCTION

X-RAY computed tomography (CT) is a popular imaging modality with applications in biology, medicine, and other fields. The extensive use of CT examination has raised concerns about the potential risks of carcinogenesis or genetic damage from X-ray radiation [1]. Low-dose CT (LDCT) can image many clinical indications to minimize radiation-related risks without significantly affecting screening or diagnostic performance. Hence, making the radiation dose as low as reasonably achievable is commonly recognized, and it has been a hot research topic during the latest three decades. However, reducing the radiation dose will increase data noise and introduce artifacts into the reconstructed images, which adversely affects its diagnostic performance if these issues are not addressed [2].

Later than the classical filtered back-projection (FBP) algorithm, iterative reconstruction methods become the mainstream in the past few decades [3]–[7]. Incorporating photon statistics and prior information of the target image, these methods have great potentials in noise reduction and information preservation. Specifically, most of the priors are manually designed under a set of neighboring pixels in the image or transform domain, emphasizing on enhancing image smoothness while maintaining edges. Total variation (TV), sparsity in wavelet transform (WT), and nonlocal patch-based priors have shown promising results in LDCT [5]–[7]. Tirer and Giryes [8] proposed a plug-and-play algorithm that achieves good results by iterating between back projections and strong denoisers. Nevertheless, these reconstruction approaches may still lose some image details and suffer from remaining artifacts.

Deep learning (DL) techniques, particularly convolutional neural networks (CNNs), have been actively developed and applied to various fields recently. The explosive development of them suggests new thinking and huge potential for the medical imaging field. Broadly speaking, these approaches can be categorized into two types: 1) one is using end-to-end supervised DL techniques as a postprocessing method [9]–[15] and 2) the other is integrating DL-driven prior techniques into an iterative scheme. In the first class, Chen et al. [9] and Kang et al. [10] were the first attempt to study the deep CNN for LDCT. By recognizing that the WT operator is able to improve the denoising efficiency and preserves or even enhances the edge features, the results in [11] demonstrated that the directional wavelet utilizing deep CNN was more effective in getting rid of low-dose related noises. Although these algorithms attained promising results, they were usually designed for a particular task with specialized architectures or loss functions, and trained with paired data by taking one image as input and the other as supervision. An efficient fashion to extend the usage of supervised DL is the self-supervised method. It can be used in more practical environment that the ground truth is not available [16], [17].

In another class of methods, it reuses the knowledge in learned priors to tackle various tasks without retraining or
modification. As expected, it is possible to model the non-linear manifold, so that knowledge of normal-dose image can be learned more precisely. Subsequently, the goal of improved reconstruction quality is achieved. Specifically, Baguer et al. [18] used deep image priors for CT reconstruction to achieve promising results in the low-data regime. Meanwhile, based on the feature learning and mapping ability of the generative models, many CT reconstruction methods have been proposed from the perspectives of network structure and objective function. Recent progress is mainly driven by two approaches: 1) likelihood-based methods [19]–[22] and 2) generative adversarial network (GAN) [23]–[25]. The former uses log likelihood (or a suitable surrogate) as the training objective, while the latter uses adversarial training to minimize $f$-divergences [26] or integral probability metrics between model and data distribution [27], [28]. For example, Adler and Öktem [29] employed a Wasserstein GAN to draw samples from the conditional distribution. Zeng et al. [30] considered the noise-generating mechanism in CT imaging and using pairs of noisy images to reduce the noise. Regrettably, despite the success of generative models in tackling natural images, there have been few studies in the field of medical imaging, especially in LDCT.

In this work, to boost the effectiveness of LDCT reconstruction, we focus on exploring deep gradient priors of generative model (EASEL). As a newly developed unsupervised learning, score-based diffusion generative models provide a powerful way to model images using the gradient of the data distribution [31]–[33]. The score-based generative model has exhibited great potential for diverse image representation [34]. Viewing the noisy observation as a conditional variable, a conditional score-based generative model is introduced into LDCT reconstruction. Since EASEL is implemented in a fully unsupervised way, it has more flexibility and fewer requirements on training data [35].

The major contributions of this article are twofold, which are described as follows.

1) Estimating the Gradient of Data Density as a Prior: Different from the traditional regularization methods (e.g., regularization by denoising [36]) that employ the penalty term as a prior first and then calculate its corresponding gradient in the iterative procedure, in this study, we utilize the gradient of data density as a prior directly. The learned deep gradient prior is directly incorporated into the iterative procedure (i.e., Langevin dynamics).

2) Conducting the Conditional Prior Generation as a Reconstruction Procedure: Different from the conventional regularization methods that directly employ the plug-and-play scheme to update the data consistency and regularization term in alternative manner, in this work, the data consistency is enforced as a conditional term in the prior generation procedure. More specifically, the ideal solution is generated by annealed Langevin dynamics, which forms the outer loops. At each outer loop, the intermediate solutions are updated alternatively by constraints in data consistency and prior generation, forming several inner loops.

The remainder of this article is organized as follows. Section II provides a brief description of the preliminary work with regard to the generative model and the gradient of the generative model. Section III presents the forward formulation of LDCT and the corresponding iterative solver. Extensive experimental comparisons between the proposed EASEL and state of the arts are conducted in Section IV. Finally, concluding remarks and directions for future research are given in Section V.

II. PRELIMINARIES

A. Deep Generative Models

Recently, advances with deep generative networks have shown promising results in modeling complex distributions, such as images [37], audios [38], and texts [39]. As shown in Fig. 1, the popular deep generative models can be primarily categorized into two groups: 1) explicit generative model and 2) implicit generative model. The former model provides an explicit parametric specification of the data distribution, specifying a log-likelihood function $\log p(x)$, including autoencoders (AEs) and its variants [20], [40], flow-based generative models [41], [42], score-based models [43]–[49], and deep Boltzmann machine [50]. Especially, score-based models train the parametric network to approximate the likelihood gradient $\nabla_x \log p(x)$, which have tractable likelihood estimation. Alternatively, we can specify implicit probabilistic models that define a stochastic procedure to directly generate data. GANs [19] are the well-known implicit likelihood models. They optimize the objection function through adversarial learning and have been shown to produce high-quality images [23], [37], [51]. Despite its success, GANs still suffer from a remarkable difficulty in training and interpretability due to the lack of theoretical guarantees.

B. Deep Gradient Priors of Generative Model

Recent works show that successful image generation can be achieved by score-based generative models [43]–[49], [31]–[33]. They represent probability distributions through score functions $\nabla_x \log p(x)$—a vector field pointing in the direction where the likelihood of data $p(x)$ increases the most rapidly. Remarkably, these score functions can be learned from data without requiring
adversarial optimization, and can produce realistic image samples that rival GAN [43]. To facilitate the description of the present method, we tabulate the notations used hereafter in Table I.

The whole procedure consists of two steps: 1) first, the denoising score matching (DSM) is used to train a network $s_0(x)$ to approximate $\nabla_x \log \rho_0(x)$ with different scale magnitude $\sigma_i |_{i = 1}$ and 2) second, via annealing strategy of deceasing $\sigma_i$, the Langevin dynamics is executed to draw sampling from a sequence of $\tilde{s}_l(x; \sigma_i)$ and approaches to the final estimation of $\log \rho(x)$.

**DSM:** In general, using score matching [47], a score network $s_0(x)$ can be trained to estimate $\nabla_x \log (p(x))$ by minimizing the objective function

$$E_{p(x)} \left[ \| s_0(x) - \nabla_x \log p(x) \|_2^2 \right].$$  \hspace{1cm} (1)

Subsequently, Vincent [48] used DSM $s_0(x)$ to match a non-parametric kernel density estimator

$$E_{p_\sigma(x)} \left[ \| s_0(\tilde{x}) - \nabla_x \log p_\sigma(\tilde{x}) \|_2^2 \right]$$  \hspace{1cm} (2)

where the corresponding perturbed data distribution is $p_\sigma(\tilde{x}) = \int p_\sigma(x)p(x)\ dx$. Crucially, one caveat of DSM is that the optimal score $s_0(x) = \nabla_x \log p_\sigma(x) \approx \nabla_x \log p(x)$ is true only when the noise $\sigma$ is small enough. Whereas, learning the score function with the single-noise perturbed data distribution will lead to inaccurate score estimation in the low data density region on high-dimension data space, which could be severe due to the low-dimensional manifold assumption. Thus, Song and Ermon [34] proposed learning a single neural network based on multiple perturbed data distributions with Gaussian noises of varying magnitudes

$$\frac{1}{2L} \sum_{l = 1}^L \sigma_l^2 E_{p_\sigma(\tilde{x})|x} \left[ \| s_0(\tilde{x}; \sigma_l) - \nabla_x \log p_\sigma(\tilde{x}|x) \|_2^2 \right]$$  \hspace{1cm} (3)

where the target $\nabla_x \log p_\sigma(\tilde{x}|x) = (x-\tilde{x})/\sigma_l^2$ has a simple closed form and the empirical average is utilized to estimate all expectations.

**Annealed Langevin Dynamics:** In many situations, score function is easier to model and estimate than the original density function [49]. For example, for an unnormalized density, it does not depend on the partition function. Once the score function is known, we can employ Langevin dynamics to sample from the corresponding distribution.

The Langevin dynamic algorithm is a family of gradient-based Markov chain Monte Carlo (MCMC) sampling method.

More precisely, given a step size $\epsilon > 0$, a total number of iterations $T$, and an initial sample $x_0$, it evaluates the gradient of the negative log probability in iterative and stochastic manner

$$x^t = x^{t-1} + \frac{\epsilon}{2} \nabla_x \log \rho_\sigma(x^{t-1}) + \sqrt{\epsilon} \epsilon^{t-1}$$

$$= x^{t-1} + \frac{\epsilon}{2} \nabla_x \log \rho_\sigma(x^{t-1} - \epsilon^{t-1}) + \sqrt{\epsilon} \epsilon^{t-1}$$  \hspace{1cm} (4)

where $z \sim N(0, I)$ is a random term and $t = 1, \ldots, T$.

However, when two modes of the data distribution are separated by low density regions, Langevin dynamics will not be able to correctly recover the relative weights of these two modes in reasonable time, and therefore, the sampling model might not converge to the true distribution. In addition, since Langevin dynamics is usually initialized in low-density regions of the data distribution, inaccurate score estimation in these regions will negatively affect the sampling process. Hence, mixing can be difficult because of the need of traversing low-density regions to transition between modes of the distribution. To tackle these two challenges, Song and Ermon [34] proposed an annealed version of Langevin dynamics, where they initially used scores corresponding to the highest noise level, and gradually annealed down the noise level $\sigma_i \rightarrow 0 |_{i=1}$ until it was small enough to be indistinguishable from the original data distribution. As a remedy, large noise levels will produce samples in low density regions of the original data distribution, which can improve score estimation.

### III. METHODOLOGY

At the previous section, how to learn the deep gradient prior of generative model and how to make use of it to generate data samples are introduced. In this section, the derivation of how to adopt the generation process to the special LDCT reconstruction problem will be given.

#### A. LDCT Imaging Model

The statistics of CT measurement data are often described by a Poisson distribution [52]. Specifically, a Poisson model for the intensity measurement is

$$I_i \sim \text{Poisson}\left[b_i e^{-Ax_i} + r_i \right], i = 1, \ldots, N_m$$  \hspace{1cm} (5)

where $I_i$ is the number of transmitted photons, $A$ is a system matrix (projection operator), $b_i$ denotes the X-ray source intensity of the $i$th ray, and $r_i$ denotes the background contributions of scatter and electrical noise. $x$ is a vector for the representation of attenuation coefficients with units of inverse length.
\( N_m \) is the number of measurements, and \( N_v \) is the number of image voxels.

After taking the logarithm operation, the sinogram data are often approximated as a weighted Gaussian \([10]\)
\[
y_i \sim N \left( [Ax]_i, \tilde{l}_i / (\tilde{l}_i - r_i)^2 \right)
\]
where \( \tilde{l}_i = E[I_i] \). In fact, the LDCT problem can be formulated as an inverse problem \( y = Ax + \mu \). In order to cover the uncertainties that occur especially with ill-posed problems, the theory of Bayesian inversion considers the posterior distribution \( p(x|y) \) \([53]\). This posterior is the conditional distribution of the image \( x \) conditioned on the measurements \( y \).

Rather than the conventional regularization methods that employ the plug-and-play scheme to update the data consistency and regularization term alternatively in an equal manner \([36]\), in the proposed EASEL, the data consistency is enforced as a conditional term in the procedure of prior generation. The whole EASEL consists of two-level loops: 1) during the outer loops, the ideal solution is generated by conditional and 2) annealed Langevin dynamics, which requires finite states to gradually approach to the satisfied estimate. At each outer loop, the intermediate solutions are alternatively updated in data consistency and prior generation constraints, forming several inner loops.

**B. Outer Loop of EASEL**

In this study, we are devoted to using the generative model to tackle the LDCT problem \([54]\). More specifically, the forward formulation of LDCT is
\[
y = Ax + \mu
\]
the specification of (7) is detailed in Section III-A.

Owing to the observation \( y \), the sampling object is not directly from \( p(x) \), but the posterior distribution \( p(x|y) \). Therefore, our goal is to employ the deep generative model in Section II to the distribution \( p(x|y) \). An intuitive solution is that the data consistency in (7) can be viewed as a conditional term and incorporated into the sampling procedure of (4). To be more precisely, it yields
\[
x' = x^{-1} + \frac{\varepsilon}{2} \nabla_x \left( \log p(x^{-1}|y) \right) + \sqrt{\varepsilon} \xi^{-1}
\]
(8)

Due to the Bayesian rule \( p(x|y) = p(y|x)p(x)/p(y) \), it becomes
\[
x' = x^{-1} + \frac{\varepsilon}{2} \nabla_x \left( \log p(x^{-1}) + \log p(y|x^{-1}) \right) + \sqrt{\varepsilon} \xi^{-1}
\]
\[
= x^{-1} + \frac{\varepsilon}{2} \nabla_x \left( \log p(x^{-1}) \right) + \frac{\varepsilon x}{2} \nabla_x \left[ \| y - Ax^{-1} \|^2 \right] + \sqrt{\varepsilon} \xi^{-1}
\]
(9)

Here, \( \log p(y|x) \) is given by the data model that is derived from the knowledge of data and \( \log(p(x)) \) is given by the prior model that represents information known beforehand about the true model parameter. The hyperparameter \( \lambda \) balances the tradeoff between priors and data fidelity. It has to be estimated if we do not know the standard deviation of the measurement noise.

Given the gradient of data density \( \nabla_x (\log p(x)) \) to be learned in advance, (9) is essentially a stochastic gradient descent scheme with the conditional strategy. Besides the conditional strategy, in this study, we simultaneously use an annealed strategy along the iterative procedure, i.e., \( \nabla_x (\log p(x)) \to \nabla_x (\log p(x)) \) with artificially decreasing \( \sigma \)-value (in Section III-C). A geometric interpretation of the whole iterative procedure is visualized in Fig. 2. Overall, the conditional strategy guarantees that the intermediate solution will approach to the data-feasible domain. At the meantime, the annealing strategy tries its best to pursue the optimal prior knowledge in a solid manner \([55]\). It is predictable that global-scale structures discovered at high noise level will be preserved as the \( \sigma \)-value drops, meanwhile, the occurrence of fine-detail structures is still admitted.

We use the annealing strategy to approximate (9). Let \( p_{\sigma}(x) \) denote the distribution of \( x + \eta \) for \( x \sim \sigma \) and \( \eta \sim N(0, \sigma^2 I) \). At high noise levels \( \sigma_t, p_{\sigma}(x) \) are approximately Gaussian and irreducible, so the Langevin dynamics (9) will mix quickly.

The modified Langevin dynamics is as follows:
\[
x' = x^{-1} + \frac{\varepsilon}{2} \nabla_x \left( \log p_{\sigma}(x^{-1}) \right) + \frac{\varepsilon x}{2} \nabla_x \left[ \| y - Ax^{-1} \|^2 \right] + \sqrt{\varepsilon} \xi^{-1}
\]
(10)

**C. Inner Loop of EASEL**

In the following, we elaborate on the details for implementing (10). As done in \([29]\) and \([56]\), we seek to an alternating and separable quadratic surrogates (SQSs) algorithm \([58], [59]\) is adopted to minimize the convex problem (12) to get the reconstructed CT volume
\[
x^k = x^{k-1} - \sum_{i=1}^M \left( \frac{A^T (A x^{k-1} - y_i)}{A^T A x^{k-1} + \beta} y_i \right)
\]
(13)
Algorithm 1: EASEL for LDCT Reconstruction

Initialization: $x^0$ and $w^0$, $\sigma \in \{\sigma_l\}_{l=1}^L$, $\beta$, $\gamma$, $\tau$

For $l = 1, 2, \ldots, L$ (Outer loop) do

\[ \epsilon_l = \tau \sigma_l^2 / \sigma_1^2 \]

For $t = 1, 2, \ldots, T$ (Inner loop) do

\[ z^{t-1} \sim N(0, 1) \]

\[ u^t = x^{t-1} + \frac{\epsilon_l}{\sqrt{\sigma_l}} \mathcal{S}_\theta(x^{t-1}, \sigma_l) + \sqrt{\epsilon_l} z^{t-1} \]

\[ x^t = w^{t-1} - \frac{A^T(Ax^{t-1} + \beta(x^{t-1} - u^t))}{A^TA1 + \beta1} \]

\[ w^t = x^t + \gamma(x^t - x^{t-1}) \]

End For

$\lambda^0 \leftarrow w^T$

End For

where $A^T$ is the transposition of $A$ and refers to back projection, $1$ is an all-ones vector. Especially, the division is a componentwise operator, and it makes the algorithm very fast and monotonous in the iterative process.

We begin by initializing $z^0$ as the standard Gaussian vector, and then take a gradient step in one of these while fixing the other. Additionally, for additional acceleration, we apply Nesterov’s momentum [60] that exploits the previous descent directions. A momentum term can be $w^{t+1} = x^{t+1} + \gamma (x^{t+1} - x^t)$, where $\gamma$ is a relaxation factor that lies between 0 and 1.

In summary, the visual flowchart of the training phase and iterative reconstruction phase for LDCT is shown in Fig. 3. Furthermore, Algorithm 1 explains the reconstruction algorithm in detail. EASEL consists of two loops. The outer loop is composed of several stages to enable that $p_{\lambda_l}(x)$ tends to $p(x)$ with decreasing $\sigma_l$-value. The inner loop conducts the conditional Langevin dynamics with alternately updating of data consistency and deep gradient priors. Since both the Langevin dynamics updating [61], [62] and SQS updating [58], [59] have convergence guarantee, the overall EASEL algorithm will come to the convergence region after finite iterations.

D. Network Architecture of $s_\theta(x, \sigma)$

As previously mentioned, the adaption of the unsupervised network to the general LDCT reconstruction is the key contribution of EASEL. Besides, the architecture of the score-based network $s_\theta(x, \sigma)$ is also an important factor that impacts the algorithm performance. In this section, following the idea of high-dimensional embedded denoising network in our previous work [54], [63], we present a channel-copy guided RefineNet as $s_\theta(x, \sigma)$.

Originally, the RefineNet [64] is a modern variant of U-Net [65] that also incorporates ResNet designs [66]. On this basis, Song and Ermon [34] replaced the max pooling layers in refine blocks with a multipath block, as the multipath block is reported to produce smoother and more diversity features for image generation tasks such as image reconstruction. More importantly, they used a convolutional operator to produce high-dimensional features before the multipath block, as seen in Fig. 4(a), such as to obtain flexible representation and excellent robustness abilities.

Alternatively, in Fig. 4(b), we choose the channel-copy strategy to replace the convolutional operator in naïve RefineNet to form the channel-copy guided RefineNet. As seen in Fig. 4, both the naïve RefineNet and channel-copy guided RefineNet extend the generative ability via extending the information/representation dimension, such as increasing the channel dimension of the input. However, the latter strategy
Fig. 5. Architecture of channel-copy guided RefineNet used in $x_\theta(x; \sigma)$ of EASEL. A distinct difference of the channel-copy guided RefineNet from the naïve RefineNet is that we use channel-copy technique to attain high-dimensional features and then inject noise into them. A more detailed visual comparison is shown in Fig. 4.

is simpler and favors to computation effectiveness. As demonstrated in Section IV-E, the naïve RefineNet needs 64 convolutional kernels, while the present channel-copy guided RefineNet only uses ten copied channels.

Fig. 5 depicts the whole architecture of the channel-copy guided RefineNet. At the prior training stage, we copy and rearrange the single-channel image to the same 10-channel images. Thus, the DSM network can be trained with these data injected with artificial Gaussian noise. At the iterative reconstruction stage, in order to pave the way to apply the trained 10-channel prior to the intermediate single-channel image from the previous iteration, it needs to conduct the channel-copy and channel-mean operators before and after the Langevin dynamics updating.

IV. EXPERIMENTAL RESULTS

In the experiments, five methods are compared with EASEL, including the classical FBP reconstruction (Ramp-filter) [67], TV-based iterative method ASD-POCS [68], dictionary learned by K-SVD algorithm [71], [74], RED-CNN network [70], and domain progressive residual network DP-ResNet [71]. The involved parameters are set following the guidelines in their original papers. For K-SVD, besides the postprocessing K-SVD method [69], the iterative-based K-SVD method that is built on the imaging forward model [56] is also recorded. For the convenience of notion, we denote them as K-SVD (postprocess) and K-SVD (Iter-recon), respectively. For more in-depth study and research, the source code is at: https://github.com/yqx7150/EASEL.

A. Data Specification

AAPM Challenge Data: To evaluate the algorithm on a clinically realistic use case, we consider reconstruction of simulated data from human abdomen CT scans as provided by Mayo Clinic for the AAPM Low Dose CT Grand Challenge [72]. The data include high-dose CT scans from ten patients, of which we use nine for training and one for evaluation. We use the 1-mm slice thickness reconstructions, resulting in 2961 training images with each $512 \times 512$ pixel in size. We use a 2-D fan-beam geometry with 1000 angles, 1000 pixels, source to axis distance 500 mm, and axis to detector distance 500 mm. The corresponding LDCT images are simulated by adding Poisson noise [whose intensities are $b_i = 5e4$ in (5)] into the sinogram data [73].

CIRS Phantom Data: A high-quality set of CT volumes ($512 \times 512 \times 100$ voxels and voxel size of $0.78 \times 0.78 \times 0.625$ mm$^3$) of an anthropomorphic CIRS phantom is obtained from a GE Discovery HD750 CT system, in which the tube current value is set to 600 mAs to guarantee a good image quality for low-dose simulation under 150 mAs. The source-to-axial distance is 573 mm, and the source-to-detector distance is 1010 mm. Fig. 6 displays some representative high-dose images, LDCT images, and the differential images.

B. Model Training and Parameter Selection

In our experiments, the model was trained by the Adam algorithm with the learning rate $10^{-3}$ and Kaiming initialization.
is used to initialize the weights. The method implemented in Python using operator discretization library (ODL) [74] and PyTorch on a personal workstation with a GPU card (GeForce RTX 2080) accelerated the calculation process. An example of model fitting when training the model is given in Fig. 7, and it can be seen that as the number of iterations increases, the model gradually converges and remains stable. In the reconstruction stage, we evaluated several parameter combinations and finalized the parameter settings as follows. The iteration number is set to \( T = 150 \) for each noise scale and the scale number \( L = 12 \), the relaxation factor \( \gamma = 0.5 \), and regularization weight \( \lambda = 150 \), \( \tau = 1.8 \times 10^{-5} \).

For a fair comparison, the parameters of K-SVD and ASD-POCS are hand tuned for the best performance. Considering the harsh parameter setting of K-SVD, we asked Chen et al. [69] to tune the hyperparameters. In this work, the discrete cosine transform dictionary was chosen to learn image features, the size of overlapping patches is set to \( 8 \times 8 \), the atom number is set to 64, and the tolerance parameter is set to 0.8. For RED-CNN model training, under the guidance of Chen et al. [70], the mean squared error (MSE) is utilized as the loss function and is optimized by Adam, the base learning rate was set to \( 10^{-4} \), and slowly decreased down to \( 10^{-5} \), after the 100th epoch, the loss function converges and stabilizes. In the ASD-POCS method, the parameters are optimized using a grid search to maximize the peak signal-to-noise ratio (PSNR), and the regularization parameter was set to 130.

The parameters in DP-ResNet were set as per the suggestions \( \hat{\mu}_4 = 5e4 \).

### C. Quantitative Indices

To evaluate the quality of the reconstructed images, four metrics, MSE, mean absolute error (MAE), PSNR, and structural similarity index (SSIM), are selected for quantitative assessment.

MSE and MAE are the sum of squares and absolute values of the distance between predicted value and true value, respectively. MSE will amplify the image error to make the comparison more obvious, while MAE will ignore the outliers as much as possible. Therefore, MSE can be used to compare the performance of different methods, and MAE can be used to make quantitative analysis of different methods and parameters. MSE and MAE are defined as

\[
\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} |x_i - \hat{x}_i|^2
\]

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |x_i - \hat{x}_i|
\]

where \( N \) is the number of pixels in the reconstructed image.

The PSNR measure describes the relationship of the maximum possible power of a signal with the power of noise corruption. Higher PSNR means better image quality. Denoting \( x \) and \( \hat{x} \) to be the reconstructed image and ground truth, \( \text{PSNR} \) is expressed as

\[
\text{PSNR}(x, \hat{x}) = 20 \log_{10} \left( \frac{\text{Max}(x)}{\sqrt{\sum (x - \hat{x})^2}} \right).
\]

The SSIM-value is used to measure the similarity between the original CT image and reconstructed images, evaluated on three aspects: 1) luminance; 2) contrast; and 3) structural correlation. SSIM is defined as

\[
\text{SSIM}(x, \hat{x}) = \frac{(2\mu_x \mu_{\hat{x}} + c_1)(2\sigma_{x\hat{x}} + c_2)}{\left(\mu_x^2 + \mu_{\hat{x}}^2 + c_1\right)\left(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2\right)}
\]

where \( \mu_x \) and \( \sigma_x^2 \) are the average and variances of \( x \); \( \sigma_{x\hat{x}} \) is the covariance of \( x \) and \( \hat{x} \); \( c_1 \) and \( c_2 \) are used to maintain a stable constant.

### D. Reconstruction Results

#### AAPM Challenge Data:
To evaluate the image quality of reconstructed results, the MSE, PSNR, and SSIM values are given in terms of Means ± STDs (average scores ± standard deviations) for the LDCT images in the test dataset. Table II presents all the results, and the best value of each metric is marked in black bold. Intuitively, our method scores the highest PSNR and MSE, and DP-ResNet scores the highest SSIM. At the same time, comparing the experimental results of ASD-POCS and K-SVD, both of them obtain worse image quality evaluation metrics. It implies that the performance of the classical algorithms may still lose some details and suffer from remaining artifacts.

To visually illustrate the reconstruction performance, we perform qualitative comparisons over the selected methods for CT images of different body parts, as shown in Fig. 8.
It is noteworthy that the results focus on content restoration, artifact suppression, and noise reduction. From the FBP reconstructions in Fig. 8(b1)–(b3), we can see that FBP reconstruction leads to severely degraded LDCT images with obviously amplified noise and artifacts. As a traditional method, K-SVD (postprocess) produces reconstruction images [Fig. 8(d1)–(d3)] with visually smoother appearances. However, some tiny structures might be smoothed out and lead to lowered tissue contrast, as indicated by the red arrows. Furthermore, it also can be seen that DL methods effectively reduce noise and remarkably outmatch ASD-POCS and K-SVD in Fig. 8(e1)–(f3). They improve the effect of noise reduction and suppress most artifacts. However, they have incomplete preservation of details and texture information. Comparing the results in Fig. 8(g1)–(g3), we can see that the proposed EASEL method achieves the best performance in terms of noise-artifact suppression and tissue feature preservation. The corresponding residual images are shown in Fig. 9. Among them, ASD-POCS performs well but tends to smooth textures and edges. Compared to the competitive reconstruction methods, EASEL has its own advantages, which can effectively reduce the noise, and its reconstruction performs better in terms of artifact reduction and detail preservation.

To better illustrate the effectiveness of EASEL, Fig. 10 plots the 1-D line intensity profile passing through the red-dashed line in Fig. 8(a1). It compares the same line intensity profiles from the CT image reconstructed by various methods. It is evident that the line intensity profile from our proposed method resembles most closely to the one from the normal-dose CT image in most areas. At the same time, DP-ResNet has the advantage of jointly learning image features in the projection domain and the image domain. It can be observed that DP-ResNet performs well in Fig. 10.

For the differences between the K-SVD (postprocess) and K-SVD (Iter-recon) methods in this dataset, we qualitatively compare the CT images of different body parts, as shown in
Fig. 9. Absolute difference images between the reference CT image and the CT images reconstructed from the different algorithms. (a) FBP. (b) ASD-POCS. (c) K-SVD (postprocess). (d) RED-CNN. (e) DP-ResNet. (f) EASEL.

Fig. 10. One-dimensional intensity profile passing through the solid red line in Fig. 8(a). All the results in Fig. 8(a1)–(g1) are compared.

Fig. 11. It can be observed that in comparison with the iterative method, the reconstruction result of the postprocessing method lacks more details. The method based on postprocessing is more inclined to smooth the image.

CIRS Phantom Data: For the CIRS phantom data, quantitative results from different reconstruction methods are tabulated in Table III. It can be observed that EASEL performs better than the other methods in a trend similar to what we have seen from the reconstruction images and produces the highest PSNR. The PSNR measure of the EASEL reconstruction image increases by 0.37 dB compared to that of DP-ResNet. In fact, the unsupervised approach is less stringent on the imaging geometry than the end-to-end DL algorithm.

To visualize the benefits of the proposed method, reconstruction images with ROIs using different methods to the ground truth (full-dose CT images) are provided in Fig. 12. Specifically, a supervised method with the network structure would cause an obvious artifact around the center of the reconstruction, whereas the similar artifact would not appear in the reconstruction of K-SVD (postprocess) and ASD-POCS. One possible reason is those end-to-end learning models are particularly trained for a certain task with the same data. By comparing the result of K-SVD, we find that the boundary of the reconstruction image is still visible, while it is blurrier in the CT images from other methods. Moreover, DP-ResNet works a bit better than RED-CNN method but some edges and small structures are over smoothed.

For attaining further perspectives of our adapted EASEL for LDCT reconstruction, the residual images between the reconstructions and the reference are presented in Fig. 13, which demonstrates that the EASEL can achieve better reconstruction accuracy than that of the other algorithms on edge preservation.

To further investigate the algorithms’ ability of reconstruction, Fig. 14 plots the image profiles for the six methods together with that of the ground-truth image. Intuitively, the bias can be observed more clearly in the profile plots. The pixel intensities for the DP-ResNet reconstruction better follow those of the “true” clinical image, while those for

| Method                  | MSE  | PSNR | SSIM  |
|-------------------------|------|------|-------|
| FBP (Ramp-filter)       | 8.62 | 40.67| 0.941 |
| ASD-POCS                | 4.81 | 42.12| 0.964 |
| K-SVD (Post-process)    | 3.48 | 42.86| 0.969 |
| K-SVD (Iter-recon)      | 3.84 | 42.53| 0.971 |
| RED-CNN                 | 30.92| 41.76| 0.975 |
| DP-ResNet               | 19.11| 42.89| 0.981 |
| EASEL                   | 3.50 | 43.26| 0.981 |
the K-SVD reconstruction are much worse than the “true” values. Moreover, the gap between the profiles of the ASD-POCS method and the ground truth shows the gigantic bias. This means that ASD-POCS produces a strong bias in the reconstruction. All in all, among the compared methods, the reconstruction contours of EASEL, RED-CNN, and DP-ResNet are closer to the real situation in most areas than traditional reconstruction methods.

To sum up, in almost existing supervised DL-based methods, the network is learned by training a large amount of data that is acquired with specific imaging geometry. Large changes in the site under low-dose scanning may lead to rapid deterioration of the reconstruction results. In contrast, the proposed EASEL method largely alleviates the deficiency. In addition, we additionally calculate the average time of each iteration to facilitate the comparison of such algorithms. In Table IV, it can be seen that the end-to-end DL methods RED-CNN and DP-ResNet have great advantages in terms of reconstruction time. On the contrary, the iterative algorithm EASEL requires a certain number of iterations to obtain the best solution. Especially, the simulated annealing method requires iterative reconstruction under different noise levels. Therefore, compared with other methods, the reconstruction process of EASEL is time consuming and there is room for improvement in reconstruction time, but its single iteration time is very effective in practical applications.

To verify the superiority of EASEL as a tool of the unsupervised learning methodology, we use EASEL and DP-ResNet to reconstruct LDCT images with different noise levels, neither of which retrained the model. The experimental numerical indicators are shown in Table VII. As seen, EASEL is a bit superior to DP-ResNet for reconstructing the noisy data under noise level $b_L = 1e5$. However, if we use the same trained DP-ResNet model to tackle the noisy data under noise level $b_L = 1e4$, the performance of DP-ResNet degrades heavily. While our method

| Method   | Cost time (s) | Iteration number |
|----------|---------------|-----------------|
| FBP      | 0.06 s (CPU)  | /               |
| ASD-POCS | 0.26 s/iter (CPU) | 200            |
| K-SVD (Post-process) | 4.77 s/iter (CPU) | 40             |
| RED-CNN  | 0.08 s (GPU)  | /               |
| DP-ResNet| 0.19 s (GPU)  | /               |
| EASEL    | 0.22 s/iter (GPU) | 800            |
alleviates the remedy much and the performance still retains good behavior, the PSNR of EASEL is 3.7 dB higher. The numerical difference can be clearly shown in the reconstruction image in Fig. 15. Overall, the proposed method EASEL can handle a broad range of noise strengths.

In addition, we compare the proposed method with the GAN-based method [75]. Table VI gives the evaluation index of the reconstruction result. From the table, one can see that EASEL is superior to the GAN-based algorithm. Besides, some visualizations of the reconstruction results are shown in Fig. 16. The results of our algorithm are more natural and realistic than those in the GAN-based method.

To compare the visual quality of different algorithms, we add a couple of medical images with uncommon pathologies in the visual comparison experiment. Fig. 17 shows that the reconstructed LDCT images of EASEL suffer less from noise and artifacts with a better tissue identification. The results of the EASEL preserve the textures of the hemangioma tumor such that a better determination of the location of the lesion is possible.

E. Convergence of the Model

To examine the convergence of EASEL, the evolution curves of MSE, PSNR, and SSIM curve versus iteration are plotted in Fig. 18. It can be seen that the fluctuation of these curves is not obvious as the iteration increases. Besides, it is evident that EASEL is effective for noise and artifact suppression of LDCT images after 1500 iterations. Therefore, EASEL has a reasonable convergence rate.

In addition, the reconstruction results with regard to the number of input channels of the network $\theta(x)$ are investigated in Table V. It is predictable that as the channel number increases, the representation ability of the prior leverages. Subsequently, the performance will be improved. Considering
TABLE VII
QUANTITATIVE RESULTS (PSNR/SSIM) ASSOCIATED WITH DIFFERENT NOISE LEVEL IN AAPM TEST DATASET

| Method   | $h_0=1e5$ | $h_1=1e4$ |
|----------|-----------|-----------|
| DP-ResNet| 42.45/0.9609 | 34.97/0.8547 |
| EASEL    | 42.57/0.9621 | 38.67/0.9228 |

Fig. 19. Loss functions across training iterations under different training parameters. All conditions of convergence behavior are fast and stable.

Fig. 20. Reconstruction results of the same object under ten different implementations. It can be observed that the finally iterative reconstruction results under various implementations have no much difference. Here, the intermediate result at each step is added by a small amount of randomly artificial noise, whose amplitude is decreasing along the iterative procedure.

Meanwhile, we also tested the convergence trend of the model loss function under different conditions, as shown in Fig. 19 below. As can be seen from the convergence trend diagram of loss function under different parameter settings, EASEL can achieve stable convergence under different parameter settings.

Similarly, EASEL is also satisfactory in terms of stability. The following Fig. 20 shows the result of EASEL’s image reconstruction of the same image for ten times. It can be seen that EASEL’s reconstruction result is very stable.

F. Variants of Hyperparameters

The hyperparameter selection is one of the most crucial factors for the image quality attainable with the proposed method. We vary one parameter at a time while keeping the others fixed at their nominal values. Additionally, the AAPM challenge data are chosen as the training and test sets. Since the $\beta$-value is the most sensitive parameter to the image quality, it is estimated by comparing the numerical indicators between the full-dose CT images and the processed LDCT images. Fig. 21 depicts the MSEs and SSIMs of the reconstructed results under various implementations have no much difference. Here, the intermediate result at each step is added by a small amount of randomly artificial noise, whose amplitude is decreasing along the iterative procedure.

Later, the difference of performance between large $\beta$ and the peak $\beta$ vanishes. Thus, we set $\beta = 150$ in our experiments.

G. Analysis of Loss Function

The loss function, despite being the effective driver of the network’s learning, has attracted little attention within the image processing community [76]. The choice of the cost function generally defaults to be the squared $L_2$ norm of the error. In this study, we bring attention to loss function for LDCT reconstruction. Specifically, the MSE $L_2$ penalizes larger errors, but it is more tolerant to small errors, regardless of the underlying structure in the image. In contrast, the loss function with $L_1$-norm does not overpenalize larger errors. Consequently, they may have different convergence properties. Inspired by this observation, we test whether a different local metric such as $L_1$ can produce better results with the state-of-the-art metrics for image quality. The performance of several losses (i.e., $L_1$, $L_2$, $L_1+L_2$) is recorded in Table VIII. As can be seen, the reconstruction quality varies scarcely with the loss functions.

$L_2$-norm is arguably the dominant error measure across very diverse fields. The main reason for its popularity is the fact that it is convex and differentiable—very convenient properties for optimization problems. Meanwhile, it provides the maximum-likelihood estimate in case of independent and identically distributed Gaussian noise, to the fact that it is additive for independent noise sources. Overall, considering the case of our work that performs noise distribution mapping, i.e., the distribution obeys normal distribution and has the same
mathematical expression as $L_2$-norm, we choose $L_2$-norm as loss function to avoid the selection of complex parameters.

V. CONCLUSION AND DISCUSSION

In this work, a novel iterative reconstruction framework EASEL was proposed. By annealing the gradient of data density, elaborate prior knowledge from generative models was incorporated into the iterative procedure. The experimental results on two public datasets demonstrated the feasibility and efficiency of EASEL for LDCT imaging problem, improving image quality and avoiding noise effect. On the one hand, to effectively extract image features at multiple scales, we self-copied the LDCT image into ten channels, and then the generative network used the ten components as input. Moreover, the modification of network architecture was used in our experiments to better combine high-dimensional information for LDCT reconstruction. On the other hand, it is different from the NGM-driven DL mode in [30], which considers the noise generation mechanism in CT imaging and combines data-driven and model-driven, EASEL uses a generative model where samples were produced via Langevin dynamics using gradients of the data distribution estimated with DSM was utilized for LDCT reconstruction. In addition, the distance between the reconstructed images and the learned manifold was minimized along with the data fidelity by the SQS algorithm during iterative reconstruction. Different from other methods, the EASEL method can handle a broad range of noise strengths. The reconstructed image is more detailed, but this also has its limitations, when the original image contains many details and different scales of information, high-precision reconstruction can still be challenging.

Based on the above-mentioned improvements, the present EASEL can effectively alleviate the issues that the generative model is difficult to fully capture complex distributions, and avoid the lack of real rare details and with hallucinated details that appeared in GAN-based approaches [77]–[79]. In future study, we will further extend our model to find the image similarity search on latent space over huge clinical image dataset and deal with more challenging tasks. The most significant requirement will be the availability of a larger multi-GPU server architecture. The experiments that use such a computational infrastructure are underway. Besides, we would also like to integrate the deep gradient priors of generative model into an iterative reconstruction pipeline to mitigate possible image feature suppression imposed by the network. Last but not least, the current EASEL still needs ground-truth data at the prior learning stage. It is desirable to absorb the spirit in self-supervised schemes [64], [65] to relax the prior learning process, so as to broaden its imaging application scenarios.

ACKNOWLEDGMENT

The authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this article. The authors sincerely thank the anonymous referees for their valuable comments that greatly improved the quality of this work.

REFERENCES

[1] D. J. Brenner and E. J. Hall, “Computed tomography—An increasing source of radiation exposure,” New Engl. J. Med., vol. 357, no. 22, pp. 2277–2286, 2007.
[2] M. K. Kalra et al., “Strategies for CT radiation dose optimization,” Radiology, vol. 230, no. 3, pp. 619–628, 2004.
[3] J. Wang, T. Li, H. Lu, and Z. Liang, “Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography,” IEEE Trans. Med. Imag., vol. 25, no. 10, pp. 1272–1283, Oct. 2006.
[4] Y. Chen et al., “Improving low-dose abdominal CT images by weighted intensity averaging over large-scale neighborhoods,” Eur. J. Radiol., vol. 80, no. 2, pp. 42–49, 2011.
[5] Z. Tian, X. Jia, K. Yuan, T. Pan, and S. B. Jiang, “Low-dose CT reconstruction via edge-preserving total variation regularization,” Phys. Med. Biol., vol. 56, no. 18, pp. 5949–5967, 2011.
[6] A. Mehranian, M. R. A. Yaj, A. Rahim, and H. Zaidi, “X-ray CT metal artifact reduction using wavelet domain L0 sparse regularization,” IEEE Trans. Med. Imag., vol. 32, no. 9, pp. 1707–1722, Sep. 2013.
[7] D. Zeng et al., “Spectral CT image restoration via an image-inducled nonlocal means filter,” IEEE Trans. Biomed. Eng., vol. 63, no. 5, pp. 1044–1057, May 2016.
[8] T. Tirer and R. Giryes, “Image restoration by iterative denoising and backward projections,” IEEE Trans. Image Process., vol. 28, pp. 1220–1234, 2019.
[9] H. Chen et al., “Low-dose CT via convolutional neural network,” Biomed. Opt. Express, vol. 8, no. 2, pp. 679–694, 2017.
[10] E. Kang, W. Chang, J. Yoo, and J. C. Ye, “Deep convolutional framelet denoising for low-dose CT via wavelet residual network,” IEEE Trans. Med. Imag., vol. 37, no. 6, pp. 1358–1369, Jun. 2018.
[11] E. Kang, J. Min, and J. C. Ye, “A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction,” Med. Phys., vol. 44, no. 10, pp. e360–e375, 2017.
[12] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, “Generative image feature suppression imposed by the network. Last but

image feature suppression imposed by the network. Last but
