Journal of Scientific Computing (2024) 101:10
https://doi.org/10.1007/s10915-024-02638-7

Provably Convergent Learned Inexact Descent Algorithm for Low-Dose CT Reconstruction

Qingchao Zhang[1] · Mehrdad Alvandipour[1] · Wenjun Xia[2] · Yi Zhang[2] · Xiaojing Ye[3] · Yunmei Chen[1]

Received: 29 July 2023 / Revised: 19 February 2024 / Accepted: 9 April 2024 /
Published online: 20 August 2024

Abstract
We propose an Efficient Inexact Learned Descent-type Algorithm (ELDA) for a class of nonconvex and nonsmooth variational models, where the regularization consists of a sparsity enhancing term and non-local smoothing term for learned features. The ELDA improves the performance of the LDA in Chen et al. (SIAM J Imag Sci 14(4), 1532–1564, 2021) by reducing the number of the subproblems from two to one for most of the iterations and allowing inexact gradient computation. We generate a deep neural network, whose architecture follows the algorithm exactly for low-dose CT (LDCT) reconstruction. The network inherits the convergence behavior of the algorithm and is interpretable as a solution of the varational model and parameter efficient. The experimental results from the ablation study and comparisons with several state-of-the-art deep learning approaches indicate the promising performance of the proposed method in solution accuracy and parameter efficiency.

Keywords Learned inexact descent algorithm · Nonconvex nonsmooth optimization · Deep learning · Image reconstruction

Mathematics Subject Classification 90C26 · 65K05 · 65K10 · 68U10

✉ Yunmei Chen
yun@ufl.edu

Qingchao Zhang
qingchaozhang3@gmail.com

Mehrdad Alvandipour
m.alvandipour@ufl.edu

Wenjun Xia
xwj90620@gmail.com

Yi Zhang
yzhang@scu.edu.cn

Xiaojing Ye
xye@gsu.edu

1 Department of Mathematics, University of Florida, Gainesville, FL 32611, USA

2 College of Computer Science, Sichuan University, Chengdu 610065, China

3 Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, USA

1 Introduction

Computed Tomography (CT) is one of the most widely used imaging technologies for medical diagnosis. CT employs X-ray measurements from different angles to generate cross-sectional images of the human body [23, 38]. As high dosage X-rays can be harmful to human body [10, 11, 78], substantial efforts have been devoted to image reconstruction using low-dose CT measurements [32, 48, 79]. There are two main strategies for dose reduction in CT scans: one is to reduce the number of views, and the other is to reduce the exposure time and the current of X-ray tube [39], both of which will introduce various degrees of noise and artifacts and then compromise the subsequent diagnosis. Here we focus on the second type however our method is not specific to a particular scanning mode. We formulate the problem as an optimization problem which will be solved with our Efficient Learned Descent Algorithm (ELDA).

The classic analytical method to reconstruct CT images from projection data, Filtered Back-Projection (FBP), leads to heavy noise and artifacts in the low dose scenario. The remedy for this problem have been sought from three different perspectives: pre-processing the sinograms [44, 65, 71], post-processing the images [88], or the hybrid approach with iterative reconstructions that encode prior information into the process [29, 98, 119].

The advent of machine learning methods and its success on various real-world applications, including the image processing [3–9, 17, 100, 110–113, 116, 121, 129], classification and detection [12, 50, 51, 53, 54, 62, 85, 86, 89–97, 114, 123], health care [41, 52, 63, 66–70, 122], chemistry [75], industry [81, 83, 84] and beyond [27, 35–37, 42, 58, 61, 64, 101, 102, 105, 106, 118, 120, 124–128], have naturally led to incorporation of deep models into all of the above approaches and produced a better performance than analytical methods [80]. For instance, CNN methods [13, 26, 31, 60, 104], that have been applied to sparse view [40, 43, 57, 104, 117] and low dose [13, 31, 46, 60, 107] data. It is also applied in projection domain synthesis [56, 57], post processing [13, 14, 43, 47, 104], and for prior learning in iterative methods [15, 99, 109, 115].

Recently, a number of learned optimization methods have been proposed and are proven very effective in CT reconstruction problem, as they are able to learn adaptive regularizer which leads to more accurate image reconstruction in a variety of medical imaging applications. However, existing works model regularizers using convolutional neural networks (CNNs) which only explore local image features. This limits the representation power of deep neural networks and is not suitable for medical imaging applications which demand high image qualities. Moreover, most of existing deep networks for image reconstruction are cast as black-boxes and can be difficult to interpret. Last but not least, deep neural networks for image reconstruction are also criticized for lacking mathematical justifications and convergence guarantee.

In this work, we leverage the framework developed in [16] and propose an improved learned descent algorithm ELDA. It further boosts image reconstruction quality using an adaptive non-local feature regularizer. More importantly, compared to [16], ELDA is more computationally efficient since the safeguard iterate is only computed when a descent condition fails to hold, which happens rarely due to allowance of inexact gradient computation in our algorithmic design. As a result, our model retains convergence guarantee and meanwhile also improves reconstruction quality over existing methods. The main contributions of this work are summarized as follows.

– We propose an efficient learned descent algorithm with inexact gradients, to solve the non-smooth non-convex optimization problem in low-dose CT reconstruction. Inexact

gradients improve our model in comparison with LDA [16] in two major ways: First, it reduces computational cost by allowing the gradient to be inexact. Second, it increases the capacity of the network and thus improves accuracy. We also provide comprehensive convergence and iteration complexity analysis of ELDA.
- ELDA adopts efficient update scheme which only computes safeguard iterate when the desired descent condition fails to hold, and hence is more computationally economical than LDA developed in [16]. In particular, we show that in 99.6% of the cases, our proximal candidate is selected and extra computations for the plain gradient descent is avoided. This is in contrast with LDA where the descent conditions holds for 86.2% of the cases and therefore extra computational cost is paid more frequently for the alternate candidate.
- We designed a novel non-local smoothing regularization which along with the sparsity enhancing regularizer, further improve image quality, and enable faster convergence. As a result our largest model uses only 19 iteration blocks and has about 20 times less parameters than the nearest competing model [103].
- We conduct comprehensive experimental analysis of ELDA and compare it to several state-of-the-art deep-learning based methods for LDCT reconstruction.

In Sect. 2, we present the related works in the literature that associate with our problem. Then in Sect. 3, we present our method by first defining our model and each of its components, and then stating the algorithm and details of network training. After that in Sect. 4, we state our lemma and theorem regarding the output of the network. Section 5 presents the numerical results including parameter study, ablation study and comparison with other competing algorithms.

2 Related Works

A natural application of neural networks in CT reconstruction, has been in noise removal in either the projection domain [56, 57] or the image domain [13, 14, 43, 47, 104]. In particular, Residual Encoder–Decoder Convolutional Neural Network (RED-CNN) proposed by Chen et al. [14], is an end-to-end mapping from low-dose CT images to normal dose which uses FBP to get low-dose CT images from projections and restrict the problem to denoising in the image domain. And yet another attempt is FBPConvNet by Jin et al. [43] which is inspired by U-net [77] and further explores CNN architectures while noting the parallels with the general form of an iterative proximal update.

Model Based Image Reconstruction (MBIR) methods attempt to model CT physics, measurement noise, and image priors in order to achieve higher reconstruction quality in LDCT. Such methods learn the regularizer and are able to improve LDCT reconstruction significantly [18, 29, 109, 119], however their convergence speed is not optimal [19]. Later, researchers adopted NNs in other aspects of the algorithm and formed a new class of methods called Iterative Neural Networks (INN). INNs seek to enjoy the best of both world of MBIR and denoising deep neural networks, by employing moderate complexity denoisers for image refining and learning better regularizers [20, 21, 82, 108].

INNs have network architectures that are *inspired* by the optimization model and algorithm and this learning capacity enables them to outperform the classical iterative solutions by learning better regularizer while also being more time efficient. For example, recently BCD-Net [21] improved the reconstruction accuracy compared to MBIR methods and NN denoisers. It showed that it generalizes better than denoisers such as FBPConvNet which lack

MBIR modules, and also its learned transforms help to outperform state-of-the-art MBIR methods. Further research in this area has been devoted to improving time efficiency of the algorithm with the image quality. Recently, Chun et al. [20] proposed Momentum-Net, as the first fast and convergent INN architecture inspired by Block Proximal Extrapolated Gradient method using a Majorizer. It also guarantees convergence to a fixed-point while improving MBIR speed, accuracy, and reconstruction quality. Momentum-Net is a general framework for inverse problems and its recent application to LDCT [108] showed it improves image reconstruction accuracy compared to a state-of-the-art noniterative image denoising NN. Convergence guarantee is one of the main challenges in the design of INNs and beside its theoretical value, it is highly desirable in medical applications. LEARN[15] is another model that unrolls an iterative reconstruction scheme while modeling the regularization by a field-of-experts. And yet another similar attempt is Learned Primal-Dual [1] which unfolds a proximal primal-dual optimization method where the proximal operator is replaced with a CNN. Their choice of iterative scheme is primal dual hybrid gradient (PDHG) which is further modified to benefit from the learning capacity of NNs and then used to solve the TV regularized CT reconstruction problem.

In all of the previous works, the architecture is *only inspired* by the optimization model, and in order to improve their performance they introduce components in the network that does not correspond to steps of the algorithm. Also, the choice of regularization limits the network to only learn local features and as we will empirically demonstrate, it limits the performance of these networks. One model that attempts to learn non-local features is MAGIC [103]. It is also a deep neural network inspired by a simple iterative reconstruction method, i.e. gradient descent. However MAGIC breaks the correspondence between architecture and algorithm in order to extract non-local features [49, 87]. They manually add a non-local corrector in iteration steps which is only *intuitively* justified, and does not directly correspond to a modified regularizer in the optimization model.

In [16], a Learned Descent Algorithm (LDA) is developed. The LDA architecture is fully determined by the algorithm and thus the network is fully interpretable. As interpretability and convergence guarantee is highly desirable in medical imaging, this framework is a promising method for inverse problems such as LDCT reconstruction. Compared to [16], the present work proposes a more efficient numerical scheme of LDA, leading to comparable network parameters, lower computational cost, and more stable convergence behavior. We achieve this by developing an efficient learned inexact descent algorithm which only computes the safeguard iterate when a prescribed descent condition fails to hold and thus substantially reduces computational cost in practice. Additionaly, we propose a novel non-local smoothing regularizer that further confirms the heuristics in optimization inspired networks such as MAGIC [103] but leads to a fully interpretable network and allows us to provide convergence guarantee of the network.

3 Method

In this section, we introduce the proposed inexact learned descent algorithm for solving the following low-dose CT reconstruction model:

$$\mathbf{x}_\theta^{(s)} = \arg\min_{\mathbf{x}} \{\phi(\mathbf{x}; \mathbf{b}^{(s)}, \theta) := f(\mathbf{x}; \mathbf{b}^{(s)}) + r(\mathbf{x}; \theta)\}, \tag{1}$$

where f is the data fidelity term that measures the consistency between the reconstructed image $\mathbf{x}$ and the sinogram measurements $\mathbf{b}$, and r is the regularization that may incorporate

prior information of $\mathbf{x}$. The regularization $r(\cdot;\theta)$ is realized as a highly structured DNN with parameter θ to be learned. The optimal parameter θ of r is then obtained by minimizing the loss function $\mathcal{L}$, where $\mathcal{L}$ measures the (averaged) difference between $\mathbf{x}_\theta^{(s)}$-the minimizer of $\phi(\cdot;\mathbf{b}^{(s)},\theta)$, and the given ground truth $\hat{\mathbf{x}}^{(s)}$ for every $s \in [N]$, where N is the number of training data pairs. For notation simplicity, we write $f(\mathbf{x})$ and $r(\mathbf{x})$ instead of $f(\mathbf{x};\mathbf{b}^{(s)})$ and $r(\mathbf{x};\theta)$ respectively hereafter. We choose $f(\mathbf{x}) = \frac{1}{2}\|A\mathbf{x}-\mathbf{b}\|^2$ as the data-fidelity term, where A is the system matrix for CT scanner. However, our proposed method can be readily extended to any smooth but (possibly) nonconvex f.

3.1 Regularization Term in Model (1)

The regularization term r in (1) consists of two parts. One of them enhances the sparsity of the solution under a learned transform and the other one smooths the feature maps non-locally:

$$r(\mathbf{x}) := \hat{r}(\mathbf{x}) + \lambda \bar{r}(\mathbf{x}), \tag{2}$$

where λ is a coefficient to balance these two terms which can be learned.

3.1.1 The Sparsity-Enhancing Regularizer

To enhance the sparsity of $\mathbf{x}$ under a learned transform $\mathbf{g}$, we propose to minimize the $l_{2,1}$ norm of $\mathbf{g}(\mathbf{x})$. If $\mathbf{g}$ is a differential operator, then the $l_{2,1}$ norm of $\mathbf{g}(\mathbf{x})$ reduces to the total variation of $\mathbf{x}$. That is,

$$\hat{r}(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|_{2,1} = \sum_{i=1}^{m} \|\mathbf{g}_i(\mathbf{x})\|, \tag{3}$$

where each $\mathbf{g}_i(\mathbf{x}) \in \mathbb{R}^d$ can be viewed as a feature descriptor vector at the position i, as depicted in Fig. 1 (up). In our experiments, we simply set the feature extraction operator $\mathbf{g}$ to a vanilla l-layer CNN with nonlinear activation function σ but no bias, as follows:

$$\mathbf{g}(\mathbf{x}) = \mathbf{w}_l * \sigma \cdots \sigma(\mathbf{w}_3 * \sigma(\mathbf{w}_2 * \sigma(\mathbf{w}_1 * \mathbf{x}))), \tag{4}$$

where $\{\mathbf{w}_q\}_{q=1}^{l}$ denote the convolution weights consisting of d kernels with identical spatial kernel size (3×3), and $*$ denotes the convolution operation. Here, the componentwise activation function σ is constructed to be the smoothed rectified linear unit as defined below

$$\sigma(x) = \begin{cases} 0, & \text{if } x \le -\delta, \\ \frac{1}{4\delta}x^2 + \frac{1}{2}x + \frac{\delta}{4}, & \text{if } -\delta < x < \delta, \\ x, & \text{if } x \ge \delta, \end{cases} \tag{5}$$

where the prefixed parameter δ is set to be 0.001 in our experiment. Besides the smooth σ, each convolution operation of $\mathbf{g}$ in (4) can be viewed as matrix multiplication, which enable $\mathbf{g}$ to be differentiable, and $\nabla\mathbf{g}$ can be easily obtained by Chain Rule where each $\mathbf{w}_q^\top$ can be implemented as transposed convolutional operation [28].

As $\hat{r}(\mathbf{x})$ defined in (3) is nonsmooth and nonconvex, we apply the Nesterov's smoothing technique [73] to get the smooth approximation and the detail is given in [16]:

$$\hat{r}_\varepsilon(\mathbf{x}) = \sum_{i\in I_0} \frac{1}{2\varepsilon}\|\mathbf{g}_i(\mathbf{x})\|^2 + \sum_{i\in I_1}\left(\|\mathbf{g}_i(\mathbf{x})\| - \frac{\varepsilon}{2}\right), \tag{6}$$

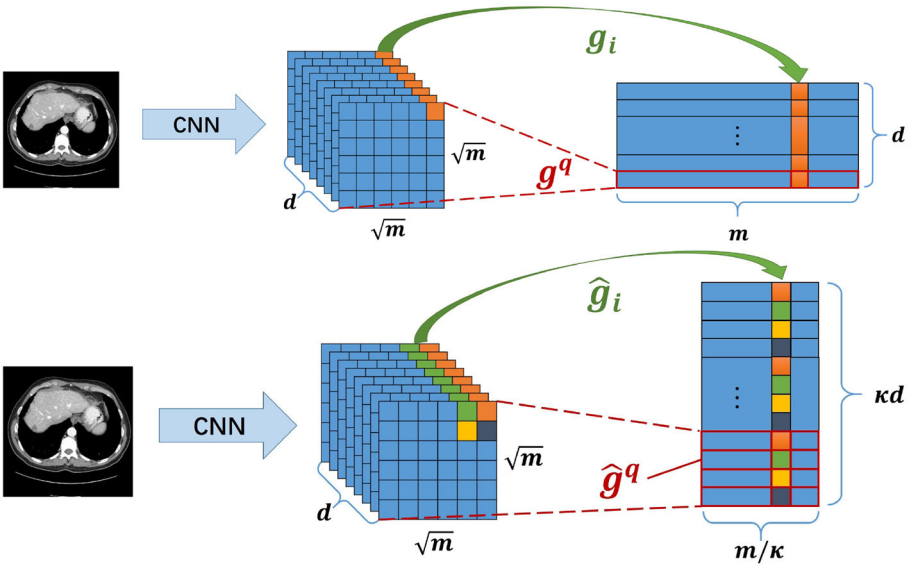

Fig. 1 The feature matrix $\mathbf{g} \in \mathbb{R}^{d\times m}$ (up) and the folded feature matrix $\hat{\mathbf{g}} \in \mathbb{R}^{\kappa d\times \frac{m}{\kappa}}$ (bottom) reshaped from the feature maps obtained from the last convolution of the CNN defined in (4). The folding rate κ is 4 for $\hat{\mathbf{g}}$ in this illustration

where $I_0 = \{i \in [m] \mid \|\mathbf{g}_i(\mathbf{x})\| \leq \varepsilon\}$, $I_1 = [m] \setminus I_0$. Here the parameter ε controls how close the smoothed $\hat{r}_\varepsilon(\mathbf{x})$ is to the original function $\hat{r}(\mathbf{x})$, and one can readily show that $\hat{r}_\varepsilon(\mathbf{x}) \leq \hat{r}(\mathbf{x}) \leq \hat{r}_\varepsilon(\mathbf{x}) + \frac{m\varepsilon}{2}$ for all $\mathbf{x}$ in $\mathbb{R}^n$. From (6) we can also derive $\nabla\hat{r}_\varepsilon(\mathbf{x})$ to be

$$\nabla\hat{r}_\varepsilon(\mathbf{x}) = \sum_{i\in I_0} \nabla\mathbf{g}_i(\mathbf{x})^\top \frac{\mathbf{g}_i(\mathbf{x})}{\varepsilon} + \sum_{i\in I_1} \nabla\mathbf{g}_i(\mathbf{x})^\top \frac{\mathbf{g}_i(\mathbf{x})}{\|\mathbf{g}_i(\mathbf{x})\|},$$

where $\nabla\mathbf{g}_i(\mathbf{x}) \in \mathbb{R}^{d\times n}$ is the Jacobian of $\mathbf{g}_i$ at $\mathbf{x}$.

3.1.2 The Nonlocal Smoothing Regularizer

Since convolution operations only extract the local information, each feature descriptor vector $\mathbf{g}_i$ can only encode the local features of a small patch of the input $\mathbf{x}$ (i.e. receptive field) [55]. So here we seek to incorporate an additional non-local smoothing regularizer $\bar{r}$ that enables capturing of the underlying long-range dependencies between the *patches* of the feature descriptor vectors. To this end we form the folded feature descriptor vectors $\{\hat{\mathbf{g}}_i\}$ as described in Fig. 1 (bottom) by folding the adjacent κ feature descriptors together, and define $\bar{r}$ by:

$$\bar{r}(\mathbf{x}) = \sum_{(i,j)} \mathcal{W}_{ij} \|\hat{\mathbf{g}}_i(\mathbf{x}) - \hat{\mathbf{g}}_j(\mathbf{x})\|^2, \tag{7}$$

where the similarity matrix $\mathcal{W}$ is defined by $\mathcal{W}_{ij} = \exp\left(-\frac{\|\hat{\mathbf{g}}_i(\mathbf{x})-\hat{\mathbf{g}}_j(\mathbf{x})\|^2}{\delta^2}\right)$, and δ is the standard deviation, which is estimated by the median of the Euclidean distances between the folded feature descriptor vectors in the model.

Additionally, $\bar{r}$ can also be written in the quadratic form [2] as $\bar{r}(\mathbf{x}) = tr(\hat{\mathbf{g}}(\mathbf{x})\mathcal{L}\,\hat{\mathbf{g}}(\mathbf{x})^\top)$, where $tr()$ is the trace operator, $\mathcal{L} = \mathcal{D} - \mathcal{W}$, and $\mathcal{D}$ is the diagonal matrix with $\mathcal{D}_{ii} = \sum_{j=1}^{\frac{m}{\kappa}} \mathcal{W}_{ij}$, and $\mathcal{L}$ is positive semidefinite. And its gradient is computed by

$$\begin{aligned}\nabla\bar{r}(\mathbf{x}) &= 2\cdot\sum_{(i,j)}\tilde{\mathcal{W}}_{ij}(\nabla\hat{\mathbf{g}}_i(\mathbf{x}) - \nabla\hat{\mathbf{g}}_j(\mathbf{x}))^\top(\hat{\mathbf{g}}_i(\mathbf{x}) - \hat{\mathbf{g}}_j(\mathbf{x}))\\ &= 2\cdot\sum_{q=1}^{\kappa d}\nabla\hat{\mathbf{g}}^q(\mathbf{x})^\top\tilde{\mathcal{L}}\,\hat{\mathbf{g}}^q(\mathbf{x}),\end{aligned}$$

where $\nabla\hat{\mathbf{g}}^q(\mathbf{x}) \in \mathbb{R}^{\frac{m}{\kappa}\times n}$ is the Jacobian of $\hat{\mathbf{g}}^q(\mathbf{x})$. And $\tilde{\mathcal{L}} = \tilde{\mathcal{D}} - \tilde{\mathcal{W}}$, $\tilde{\mathcal{D}}_{ii} = \sum_{j=1}^{\frac{m}{\kappa}}\tilde{\mathcal{W}}_{ij}$ and $\tilde{\mathcal{W}}_{ij} = \mathcal{W}_{ij}(\mathbf{x})(1 - \frac{\|\hat{\mathbf{g}}_i(\mathbf{x})-\hat{\mathbf{g}}_j(\mathbf{x})\|^2}{\delta^2})$. Each $\hat{\mathbf{g}}^q(\mathbf{x})$ represents the q-th row of the folded feature matrix $\hat{\mathbf{g}}(\mathbf{x})$ as illustrated in Fig. 1.

3.2 Inexact Learned Descent Algorithm

Now we present an inexact smoothing gradient descent type algorithm to solve the nonconvex and nonsmooth problem (1) with the smooth approximation of $r(\mathbf{x})$ defined by $r_\varepsilon(\mathbf{x}) := \hat{r}_\varepsilon(\mathbf{x}) + \lambda\bar{r}(\mathbf{x})$. The proposed algorithm is shown in Algorithm 1. In each iteration k, we solve the following smoothed problem (9) with fixed $\varepsilon = \varepsilon_k$ in Line 3–14. And Line 15 is aimed to check and update ε_k by a reduction principle.

$$\min_{\mathbf{x}}\{\phi_\varepsilon(\mathbf{x};\mathbf{b}^{(s)},\theta) := f(\mathbf{x};\mathbf{b}^{(s)}) + r_\varepsilon(\mathbf{x};\theta)\}. \tag{9}$$

As the regularization term r_ε is learned via a deep neural network (DNN), some common issues of the DNN have to be taken into consideration when designing the algorithm, such as gradient exploding and vanishing problem during training [34]. Substantial improvement in performance has been achieved by ResNet [33] which introduces residual connections to alleviate these issues. As in (9) only the second term $r_\varepsilon(\mathbf{x};\theta)$ is learned, we desire to have individual residual updates for this term in our algorithm. To this end, we use the first order proximal method to solve the smoothed problem (9) by iterating the following steps

$$\mathbf{z}_{k+1} = \mathbf{x}_k - \alpha_k\nabla f(\mathbf{x}_k), \tag{10a}$$

$$\mathbf{x}_{k+1} = \mathrm{prox}_{\alpha_k r_{\varepsilon_k}}(\mathbf{z}_{k+1}), \tag{10b}$$

where $\mathrm{prox}_{\alpha r}(\mathbf{z}) := \arg\min_{\mathbf{x}}\ \frac{1}{2\alpha}\|\mathbf{x}-\mathbf{z}\|^2 + r(\mathbf{x})$.

From our construction of r_{ε_k}, it is hard to get the close-form solution to subproblem in (10b). Here we propose to linearize the "nonsimple" term r_{ε_k} by

$$\tilde{r}_{\varepsilon_k}(\mathbf{x}) = r_{\varepsilon_k}(\mathbf{z}_{k+1}) + \langle\nabla r_{\varepsilon_k}(\mathbf{z}_{k+1}), \mathbf{x}-\mathbf{z}_{k+1}\rangle + \frac{1}{2\beta_k}\|\mathbf{x}-\mathbf{z}_{k+1}\|^2. \tag{11}$$

With this approximation, instead of solving (10b) directly, we update by the following step

$$\mathbf{u}_{k+1} = \mathrm{prox}_{\alpha_k\tilde{r}_{\varepsilon_k}}(\mathbf{z}_{k+1}), \tag{12}$$

which has a closed-form solution giving the residual update

$$\mathbf{u}_{k+1} = \mathbf{z}_{k+1} - \tau_k\nabla r_{\varepsilon_k}(\mathbf{z}_{k+1}), \tag{13}$$

where $\nabla r_{\varepsilon_k} = \nabla\hat{r}_{\varepsilon_k} + \lambda\nabla\bar{r}$ and $\tau_k = \frac{\alpha_k\beta_k}{\alpha_k+\beta_k}$.

From the optimization perspective, with the approximation in (11), if we update $\mathbf{x}_{k+1} = \mathbf{u}_{k+1}$ then the algorithm can not be guaranteed to converge. In order to ensure convergence, we check whether $\mathbf{u}_{k+1}$ satisfies

$$\|\nabla\phi_{\varepsilon_k}(\mathbf{x}_k)\| \le c\|\mathbf{u}_{k+1} - \mathbf{x}_k\| \quad \text{and} \quad \phi_\varepsilon(\mathbf{u}_{k+1}) - \phi_\varepsilon(\mathbf{x}_k) \le -\frac{\iota}{2}\|\mathbf{u}_{k+1} - \mathbf{x}_k\|^2, \tag{14}$$

where c and ι are prefixed constant numbers. If the condition (14) holds, we take $\mathbf{x}_{k+1} = \mathbf{u}_{k+1}$; otherwise, we take the standard gradient descent $\mathbf{v}_{k+1}$ coming from

$$\mathbf{v}_{k+1} = \arg\min_{\mathbf{x}} \langle \nabla f(\mathbf{x}_k), \mathbf{x} - \mathbf{x}_k\rangle + \langle \nabla r_{\varepsilon_k}(\mathbf{x}_k), \mathbf{x} - \mathbf{x}_k\rangle + \frac{1}{2\alpha_k}\|\mathbf{x} - \mathbf{x}_k\|^2, \tag{15}$$

which has the exact solution

$$\mathbf{v}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k) - \alpha_k \nabla r_{\varepsilon_k}(\mathbf{x}_k). \tag{16}$$

To ensure convergence, we need to find α_k through line search such that $\mathbf{v}_{k+1}$ satisfies

$$\phi_{\varepsilon_k}(\mathbf{v}_{k+1}) - \phi_{\varepsilon_k}(\mathbf{x}_k) \le -\tau\|\mathbf{v}_{k+1} - \mathbf{x}_k\|^2, \tag{17}$$

where τ is a prefixed constant. Lemma 2 proves the convergence of lines 3–14 Algorithm 1, including the termination of its line search (lines 9–13) in finitely many steps.

Inspired by [16], the proposed algorithm boasts numerous modifications that enhance its efficiency and suitability for deep neural networks. One key contrast lies in the handling of the two candidate updates, $\mathbf{u}_{k+1}$ and $\mathbf{v}_{k+1}$. While [16] computes both candidates at every iteration and select the one yielding a lower function value, we introduce a new criterion (14) for updating $\mathbf{x}_{k+1}$. This potentially eliminates the need to calculate $\mathbf{v}_{k+1}$ altogether, leading to significant computational savings. Furthermore, the descending condition in Line 5 of Algorithm 1 mitigates the frequent switching between the candidate updates as the algorithm progresses (see Sect. 5.2 for details). This not only contributes to enhanced stability but also allows us to leverage inexact gradient descent, augmenting the network's capacity.

The learned inexact gradient

To further increase the capacity of the network, we employ the learned transposed convolution operator, i.e. we replace $\mathbf{w}_q^\top$ by a transposed convolution

$\widetilde{\mathbf{w}}_q$ with relearned weights, where q denotes the index of convolution in (4). To approximately achieve $\widetilde{\mathbf{w}}_q \approx \mathbf{w}_q^\top$, we add the constraint term $\mathcal{L}_{constraint} = \frac{1}{N_w}\sum_{q=1}^{4}\|\widetilde{\mathbf{w}}_q - \mathbf{w}_q^\top\|_F^2$ to the loss function in training to produce the data-driven transposed convolutions. Here N_w is the number of parameters in learned transposed convolutions and $\|\cdot\|_F$ is the Frobenius norm. In effect, the consequence of this modification is only to substitute ∇r_{ε_k} by the inexact gradient $\widetilde{\nabla} r_{\varepsilon_k}$ equipped with learned transpose at Line 4 in Algorithm 1. This can further increase the capacity of the unrolled network while maintaining the convergence property.

3.3 Network Training

We allow the step sizes α_k and τ_k to vary in different phases. Moreover, all $\{\alpha_k, \tau_k\}_{k=1}^K$ and initial threshold ε_0 are designed to be learned parameters fitted by data. Here let θ stand for the set of all learned parameters of ELDA which consists of the weights of the convolutions and approximated transposed convolutions, step sizes $\{\alpha_k, \tau_k\}_{k=1}^K$ and threshold ε_0, parameter λ. Given N training data pairs $\{(\mathbf{b}^{(s)}, \hat{\mathbf{x}}^{(s)})\}_{s=1}^N$ of the ground truth data $\hat{\mathbf{x}}^{(s)}$ and its corresponding

Algorithm 1 The Efficient learned Descent Algorithm (ELDA) for the Nonsmooth Nonconvex Problem

Input: Initial $\mathbf{x}_0$, $0 < \rho, \gamma < 1$, and $\varepsilon_0, \sigma, c, \iota, \tau > 0$. Set maximum iteration K or tolerance $\epsilon_{\text{tol}} > 0$ and select $\bar{\alpha} > 0$.
for $k = 0, 1, 2, \ldots, K$ **do**
  $\mathbf{z}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k)$, (initialize α_k s.t. $\alpha_k \geq \bar{\alpha}$)
  $\mathbf{u}_{k+1} = \mathbf{z}_{k+1} - \tau_k \nabla r_{\varepsilon_k}(\mathbf{z}_{k+1})$, (possibly inexact)
  if condition (14) holds **then**
    set $\mathbf{x}_{k+1} = \mathbf{u}_{k+1}$,
  else
    $\mathbf{v}_{k+1} = \mathbf{x}_k - \alpha_k \nabla f(\mathbf{x}_k) - \alpha_k \nabla r_{\varepsilon_k}(\mathbf{x}_k)$,
    if condition (17) holds **then**
      set $\mathbf{x}_{k+1} = \mathbf{v}_{k+1}$,
    else
      update $\alpha_k \leftarrow \rho \alpha_k$, then **go to** 8,
    end if
  end if
  if $\|\nabla \phi_{\varepsilon_k}(\mathbf{x}_{k+1})\| < \sigma\gamma\varepsilon_k$, set $\varepsilon_{k+1} = \gamma\varepsilon_k$; **otherwise**, set $\varepsilon_{k+1} = \varepsilon_k$.
  if $\sigma\varepsilon_k < \epsilon_{\text{tol}}$, terminate.
end for
Output: $\mathbf{x}_{k+1}$.

measurement $\mathbf{b}^{(s)}$, the loss function $\mathcal{L}(\theta)$ is defined to be the sum of the discrepancy loss $\mathcal{L}_{discrepancy}$ and the constraint loss $\mathcal{L}_{constraint}$:

$$\mathcal{L}(\theta) = \underbrace{\frac{1}{N}\sum_{s=1}^{N} \|\mathbf{x}_\theta^K - \hat{\mathbf{x}}^{(s)}\|^2}_{\mathcal{L}_{discrepancy}} + \underbrace{\frac{\vartheta}{N_w}\sum_{q=1}^{4} \|\widetilde{\mathbf{w}}_q - \mathbf{w}_q^\top\|_F^2}_{\mathcal{L}_{constraint}}, \tag{18}$$

where $\mathcal{L}_{discrepancy}$ measures the discrepancy between the ground truth $\hat{\mathbf{x}}^{(s)}$ and $\mathbf{x}_\theta^K$ which is the output of the K-phase network. Here, the constraint coefficient ϑ is set to 10^{-2} in our experiment.

4 Convergence Analysis

According to the problem we are solving, we make a few assumptions on f and $\mathbf{g}$ throughout this section.

(A1) : f is differentiable and (possibly) nonconvex, and ∇f is L_f-Lipschitz continuous.
(A2) : Every component of $\mathbf{g}$ is differentiable and (possibly) nonconvex, $\nabla \mathbf{g}$ is L_g-Lipschitz continuous, and $\sup_{\mathbf{x}\in\mathcal{X}} \|\nabla \mathbf{g}(\mathbf{x})\| \leq M$ for some constant $M > 0$.
(A3) : ϕ is coercive, and $\phi^* = \min_{\mathbf{x}\in\mathcal{X}} \phi(\mathbf{x}) > -\infty$.

With the smoothly differentiable activation σ defined in (5) and boundedness of σ' as well as the fixed convolution weights in (4) after training, we can immediately verify that the first two assumptions hold, and typically in image reconstruction ϕ is assumed to be coercive [16].

As the objective function in (1) is nonsmooth and nonconvex, we utilize the Clarke subdifferential[22] to characterize the optimality of solutions. We denote $D(\mathbf{x}; \mathbf{v}) := \limsup_{\mathbf{z}\to\mathbf{x},\, t\downarrow 0}[(f(\mathbf{z} + t\mathbf{v}) - f(\mathbf{z}))/t]$.

Definition 1 (Clarke subdifferential) Suppose that $f : \mathbb{R}^n \to (-\infty, +\infty]$ is locally Lipschitz, the Clarke subdifferential $\partial f(\mathbf{x})$ of f at $\mathbf{x}$ is defined as

$$\partial f(\mathbf{x}) := \Big\{ \mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{x}, \mathbf{v} \rangle \leq D(\mathbf{x}; \mathbf{v}), \forall\, \mathbf{v} \in \mathbb{R}^n \Big\}.$$

Definition 2 (Clarke stationary point) For a locally Lipschitz function f, a point $\mathbf{x} \in R^n$ is called a Clarke stationary point of f if $0 \in \partial f(\mathbf{x})$.

Lemma 1 *The gradient of $\overline{r}$ is Lipschitz continuous.*

Proof From equation 6 in the paper, it follows that

$$\nabla_{(\mathbf{x})} \overline{r} = \sum_{(i,j)} 2 \exp\Big(- \frac{\|\mathbf{g}_i(\mathbf{x}) - \mathbf{g}_j(\mathbf{x})\|^2}{\sigma^2} \Big) (1 - \frac{\|\mathbf{g}_i(\mathbf{x}) - \mathbf{g}_j(\mathbf{x})\|^2}{\sigma^2}) (\nabla \mathbf{g}_i(\mathbf{x}) - \nabla \mathbf{g}_j(\mathbf{x}))^\top (\mathbf{g}_i(\mathbf{x}) - \mathbf{g}_j(\mathbf{x})).$$

We only need to show one of the terms under the sum is Lipschitz, i.e. we need $f(\mathbf{u}) = \exp\Big(- \frac{\|\mathbf{u}\|^2}{\sigma^2} \Big)(1 - \frac{\|\mathbf{u}\|^2}{\sigma^2})(\nabla \mathbf{u})^\top(\mathbf{u})$ to be Lipschitz, where $\mathbf{u} = \mathbf{g}_i(\mathbf{x}) - \mathbf{g}_j(\mathbf{x}) \in \mathbb{R}^d$. We show $f(\mathbf{u})$ is Lipschitz by showing its derivative is bounded:

$$\begin{aligned}
\|\nabla f(\mathbf{u})\| \leqslant & \|\exp\Big(- \frac{\|\mathbf{u}\|^2}{\sigma^2} \Big)\Big(- \frac{2}{\sigma^2} (\nabla \mathbf{u})^\top(\mathbf{u}) \Big)\| \, (1 - \frac{\|\mathbf{u}\|^2}{\sigma^2}) \, \|(\nabla \mathbf{u})^\top\| \|\mathbf{u}\| \\
& + \exp\Big(- \frac{\|\mathbf{u}\|^2}{\sigma^2} \Big) \, \|\Big(- \frac{2}{\sigma^2} (\nabla \mathbf{u})^\top(\mathbf{u}) \Big)\| \, \|(\nabla \mathbf{u})^\top\| \|\mathbf{u}\| \\
& + |\exp\Big(- \frac{\|\mathbf{u}\|^2}{\sigma^2} \Big) \, (1 - \frac{\|\mathbf{u}\|^2}{\sigma^2})| \, \|(\nabla^2 \mathbf{u})\| \|\mathbf{u}\| \\
& + |\exp\Big(- \frac{\|\mathbf{u}\|^2}{\sigma^2} \Big)(1 - \frac{\|\mathbf{u}\|^2}{\sigma^2})| \|\nabla \mathbf{u}\|^2.
\end{aligned}$$

Note that $\mathbf{u} = \mathbf{g}_i(\mathbf{x}) - \mathbf{g}_j(\mathbf{x})$ is Lipschitz in $\mathbf{x}$ since $\sup_{\mathbf{x} \in \mathcal{X}} \|\nabla \mathbf{g}(\mathbf{x})\| \leq M$, therefore $\|\nabla \mathbf{u}\| \leqslant M$. Also since $\nabla \mathbf{g}$ is L_g Lipschitz, we get $\|\nabla^2 \mathbf{u}\| \leqslant 2L_g$. Also, a polynomial in $\|\mathbf{u}\|$ times $\exp\Big(- \frac{\|\mathbf{u}\|^2}{\sigma^2} \Big)$ is bounded so we get:

$$\|\nabla f(\mathbf{u})\| \leqslant (4\frac{M^2}{\sigma^2} + L_g + M^2)C. \tag{19}$$

where C is some constant that bounds all the instances of polynomial in $\|\mathbf{u}\|$ times $\exp\Big(- \frac{\|\mathbf{u}\|^2}{\sigma^2} \Big)$ occurring above. Therefore $\nabla_{(\mathbf{x})} \overline{r}$ is Lipschitz with constant $L_r = 2m^2(4\frac{M^2}{\sigma^2} + L_g + M^2)C$. □

The following Lemma 2 considers the case where ε is a positive constant, which corresponds to an iterative scheme that only executes Lines 3–14 of Algorithm 1.

Lemma 2 *Let $\varepsilon, c, \iota, \tau > 0$, $0 < \rho < 1$ and arbitrary initial $\mathbf{x}_0 \in \mathbb{R}^n$. Suppose $\{\mathbf{x}_k\}$ is the sequence generated by repeating Lines 3–14 of Algorithm 1 with fixed $\varepsilon_k = \varepsilon$, $\alpha_k \geq \bar{\alpha}$, where $\bar{\alpha} > 0$ is a constant in user's choice and $\phi^* := \min_{\mathbf{x} \in \mathbb{R}^n} \phi(\mathbf{x})$. Then $\|\nabla \phi_\varepsilon(\mathbf{x}_k)\| \to 0$ as $k \to \infty$.*

Proof In each iteration, we compute $\mathbf{u}_{k+1}$ by $\mathbf{z}_{k+1} - \tau_k \nabla r_\varepsilon(\mathbf{z}_{k+1})$, and put $\mathbf{x}_{k+1} = \mathbf{u}_{k+1}$ only if the condition $\{\|\phi_\varepsilon(\mathbf{x}_k)\| \leq c\|\mathbf{u}_{k+1} - \mathbf{x}_k\|$ *and* $\phi_\varepsilon(\mathbf{u}_{k+1}) - \phi_\varepsilon(\mathbf{x}_k) \leq -\frac{\iota}{2}\|\mathbf{u}_{k+1} - \mathbf{x}_k\|^2\}$ is satisfied, from which it is easy to get

$$\|\nabla \phi_\varepsilon(\mathbf{x}_k)\|^2 \leq \frac{2c^2}{\iota}(\phi_\varepsilon(\mathbf{x}_k) - \phi_\varepsilon(\mathbf{u}_{k+1})). \tag{20}$$

If $\mathbf{u}_{k+1}$ fails to satisfy the above condition, we compute $\mathbf{v}_{k+1}$ through line search until the criteria

$$\phi_{\varepsilon_k}(\mathbf{v}_{k+1}) - \phi_{\varepsilon_k}(\mathbf{x}_k) \leq -\tau \|\mathbf{v}_{k+1} - \mathbf{x}_k\|^2 \tag{21}$$

is satisfied and then set $\mathbf{x}_{k+1} = \mathbf{v}_{k+1}$.

We will now demonstrate that the line search requires only a finite number of steps. To achieve this, we reformulate the scheme for computing $\mathbf{v}_{k+1}$ with line search. Let ℓ_k denote the number of line search steps required to satisfy the condition in Eq. (21) with the initial step size α_k, then

$$v_{k+1} = x_k - \alpha_k \rho^{\ell_k} \nabla \phi_\epsilon(x_k). \tag{22}$$

From Lemma 3.2 in [16] we have that the gradient ∇r_ε is Lipschitz continuous with constant $\sqrt{m}L_g + \frac{M^2}{\varepsilon}$. And $\nabla \bar{r}$ defiend in (7) is also Lipschitz continuous with constant L_r. Furthermore, in (A1) we assumed ∇f is L_f-Lipschitz continuous. Hence putting $L_\varepsilon = L_f + \sqrt{m}L_g + \frac{M^2}{\varepsilon} + L_r$, we get that $\nabla \phi_\varepsilon$ is L_ε-Lipschitz continuous, which implies

$$\phi_\varepsilon(\mathbf{v}_{k+1}) \leq \phi_\varepsilon(\mathbf{x}_k) + \langle \nabla \phi_\varepsilon(\mathbf{x}_k), \mathbf{v}_{k+1} - \mathbf{x}_k \rangle + \frac{L_\varepsilon}{2} \|\mathbf{v}_{k+1} - \mathbf{x}_k\|^2. \tag{23}$$

Then, the combination of (22) and (23) gives

$$\phi_\varepsilon(\mathbf{v}_{k+1}) - \phi_\varepsilon(\mathbf{x}_k) \leq -\left(\frac{1}{\rho^{\ell_k}\alpha_k} - \frac{L_\varepsilon}{2} \right) \|\mathbf{v}_{k+1} - \mathbf{x}_k\|^2. \tag{24}$$

Hence, for any $k = 1, 2, \ldots$, if $\frac{1}{\rho^{\ell_k}\alpha_k} - \frac{L_\varepsilon}{2} \geq \tau$, the (21) is met. This shows that there is a finite upper bound of the maximum search steps ℓ_{max} required for having (21) satisfy $\rho^{\ell_{max}}\alpha_k \leqslant \frac{1}{\tau + L_\epsilon/2}$. From the discussion above, we can have

$$\phi_\varepsilon(\mathbf{v}_{k+1}) - \phi_\varepsilon(\mathbf{x}_k) \leq -\tau \|\mathbf{v}_{k+1} - \mathbf{x}_k\|^2 = -\tau(\rho^{\ell_k}\alpha_k)^2 \|\nabla \phi_\varepsilon(\mathbf{x}_k)\|^2 \leq -\tau(\rho^{\ell_{max}}\bar{\alpha})^2 \|\nabla \phi_\varepsilon(\mathbf{x}_k)\|^2, \tag{25}$$

here the first inequality is from (21), the equality is from (22), and the last inequality uses the fact that $\ell_k \leq \ell_{max}$ and $0 < \rho < 1$ and $\alpha_k \geq \bar{\alpha}$. Rearranging this inequality, we get

$$\|\nabla \phi_\varepsilon(\mathbf{x}_k)\|^2 \leq \frac{1}{\tau(\rho^{\ell_{max}}\bar{\alpha})^2} (\phi_\varepsilon(\mathbf{x}_k) - \phi_\varepsilon(\mathbf{v}_{k+1})). \tag{26}$$

Hence, in either case $\mathbf{x}_{k+1} = \mathbf{u}_{k+1}$ or $\mathbf{x}_{k+1} = \mathbf{v}_{k+1}$, from (20) and (26) we have

$$\|\nabla \phi_\varepsilon(\mathbf{x}_k)\|^2 \leq C(\phi_\varepsilon(\mathbf{x}_k) - \phi_\varepsilon(\mathbf{x}_{k+1})), \tag{27}$$

where $C = \max\{\frac{1}{\tau(\rho^{\ell_{max}}\bar{\alpha})^2}, \frac{2c^2}{\iota}\}$, which is independent of k.

Summing up (27) for $k = 0, \ldots, K$, we have

$$\sum_{k=0}^{K} \|\nabla \phi_\varepsilon(\mathbf{x}_k)\|^2 \leq C(\phi_\varepsilon(\mathbf{x}_0) - \phi_\varepsilon(\mathbf{x}_{K+1})). \tag{28}$$

Combining with the fact $\phi_\varepsilon(\mathbf{x}) \geq \phi(\mathbf{x}) - \frac{m\varepsilon}{2} \geq \phi^* - \frac{m\varepsilon}{2}$, we have

$$\sum_{k=0}^{K} \|\nabla \phi_\varepsilon(\mathbf{x}_k)\|^2 \leq C(\phi_\varepsilon(\mathbf{x}_0) - \phi^* + \frac{m\varepsilon}{2}). \tag{29}$$

The right hand side is finite, and thus by letting $K \to \infty$ we conclude $\|\nabla \phi_\varepsilon(\mathbf{x}_k)\| \to 0$, which proves the lemma. □

Next we consider the case where ε varies in Theorem 1. More precisely, we focus on the subsequence $\{\mathbf{x}_{k_l+1}\}$ which selects the iterates when the reduction criterion in Line 15 is satisfied for $k = k_l$ and ε_k is reduced.

Lemma 3 *Suppose that the sequence $\{\mathbf{x}_k\}$ is generated by Algorithm 1 and any initial $\mathbf{x}_0$. Then for any $k \geq 0$ we have*

$$\phi_{\varepsilon_{k+1}}(\mathbf{x}_{k+1}) + \frac{m\varepsilon_{k+1}}{2} \leq \phi_{\varepsilon_k}(\mathbf{x}_{k+1}) + \frac{m\varepsilon_k}{2} \leq \phi_{\varepsilon_k}(\mathbf{x}_k) + \frac{m\varepsilon_k}{2}. \tag{30}$$

Proof The proof of this lemma is similar to Lemma 3.4 of [16]. The second inequality is immediate from (27). So we focus on the first inequality. For any $\varepsilon > 0$ and $\mathbf{x}$, denote

$$r_{\varepsilon,i}(\mathbf{x}) := \begin{cases} \frac{1}{2\varepsilon}\|\mathbf{g}_i(\mathbf{x})\|^2, & \text{if } \|\mathbf{g}_i(\mathbf{x})\| \leq \varepsilon, \\ \|\mathbf{g}_i(\mathbf{x})\| - \frac{\varepsilon}{2}, & \text{if } \|\mathbf{g}_i(\mathbf{x})\| > \varepsilon. \end{cases} \tag{31}$$

Since $\phi_\varepsilon(\mathbf{x}) = \sum_{i=1}^m r_{\varepsilon,i}(\mathbf{x}) + \lambda \bar{r}(\mathbf{x}) + f(\mathbf{x})$, to prove the first inequality it suffices to show that

$$r_{\varepsilon_{k+1},i}(\mathbf{x}_{k+1}) + \frac{\varepsilon_{k+1}}{2} \leq r_{\varepsilon_k,i}(\mathbf{x}_{k+1}) + \frac{\varepsilon_k}{2}. \tag{32}$$

If $\varepsilon_{k+1} = \varepsilon_k$, then the two quantities above are identical and the first inequality holds. Now suppose $\varepsilon_{k+1} = \gamma\varepsilon_k < \varepsilon_k$. We then consider the relation between $\|\mathbf{g}_i(\mathbf{x}_{k+1})\|$, ε_{k+1} and ε_k in three cases:

1. If $\|\mathbf{g}_i(\mathbf{x}_{k+1})\| > \varepsilon_k > \varepsilon_{k+1}$, then by the definition in (31), there is
$$r_{\varepsilon_{k+1},i}(\mathbf{x}_{k+1}) + \frac{\varepsilon_{k+1}}{2} = \|\mathbf{g}_i(\mathbf{x}_{k+1})\| = r_{\varepsilon_k,i}(\mathbf{x}_{k+1}) + \frac{\varepsilon_k}{2}.$$
2. If $\varepsilon_k \geq \|\mathbf{g}_i(\mathbf{x}_{k+1})\| > \varepsilon_{k+1}$, then (31) implies
$$\begin{aligned} r_{\varepsilon_{k+1},i}(\mathbf{x}_{k+1}) + \frac{\varepsilon_{k+1}}{2} = \|\mathbf{g}_i(\mathbf{x}_{k+1})\| &= \frac{\|\mathbf{g}_i(\mathbf{x}_{k+1})\|}{2} + \frac{\|\mathbf{g}_i(\mathbf{x}_{k+1})\|}{2} = \frac{\|\mathbf{g}_i(\mathbf{x}_{k+1})\|^2}{2\|\mathbf{g}_i(\mathbf{x}_{k+1})\|} + \frac{\|\mathbf{g}_i(\mathbf{x}_{k+1})\|}{2} \\ &\leqslant \frac{\|\mathbf{g}_i(\mathbf{x}_{k+1})\|^2}{2\varepsilon_k} + \frac{\varepsilon_k}{2} = r_{\varepsilon_k,i}(\mathbf{x}_{k+1}) + \frac{\varepsilon_k}{2}. \end{aligned}$$
The second lines follows from the fact that $\frac{\|\mathbf{g}_i(\mathbf{x}_{k+1})\|^2}{2\varepsilon} + \frac{\varepsilon}{2}$—as a function of ε—is non-decreasing for all $\varepsilon \geq \|\mathbf{g}_i(\mathbf{x}_{k+1})\|$
3. If $\varepsilon_k > \varepsilon_{k+1} \geq \|\mathbf{g}_i(\mathbf{x}_{k+1})\|$, then again the previous fact and (31) imply (32).

Therefore, in either of the three cases, (32) holds and hence

$$r_{\varepsilon_{k+1}}(\mathbf{x}_{k+1}) + \frac{m\varepsilon_{k+1}}{2} = \sum_{i=1}^m \left(r_{\varepsilon_{k+1},i}(\mathbf{x}_{k+1}) + \frac{\varepsilon_{k+1}}{2}\right) \leq \sum_{i=1}^m \left(r_{\varepsilon_k,i}(\mathbf{x}_{k+1}) + \frac{\varepsilon_k}{2}\right) = r_{\varepsilon_k}(\mathbf{x}_{k+1}) + \frac{m\varepsilon_k}{2},$$

which implies the first inequality of (30). □

Theorem 1 *Suppose that $\{\mathbf{x}_k\}$ is the sequence generated by Algorithm 1 with any initial $\mathbf{x}_0$ and $\alpha_k \geq \bar{\alpha}$ as in Lemma 2. Let $\{\mathbf{x}_{k_l+1}\}$ be the subsequence where the reduction criterion $\varepsilon_{k+1} = \gamma\varepsilon_k$ in line 15 is met for $k = k_l$ and $l = 1, 2, \ldots$. Then, $\{\mathbf{x}_{k_l+1}\}$ has at least one accumulation point, and every accumulation point of $\{\mathbf{x}_{k_l+1}\}$ is a Clarke stationary point of (1).*

Proof By the Lemma 4.2 and the definition of Clarke subdifferential, this theorem can be easily proved in the similar way to Theorem 3.6 in [16].

Due to Lemma 3 and $\phi(\mathbf{x}) \le \phi_\varepsilon(\mathbf{x}) + \frac{m\varepsilon}{2}$ for all $\varepsilon > 0$ and $\mathbf{x} \in \mathcal{X}$, we know that

$$\phi(\mathbf{x}_k) \le \phi_{\varepsilon_k}(\mathbf{x}_k) + \frac{m\varepsilon_k}{2} \le \cdots \le \phi_{\varepsilon_0}(\mathbf{x}_0) + \frac{m\varepsilon_0}{2} < \infty.$$

Since ϕ is coercive, we know that $\{\mathbf{x}_k\}$ is bounded. Hence $\{\mathbf{x}_{k_l+1}\}$ is also bounded and has at least one accumulation point.

Note that $\mathbf{x}_{k_l+1}$ satisfies the reduction criterion in Line 15 of Algorithm 1, i.e., $\|\nabla\phi_{\varepsilon_{k_l}}(\mathbf{x}_{k_l+1})\| \le \sigma\gamma\varepsilon_{k_l} = \sigma\varepsilon_0\gamma^{l+1} \to 0$ as $l \to \infty$. For notation simplicity, let $\{\mathbf{x}_{j+1}\}$ denote any convergent subsequence of $\{\mathbf{x}_{k_l+1}\}$ and ε_j the corresponding ε_k used in the iteration to generate $\mathbf{x}_{j+1}$. Then there exists $\hat{\mathbf{x}} \in \mathcal{X}$ such that $\mathbf{x}_{j+1} \to \hat{\mathbf{x}}$, $\varepsilon_j \to 0$, and $\nabla\phi_{\varepsilon_j}(\mathbf{x}_{j+1}) \to 0$ as $j \to \infty$.

Note that the Clarke subdifferential of ϕ at $\hat{\mathbf{x}}$ is given by $\partial\phi(\mathbf{x}) = \partial\hat{r}(\mathbf{x}) + \lambda\nabla\overline{r}(\mathbf{x}) + \nabla f(\mathbf{x})$:

$$\begin{aligned}\partial\phi(\hat{\mathbf{x}}) = \Bigg\{ &\sum_{i\in I_0} \nabla\mathbf{g}_i(\hat{\mathbf{x}})^\top \mathbf{w}_i + \sum_{i\in I_1} \nabla\mathbf{g}_i(\hat{\mathbf{x}})^\top \frac{\mathbf{g}_i(\hat{\mathbf{x}})}{\|\mathbf{g}_i(\hat{\mathbf{x}})\|} \\ &+\lambda\nabla\overline{r}(\hat{\mathbf{x}}) + \nabla f(\hat{\mathbf{x}}) \;\Big|\; \|\Pi(\mathbf{w}_i; \mathcal{C}(\nabla\mathbf{g}_i(\hat{\mathbf{x}}))) \le 1,\ \forall\, i \in I_0\},\end{aligned} \tag{33}$$

where $I_0 = \{i \in [m] \mid \|\mathbf{g}_i(\hat{\mathbf{x}})\| = 0\}$ and $I_1 = [m] \setminus I_0$. Then we know that there exists J sufficiently large, such that

$$\varepsilon_j < \frac{1}{2}\min\{\|\mathbf{g}_i(\hat{\mathbf{x}})\| \mid i \in I_1\} \le \frac{1}{2}\|\mathbf{g}_i(\hat{\mathbf{x}})\| \le \|\mathbf{g}_i(\mathbf{x}_{j+1})\|, \quad \forall\, j \ge J, \quad \forall\, i \in I_1,$$

where we used the facts that $\min\{\|\mathbf{g}_i(\hat{\mathbf{x}})\| \mid i \in I_1\} > 0$ and $\varepsilon_j \to 0$ in the first inequality, and $\mathbf{x}_{j+1} \to \hat{\mathbf{x}}$ and the continuity of $\mathbf{g}_i$ for all i in the last inequality. Furthermore, we denote

$$\mathbf{s}_{j,i} := \begin{cases} \frac{\mathbf{g}_i(\mathbf{x}_{j+1})}{\varepsilon_j}, & \text{if } \|\mathbf{g}_i(\mathbf{x}_{j+1})\| \le \varepsilon_j, \\ \frac{\mathbf{g}_i(\mathbf{x}_{j+1})}{\|\mathbf{g}_i(\mathbf{x}_{j+1})\|}, & \text{if } \|\mathbf{g}_i(\mathbf{x}_{j+1})\| > \varepsilon_j. \end{cases}$$

Then we have

$$\nabla\phi_{\varepsilon_j}(\mathbf{x}_{j+1}) = \sum_{i\in I_0} \nabla\mathbf{g}_i(\mathbf{x}_{j+1})^\top \mathbf{s}_{j,i} + \sum_{i\in I_1} \nabla\mathbf{g}_i(\mathbf{x}_{j+1})^\top \frac{\mathbf{g}_i(\mathbf{x}_{j+1})}{\|\mathbf{g}_i(\mathbf{x}_{j+1})\|} + \lambda\nabla\overline{r}(\mathbf{x}_{j+1}) + \nabla f(\mathbf{x}_{j+1}). \tag{34}$$

Comparing (33) and (34), we can see that the last two terms on the right hand side of (34) converge to those of (33), respectively, due to the facts that $\mathbf{x}_{j+1} \to \hat{\mathbf{x}}$ and the the continuity of $\mathbf{g}_i, \nabla\mathbf{g}_i, +\nabla\overline{r}, \nabla f$. Moreover, noting that $\|\Pi(\mathbf{s}_{j,i}; \mathcal{C}(\nabla\mathbf{g}_i(\hat{\mathbf{x}})))\| \le \|\mathbf{s}_{j,i}\| \le 1$, we can see that the first term on the right hand side of (34) also converges to the set formed by the first term of (33) due to the continuity of $\mathbf{g}_i$ and $\nabla\mathbf{g}_i$. Hence we know that

$$\operatorname{dist}(\nabla\phi_{\varepsilon_j}(\mathbf{x}_{j+1}), \partial\phi(\hat{\mathbf{x}})) \to 0,$$

as $j \to 0$. Since $\nabla\phi_{\varepsilon_j}(\mathbf{x}_{j+1}) \to 0$ and $\partial\phi(\hat{\mathbf{x}})$ is closed, we conclude that $0 \in \partial\phi(\hat{\mathbf{x}})$. □

From Theorem 1, we conclude that the output of our network converges to a (local) minimizer of the original regularized problem (1).

5 Experiments and Results

Here we present our experiments on LDCT image reconstruction problems with various dose levels and compare with existing state-of-the-art algorithms in terms of image quality, run time and the number of parameters etc. We adopt a warm start training strategy which imitates the iterating of optimization algorithm. More precisely, first we train the network with $K = 3$ phases, where each phase in the network corresponds to an iteration in optimization algorithm. After it converges, we add 2 more phases and we continue training the 5-phase network until it converges. We continue adding 2 more phases until there is no noticeable improvement.

As computing the similarity weight matrix $\mathcal{W}$ is high-cost in time and memory, we also experiment with approximation of $\mathcal{W}$ computed on the initial reconstruction $\mathbf{x}_0$ without updating in each iteration, i.e. $\mathcal{W}_{ij} \approx \exp\big(-\frac{\|\hat{\mathbf{g}}_i(\mathbf{x}_0)-\hat{\mathbf{g}}_j(\mathbf{x}_0)\|^2}{\delta^2}\big)$. Thus $\bar{r}(\mathbf{x})$ can be differentiated as $\nabla\bar{r}(\mathbf{x}) = 2\cdot\sum_{q=1}^{d}\nabla\hat{\mathbf{g}}^q(\mathbf{x})^\top\mathcal{L}\,\hat{\mathbf{g}}^q(\mathbf{x})$. In the approximation scenario we compute $\mathcal{L}$ once and use it for all the phases, but in the other case we have an extra $\mathcal{O}(m^2)$ computation in each phase, every time we compute $\nabla\phi$. As shown in the Sect. 5.2, this approximation does not exacerbate the network performance much but can increase the running speed.

All the experiments are performed on a computer with Intel i7-6700K CPU at 3.40 GHz, 16 GB of memory, and a Nvidia GTX-1080Ti GPU of 11GB graphics card memory, and implemented with the PyTorch toolbox [76] in Python. The initial $\mathbf{x}_0$ is obtained by FBP algorithm. The spatial kernel size of the convolution and transposed convolution is set to be 3×3 and the channel number is set to 48 with layer number $l = 4$ as default. The learned weights of convolutions and transposed convolutions are initialized by Xavier Initializer [30] and the starting ε_0 is initialized to be 0.001. All the learnable parameters are trained by the Adam Optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The network is trained with learning rate 1e-4 for 200 epochs when phase number $K = 3$, followed by 100 epochs when adding more phases.

We test the performance of ELDA on the "*2016 NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge*" [72] which contains 5936 full-dose CT (FDCT) data from 10 patients, from which we randomly select 500 images and resize them to the size 256×256. Then we randomly divide the dataset into 400 images for training and 100 for testing. Distance-driven algorithm [24, 25] is applied to simulate the projections in fan-beam geometry. The source-to-rotation center and detector-to-rotation center distances are both set to 250 mm. The physical image region covers 170 mm $\times$ 170 mm. On detector there are 512 detector elements each with width 0.72 mm. There are 1024 projection views in total with the projection angles are evenly distributed over a full scan range. Similar to [59], the simulated noisy transmission measurement I is generated by adding Poisson and electronic noise as

$$I = Possion(I_0 \exp(-\hat{b})) + Normal(0, \sigma_e^2), \tag{35}$$

where I_0 is the incident X-ray intensity and σ_e^2 is the variance the background electronic noise. And $\hat{b}$ represents the noise-free projection. In this simulation, I_0 is set to 1.0×10^6 for normal dose and σ_e^2 is prefixed to be 10 for all dose cases. Then the noisy projection $\mathbf{b}$ is calculated by taking the logarithm transformation on $\frac{I_0}{I}$. In low dose case, in order to investigate the robustness of all compared algorithms, we generated three sets of different low dose projections with $I_0 = 1.0\times10^5$, 5.0×10^4 and 2.5×10^4 which correspond to 10%, 5% and 2.5% of the full dose incident accordingly. We use peak signal to noise ratio (PSNR) and structureal similarity index measure (SSIM) to evaluate the quality of the reconstructed images.

Table 1 The reconstruction results associated with different depths of convolution kernels and different number of convolutions in each phase with dose level 10%

Depth of conv. kernels	16	32	48	64
PSNR (dB)	47.11	47.51	47.73	47.75
Number of parameters	14,152	55,912	125,320	222,376
Average testing time (s)	1.139	1.231	1.411	1.539
Number of convolutions	**2**	**3**	**4**	**5**
PSNR (dB)	47.27	47.60	47.73	47.77
Number of parameters	42,376	83,848	125,320	167,656
Average testing time (s)	1.117	1.258	1.411	1.530

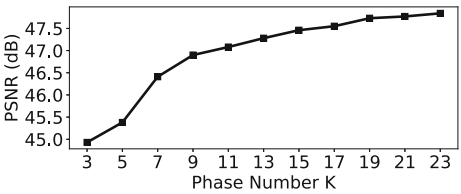

Fig. 2 The reconstruction PSNR of ELDA across various phase numbers on dose level 10%

Table 2 Comparison of different algorithms on the LDCT reconstruction performance with dose level 10%

Algorithms	Plain-GD	AGD	LDA	ELDA
PSNR (dB)	45.47	45.95	46.29	46.35
Average testing time (s)	0.602	0.605	1.338	1.239

5.1 Parameter Study

The regularization term of our model is learned from training samples, yet there are still a few key network hyperparameters need to be set manually. Specifically, we investigate the impacts of some parameters of the architecture, which includes the number of convolutions (l), the depth of the convolution kernels (d) and the phase number (K). The influence of each hyperparameter is sensed by perturbing it with others fixed at $d = 48, l = 4$ and $K = 19$. The setting includes all factors/components listed in Table 3 and all following results are trained and tested with dose level 10%.

5.1.1 Depth of the Convolution Kernels

We evaluate the instances of $d = 16, 32, 48$ and 64. The results are listed in Table 1.

It is evident that the PSNR score raises with growing depth of the kernels, but the profit gradually reduces. On the contrary, the number of parameters and running time grow greatly in the meantime.

5.1.2 Number of Convolutions

We evaluate the cases of different number of convolutions $l = 2, 3, 4$ and 5. The corresponding results are reported in Table 1. We can observe that more convolutions contribute

Table 3 Comparison of the influence of various design factors or components on the LDCT reconstruction performance with dose level 10%

ELDA Factors/Components				
Inexact Transpose?	✘	✔	✔	✔
Non-local regularizer?	✘	✘	✔	✔
Approximated Matrix $\mathcal{W}$?	✘	✘	✘	✔
PSNR (dB)	46.35	46.74	47.75	47.73
Average testing time (s)	1.239	1.247	3.703	1.411

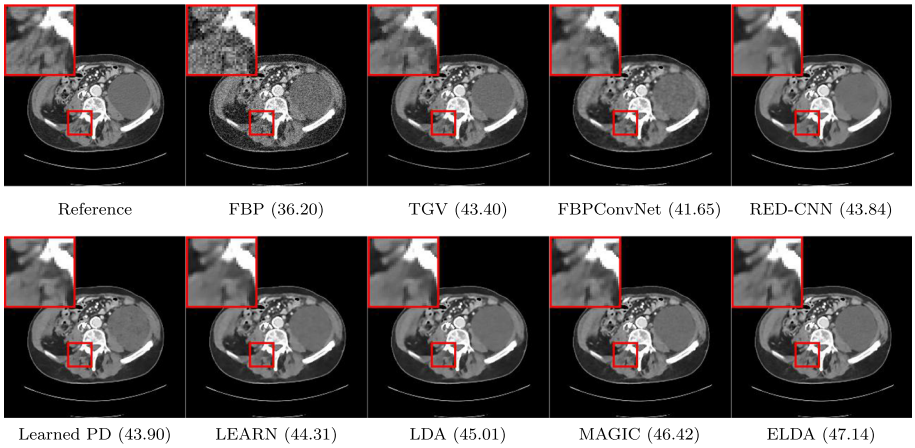

Fig. 3 Representative CT images of AAPM-Mayo data reconstructed by various methods with dose level 5%. The display window is [-160, 240] HU. PSNRs (dB) are shown in the parentheses. The regions of interest are magnified in red boxes for better visualization

to better reconstructed image quality. But the increase of PSNR score is insignificant from 4 convolutions to 5 while the parameter number and the test time rise significantly.

5.1.3 Phase Number K

As shown in Fig. 2, the PSNR increases with the phase number K. And the plot approaches nearly flat after 19 phases.

To balance the trade-off between reconstruction performance and network complexity, we take $d = 48, l = 4, K = 19$ when comparing with other methods.

5.2 Ablation Study

In this section, we first investigate the effectiveness of the proposed algorithm in ELDA. To this end, we compare ELDA with unrolling the standard gradient descent iteration of (1), and an accelerated inertial version by setting $\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \nabla \phi(\mathbf{x}_k) + \theta_k(\mathbf{x}_k - \mathbf{x}_{k-1})$ where θ_k is also learned. Here these two algorithms are named as Plain-GD and AGD, respectively. In addition, to show the superiority of the new descending condition (14)&(17) over the competition strategy in LDA [16], we compare the result with LDA here as well. The comparison of different algorithms are shown in Table 2, where the experiments follow the default parameter configuration as Sect. 5.1 without the additional components listed in

Table 4 Quantitative results (Mean ± Standard Deviation) of the LDCT reconstructions of AAPM-Mayo data obtained by various algorithms and different dose levels. Average run time (per image) is measured in second

Dose Level	1.0×10^5		5.0×10^4		2.5×10^4		Run Time	Parameters
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM		
FBP [45]	38.03±0.68	0.9084±0.0128	35.25±0.74	0.8459±0.0209	32.35±0.77	0.7556±0.0295	1.0	N/A
TGV [74]	43.60±0.50	0.9845±0.0020	42.80±0.52	0.9812±0.0027	41.10±0.62	0.9698±0.0055	$9.4 \cdot 10^2$	N/A
FBPConvNet [43]	42.49±0.47	0.9764±0.0029	41.14±0.49	0.9698±0.0041	39.62±0.48	0.9589±0.0052	$7.4 \cdot 10^{-2}$	$1.0 \cdot 10^7$
RED-CNN [14]	44.58±0.58	0.9848±0.0026	43.25±0.60	0.9808±0.0033	41.61±0.63	0.9737±0.0048	$\mathbf{2.6 \cdot 10^{-3}}$	$1.8 \cdot 10^6$
Learned PD [1]	44.81±0.59	0.9863±0.0025	43.28±0.61	0.9813±0.0034	41.74±0.62	0.9745±0.0048	1.8	$2.5 \cdot 10^5$
LEARN [15]	45.19±0.60	0.9868±0.0023	43.86±0.61	0.9840±0.0027	42.06±0.61	0.9776±0.0037	4.8	$1.9 \cdot 10^6$
LDA [16]	46.29±0.60	0.9896±0.0019	44.64±0.61	0.9858±0.0025	42.97±0.64	0.9802±0.0035	1.3	$\mathbf{6.2 \cdot 10^4}$
MAGIC [103]	47.54±0.66	0.9919±0.0017	45.83±0.64	0.9888±0.0022	44.37±0.63	0.9853±0.0027	5.3	$2.1 \cdot 10^6$
ELDA (Ours)	**47.73±0.69**	**0.9923±0.0017**	**46.40±0.64**	**0.9899±0.0021**	**44.86±0.65**	**0.9859±0.0031**	1.4	$1.2 \cdot 10^5$

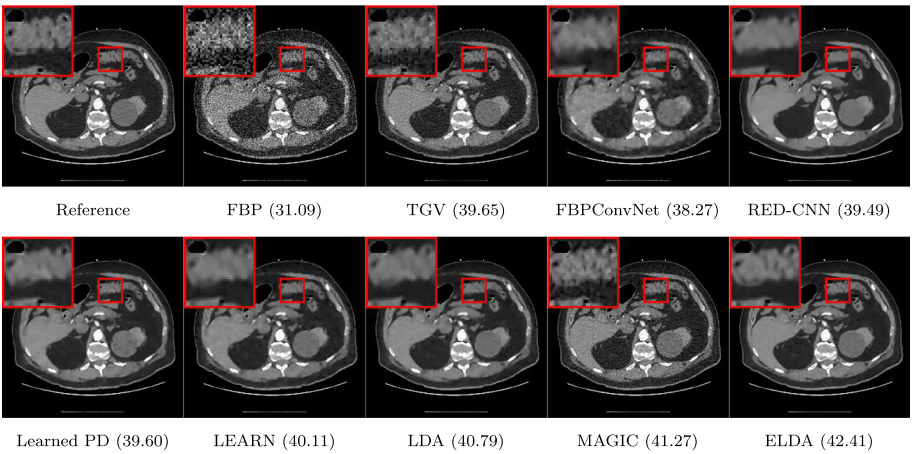

Fig. 4 Representative CT images of NBIA data reconstructed by various methods with dose level 2.5%. The display window is [-160, 240] HU. PSNRs (dB) are shown in the parentheses. The regions of interest are magnified in red boxes for better visualization

Table 3. It is quite obvious that all AGD, LDA and ELDA achieve higher PNSRs than Plain-GD, where ELDA achieves the best. With the new descending condition it achieves average 0.06 dB PSNR better than LDA and about 0.1 s faster. Furthermore, we have also computed the ratios that the candidate $\mathbf{u}_{k+1}$ is taken instead of $\mathbf{v}_{k+1}$, and found it to be 86.2% for LDA and 99.6% for ELDA respectively. That indicates that the proposed descending condition (14) can effectively avoid the frequent candidate alternating compared to the competition strategy used in LDA.

Moreover, we check the influence of some essential factors/components of our ELDA model, i.e. the inexact transpose, the nonlocal smoothing regularizer and the approximated weight matrix $\mathcal{W}$. The results are summarized in Table 3. It is remarkable that the inexact transpose and the nonlocal smoothing regularizer can effectively increase the network performance by a large margin. And with the approximated weight matrix $\mathcal{W}$ there is no significant decreasing of the PSNR. As the initial $\mathbf{x}_0$ obtained by FBP is not far from $\mathbf{x}_k$ in each iteration, the $\mathcal{W}$ approximated by $\mathbf{x}_0$ can provide a good estimation to the true one. Thus, in the following sections we will keep all these features in Table 3 when comparing ELDA with other methods.

5.3 Comparison with the State-of-the-Art

In this section, we benchmark the ELDA against several state-of-the-art methods using two widely recognized datasets: the AAPM-Mayo and the National Biomedical Imaging Archive (NBIA).

5.3.1 AAPM-Mayo

In this section, we compare our reconstruction results on the 100 AAPM-Mayo testing images with several existing algorithms: two classic reconstruction methods, i.e., FBP [45] and TGV [74] and six approaches based on deep learning, i.e., FBPConvNet [43], RED-CNN [14], Learned Primal-Dual [1], LDA [16], LEARN [15] and MAGIC [103]. For fair comparison,

Table 5 Quantitative Results (Mean ± Standard Deviation) of the LDCT Reconstructions Obtained by Various Algorithms and Different Dose Levels on NBIA Data

Dose Level	1.0×10^5		5.0×10^4		2.5×10^4	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
FBP	37.36±0.43	0.9125±0.0145	34.61±0.51	0.8533±0.0240	31.74±0.55	0.7682±0.0339
FBPConvNet	40.86±0.25	0.9693±0.0026	39.27±0.29	0.9587±0.0033	37.69±0.38	0.9449±0.0036
RED-CNN	41.74±0.32	0.9751±0.0022	40.32±0.38	0.9683±0.0026	38.76±0.46	0.9578±0.0029
Learned PD	41.86±0.31	0.9760±0.0025	40.39±0.39	0.9687±0.0031	38.89±0.43	0.9594±0.0033
LEARN	42.44±0.35	0.9780±0.0021	41.06±0.42	0.9738±0.0021	39.33±0.48	0.9638±0.0024
TGV	42.53±0.52	0.9818±0.0017	41.44±0.32	0.9762±0.0032	39.60±0.22	0.9601±0.0080
LDA	43.37±0.38	0.9829±0.0017	41.64±0.47	0.9763±0.0023	39.99±0.55	0.9678±0.0021
MAGIC	44.58±0.37	0.9866±0.0017	43.10±0.28	0.9821±0.0024	41.12±0.20	0.9707±0.0058
ELDA (Proposed)	**44.65±0.53**	**0.9867±0.0016**	**43.33±0.45**	**0.9826±0.0017**	**41.61±0.55**	**0.9763±0.0017**

Table 6 Comparison of the LDCT reconstruction performance on different Gaussian noise levels

Noise Level σ_e^2	5	10	20	30	40
PSNR (dB)	41.611	41.606	41.602	41.598	41.595

all deep learning algorithms compared here are trained and evaluated on the same dataset, dose levels and evaluation metrics. The experimental results on various dose levels are summarized in Table 4 and the representative qualitative results on dose level 5% are shown in Fig. 3. These results show that ELDA reconstructs more accurate images using relatively much fewer network parameters and decent running time.

5.3.2 NBIA Data

To demonstrate the generalizability of the proposed method, we validate our model on another dataset NBIA. We randomly sampled 80 images from the NBIA dataset with various parts of the human body for diversity. For fair comparison, all deep learning based reconstruction models compared here are trained on the same dataset identical to Sect. 5.3.1. Figure 4 visualizes the reconstructed images obtained by different methods under dose level 2.5%. It can be seen that ELDA preserves the details well, avoids over-smoothing and reduces artifacts, which gives the promising reconstruction quality in Fig. 4. The quantitative results are provided in Table 5.

5.4 Examination of Robustness Against Noise

To assess the proposed ELDA model's robustness, we conducted tests by introducing Gaussian noise at various levels ($\sigma_e^2 \in \{5; 10; 20; 30; 40\}$ in Eq. (35) with $I_0 = 2.5 \times 10^4$). The model was trained on the 400 images mentioned in Sect. 5, with a single Gaussian noise level of $\sigma_e^2 = 10$. We used the NBIA data mentioned in Sect. 5.3.2 as the test set, applying diverse levels of Gaussian noise. Table 6 presents the results, demonstrating that the LDCT performance remains consistent across various noise levels, highlighting ELDA's robustness.

6 Conclusion

In brief, we propose an efficient inexact learned descent algorithm for low-dose CT reconstruction. With incorporating the sparsity enhancing and non-local smoothing modules in the regularizer, the proposed ELDA outperforms several existing state-of-the-art reconstruction methods in accuracy and efficiency on two widely known datasets and retains convergence property.

Funding Funding was provided by National Science Foundation (Grant No. 2152961) and Division of Mathematical Sciences (Grant Nos. 1818886, 1925263, 2152960, 2152961).

Data Availability Enquiries about data availability should be directed to the authors.

Declarations

Competing interests The authors have not disclosed any competing interests.

References

1. Adler, J., Öktem, O.: Learned primal–dual reconstruction. IEEE Trans. Med. Imaging **37**(6), 1322–1332 (2018)
2. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. **15**(6), 1373–1396 (2003). https://doi.org/10.1162/089976603321780317
3. Bian, W.: Optimization-based deep learning methods for magnetic resonance imaging reconstruction and synthesis (2023)
4. Bian, W., Chen, Y., Ye, X.: An optimal control framework for joint-channel parallel mri reconstruction without coil sensitivities. Magn. Reson. Imaging **89**, 1–11 (2022). https://doi.org/10.1016/j.mri.2022.01.011
5. Bian, W., Chen, Y., Ye, X.: Deep Parallel MRI Reconstruction Network Without Coil Sensitivities, pp. 17–26 (MLMIR@MICCAI 2020). https://doi.org/10.1007/978-3-030-61598-7_2
6. Bian, W., Chen, Y., Ye, X., Zhang, Q.: An optimization-based meta-learning model for mri reconstruction with diverse dataset. J. Imaging **7**(11), 231 (2021)
7. Bian, W., Jang, A., Liu, F.: Diffusion modeling with domain-conditioned prior guidance for accelerated mri and qmri reconstruction (2023)
8. Bian, W., Jang, A., Liu, F.: Multi-task magnetic resonance imaging reconstruction using meta-learning (2024)
9. Bian, W., Zhang, Q., Ye, X., Chen, Y.: A learnable variational model forÂ joint multimodal mri reconstruction andÂ synthesis. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention - MICCAI 2022, pp. 354–364. Springer, Cham (2022)
10. Brenner, D.J., Elliston, C.D., Hall, E.J., Berdon, W.E.: Estimated risks of radiation-induced fatal cancer from pediatric ct. Am. J. Roentgenol. **176**(2), 289–296 (2001)
11. Brody, A.S., Frush, D.P., Huda, W., Brent, R.L., et al.: Radiation risk to children from computed tomography. Pediatrics **120**(3), 677–682 (2007)
12. Chen, H., Huang, W., Ni, Y., Yun, S., Wen, F., Latapie, H., Imani, M.: Taskclip: Extend large vision-language model for task oriented object detection (2024)
13. Chen, H., et al.: Low-dose ct via convolutional neural network. Biomed. Opt. Express **8**(2), 679–694 (2017)
14. Chen, H., et al.: Low-dose ct with a residual encoder-decoder convolutional neural network. IEEE Trans. Med. Imaging **36**(12), 2524–2535 (2017)
15. Chen, H., et al.: Learn: Learned expertsâ€™ assessment-based reconstruction network for sparse-data ct. IEEE Trans. Med. Imaging **37**(6), 1333–1347 (2018)
16. Chen, Y., Liu, H., Ye, X., Zhang, Q.: Learnable descent algorithm for nonsmooth nonconvex image reconstruction. SIAM J. Imag. Sci. **14**(4), 1532–1564 (2021). https://doi.org/10.1137/20M1353368
17. Chen, Y., Ye, X., Zhang, Q.: Variational Model-Based Deep Neural Networks for Image Reconstruction, pp. 1–29. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-03009-4_57-1
18. Chun, I.Y., Fessler, J.A.: Convolutional dictionary learning: acceleration and convergence. IEEE Trans. Image Process. **27**(4), 1697–1712 (2017)
19. Chun, I.Y., Fessler, J.A.: Convolutional analysis operator learning: acceleration and convergence. IEEE Trans. Image Process. **29**(1), 2108–2122 (2019)
20. Chun, I.Y., Huang, Z., Lim, H., Fessler, J.: Momentum-net: fast and convergent iterative neural network for inverse problems. IEEE Trans. Pattern Anal. Mach. Intell. (2020). https://doi.org/10.1109/tpami.2020.3012955
21. Chun, I.Y., Zheng, X., Long, Y., Fessler, J.A.: Bcd-net for low-dose ct reconstruction: Acceleration, convergence, and generalization. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 31–40. Springer (2019)
22. Clarke, F.H.: Optimization and Nonsmooth Analysis, vol. 5. SIAM, Philadelphia (1990)
23. Cormack, A.M.: Representation of a function by its line integrals, with some radiological applications. ii. J. Appl. Phys. **35**(10), 2908–2913 (1964)
24. De Man, B., Basu, S.: Distance-driven projection and backprojection. In: 2002 IEEE Nuclear Science Symposium Conference Record, vol. 3, pp. 1477–1480. IEEE (2002)
25. De Man, B., Basu, S.: Distance-driven projection and backprojection in three dimensions. Phys. Med. Biol. **49**(11), 2463 (2004)
26. Ding, C., Zhang, Q., Wang, G., Ye, X., Chen, Y.: Learned alternating minimization algorithm forÂ dual-domain sparse-view ct reconstruction. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention - MICCAI 2023, pp. 173–183. Springer, Cham (2023)

27. Dong, X., Wong, R., Lyu, W., Abell-Hart, K., Deng, J., Liu, Y., Hajagos, J.G., Rosenthal, R.N., Chen, C., Wang, F.: An integrated lstm-heterorgnn model for interpretable opioid overdose risk prediction. Artif. Intell. Med. **135**, 102439 (2023). https://doi.org/10.1016/j.artmed.2022.102439
28. Dumoulin, V., Visin, F.: A guide to convolution arithmetic for deep learning. arXiv preprint arXiv:1603.07285 (2016)
29. Geyer, L.L., et al.: State of the art: iterative ct reconstruction techniques. Radiology **276**(2), 339–357 (2015)
30. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: In Proceedings of the International Conference on Artificial Intelligence and Statistics. Society for Artificial Intelligence and Statistics (2010)
31. Han, Y., Ye, J.C.: Framing u-net via deep convolutional framelets: application to sparse-view ct. IEEE Trans. Med. Imaging **37**(6), 1418–1429 (2018)
32. den Harder, A.M., et al.: Radiation dose reduction in pediatric great vessel stent computed tomography using iterative reconstruction: a phantom study. PLoS ONE **12**(4), e0175714 (2017)
33. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2016). https://doi.org/10.1109/CVPR.2016.90
34. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European Conference on Computer Vision, pp. 630–645. Springer (2016)
35. He, W., Jiang, Z., Zhang, C., Sainju, A.M.: Curvanet: Geometric deep learning based on directional curvature for 3d shape analysis. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 2214–2224 (2020)
36. He, W., Sainju, A.M., Jiang, Z., Yan, D.: Deep neural network for 3d surface segmentation based on contour tree hierarchy. In: Proceedings of the 2021 SIAM International Conference on Data Mining (SDM), pp. 253–261. SIAM (2021)
37. He, W., Sainju, A.M., Jiang, Z., Yan, D., Zhou, Y.: Earth imagery segmentation on terrain surface with limited training labels: a semi-supervised approach based on physics-guided graph co-training. ACM Trans. Intell. Syst. Technol. **13**(2), 1–22 (2022)
38. Hounsfield, G.N.: Computerized transverse axial scanning (tomography): Part 1: description of system. Br. J. Radiol. **46**(552), 1016–1022 (1973). https://doi.org/10.1259/0007-1285-46-552-1016
39. Hsieh, C.J., Jin, S.C., Chen, J.C., Kuo, C.W., Wang, R.T., Chu, W.C.: Performance of sparse-view ct reconstruction with multi-directional gradient operators. PLoS ONE **14**(1), e0209674 (2019)
40. Hu, D., et al.: Hybrid-domain neural network processing for sparse-view ct reconstruction. IEEE Trans. Radiat. Plasma Med. Sci. **5**(1), 88–98 (2021). https://doi.org/10.1109/TRPMS.2020.3011413
41. Huang, X., Zhang, Z., Guo, F., Wang, X., Chi, K., Wu, K.: Research on older adults' interaction with e-health interface based on explainable artificial intelligence. In: Gao, Q., Zhou, J. (eds.) Human Aspects of IT for the Aged Population, pp. 38–52. Springer, Cham (2024)
42. Jiang, Z., He, W., Kirby, M.S., Sainju, A.M., Wang, S., Stanislawski, L.V., Shavers, E.J., Usery, E.L.: Weakly supervised spatial deep learning for earth image segmentation based on imperfect polyline labels. ACM Trans. Intell. Syst. Technol. **13**(2), 1–20 (2022)
43. Jin, K.H., McCann, M.T., Froustey, E., Unser, M.: Deep convolutional neural network for inverse problems in imaging. IEEE Trans. Image Process. **26**(9), 4509–4522 (2017)
44. Wang, Jing, Li, Tianfang, Hongbing, Lu., Liang, Zhengrong: Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography. IEEE Trans. Med. Imaging **25**(10), 1272–1283 (2006)
45. Kak, A.C., Slaney, M., Wang, G.: Principles of computerized tomographic imaging. Med. Phys. **29**(1), 107 (2002)
46. Kang, E., Chang, W., Yoo, J., Ye, J.C.: Deep convolutional framelet denosing for low-dose ct via wavelet residual network. IEEE Trans. Med. Imaging **37**(6), 1358–1369 (2018)
47. Kang, E., Min, J., Ye, J.C.: A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. Med. Phys. **44**(10), e360–e375 (2017)
48. Keller, E.J., et al.: Reinforcing the importance and feasibility of implementing a low-dose protocol for ct-guided biopsies. Acad. Radiol. **25**(9), 1146–1151 (2018)
49. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017)
50. Lai, Z., Chauhan, J., Chen, D., Dugger, B.N., Cheung, S.C., Chuah, C.N.: Semi-path: an interactive semi-supervised learning framework for gigapixel pathology image analysis. Smart Health **32**, 100474 (2024). https://doi.org/10.1016/j.smhl.2024.100474
51. Lai, Z., Guo, R., Xu, W., Hu, Z., Mifflin, K., Dugger, B.N., Chuah, C.N., Cheung, S.c.: Automated grey and white matter segmentation in digitized aI^{2} human brain tissue slide images. In: 2020 IEEE

International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1–6 (2020). https://doi.org/10.1109/ICMEW46912.2020.9105974
52. Lai, Z., Vadlaputi, P., Tancredi, D.J., Garg, M., Koppel, R.I., Goodman, M., Hogan, W., Cresalia, N., Juergensen, S., Manalo, E., Lakshminrusimha, S., Chuah, C.N., Siefkes, H.: Enhanced critical congenital cardiac disease screening by combining interpretable machine learning algorithms. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1403–1406 (2021). https://doi.org/10.1109/EMBC46164.2021.9630111
53. Lai, Z., Wang, C., Hu, Z., Dugger, B.N., Cheung, S.C., Chuah, C.N.: A semi-supervised learning for segmentation of gigapixel histopathology images from brain tissues. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1920–1923 (2021). https://doi.org/10.1109/EMBC46164.2021.9629715
54. Lai, Z., Wang, C., Oliveira, L.C., Dugger, B.N., Cheung, S.C., Chuah, C.N.: Joint semi-supervised and active learning for segmentation of gigapixel pathology images with cost-effective labeling. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, pp. 591–600 (2021)
55. Le, H., Borji, A.: What are the receptive, effective receptive, and projective fields of neurons in convolutional neural networks? CoRR arXiv:1705.07049 (2017)
56. Lee, H., Lee, J., Cho, S.: View-interpolation of sparsely sampled sinogram using convolutional neural network. In: Medical Imaging 2017: Image Processing, vol. 10133, p. 1013328. International Society for Optics and Photonics (2017)
57. Lee, H., Lee, J., Kim, H., Cho, B., Cho, S.: Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction. IEEE Trans. Radiat. Plasma Med. Sci. **3**(2), 109–119 (2018)
58. Li, M., Ling, P., Wen, S., Chen, X., Wen, F.: Bubble-wave-mitigation algorithm and transformer-based neural network demodulator for water-air optical camera communications. IEEE Photonics J. **15**(5), 1–10 (2023)
59. Li, T., Lu, H., Liang, Z.: Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography. IEEE Trans. Med. Imaging **25**, 1272–83 (2006). https://doi.org/10.1109/TMI.2006.882141
60. Liang, K., Yang, H., Kang, K., Xing, Y.: Improve angular resolution for sparse-view ct with residual convolutional neural network. In: Medical Imaging 2018: Physics of Medical Imaging, vol. 10573, p. 105731K. International Society for Optics and Photonics (2018)
61. Liao, D., Liu, C., Christensen, B.W., Tong, A., Huguet, G., Wolf, G., Nickel, M., Adelstein, I., Krishnaswamy, S.: Assessing neural network representations during training using noise-resilient diffusion spectral entropy. In: 2024 58th Annual Conference on Information Sciences and Systems (CISS), pp. 1–6 (2024). https://doi.org/10.1109/CISS59072.2024.10480166
62. Liu, C.: Fourier transform approximation as an auxiliary task for image classification (2021)
63. Liu, C., Zhu, N., Sun, H., Zhang, J., Feng, X., Gjerswold-Selleck, S., Sikka, D., Zhu, X., Liu, X., Nuriel, T., Wei, H.J., Wu, C.C., Vaughan, J.T., Laine, A.F., Provenzano, F.A., Small, S.A., Guo, J.: Deep learning of mri contrast enhancement for mapping cerebral blood volume from single-modal non-contrast scans of aging and Alzheimer's disease brains. Fron. Aging Neurosci. **14**, 673 (2022). https://doi.org/10.3389/fnagi.2022.923673
64. Liu, S., Wu, K., Jiang, C., Huang, B., Ma, D.: Financial time-series forecasting: Towards synergizing performance and interpretability within a hybrid machine learning approach (2023). arXiv:2401.00534
65. Lu, H., li, T., Liang, Z.: Sinogram noise reduction for low-dose ct by statistics-based nonlinear filters. Proceedings of SPIE - The International Society for Optical Engineering **5747** (2005). https://doi.org/10.1117/12.595662
66. Lyu W Dong X, W.R.Z.S.A.H.K.W.F.C.C.: A multimodal transformer: Fusing clinical notes with structured ehr data for interpretable in-hospital mortality prediction. AMIA Annu Symp Proc pp. 719–728 (2023)
67. Ma, H., Liu, Y., Wu, G.: Elucidating multi-stage progression of neuro-degeneration process in Alzheimerâ€™s disease. Alzheimer's Dementia **18**, e068774 (2022)
68. Ma, H., Shi, Z., Kim, M., Liu, B., Smith, P.J., Liu, Y., Wu, G., (ADNI, A.D.N.I., et al.: Disentangling sex-dependent effects of apoe on diverse trajectories of cognitive decline in Alzheimer's disease. NeuroImage 120609 (2024)
69. Ma, H., Zeng, D., Liu, Y.: Learning individualized treatment rules with many treatments: a supervised clustering approach using adaptive fusion. Adv. Neural. Inf. Process. Syst. **35**, 15956–15969 (2022)
70. Ma, H., Zeng, D., Liu, Y.: Learning optimal group-structured individualized treatment rules with many treatments. J. Mach. Learn. Res. **24**(102), 1–48 (2023)
71. Manduca, A., et al.: Projection space denoising with bilateral filtering and ct noise modeling for dose reduction in ct. Med. Phys. **36**(11), 4911–4919 (2009). https://doi.org/10.1118/1.3232004

72. McCollough, C.: Tu-fg-207a-04: overview of the low dose ct grand challenge. Med. Phys. **43**, 3759–3760 (2016). https://doi.org/10.1118/1.4957556
73. Nesterov, Y.: Smooth minimization of non-smooth functions. Math. Program. **103**(1), 127–152 (2005)
74. Niu, S., et al.: Sparse-view x-ray ct reconstruction via total generalized variation regularization. Phys. Med. Biol. **59**(12), 2997 (2014)
75. Pang, N., Qian, L., Lyu, W., Yang, J.D.: Transfer learning for scientific data chain extraction in small chemical corpus with bert-crf model (2019)
76. Paszke, A., et al.: Pytorch: an imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems 32, pp. 8024–8035. Curran Associates, Inc. (2019)
77. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer (2015)
78. Saltybaeva, N., Martini, K., Frauenfelder, T., Alkadhi, H.: Organ dose and attributable cancer risk in lung cancer screening with low-dose computed tomography. PLoS ONE **11**(5), e0155722 (2016)
79. Sauter, A., et al.: Ultra low dose ct pulmonary angiography with iterative reconstruction. PLoS ONE **11**(9), e0162716 (2016)
80. Shan, H., et al.: Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose ct image reconstruction. Nat. Mach. Intell. **1**(6), 269–276 (2019)
81. Sun, J., Deep, A., Zhou, S., Veeramani, D.: Industrial system working condition identification using operation-adjusted hidden Markov model. J. Intell. Manuf. **34**(6), 2611–2624 (2023)
82. Sun, J., Li, H., Xu, Z., et al.: Deep admm-net for compressive sensing mri. In: Advances in neural information processing systems, pp. 10–18 (2016)
83. Sun, J., Zhou, S., Veeramani, D.: A neural network-based control chart for monitoring and interpreting autocorrelated multivariate processes using layer-wise relevance propagation. Qual. Eng. **35**(1), 33–47 (2023)
84. Sun, J., Zhou, S., Veeramani, D., Liu, K.: Prediction of condition monitoring signals using scalable pairwise gaussian processes and Bayesian model averaging. IEEE Trans. Autom. Sci. Eng. (2024)
85. Sun, S., Ren, W., Li, J., Wang, R., Cao, X.: Logit standardization in knowledge distillation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15731–15740 (2024)
86. Sun, S., Ren, W., Wang, T., Cao, X.: Rethinking image restoration for object detection. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (eds.) Advances in Neural Information Processing Systems, vol. 35, pp. 4461–4474. Curran Associates Inc, New York (2022)
87. Tian, H., Jiang, X., Tao, P.: PASSer: prediction of allosteric sites server. Mach. Learn. Sci. Technol. **2**(3), 035015 (2021). https://doi.org/10.1088/2632-2153/abe6d6
88. Tipnis, S., et al.: Iterative reconstruction in image space (iris) and lesion detection in abdominal ct. In: Medical Imaging 2010: Physics of Medical Imaging, vol. 7622, p. 76222K. International Society for Optics and Photonics (2010)
89. Wang, Z., Li, T., Zheng, J.Q., Huang, B.: When cnn meet with;vit: Towards semi-supervised learning for;multi-class medical image semantic segmentation. In: Computer Vision – ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VII, p. 424–441. Springer-Verlag, Berlin, Heidelberg (2023). https://doi.org/10.1007/978-3-031-25082-8_28
90. Wang, Z., Ma, C.: Dual-contrastive dual-consistency dual-transformer: A semi-supervised approach to medical image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 870–879 (2023)
91. Wang, Z., Su, M., Zheng, J.Q., Liu, Y.: Densely connected swin-unet for multiscale information aggregation in medical image segmentation. In: 2023 IEEE International Conference on Image Processing (ICIP), pp. 940–944 (2023).https://doi.org/10.1109/ICIP49359.2023.10222451
92. Wang, Z., Yang, C.: Mixsegnet: Fusing multiple mixed-supervisory signals with multiple views of networks for mixed-supervised medical image segmentation. Eng. Appl. Artif. Intell. **133**, 108059 (2024)
93. Wang, Z., Zhao, W., Ni, Z., Zheng, Y.: Adversarial vision transformer for medical image semantic segmentation with limited annotations. In: BMVC, p. 1002 (2022)
94. Wei, Y., Gao, M., Xiao, J., Liu, C., Tian, Y., He, Y.: Research and implementation of cancer gene data classification based on deep learning. J. Softw. Eng. Appl. **16**(6), 155–169 (2023)
95. Wei, Y., Gao, M., Xiao, J., Liu, C., Tian, Y., He, Y.: Research and implementation of traffic sign recognition algorithm model based on machine learning. J. Softw. Eng. Appl. **16**(6), 193–210 (2023)
96. Wei, Y., Zhang, D., Gao, M., Mulati, A., Zheng, C., Huang, B.: Skin cancer detection based on machine learning. J. Knowl. Learn. Sci. Technol. **3**(2), 72–86 (2024)
97. Wei, Y., Zhang, D., Gao, M., Tian, Y., He, Y., Huang, B., Zheng, C.: Breast cancer prediction based on machine learning. J. Softw. Eng. Appl. **16**(8), 348–360 (2023)

98. Willemink, M.J., et al.: Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality. Eur. Radiol. **23**(6), 1632–1642 (2013)
99. Wu, D., Kim, K., El Fakhri, G., Li, Q.: Iterative low-dose ct reconstruction with priors trained by artificial neural network. IEEE Trans. Med. Imaging **36**(12), 2479–2486 (2017)
100. Wu, K.: Creating panoramic images using orb feature detection and ransac-based image alignment*. Adv. Comput. Commun. **4**(4), 220–224 (2023)
101. Wu, K., Chen, J.: Cargo operations of express air. Eng. Adv. **3**(4), 337–341 (2023)
102. Wu, K., Chi, K.: Enhanced e-commerce customer engagement: a comprehensive three-tiered recommendation system. J. Knowl. Learn. Sci. Technol. **2**(3), 348–359 (2024)
103. Xia, W., Lu, Z., Huang, Y., Shi, Z., Liu, Y., Chen, H., Chen, Y., Zhou, J., Zhang, Y.: Magic: Manifold and graph integrative convolutional network for low-dose ct reconstruction. IEEE Trans. Med. Imaging (2021)
104. Xie, S., Yang, T.: Artifact removal in sparse-angle ct based on feature fusion residual network. IEEE Trans. Radiat. Plasma Med. Sci. **5**(2), 261–271 (2021). https://doi.org/10.1109/TRPMS.2020.3000789
105. Xu, Z., Xiao, T., He, W., Wang, Y., Jiang, Z.: Spatial knowledge-infused hierarchical learning: An application in flood mapping on earth imagery. In: Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems, pp. 1–10 (2023)
106. Xu, Z., Xiao, T., He, W., Wang, Y., Jiang, Z., Chen, S., Xie, Y., Jia, X., Yan, D., Zhou, Y.: Spatial-logic-aware weakly supervised learning for flood mapping on earth imagery. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 22457–22465 (2024)
107. Yang, Q., et al.: Low-dose ct image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. IEEE Trans. Med. Imaging **37**(6), 1348–1357 (2018)
108. Ye, S., Long, Y., Chun, I.Y.: Momentum-net for low-dose ct image reconstruction. arXiv preprint arXiv:2002.12018 (2020)
109. Ye, S., Ravishankar, S., Long, Y., Fessler, J.A.: Spultra: Low-dose ct image reconstruction with joint statistical and learned image models. IEEE Trans. Med. Imaging **39**(3), 729–741 (2019)
110. Zhang, D., Zhou, F.: Self-supervised image denoising for real-world images with context-aware transformer. IEEE Access **11**, 14340–14349 (2023)
111. Zhang, D., Zhou, F., Jiang, Y., Fu, Z.: Mm-bsn: Self-supervised image denoising for real-world with multi-mask based on blind-spot network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4188–4197 (2023)
112. Zhang, D., Zhou, F., Wei, Y., Yang, X., Gu, Y.: Unleashing the power of self-supervised image denoising: A comprehensive review. arXiv preprint arXiv:2308.00247 (2023)
113. Zhang, Q.: Learnable Nonconvex Nonsmooth Optimization Algorithms and Theories for Variational Neural Networks in Solving Inverse Problems. University of Florida (2022). https://books.google.com.mx/books?id=M-LdzwEACAAJ
114. Zhang, Q., Heldermon, C.D., Toler-Franklin, C.: Multiscale detection of cancerous tissue in high resolution slide scans. In: Bebis, G., Yin, Z., Kim, E., Bender, J., Subr, K., Kwon, B.C., Zhao, J., Kalkofen, D., Baciu, G. (eds.) Advances in Visual Computing, pp. 139–153. Springer, Cham (2020)
115. Zhang, Q., Ye, X., Chen, Y.: Nonsmooth nonconvex LDCT image reconstruction via learned descent algorithm. In: B. Müller, G. Wang (eds.) Developments in X-Ray Tomography XIII, vol. 11840, p. 1184013. International Society for Optics and Photonics, SPIE (2021). https://doi.org/10.1117/12.2597798
116. Zhang, Q., Ye, X., Chen, Y.: Extra proximal-gradient network with learned regularization for image compressive sensing reconstruction. J. Imaging **8**(7), 178 (2022)
117. Zhang, Z., Liang, X., Dong, X., Xie, Y., Cao, G.: A sparse-view ct reconstruction method based on combination of densenet and deconvolution. IEEE Trans. Med. Imaging **37**(6), 1407–1417 (2018)
118. Zheng, S., Zhang, Y., Lyu, W., Goswami, M., Schneider, A., Nevmyvaka, Y., Ling, H., Chen, C.: On the existence of a trojaned twin model (2023). https://openreview.net/forum?id=w48XN5HwpV8
119. Zheng, X., Ravishankar, S., Long, Y., Fessler, J.A.: Pwls-ultra: an efficient clustering and learning-based approach for low-dose 3d ct image reconstruction. IEEE Trans. Med. Imaging **37**(6), 1498–1510 (2018)
120. Zhou, C., Zhao, Y., Cao, J., Shen, Y., Cui, X., Cheng, C.: Optimizing search advertising strategies: Integrating reinforcement learning with generalized second-price auctions for enhanced ad ranking and bidding (2024)
121. Zhou, F., Fu, Z., Zhang, D.: High dynamic range imaging with context-aware transformer. In: 2023 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2023)
122. Zhu, N., Liu, C., Feng, X., Sikka, D., Gjerswold-Selleck, S., Small, S.A., Guo, J.: Deep learning identifies neuroimaging signatures of alzheimerâ€™s disease using structural and synthesized functional mri data. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), pp. 216–220 (2021). https://doi.org/10.1109/ISBI48211.2021.9433808

123. Zhu, N., Liu, C., Forsyth, B., Singer, Z.S., Laine, A.F., Danino, T., Guo, J.: Segmentation with residual attention u-net and an edge-enhancement approach preserves cell shape features. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 2115–2118 (2022). https://doi.org/10.1109/EMBC48229.2022.9871026
124. Zhu, N., Liu, C., Laine, A.F., Guo, J.: Understanding and modeling climate impacts on photosynthetic dynamics with fluxnet data and neural networks. Energies **13**(6), 1322 (2020)
125. Zhuang, J., Al Hasan, M.: Non-exhaustive learning using gaussian mixture generative adversarial networks. In: Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part II 21, pp. 3–18. Springer (2021)
126. Zhuang, J., Gao, M., Hasan, M.A.: Lighter u-net for segmenting white matter hyperintensities in mr images. Proceedings of the 16th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (2019)
127. Zhuang, J., Hasan, M.A.: Robust node representation learning via graph variational diffusion networks. arXiv preprint arXiv:2312.10903 (2023)
128. Zhuang, J., Kennington, C.: Understanding survey paper taxonomy about large language models via graph representation learning. arXiv preprint arXiv:2402.10409 (2024)
129. Zhuang, J., Wang, D.: Geometrically matched multi-source microscopic image synthesis using bidirectional adversarial networks. In: Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021) Medical Imaging and Computer-Aided Diagnosis, pp. 79–88. Springer (2022)