Computerized Medical Imaging and Graphics 123 (2025) 102530

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Deep Guess acceleration for explainable image reconstruction in sparse-view CT

Elena Loli Piccolomini [a],*, Davide Evangelista [a], Elena Morotti [b]

[a] Department of Computer Science and Engineering, University of Bologna, Via Mura Anteo Zamboni 7, 40126 Bologna, Italy
[b] Department of Political and Social Sciences, University of Bologna, Strada Maggiore 45, 40125 Bologna, Italy

ARTICLE INFO

Keywords:
Non-convex optimization
Sparse view computed tomography
Model-based iterative reconstruction
Deep neural networks
Interpretable reconstruction

ABSTRACT

Sparse-view Computed Tomography (CT) is an emerging protocol designed to reduce X-ray dose radiation in medical imaging. Reconstructions based on the traditional Filtered Back Projection algorithm suffer from severe artifacts due to sparse data. In contrast, Model-Based Iterative Reconstruction (MBIR) algorithms, though better at mitigating noise through regularization, are too computationally costly for clinical use. This paper introduces a novel technique, denoted as the Deep Guess acceleration scheme, using a trained neural network both to quicken the regularized MBIR and to enhance the reconstruction accuracy. We integrate state-of-the-art deep learning tools to initialize a clever starting guess for a proximal algorithm solving a non-convex model and thus computing a (mathematically) interpretable solution image in a few iterations. Experimental results on real and synthetic CT images demonstrate the Deep Guess effectiveness in (very) sparse tomographic protocols, where it overcomes its mere variational counterpart and many data-driven approaches at the state of the art. We also consider a ground truth-free implementation and test the robustness of the proposed framework to noise.

1. Introduction

Computed Tomography (CT) reconstruction from sparse views has emerged as a focal point in medical imaging research. A key challenge lies in diminishing the total X-ray radiation dose administered to patients while maintaining reconstruction precision. The sparse-view methodology revolves around decreasing the number of rotation angles utilized, thereby eliciting compromised image quality due to the introduction of numerous streaking artifacts by the classical analytical algorithms, such as the Filtered Back Projection (FBP).

Model-Based Iterative Reconstruction (MBIR) methods are extensively employed, relying on mathematical models to represent the problem being solved. In this case, sparse CT image reconstruction can be addressed as a linear problem of the form:

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{e}^{\nu}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^m$ is the sinogram, affected by noise $\boldsymbol{e}^{\nu} \in \mathbb{R}^m$ with intensity depending on the positive parameter ν, and $A \in \mathbb{R}^{m\times n}$ is the forward operator describing the CT system geometry. Due to the sparse sampling, $m < n$ and the system (1) has infinitely many solutions. Additionally, owing to the typical ill-posedness of inverse problems, it is beneficial to introduce a prior term.

MBIR methods allow the inverse problem to be reformulated as a minimization problem involving a fit-to-data function and a regularization (or prior) function. In this paper, we consider the popular formulation given by:

$$\min_{\boldsymbol{x}\geq 0} \|A\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda \mathcal{R}(\boldsymbol{x}), \tag{2}$$

where the least squares term represents the data-fitting function, $\mathcal{R}$ is the prior term, and $\lambda > 0$ is the regularization parameter. Compressed sensing theory plays a crucial role in this framework, enabling the reconstruction of high-quality images from a limited amount of data, through a suitable choice of the prior $\mathcal{R}(\boldsymbol{x})$. Specifically, forcing gradient sparsity through the prior has been proven to be extremely useful in fields such as medical imaging, where preserving boundaries of low-contrast objects is essential for accurate diagnosis and treatment. Given a fixed $0 < p \leq 1$, the gradient sparsity-inducing prior $\mathcal{R}$ is defined as:

$$\mathcal{R}(\boldsymbol{x}) = \|D\boldsymbol{x}\|_{2,p}^p := \| \, |D\boldsymbol{x}| \, \|_p^p, \tag{3}$$

where $|D\boldsymbol{x}| \in \mathbb{R}^n$ represents the gradient-magnitude image of $\boldsymbol{x}$, given by $(|D\boldsymbol{x}|)_i := \sqrt{(D_h\boldsymbol{x})_i^2 + (D_v\boldsymbol{x})_i^2}$ with the operators D_h and D_v, denoting the differences among two adjacent pixels in the horizontal and vertical directions, respectively.

* Corresponding author.
E-mail address: elena.loli@unibo.it (E. Loli Piccolomini).

https://doi.org/10.1016/j.compmedimag.2025.102530
Received 15 August 2024; Received in revised form 3 March 2025; Accepted 13 March 2025
Available online 21 March 2025
0895-6111/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

The ℓ_0 quasi-norm (given by Eq. (3) when $p \to 0$) quantifies the cardinality of the non-zero entries in its argument, and it is widely acknowledged that it represents the optimal sparsifying prior. Unfortunately, minimizations involving an ℓ_0 prior pose significant computational challenges and are seldom addressed directly. Luckily, signal recovery remains feasible by substituting the ℓ_1 norm (with $p = 1$ in Eq. (3)) for the ℓ_0 quasi-norm under appropriate assumptions (Candès et al., 2006; Donoho, 2006). In this case, the regularization function is denoted as Total Variation (TV), and it is very popular in CT applications. Many scholars also conducted a large amount of research on TV regularization, proposing numerous improved convex TV-based models, as we will detail in the next Section. Nonetheless, the TV prior exhibits limitations such as its tendency to excessively smooth regions containing fine details or textures.

Setting $0 < p < 1$ in Eq. (3) defines $\mathcal{R}$ as the isotropic Total p-norm Variation (TpV), which represents a better alternative to the TV prior for approximating the ℓ_0 quasi-norm. The TpV has already been successfully used in a number of imaging applications, including CT reconstruction, but its main drawback lies in its non-convex nature. Thus, ensuring the attainment of the global minimum by an MBIR may not be feasible, and MBIR approaches are prone to fall in one of the local minima, which could be significantly distant from the desired global solution. Consequently, non-convex minimization presents two practical disadvantages: the extensive computational effort required by the MBIR solver to converge and the unreliable dependence of the reconstructed image on the algorithm initialization. Nevertheless, the identification of a *good local* minimum can yield accurate reconstructions of sparse signals using significantly fewer measurements compared to the $p = 1$ case (Chartrand, 2007).

Given the urgent need for rapid and accurate images in clinical applications, Deep Learning (DL) has emerged as a powerful alternative to MBIR. Many different approaches have been designed in the last years to address CT reconstruction with neural networks, and they have demonstrated the ability of data-driven methods to learn complex patterns from vast datasets, enabling the computation of high-quality images swiftly. However, three main drawbacks must be considered to make deep learning-based approaches usable in practice. Firstly, the black box structure of the trained networks makes it impossible to characterize the final outputs as numerical solutions of any mathematical inverse problem (Sidky et al., 2021). Secondly, the training of the neural networks typically requires extensive training samples, which may not be available in many real-world medical scenarios. As we will discuss in Section 2, only a few works tackle this disadvantage in the literature. Lastly, while highly accurate when used on images consistent with the training samples, neural networks may manifest untrustworthy effects when applied to slightly different data. This unstable behavior is the subject of study in some recent research (Antun et al., 2020; Morotti et al., 2021; Evangelista et al., 2023b).

Our proposal and goals

This research proposes to the imaging community a fast and reliable tool for CT image reconstruction, taking into consideration the pros and cons of the aforementioned state-of-the-art approaches and directly addressing the scientific challenges for the implementation of software in clinical routines. Specifically, we combine the advantages of both MBIR and DL-based best tools and design a unified solution of practical interest for medical imaging. The core idea is to use a neural network to initialize the starting iterate in the algorithm addressing the TpV-regularized problem. We refer to our strategy as the *Deep Guess acceleration* to highlight that we exploit the effectiveness of deep learning to define the initial guess of an MBIR solver and to make its numerical convergence faster. The scheme of the Deep Guess acceleration framework is visually illustrated in Fig. 1.

The benefits of the proposed approach are manifold.

- Fast execution: by using a pre-trained neural network, the iterative solver can start from a position much closer to its solution, reducing the number of needed iterations.
- Reduced risk of bad local minima: with better initial guess, the likelihood of the solver falling into a bad image is significantly decreased.
- Mathematical explainability of the solution: running an MBIR to its convergence, we fully preserve convergence properties and the numerical characterization of the final solution, despite the use of a neural network. This property is desirable, especially in the medical field, as also highlighted in the literature (see, for example, Tang et al. (2023)).

At last, we highlight that we have also implemented neural networks using a ground truth-free supervised learning approach. Given the need for a network to be trained on a representative dataset, we propose incorporating into the Deep Guess step a network trained on images generated from available tomographic data. This offers a viable solution in scenarios where ground truth images are not available for standard training of a neural network.

The paper is organized as follows. In Section 2 we present a state-of-art on both MBIR and DL-based algorithms for sparse-view CT image reconstruction; in Section 3 we describe in detail the implementation of the proposed framework. We report in Section 4 the results of several numerical experiments on both a real and a synthetic data set, demonstrating that our framework achieves good solutions, outperforming iterative non-convex algorithms and the data-driven methods considered for comparison. Lastly, in Section 5 we state some conclusions.

2. State of the art

CT image reconstruction techniques have been evolving through two significant classes of tools: model-based iterative algorithms and deep learning-based tools. Both methodologies seek to enhance the precision and efficiency of computed tomography (CT) image reconstruction. MBIR emphasizes iterative optimization processes and the incorporation of prior knowledge. In contrast, Deep Learning-based techniques exploit data-driven learning to achieve superior performance. Here, we review some of the most impactful techniques proposed over the years.

The first class we focus on is the one composed of MBIR approaches, which iteratively solve optimization problems as the one considered in (2). To produce high-quality images from sparse and noisy data, effectively reduce artifacts, and enhance edge preservation, MBIR schemes incorporate specific regularizers $\mathcal{R}(\boldsymbol{x})$ so that the image processing is performed according to precise mathematically-grounded image priors. In medical imaging, priors that enforce gradient sparsity have become a well-established de facto choice because most parts of medical images are relatively smooth and uniform, with significant information concentrated at the edges.

Total Variation (TV) prior, minimizing the ℓ_1 norm of the gradient-magnitude image, has been extensively utilized since its seminal paper by Rudin et al. (1992), to effectively denoise images (see Vogel and Oman (1996), Chan et al. (2005) and Rodríguez (2013), to cite a few). These formulations effectively differentiate between artifacts (small, random intensity changes) and significant image features (sharp edges), making TV prior interesting for tomographic applications and particularly beneficial in sparse data scenarios, where sparse-view artifacts and noise must be removed. This success has been proved in many works, such as Persson et al. (2001), Sidky and Pan (2008), Tian et al. (2011), Ritschl et al. (2011), Caselles et al. (2015), Loli Piccolomini et al. (2016), Lv et al. (2020) and Loli Piccolomini and Morotti (2021).

The TV models mentioned above are all convex and relatively easy to handle numerically. At the same time, it is well known that non-convex norms are more suitable for measuring sparsity than the corresponding convex norms, since they are much closer to the ℓ_0 norm

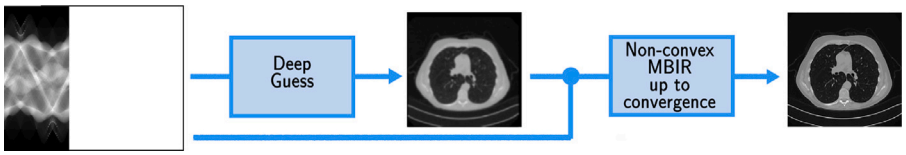

Fig. 1. Graphical representation of the proposed data-driven approach for solving the non-convex imaging problem. As non-convex model we fix a TpV regularizer as in (3).

(that is exactly the measure of sparsity) than their convex counterparts (Chartrand, 2007; Wang et al., 2018). For instance, the TpV prior, which uses the p-norm of the gradient magnitude with $0 < p < 1$, allows for more flexible and adaptive regularization than TV. It, in fact, tends to avoid staircasing effects better than TV, because it adjusts the degree of sparsity imposed on the gradients. For this reason it has been successfully used in medical imaging (Tong et al., 2018; Wang, 2021) and for CT reconstruction (Sidky et al., 2014; Demirel et al., 2021). The wavelets have been used in tomographic imaging in Borsdorf et al. (2008), Beck and Teboulle (2009b), Mehranian et al. (2013), Xu et al. (2016) and Purisha et al. (2017).

The second class of algorithms to tackle CT reconstruction is the one involving Deep Learning. DL-based methods offer faster reconstructions and can capture intricate patterns and structures, often outperforming traditional techniques in terms of speed and image quality. Nowadays, there is a plethora of papers exploring different ways in which neural networks can be used for medical image reconstruction.

One approach is to learn the reconstruction operator/solver, as the mapping between projection and image domain, from an appropriate corpus of training data (Zhu et al., 2018; Boink and Brune, 2019). A further approach implements two-step frameworks. In the first step, a reconstruction algorithm is applied offline to quickly compute the reconstruction from the projection data (typically, an analytic solver as the FBP), and then a trained neural network is applied as a post-processing operator to remove noise and artifacts (Jin et al., 2017; Chen et al., 2017; Schnurr et al., 2019; Morotti et al., 2021). This approach is often referred to as Learnt Post Processing (LPP). Both these approaches are typically very fast since they do not need the computation of the projector operator A and any related matrix–vector multiplications (that are computationally heavy steps in MBIR), but the mathematical characterization of the solution is questionable (Sidky et al., 2021). Alternatively, there are hybrid approaches where DL and MBIR are merged together. Examples are the unrolling algorithms, which implement iterative solvers (as classical optimization algorithms) unrolling each iteration in a layer of a deep neural network (Adler and Öktem, 2018; Monga et al., 2021; Xiang et al., 2021). Here, the optimal model parameters are learned during the training phase so that adaptive parameters allow high-quality reconstructions, and the convergence of the MBIR solver is accelerated too. A popular option is represented by the class of Plug-and-Play (PnP) hybrid methods, where a learned regularizer is plugged inside a classical variational scheme (Venkatakrishnan et al., 2013; Cascarano et al., 2022). Specifically, the network plays the role of a denoiser in one inner step of each iteration of the MBIR, which is executed in a classic way up to its convergence. A further hybrid approach is represented by Deeply Learned Regularizers, which employ a pre-trained neural network to learn the $\mathcal{R}$ prior operator. An example is represented by the NETT algorithm (Li et al., 2020). An additional class of hybrid methods is represented by multi-step schemes. To set some examples, in Evangelista et al. (2023a), the MBIR method is incorporated into a ground truth-free Learned Post Processing scheme. Here, a coarse image is initially reconstructed from projections using an early interrupted MBIR solver and then processed by a network trained on accurate MBIR reconstructions. In a second example (Morotti et al., 2024), a neural network adjusts an adaptive space-variant total variation (TV)-based regularizer, and the resulting model-based formulation is subsequently solved efficiently.

The aforementioned papers use supervised learning, heavily relying on the availability of a set of consistent and clean ground truth images, which is not always guaranteed in real medical scenarios. To the best of our knowledge, few researches have tackled it so far. The authors of Baguer et al. (2020) and Shu and Entezari (2022) propose to use the Deep Image Prior approach, which is a scheme where the architecture of an untrained neural network itself serves as a regularizer for imaging tasks. The network is initialized with random weights and iteratively fitted to the observed datum $\boldsymbol{y}$ so that the network naturally captures the image's features and promotes accurate reconstructions. Alternatively, in Ye et al. (2021), an unsupervised hybrid reconstruction is coupled with a supervised denoiser, which does not need medical ground truth images, whereas in Evangelista et al. (2023a), the authors generate the target images for the network training from the available set of projection data, by running MBIR accurate reconstructions offline.

3. The proposed Deep Guess acceleration framework

The proposed framework comprises two primary steps, as illustrated in Fig. 1.

1. The initial step involves calculating a robust approximation of the reconstructed image using a neural network within a supervised post processing scheme. The output of this step, referred to as Deep Guess (DG), is a reconstruction scheme already proposed in the literature (see Section 2). Here, however, it serves as the input for the subsequent phase, by computing the initial iterate $\boldsymbol{x}_0$ of the non-convex MBIR solver.
2. The second step refines the quality of image computed by the DG by applying to it some iterations of an iterative optimization algorithm solving the non-convex minimization problem (2) with regularization prior given by (3). We focus on the non-convex case with $0 < p < 1$, leading to the so called TpV regularization.

We provide a precise description of the implementation of the two steps, beginning with the second step. Subsequently, we present a detailed analysis of the neural network utilized, including its architectural design and hyperparameter configuration.

3.1. A non-convex Model-Based Iterative Reconstruction method utilizing Chambolle–Pock iterations

The minimization (2) with non-convex TpV prior is solved by the iterative reweighting ℓ_1-norm (IRℓ_1) strategy introduced in Candès et al. (2008) and Daubechies et al. (2010). It is an iterative process, where each iterate is obtained by solving a convex optimization problem, i.e.:

$$\boldsymbol{x}^{(k+1)} = \arg\min_{\boldsymbol{x}\geq 0} \|A\boldsymbol{x} - \boldsymbol{y}\|_2^2 + \lambda \| \boldsymbol{w}(\boldsymbol{x}^{(k)}) \odot |D\boldsymbol{x}| \|_1, \tag{4}$$

where the weights $\boldsymbol{w}(\boldsymbol{x}^{(k)}) \in \mathbb{R}^n$ are defined as:

$$\boldsymbol{w}(\boldsymbol{x}^{(k)})_i = \left(\frac{\sqrt{\eta^2 + |D\boldsymbol{x}^{(k)}|_i^2}}{\eta} \right)^{p-1},$$

with a predefined smoothing parameter $\eta > 0$.

Each problem (4) is approximately solved using a limited number of iterations of the Chambolle–Pock (CP) algorithm, a primal–dual method originally introduced for the minimization of convex functional appearing in image processing applications in Chambolle and Pock

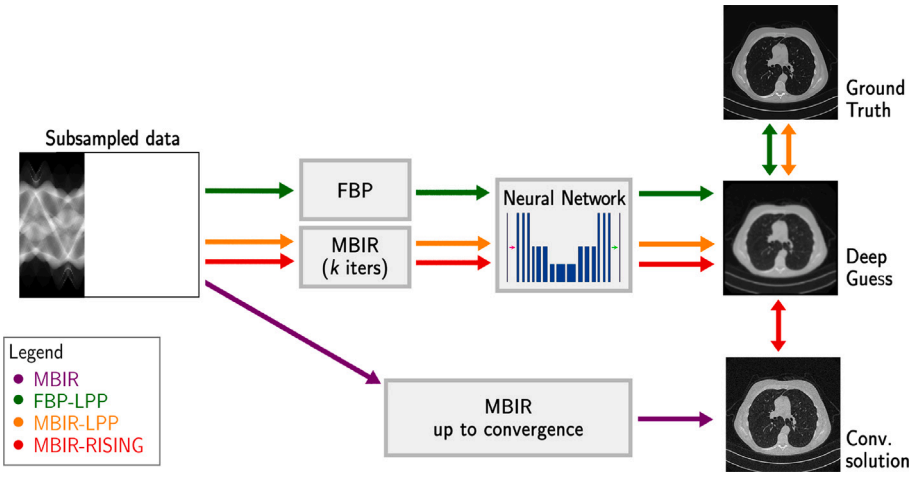

Fig. 2. The Deep Guess block. It can be computed by following the red, orange, or green path, all exploiting a neural network. The input to the network is a coarse reconstructed image, obtained either through FBP or after K iterations of an MBIR solver; the output is the Deep Guess image. The double arrows link the output to the desired targets, which can be either the ground truth or the images obtained at convergence of the MBIR method.

(2011). The CP algorithm has been effectively employed for total variation (TV)-based image reconstruction from undersampled tomographic data in Sidky et al. (2012) and Loli Piccolomini and Morotti (2021). In Sidky et al. (2014), the authors suggest applying the CP algorithm in a manner consistent with the approach used in this study.

In its original formulation, the CP algorithm was proposed to minimize an objective function of the form:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} F(M\boldsymbol{x}) + G(\boldsymbol{x}), \tag{5}$$

where both F and G are real-valued, proper, convex, lower semi-continuous functions, and M is a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^s$. Since there are no constraints on the smoothness of either F and G, the CP method can be applied to our problem by setting:

$$\begin{cases} G(\boldsymbol{x}) = \iota_{\mathcal{X}}(\boldsymbol{x}), \\ F(M\boldsymbol{x}) = \frac{1}{2}\|A\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \lambda\| \boldsymbol{w}(\boldsymbol{x}^{(k)}) \odot |D\boldsymbol{x}| \|_1, \end{cases}$$

where $\iota_{\mathcal{X}}(\boldsymbol{x})$ is the indicator function of the feasible set $\mathcal{X} := \{\boldsymbol{x} \in \mathbb{R}^n; \boldsymbol{x}_i \geq \boldsymbol{0}\ \forall i = 1,\dots n\}$, and the linear operator $M \in \mathbb{R}^{s\times n}$ is given by the row-wise concatenation of A and D, such that $M = [A; D]$.

We provide a concise description of the computations performed in each iteration $j, j = 1,\dots,J$, of the CP algorithm. To preserve a smooth and readable flow, we omit some detailed calculations that can be found in Morotti et al. (2024). We only remark that we need to define $\boldsymbol{z}^{(j)} = (p^{(j)}, q^{(j)})$ with two dual variables $p \in \mathbb{R}^m$ and $q \in \mathbb{R}^{2n}$, to explicitly derive the $\text{prox}_{\sigma F^*}$ term, so that the CP algorithm is defined by the following update rules:

$$\begin{cases} p^{(j+1)} = \frac{p^{(j)}+\sigma\left(A\bar{\boldsymbol{x}}^{(j)}-\boldsymbol{y}^{\delta}\right)}{1+3\sigma}, \\ q^{(j+1)} = \frac{\lambda \boldsymbol{w}(\bar{\boldsymbol{x}}^{(j)})\odot\left(q^{(j)}+\sigma|D\bar{\boldsymbol{x}}^{(j)}|\right)}{\max\left(\lambda \boldsymbol{w}(\bar{\boldsymbol{x}}^{(j)}),\ q^{(j)}+\sigma|D\bar{\boldsymbol{x}}^{(j)}|\right)}, \\ \boldsymbol{x}^{(j+1)} = \mathcal{P}_+\left(\bar{\boldsymbol{x}}^{(j)} - \tau M^T \begin{bmatrix} p^{(j+1)} \\ q^{(j+1)} \end{bmatrix}\right), \\ \bar{\boldsymbol{x}}^{(j+1)} = \boldsymbol{x}^{(j+1)} + \alpha\left(\boldsymbol{x}^{(j+1)} - \boldsymbol{x}^{(j)}\right). \end{cases} \tag{6}$$

In the following, we define the number of CP iterations as the total count of all iterations executed within the IRℓ_1 loop.

3.2. The Deep Guess step

The Deep Guess step essentially uses a convolutional neural network to refine a coarse reconstruction obtained by a fast algorithm, following a supervised learning approach.

Fig. 2 illustrates the main possible implementations of the Deep Guess step, that vary in the input and/or target images used to generate the starting guess image $\boldsymbol{x}^{(0)}$ for the non-convex MBIR. The three general versions are labeled as FBP-LPP, MBIR-LPP and MBIR-RISING, and they are respectively represented by the green, orange and red paths in the figure.

In all cases, the sinogram data are used to derive the coarse input image of the neural network. It is computed through a Filtered Back Projection (FBP) (first green arrow) or through a limited number K of iterations of an MBIR method (first orange and red arrows). The choice of MBIR method used in the DG step is independent of the one used in the final step of Fig. 1. In this paper, we have tested the two following MBIR methods in the Deep Guess step.

- A Scaled Gradient Projection (SGP) method (Bonettini et al., 2008; Porta et al., 2015; Loli Piccolomini et al., 2018) applied to the convex minimization problem:

  $$\min_{\boldsymbol{x}\geq 0} \|A\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \mu TV_{\beta}(\boldsymbol{x}) \tag{7}$$

  where TV_{β} is the differentiable TV prior defined as:

  $$TV_{\beta}(\boldsymbol{x}) := \sum_{i=1}^{n} \sqrt{\left(D_h\boldsymbol{x}\right)_i^2 + \left(D_v\boldsymbol{x}\right)_i^2 + \beta^2}, \tag{8}$$

  for some small scalar $\beta > 0$ and μ is a positive regularization parameter.
- The IRℓ_1 method utilizing CP iterations described in Section 3.1, applied to the non-convex TpV minimization:

  $$\min_{\boldsymbol{x}\geq 0} \|A\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \mu\|D\boldsymbol{x}\|_{2,p}^p, \tag{9}$$

  where the regularization parameter is denoted by μ to differentiate it from the parameter λ used in the final step of the framework.

Focusing on the network training, we first remark that the double-headed arrows in Fig. 2 denote the kind of target images used during the training. If available, the ground truth images can be used as target samples (see the green and orange double arrows). Instead, in the absence of ground truth images which is very common in real medical imaging, we implement the RISING scheme proposed in Evangelista et al. (2023a). In this case (red double-headed arrow), the targets are the images generated by the same MBIR method used to compute the input images but now running until convergence (purple path).

At last, in Table 1, we detail the five specific versions of the Deep Guess block that we use in the experimental part of the study. They all derive from the possible combinations of the aforementioned possible choices for the inner modules of the DG block.

Table 1
Recap of the Deep Guess blocks implemented and tested in the following experimental part. The first column refers to the general notation introduced in Fig. 2.

	Label	Input to the DG network	Target for the DG network
FBP-LPP	FBP-LPP	FBP	GT samples
MBIR-LPP	TV-LPP	K iterations of the SGP for solving (7).	GT samples.
	TpV-LPP	K iterations of the CP for solving (9).	GT samples.
MBIR-RISING	TV-RISING	K iterations of the SGP for solving (7).	Solutions at convergence of SGP for solving (7).
	TpV-RISING	K iterations of the CP for solving (9).	Solutions at convergence of CP for solving (9).

3.3. The neural network architecture

The deep neural network used in the Deep Guess step acts as an image-to-image operator in the image domain, as it takes as input a coarse reconstruction computed by the FBP or an early-stopped MBIR, and it outputs the initial guess $\boldsymbol{x}^{(0)}$ for the final non-convex MBIR solver.

Here, we use an architecture inspired by the popular UNet one, popularized by Ronneberger et al. (2015). Recent advancements in the field of image processing have shown that the UNet architecture performs well when the input image contains global artifacts because the pooling and unpooling mechanisms utilized in UNet enable an expansion of the convolutional filters' receptive field (Goodfellow et al., 2016). This allows the architecture to capture global information in the deeper layers of the network, while processing primarily local features in the higher layers. Consequently, the UNet scheme has become the standard choice for sparse-view tomographic imaging tasks, where streaking artifacts are present across the entire image, and the UNet scheme has been successfully applied for medical imaging, working on the image or on a wavelet-based image transformation (Kang et al., 2017; Ye et al., 2018).

In this paper, specifically, we utilize the residual UNet (ResUNet) architecture. As demonstrated in the theoretical study by Han et al. (2016) addressing sparse-view CT with FBP reconstructions, it can be shown that the residual manifold (which encapsulates the artifacts) is more readily learnable than the manifold representing the true image. In other words, for artifact suppression tasks, it could be more effective to learn the residual map:

$$\mathcal{M}_R : y \longmapsto y + \bar{x}, \tag{10}$$

than the correction map:

$$\mathcal{M}_C : y \longmapsto \bar{x}. \tag{11}$$

Residual deep networks have been already applied in the LPP step, to remove artifacts from low-dose or sparse-view FBP reconstructions; see for example (Han and Ye, 2018; Zhang et al., 2019; Morotti et al., 2021; Evangelista et al., 2023a).

More in details, we adopt the residual UNet architecture illustrated in Fig. 3, which is inspired by the design proposed in Wang et al. (2019). More in details, the ResUNet is a fully convolutional neural network with a symmetric encoder–decoder structure and pooling/unpooling operators to enlarge its receptive field. The pooling operations in the encoder naturally divides the network in distinct levels of resolution, to which we will refer as l, $l = 0, \dots, L$, where $L+1$ is the total number of levels in the network. At each level, a fixed number n_l of convolutional filters is applied, each one with the same number of channels c_l, which is constant along the level. Given a baseline number of convolutional channels c_0, corresponding to the number of channels at the first level, the number of channels at subsequent levels, c_l, is computed using the recursive formula $c_{l+1} = 2c_l$, $l = 0, \dots, L-1$. In our implementation, we set $L = 4$, $n_0 = \cdots = n_3 = 3$ and $c_0 = 64$. As previously mentioned, the decoder is designed to be symmetric to the encoder, with upsampling layers replacing pooling layers. Additionally, to preserve high-frequency information, skip connections are incorporated between the final layer of each encoder level and the initial layer of the corresponding decoder level. To reduce the number of parameters compared to the original architecture proposed in Wang et al. (2019), these skip connections are implemented as element-wise additions rather than the more commonly used concatenations.

The network is trained using a loss function that represents the Mean Squared Error (MSE) between the network predictions and the corresponding target values. The Adam optimizer is utilized with a learning rate of 0.001, and the training process is conducted over a total of 50 epochs.

It is noteworthy that the forward pass of a pre-trained neural network is characterized by very short computational times. Consequently, the entire Deep Guess step can be executed rapidly. At last, we remark again the described post-processing approaches are not new in the literature, but so far, they have been used as solvers for the CT inverse problem. With their usage in the Deep Guess acceleration scheme, we leverage the state-of-the-art literature strongly relying on deep learning as an efficient tool inside a more interpretable mathematical framework, providing a variational solution as the final reconstruction.

4. Numerical experiments and discussion

In this Section we present the numerical experiments performed on test problems with both real and synthetic medical images. The code to replicate the experiments can be found at: https://github.com/devangelista2/DeepGuess. Here, we first describe the datasets and test problems utilized in our experiments. Subsequently, we analyze the results obtained on real medical images using the proposed approach, comparing them to other methods. Finally, we conduct a more detailed examination of the possible implementations of the Deep Guess step and its stability properties. The labels used to denote the different DG steps are those defined in Table 1.

We also clarify that the number of iterations reported for the CP method represents the total number of iterations carried out during the IRℓ_1 scheme, calculated as the sum of CP iterations across all iterations of the IRℓ_1 procedure.

The metrics used for comparison are the Relative Error (RE) between the computed reconstruction and the ground truth and the Structural Similarity Index (SSIM) (Wang et al., 2003), whose values are scaled in the range $[0, 100]$.

4.1. The datasets

We considered both a dataset of real CT images and synthetic images.

For each ground truth image $\boldsymbol{x}^{GT}$ in a dataset, we sampled a noise realization $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, I)$, and we computed:

$$\boldsymbol{y} = A\boldsymbol{x}^{GT} + \nu \frac{\|A\boldsymbol{x}^{GT}\|}{\|\boldsymbol{e}\|}\boldsymbol{e}, \tag{12}$$

where $\nu \geq 0$ is the noise level. The matrix A has been computed with Astra Toolbox (Van Aarle et al., 2015, 2016) functions, simulating a sparse fan beam geometry with n_{angle} projections uniformly distributed within the angular range $(0, n_{\text{range}})$ degrees, with a detector resolution of $2\sqrt{n}$ pixels. In the following, we indicate this setup as the tomographic geometry $\mathcal{G}_{n_{range}, n_{angle}}$.

The dataset of real CT images considered is the Mayo Clinic dataset (McCollough, 2016), constituted of 512 × 512 pixel real images of human abdomen. We used 3306 images for training and 357 for testing. We generated projections as in (12) on a flat detector composed of 1024 pixels, using different geometries. In all the experiments, we

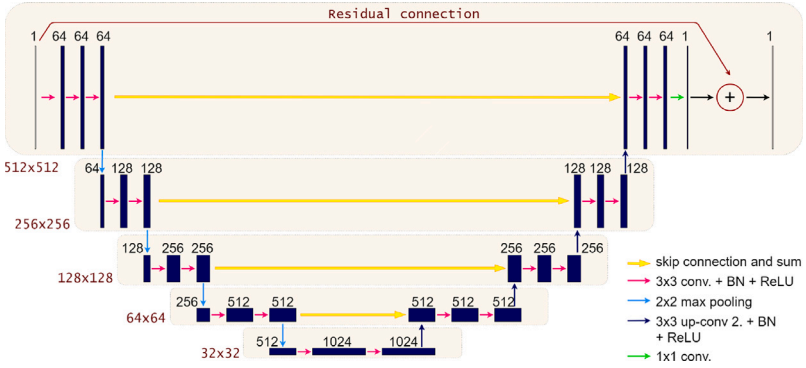

Fig. 3. Draft of the ResUNet architecture. It is a fully-convolutional neural network with a symmetric encoder–decoder structure and pooling–unpooling operators to enlarge the receptive field.

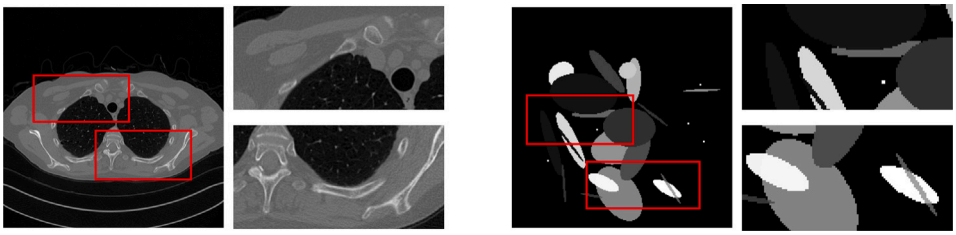

Fig. 4. Ground truth images from the test set of Mayo Clinic (left) and COULE (right), with two zooms-in on the regions depicted with red rectangles on the corresponding whole images.

added noise with $\nu = 0.001$. An example image is presented in Fig. 4, accompanied by two zoomed-in views of regions of interest.

The considered synthetic data set is the Contrasted Overlapping Uniform Lines and Ellipses (COULE) dataset, downloadable from www.kaggle.com/loiboresearchgroup/coule-dataset. It contains 430 synthetic sparse-gradient gray-scale images of size 256 × 256 with many overlying objects, varying in size and contrast with respect to the background. Synthetic images are ideal representations of objects, free from any noise or artifacts. An example of a COULE image is displayed in Fig. 4, together with two zoomed crops of interest. We generated projections as in (12) on a flat detector composed of 512 pixels, using the geometry $\mathcal{G}_{180,180}$. Noise has been added with $\nu = 0.01$. We divided the dataset into a training set comprising 400 images and a test set containing 30 images.

Both with the Mayo Clinic and the COULE datasets, we used the available ground truth images for training the neural network and getting the LPP Deep Guess images (corresponding to the orange and green final arrows, in Fig. 2). We also generated the Mayo Clinic and COULE datasets of MBIR convergent solutions, to train the DG network with the RISING principle (corresponding to the final red arrow in Fig. 2). In this case, the regularization parameters have been set by trial and error on the training samples to minimize the mean of the Relative Errors of the training set.

4.2. Preliminary results with different MBIR approaches

In our preliminary experiments, we addressed the CT image reconstruction problem using various MBIR approaches and different geometries, to motivate the selection of the non-convex T*p*V model in all subsequent tests. We denote our T*p*V-regularized model solved by the CP algorithm with T*p*V(CP), in the following.

The first alternative MBIR solver that we consider is the Scaled Gradient Projection (SGP) algorithm applied to a TV_β-regularized model (with $\beta = 10^{-3}$), as previously discussed in Section 3.2. It is named as TV(SGP) in the following.

The second choice is the Fast Iterative Shrink-Thresholding Algorithm (FISTA) applied to compute the optimal sparse wavelet (W) coefficients by solving the following minimization problem (Beck and Teboulle, 2009a,b):

$$\min_{\boldsymbol{c}\in\mathbb{R}^n} \|\bar{A}\boldsymbol{c} - \boldsymbol{y}\|_2^2 + \lambda\|c\|_1, \tag{13}$$

where the matrix $\bar{A} = A\mathcal{W}^T$ is given by the product of the projection matrix A and of the matrix $\mathcal{W}^T \in \mathbb{R}^{n\times n}$, representing the synthesis operator of a wavelet transform, and $\boldsymbol{c}$ represents the wavelet coefficients of the solution $\boldsymbol{x} = \mathcal{W}^T\boldsymbol{c}$. In the following, we will denote this approach as W(FISTA).

A further alternative is the state-of-the-art Enhanced ASD-POCS (Adaptive Steepest Descent with Projections Onto ConvexSets) recently proposed in Bappy et al. (2024). It boosts the popular ASD-POCS algorithm, which integrates ART iterations with gradient descent steps targeting the TV prior, with an unsharp masking strategy to smooth out artifacts from the filtered ART reconstruction. We denote it Mask(EASD-POCS).

All the algorithms are stopped after a prefixed number of iterations, having ascertained the reconstruction errors did not change significantly anymore. The regularization parameters have been set to minimize the RE on the test image, for each method. The results obtained on the test image previously displayed are shown in Table 2, where we also report for comparison the metrics of the traditional FBP method. From the table, it is immediately apparent that FBP performs very poorly, particularly for the sparsest geometry, denoting the necessity for MBIR approaches in these scenarios. In the settings of $\mathcal{G}_{180,180}$, the MBIR methods exhibit very similar performance. However, in cases with very few views, the T*p*V model significantly outperforms the others, exceeding approximately the TV model by 10%, the masked approach by 10% and the ℓ_1 wavelet by about 30%, in terms of both RE and SSIM. Moreover, it is notable that all the MBIR approaches are quite slow, necessitating hundreds of iterations to achieve the final reconstructions. Therefore, to make these methods applicable in real-world cases, it becomes of primary importance to reduce the number of iterations and, consequently, the computational time.

Fig. 5 displays the reconstructions of the chosen image obtained in the quite sparse case $\mathcal{G}_{180,60}$ and computed, from left to right, by

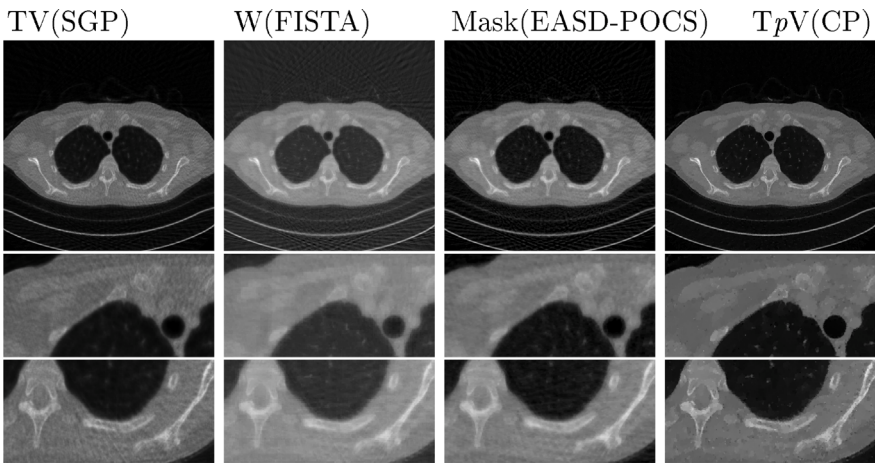

Fig. 5. Results on the Mayo test image shown in Fig. 4, computed using the TV prior (first column), the W regularizer (second column), the enhanced masked model (third column), and the T*p*V prior (fourth column), in case of $\mathcal{G}_{180,60}$ sparse protocol. Their SSIM and RE values are reported in Table 2.

Table 2
Values of the quantitative metrics, computed on the Mayo Clinic test image reconstructions, for different analytical and variational methods.

		SSIM	RE	iters
$\mathcal{G}_{180,180}$	FBP	58.84	0.1466	–
	TV (SGP)	**87.99**	**0.0662**	300
	W (FISTA)	86.43	0.0726	500
	Mask (EASD-POCS)	85.14	0.0778	300
	T*p*V (CP)	86.60	0.0699	500
$\mathcal{G}_{180,120}$	FBP	47.99	0.1855	–
	TV (SGP)	83.43	0.0835	300
	W (FISTA)	79.05	0.0981	500
	Mask (EASD-POCS)	79.05	0.0990	300
	T*p*V (CP)	**83.95**	**0.0803**	500
$\mathcal{G}_{180,60}$	FBP	30.56	0.2955	–
	TV (SGP)	74.81	0.1190	300
	W (FISTA)	63.25	0.1517	500
	Mask (EASD-POCS)	67.46	0.1401	300
	T*p*V (CP)	**78.08**	**0.1039**	500
$\mathcal{G}_{180,30}$	FBP	18.12	0.4698	–
	TV (SGP)	66.65	0.1646	300
	W (FISTA)	51.70	0.2093	500
	Mask (EASD-POCS)	58.28	0.1909	300
	T*p*V (CP)	**71.11**	**0.1464**	500

the TV(SGP), W(FISTA), Mask(EASD-POCS) and T*p*V(CP) MBIR approaches. The T*p*V image is the most accurate and clean, whereas artifacts are clearly visible in the TV and wavelet ones, according to the metrics in Fig. 5.

4.3. Results with the proposed Deep Guess acceleration framework

Now, we present the results obtained by the proposed Deep Guess accelerated framework on the Mayo Clinic test image. We consider the $\mathcal{G}_{180,60}$ geometry and set $K = 15$ in the DG step, to compute the input of the neural network with an early-stopped MBIR. Concerning the regularization parameter μ in (7) and (9), it is chosen to optimize the CP Relative Error when the initial estimate, $\boldsymbol{x}^{(0)}$, is set to the null image, and its performance remains unaffected by the choice of a different initial guess. The following comparison includes results obtained using different initial guesses $\boldsymbol{x}^{(0)}$ for the T*p*V(CP) method, specifically: a null vector, the analytical FBP solution, the output of K iterations of SGP applied as in TV(SGP), and two implementations of the Deep Guess. Additionally, we compare our DG accelerated framework with some state-of-the-art data-driven approaches for CT image reconstruction. Firstly, we consider the FBP-LPP and TV-RISING approaches themselves, to be conceived not only as DG accelerators but also as final reconstructions. Furthermore, among the various methods available in the literature, we also consider the NETT algorithm (Li et al., 2020), which is popular for its mathematically explainable model combining deep neural networks with a Tikhonov regularization strategy within a convergent scheme. At last, we test the Learned Primal Dual (LPD) unrolling approach (Adler and Öktem, 2018), implementing a proximal primal–dual optimization method where the proximal operators have been replaced with convolutional neural networks.

Focusing on the experiments performed with sparse tomographic acquisitions relative to the $\mathcal{G}_{180,60}$ geometry, we report in Table 3 the metrics obtained by TpV with both $p = 0.5$ and $p = 0.2$. At first glance, the value of p does not have a significant impact on the results. In all cases, initializing the CP iterations with the proposed Deep Guess substantially reduces the total number of iterations required and consequently decreases the computational time by a factor of 200–400. Additionally, it enhances the accuracy of our comparing methods FBP-LPP, TV-RISING, NETT and LPD.

As shown in Fig. 6, depicting the results by $p = 0.2$, the CP reconstruction obtained with null initial guess starting still contains artifacts and noise, and the low-contrast regions appear blurred. The images generated using the Deep Guess approach are already of high visual quality. However, analysis of the residual maps from the zoomed-in regions reveals that the final reconstruction further enhances the Deep Guess images. Notably, in low-contrast regions, errors in object contours observed in the Deep Guess images are absent in the final reconstructions. Even if there are no guarantees that T*p*V(CP) has ever reached a global minimum in these experiments, the solutions achieved from the DG starting iterates are for sure much closer to the global optimum than the other solutions and, at least, good local minima.

Finally, we have reported in the last row of the figure the Deep Guess images computed by the neural network. We underline that these images are generated by state-of-the-art methods for CT reconstruction, and our proposal demonstrates that these methods can be successfully exploited to achieve higher-quality images with limited computational effort.

To better validate our proposal with respect to the data-driven state-of-the-art methods, in Fig. 7 we show further results achieved on the Mayo Clinic test images by the considered approaches. As visible from the ground truth images in the first row, these test samples show different anatomical regions. The first row also highlights zoomed-in regions of interest with the red rectangles on the ground truth images. The subsequent rows display results obtained using the NETT, LPD, and FBP-LPP data-driven techniques, followed by the proposed approach in the last row. The SSIM values, reported above each image, quantitatively assess reconstruction quality.

Table 3
Values of the quantitative metrics, computed on the Mayo Clinic test image reconstructions for the $\mathcal{G}_{180,60}$ protocol, with different initial guesses $\boldsymbol{x}^{(0)}$ for the CP and for $p = 0.5$ and $p = 0.2$, and with state-of-the-art method.

	$\boldsymbol{x}^{(0)}$			Output		
		SSIM	RE	SSIM	RE	iters
	zeros	–	–	78.08	0.1039	500
TpV(CP)	FBP	30.56	0.2955	77.01	0.1076	400
with	TV ($K = 15$)	67.18	0.1906	75.36	0.1147	400
$p = 0.5$	*DG* by FBP-LPP	**80.74**	**0.0876**	**83.16**	**0.0792**	12
	DG by TV-RISING	76.99	0.1059	78.79	0.0989	19
	zeros	–	–	77.03	0.1071	500
TpV(CP)	FBP	30.56	0.2955	76.35	0.1102	400
with	TV ($K = 15$)	67.18	0.1906	76.43	0.1093	400
$p = 0.2$	*DG* by FBP-LPP	**80.74**	**0.0876**	**82.75**	**0.0794**	18
	DG by TV-RISING	76.99	0.1059	79.65	0.0958	36
NETT (Li et al., 2020)	–	–	–	61.12	0.1576	300
LPD (Adler and Öktem, 2018)	–	–	–	79.53	0.1033	10

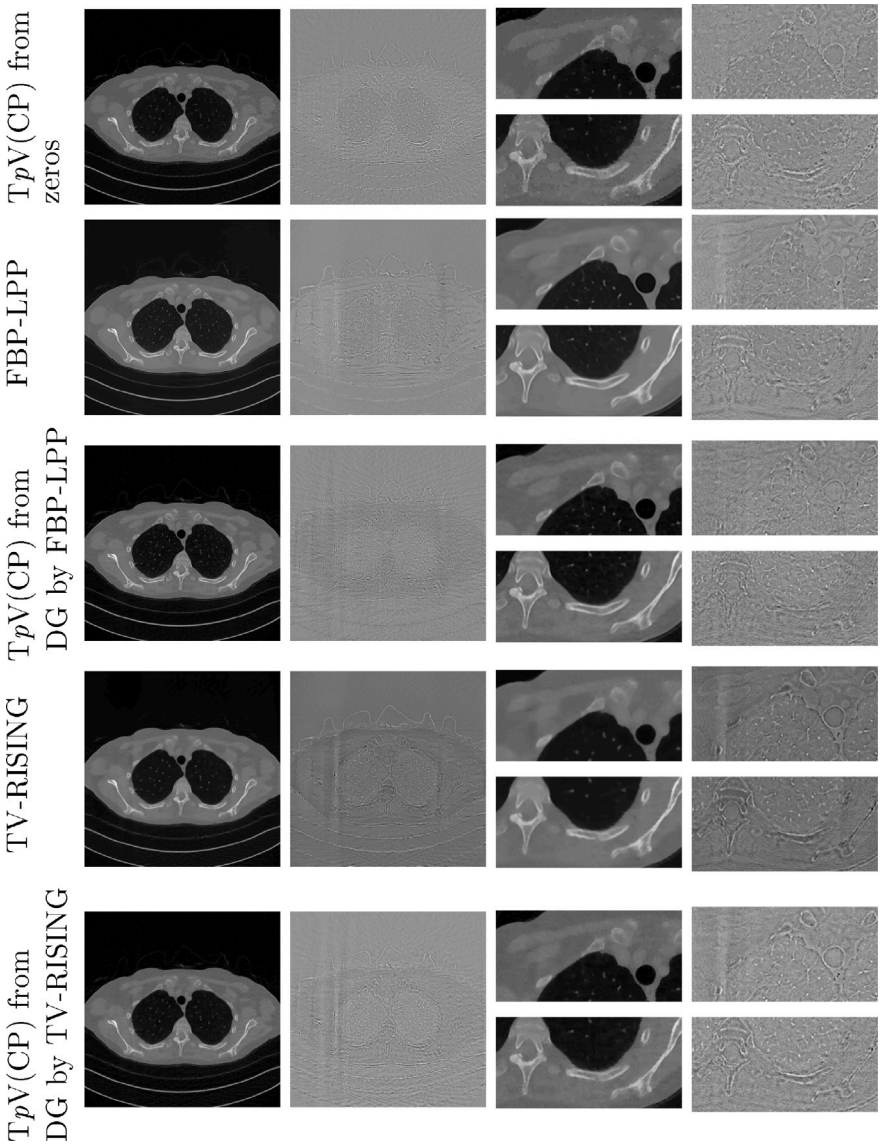

Fig. 6. Reconstructions and residual maps (obtained as the difference between the reconstructed and the ground truth images), computed using the TpV(CP) ($p = 0.2$, $\mathcal{G}_{180,60}$) algorithm with different initial guess: null (first row) and the proposed FBP-LPP and TV-RISING Deep Guess approaches (third and fifth row), together with the correspondent Deep Guess images (second and fourth row). The SSIM and RE values of the reconstructed images are reported in Table 3. All the residual maps are depicted using the same colormap.

Among the tested methods, NETT still exhibits visible artifacts and reduced structural fidelity in certain regions. The LPD and FBP-LPP methods enhance reconstruction accuracy, as reflected in increased SSIM scores. Our proposed approach, here implemented as the TpV(CP) using $p = 0.5$ and with DG image by the FBP-LPP procedure, achieves the highest SSIM values across all cases, demonstrating superior image

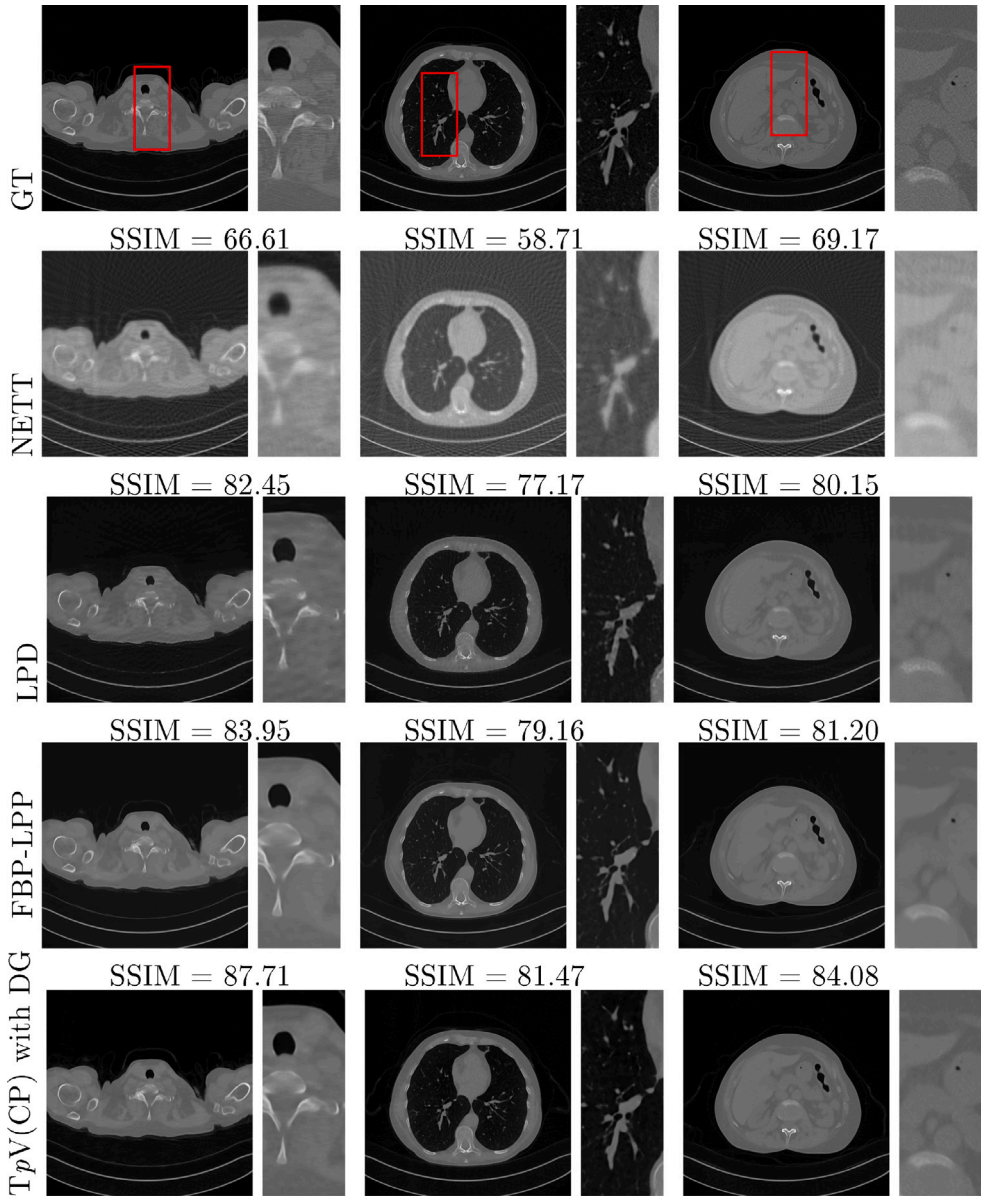

Fig. 7. Results on further Mayo Clinic test images, computed within the sparse $G_{180,60}$ tomographic protocol. The ground truth images are on the first row, with red rectangles depicting the zoomed regions. The solutions provided by state-of-the-art data-driven methods are reported in each row. The solutions by our proposal (bottom row) have been achieved by the TpV(CP) method with $p = 0.5$ and Deep Guess by the FBP-LPP solver (fourth row). The SSIM metric is reported for all images.

quality with SSIM values always higher than 81. The zoomed-in regions confirm that our method avoids sparse-view typical artifacts and pattern, preserves fine details and improves the edge detectability of small contrast structures more effectively than the alternatives. These results validate the effectiveness of our approach in reducing reconstruction artifacts while maintaining high structural similarity with the ground truth, making it a promising solution for sparse-view CT imaging.

4.4. Analysis of different Deep Guess implementations

In this Section, we analyze and compare various possible implementations of the Deep Guess step on the COULE test set. Here, we set $p = 0.5$ and $K = 10$ for both TV- and TpV-based Deep Guess accelerations. The regularization parameter $\lambda = 10^{-3}$ was heuristically determined to minimize the Relative Error when the CP solver is applied from the null $\boldsymbol{x}^{(0)}$ image. This parameter remains unchanged for all the experiments reported in this Section.

Table 4 presents the results obtained on the COULE test set. All SSIM and RE values are averages calculated across the test samples, with standard deviations provided in parentheses. The first two columns pertain to the Deep Guess image $\boldsymbol{x}^{(0)}$, the subsequent metrics correspond to the final reconstructions, and the last column indicates the range of iterations required for the CP method to converge in the second step of our scheme.

We observe that the ground truth-free RISING strategies for computing Deep Guess perform very well, losing, in the final output, only one point of SSIM compared to the corresponding LPP approaches (96.17 against 97.32 for the TV approach and 93.01 against 94.54 for the TpV one). Additionally, the TV-LPP and TV-RISING frameworks outperform their TpV counterparts, offering higher metrics and, more importantly, smaller standard deviations, making them more reliable. Regarding the results limited to the Deep Guess step to compute $\boldsymbol{x}^{(0)}$ (in the left part of the table), we note that TpV-RISING performs significantly worse than the others. We attribute this behavior to the potential use of poor local minima as targets during network training.

In Fig. 8, we present the reconstruction (together with the two considered crops) of the test image displayed in Fig. 4. The considered

Table 4
Mean values (standard deviations) of the quantitative metrics, computed on the COULE test images. The first two columns present the metric values for the computed image $\boldsymbol{x}^{(0)}$ (DG step); on the right, the metrics for the final reconstructed image by the DG accelerated T*p*V(CP) are shown.

	$\boldsymbol{x}^{(0)}$		Output		
	SSIM	RE	SSIM	RE	iters
zeros	–	–	88.70 (3.37)	0.0851 (0.0073)	200
DG by FBP-LPP	90.02 (4.19)	0.0642 (0.0163)	96.20 (2.34)	0.0321 (0.0057)	45-55
DG by TV-LPP	87.45 (3.28)	0.0527 (0.0074)	97.32 (1.03)	0.0290 (0.0025)	20-32
DG by TV-RISING	90.79 (2.60)	0.0619 (0.0105)	96.17 (2.49)	0.0341 (0.0047)	47-55
DG by T*p*V-LPP	93.80 (2.07)	0.1372 (0.0517)	94.54 (4.23)	0.0530 (0.0120)	75-80
DG by T*p*V-RISING	83.78 (10.84)	0.1434 (0.0515)	93.01 (4.49)	0.0538 (0.0142)	73-85

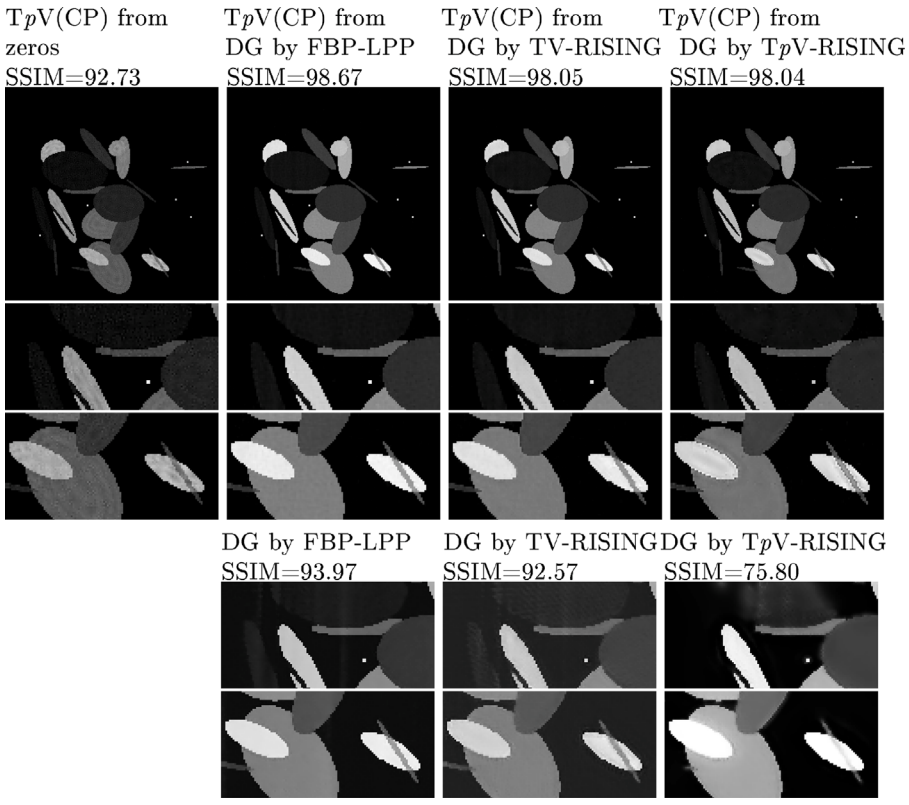

Fig. 8. Results on the COULE test image shown in Fig. 4, computed using the CP algorithm from the null image (first column) and some of the proposed Deep Guess acceleration approaches. The last row displays the zooms of the images generated by the Deep Guess neural networks within the proposed framework. The SSIM metric is reported for all images.

solutions are those computed from the zero-image (first column) and from the Deep Guess computed by the FBP-LPP (second column), the TV-RISING (third column) and the T*p*V-RISING (last column) networks. The corresponding Deep Guess images are displayed at the bottom.

The left image exhibits ringing artifacts, particularly in the homogeneous regions, as clearly shown in the two magnified zooms-in. In contrast, all the other final solutions computed from the DG images are clear and accurate. In the context of the DG images $\boldsymbol{x}^{(0)}$ displayed in the final rows, the T*p*V-RISING method exhibits the lowest accuracy, evidenced by blurred ellipses and the loss of low-contrast objects, aligning with the metrics reported in Table 4. Although the FBP-LPP image achieves the highest SSIM value, it displays streaking artifacts in low-contrast regions, a common issue associated with analytical reconstructions in sparse CT.

To assess the stability with respect to the noise of the DG blocks and their corresponding final solutions, we consider sinograms affected by noise at varying levels. Now, we only focus on the FBP-LPP, TV-LPP, and the TV-RISING schemes. In Fig. 9, we plot the SSIM values as a function of the noise level ν ranging from 0 (noiseless data) to 0.05. The results are relative to the previously considered test image, and we use the same parameter settings tuned for the previous experiments so that the tests shown in Fig. 8 correspond to the case where the noise level is 0.01 in this plot. We plot the SSIM both for the DG image (dotted lines with star markers) and for the final reconstructions (dashed lines with circle markers). As observed in Evangelista et al. (2023a), the RISING network (red dotted line) is slightly less accurate, when tested on the same noise used for training, but more stable than the networks trained on the ground truth images (FBP-LPP and TV-LPP networks described by green and yellow dotted lines, respectively). In fact, when the noise level is similar to that analyzed during the training phase (i.e. $\nu = 0.01$), the LPP generally provides higher values. However, there is a turning point around the noise level of $\nu = 0.03$, where the noise perturbation significantly impacts the data, and RISING outperforms LPP (as the red dotted line surpasses the yellow and orange ones). Regarding the final reconstructions, it is evident that the DG accelerated framework consistently outperforms other methods across all the experiments considered (indicated by the red and orange dashed lines). Notably, even though the CP solver initialized from a zero image (represented by the purple dashed line) achieves SSIM values around 90 for low noise levels, its performance does not significantly degrade under conditions of high noise.

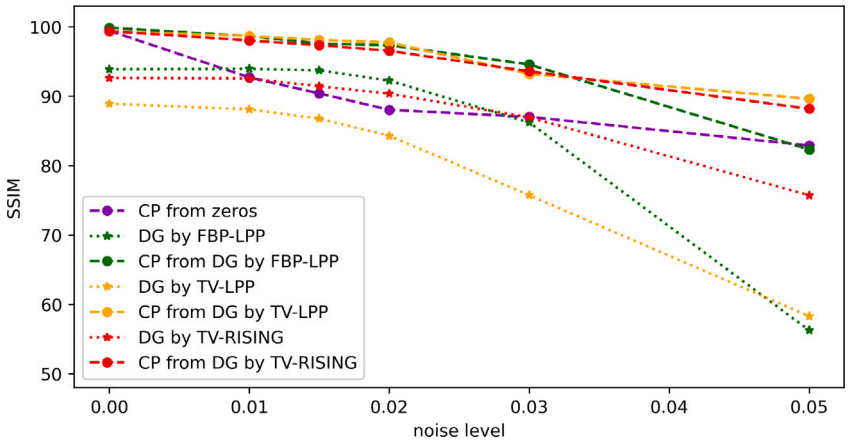

Fig. 9. Plot of the SSIM values for the COULE test image reconstructions at different noise levels. Dotted lines with star markers represent the outputs of the Deep Guess neural networks, while dashed lines with circle markers correspond to the final images produced by the proposed framework and its competitors.

5. Conclusion

This paper introduces a novel hybrid framework integrating a deep learning tool in an iterative model-based regularized approach for reconstructing CT images from sparse views. The framework addresses a non-convex model that enforces sparsity in the image gradient domain, and employs deep learning to achieve accurate image restoration faster, with minimal computational effort.

The proposed approach is composed of two steps. The first one is the Deep Guess one, computing a clever initial guess for the second step. The Deep Guess block builds upon state-of-the-art efficient methods, and it generates output images using a convolutional neural network applied to a very coarse reconstruction. In the second step, we run an IRℓ_1 scheme based on CP iterations for a TpV-based regularized model, to compute the final reconstruction ensuring data coherence.

We propose several implementations of Deep Guess, including those that do not rely on ground-truth data, thereby improving the practical applicability of our framework for real-world scenarios.

Experiments conducted on test problems using real CT images across various sparse geometries demonstrate the effectiveness of our approach. It surpasses other methods based on the same algorithm with different initial iterates, as well as the state-of-the-art hybrid methods (that merge variational and data-driven approaches) in terms of both accuracy and computational complexity, being 15 to 25 times faster. Since the considered model is non-convex, we have no guarantee that the obtained solution is a global minimum. However, since the solution computed using Deep Guess is highly accurate, we can reasonably assume that the (possible local) minimum reached by the algorithm, when initialized with Deep Guess, corresponds to an image very close to the global minimum. A thorough analysis of the stability of the proposed framework with respect to data noise demonstrates that the ground-truth-free implementations maintain strong generalization performance even as noise levels increase.

We emphasize the versatility of our framework, which can be tailored to specific tasks in both Deep Guess computation and the selection of MBIR methods for final image reconstruction, as depicted in Fig. 1. Additionally, our study demonstrates that deep learning tools can serve as a useful foundation for model-based approaches. By integrating these tools, we can accelerate computations while preserving the mathematical properties and characterizations of the final reconstructions.

CRediT authorship contribution statement

Elena Loli Piccolomini: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Davide Evangelista:** Writing – review & editing, Validation, Software, Formal analysis. **Elena Morotti:** Writing – review & editing, Writing – original draft, Validation, Methodology, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Elena loli piccolomini reports financial support was provided by European Commission under the NextGeneration EU programme. 24. Elena Morotti reports financial support was provided by Francesco Severi National Institute of Higher Mathematics National Group of Scientific Calculations. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements and funding

This work has been partially supported by Gruppo Nazionale per il Calcolo Scientifico (GNCS-INdAM), Italy within the project “Deep Variational Learning: un approccio combinato per la ricostruzione di immagini”, and by the PRIN 2022 project “STILE: Sustainable Tomographic Imaging with Learning and rEgularization”, project code: 20225STXSB, funded by the European Commission under the NextGeneration EU programme.

Data availability

I have shared my code and data on a GitHub repository indicated in the paper.

References

Adler, J., Öktem, O., 2018. Learned primal–dual reconstruction. IEEE Trans. Med. Imaging 37 (6), 1322–1332.

Antun, V., Renna, F., Poon, C., Adcock, B., Hansen, A.C., 2020. On instabilities of deep learning in image reconstruction and the potential costs of AI. Proc. Natl. Acad. Sci. 117 (48), 30088–30095.

Baguer, D.O., Leuschner, J., Schmidt, M., 2020. Computed tomography reconstruction using deep image prior and learned reconstruction methods. Inverse Problems 36 (9), 094004.

Bappy, D., Kang, D., Lee, J., Baek, H., 2024. Prior-based enhanced asd-pocs for artifact suppression and structural preservation in sparse-view cbct. In: Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing. pp. 496–505.

Beck, A., Teboulle, M., 2009a. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2 (1), 183–202.

Beck, A., Teboulle, M., 2009b. A fast iterative shrinkage-thresholding algorithm with application to wavelet-based image deblurring. In: 2009 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 693–696.

Boink, Y.E., Brune, C., 2019. Learned svd: solving inverse problems via hybrid autoencoding. arXiv preprint arXiv:1912.10840.
Bonettini, S., Zanella, R., Zanni, L., 2008. A scaled gradient projection method for constrained image deblurring. Inverse Problems 25 (1), 015002.
Borsdorf, A., Raupach, R., Flohr, T., Hornegger, J., 2008. Wavelet based noise reduction in ct-images using correlation analysis. IEEE Trans. Med. Imaging 27 (12), 1685–1703.
Candès, E.J., Romberg, J., Tao, T., 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory.
Candès, E.J., Wakin, M.B., Boyd, S.P., 2008. Enhancing sparsity by reweighted L1 minimization. J. Fourier Anal. Appl..
Cascarano, P., Loli Piccolomini, E., Morotti, E., Sebastiani, A., 2022. Plug-and-play gradient-based denoisers applied to CT image enhancement. Appl. Math. Comput..
Caselles, V., Chambolle, A., Novaga, M., 2015. Total variation in imaging.. In: Handbook of Mathematical Methods in Imaging.
Chambolle, A., Pock, T., 2011. A first-order primal–dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vision.
Chan, T., et al., 2005. Recent developments in total variation image restoration. Math. Model. Comput. Vis..
Chartrand, R., 2007. Exact reconstruction of sparse signals via nonconvex minimization. IEEE Signal Process. Lett..
Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J., Wang, G., 2017. Low-dose ct with a residual encoder–decoder convolutional neural network. IEEE Trans. Med. Imaging 36 (12), 2524–2535.
Daubechies, I., DeVore, R., Fornasier, M., Güntürk, C.S., 2010. Iteratively reweighted least squares minimization for sparse recovery. Commun. Pure Appl. Math.: J. Issued By Courant Inst. Math. Sci. 63 (1), 1–38.
Demirel, G., Yildirim, İ., Ertaş, M., 2021. P norm regularized breast tomosynthesis image reconstruction. In: 2021 Medical Technologies Congress. TIPTEKNO, IEEE, pp. 1–4.
Donoho, D.L., 2006. Compressed sensing. In: IEEE Transactions on Information Theory.
Evangelista, D., Morotti, E., Loli Piccolomini, E., 2023a. RISING: A new framework for model-based few-view CT image reconstruction with deep learning. Comput. Med. Imaging Graph..
Evangelista, D., Morotti, E., Piccolomini, E.L., Nagy, J., 2023b. Ambiguity in solving imaging inverse problems with deep-learning-based operators. J. Imaging 9 (7), 133.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
Han, Y., Ye, J.C., 2018. Framing U-Net via deep convolutional framelets: Application to sparse-view CT. IEEE Trans. Med. Imaging 37 (6), 1418–1429.
Han, Y.S., Yoo, J., Ye, J.C., 2016. Deep residual learning for compressed sensing ct reconstruction via persistent homology analysis. arXiv preprint arXiv:1611.06391.
Jin, K.H., McCann, M.T., Froustey, E., Unser, M., 2017. Deep convolutional neural network for inverse problems in imaging. IEEE Trans. Image Process. 26 (9), 4509–4522.
Kang, E., Min, J., Ye, J.C., 2017. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. Med. Phys. 44 (10), e360–e375.
Li, H., Schwab, J., Antholzer, S., Haltmeier, M., 2020. Nett: Solving inverse problems with deep neural networks. Inverse Problems 36 (6), 065005.
Loli Piccolomini, E., Coli, V.L., Morotti, E., Zanni, L., 2018. Reconstruction of 3d X-ray CT images from reduced sampling by a scaled gradient projection algorithm. Comput. Optim. Appl..
Loli Piccolomini, E., Morotti, E., 2021. A model-based optimization framework for iterative digital breast tomosynthesis image reconstruction. J. Imaging.
Loli Piccolomini, E., et al., 2016. A fast total variation-based iterative algorithm for digital breast tomosynthesis image reconstruction. J. Algorithms Comput. Technol..
Lv, D., Zhou, Q., Choi, J.K., Li, J., Zhang, X., 2020. Nonlocal tv-gaussian prior for bayesian inverse problems with applications to limited ct reconstruction.. Inverse Probl. Imaging 14 (1).
McCollough, C., 2016. Tu-fg-207a-04: Overview of the low dose ct grand challenge. Med. Phys. 43 (6Part35), 3759–3760.
Mehranian, A., Ay, M.R., Rahmim, A., Zaidi, H., 2013. X-ray ct metal artifact reduction using wavelet domain l_{0} sparse regularization. IEEE Trans. Med. Imaging 32 (9), 1707–1722.
Monga, V., Li, Y., Eldar, Y.C., 2021. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. IEEE Signal Process. Mag..
Morotti, E., Evangelista, D., Sebastiani, A., Piccolomini, E.L., 2024. Space-variant total variation boosted by learning techniques in few-view tomographic imaging. arXiv preprint arXiv:2404.16900.
Morotti, E., et al., 2021. A green prospective for learned post-processing in sparse-view tomographic reconstruction. J. Imaging.
Persson, M., Bone, D., Elmqvist, H., 2001. Total variation norm for three-dimensional iterative reconstruction in limited view angle tomography. Phys. Med. Biol. 46 (3), 853.
Porta, F., Prato, M., Zanni, L., 2015. A new steplength selection for scaled gradient methods with application to image deblurring. J. Sci. Comput. 65 (3), 895–919.
Purisha, Z., Rimpeläinen, J., Bubba, T., Siltanen, S., 2017. Controlled wavelet domain sparsity for x-ray tomography. Meas. Sci. Technol. 29 (1), 014002.
Ritschl, L., et al., 2011. Improved total variation-based CT image reconstruction applied to clinical data. Phys. Med. Biol..
Rodríguez, P., 2013. Total variation regularization algorithms for images corrupted with different noise models: a review. J. Electr. Comput. Eng. 2013 (1), 217021.
Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241.
Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. Phys. D: Nonlinear Phenom. 60 (1–4), 259–268.
Schnurr, A.-K., Chung, K., Russ, T., Schad, L.R., Zöllner, F.G., 2019. Simulation-based deep artifact correction with convolutional neural networks for limited angle artifacts. Z. Med. Phys. 29 (2), 150–161.
Shu, Z., Entezari, A., 2022. Sparse-view and limited-angle ct reconstruction with untrained networks and deep image prior. Comput. Methods Programs Biomed. 226, 107167.
Sidky, E.Y., Jørgensen, J.H., Pan, X., 2012. Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm. Phys. Med. Biol..
Sidky, E.Y., Pan, X., 2008. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Phys. Med. Biol. 53 (17), 4777.
Sidky, E.Y., et al., 2014. Constrained TpV minimization for enhanced exploitation of gradient sparsity: Application to CT image reconstruction. IEEE J. Transl. Eng. Heal. Med..
Sidky, E.Y., et al., 2021. Do CNNs solve the CT inverse problem? IEEE Trans. Biomed. Eng..
Tang, X., Huang, B., Xiao, G., 2023. Enhancing precision and interpretability of ct image reconstruction via self-supervised adaptive domain transformation. In: 2023 IEEE International Conference on Bioinformatics and Biomedicine. BIBM, IEEE, pp. 1451–1458.
Tian, Z., et al., 2011. Low-dose CT reconstruction via edge-preserving total variation regularization. Phys. Med. Biol..
Tong, S., Han, B., Tang, J., 2018. Edge-guided tvp regularization for diffuse optical tomography based on radiative transport equation. Inverse Problems 34 (11), 115009.
Van Aarle, W., Palenstijn, W.J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., De Beenhouwer, J., Joost Batenburg, K., Sijbers, J., 2016. Fast and flexible x-ray tomography using the astra toolbox. Opt. Express 24 (22), 25129–25147.
Van Aarle, W., Palenstijn, W.J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K.J., Sijbers, J., 2015. The astra toolbox: A platform for advanced algorithm development in electron tomography. Ultramicroscopy 157, 35–47.
Venkatakrishnan, S.V., Bouman, C.A., Wohlberg, B., 2013. Plug-and-play priors for model based reconstruction. In: 2013 IEEE Global Conference on Signal and Information Processing. IEEE, pp. 945–948.
Vogel, C.R., Oman, M.E., 1996. Iterative methods for total variation denoising. SIAM J. Sci. Comput..
Wang, J., 2021. Non-convex lp regularization for sparse reconstruction of electrical impedance tomography. Inverse Probl. Sci. Eng. 29 (7), 1032–1053.
Wang, S., Selesnick, I., Cai, G., Feng, Y., Sui, X., Chen, X., 2018. Nonconvex sparse regularization and convex optimization for bearing fault diagnosis. IEEE Trans. Ind. Electron. 65 (9), 7332–7342.
Wang, Z., Simoncelli, E.P., Bovik, A.C., 2003. Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers. Vol. 2, IEEE, pp. 1398–1402.
Wang, J., Zeng, L., Wang, C., Guo, Y., 2019. Admm-based deep reconstruction for limited-angle ct. Phys. Med. Biol. 64 (11), 115011.
Xiang, J., Dong, Y., Yang, Y., 2021. Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. IEEE Trans. Med. Imaging 40 (5), 1329–1339.
Xu, Q., Yang, D., Tan, J., Sawatzky, A., Anastasio, M.A., 2016. Accelerated fast iterative shrinkage thresholding algorithms for sparsity-regularized cone-beam ct image reconstruction. Med. Phys. 43 (4), 1849–1872.
Ye, J.C., Han, Y., Cha, E., 2018. Deep convolutional framelets: A general deep learning framework for inverse problems. SIAM J. Imaging Sci. 11 (2), 991–1048.
Ye, S., Li, Z., McCann, M.T., Long, Y., Ravishankar, S., 2021. Unified supervised-unsupervised (super) learning for x-ray ct image reconstruction. IEEE Trans. Med. Imaging 40 (11), 2986–3001.
Zhang, T., Gao, H., Xing, Y., Chen, Z., Zhang, L., 2019. Dualres-unet: Limited angle artifact reduction for computed tomography. In: 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference. NSS/MIC, IEEE, pp. 1–3.
Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., Rosen, M.S., 2018. Image reconstruction by domain-transform manifold learning. Nature 555 (7697), 487–492.