scientific reports

OPEN



Dual-Domain deep prior guided sparse-view CT reconstruction with multi-scale fusion attention

Jia Wu^{1,2}, Jinzhao Lin¹, Xiaoming Jiang³, Wei Zheng⁴, Lisha Zhong², Yu Pang⁵, Hongying Meng⁶ & Zhangyong Li^{3⊠}

Sparse-view CT reconstruction is a challenging ill-posed inverse problem, where insufficient projection data leads to degraded image quality with increased noise and artifacts. Recent deep learning approaches have shown promising results in CT reconstruction. However, existing methods often neglect projection data constraints and rely heavily on convolutional neural networks, resulting in limited feature extraction capabilities and inadequate adaptability. To address these limitations, we propose a Dual-domain deep Prior-guided Multi-scale fusion Attention (DPMA) model for sparse-view CT reconstruction, aiming to enhance reconstruction accuracy while ensuring data consistency and stability. First, we establish a residual regularization strategy that applies constraints on the difference between the prior image and target image, effectively integrating deep learning-based priors with model-based optimization. Second, we develop a multi-scale fusion attention mechanism that employs parallel pathways to simultaneously model global context, regional dependencies, and local details in a unified framework. Third, we incorporate a physics-informed consistency module based on range-null space decomposition to ensure adherence to projection data constraints. Experimental results demonstrate that DPMA achieves improved reconstruction quality compared to existing approaches, particularly in noise suppression, artifact reduction, and fine detail preservation.

Keywords Sparse-view CT reconstruction, Deep prior, Multi-scale fusion attention, Model-based optimization, Physics-informed consistency

X-ray Computed Tomography (CT) remains vital in clinical radiology for generating detailed cross-sectional anatomical imaging, supporting accurate diagnosis and treatment strategies across diverse medical conditions. While offering substantial diagnostic value, standard CT protocols involve elevated X-ray radiation levels, with studies demonstrating associated biological risks such as DNA alterations and elevated cancer probability, especially in younger patient groups¹. These findings drive ongoing efforts to refine low-dose CT systems that reduce ionizing radiation while sustaining clinical imaging standards. Sparse-view CT has gained attention as an effective dose-reduction strategy by acquiring fewer angular projections during scans. This sampling reduction, however, creates mathematically unstable reconstruction conditions, frequently producing images corrupted by noise interference and structural artifacts². Resolving these challenges requires algorithmic solutions that balance artifact removal, noise suppression, and anatomical integrity maintenance. Modern reconstruction research concentrates on two core pathways-physics-inspired iterative methods integrating system modeling and regularization, and deep neural networks trained on extensive medical imaging data to enhance reconstruction accuracy³.

Model-based iterative reconstruction algorithms for sparse-view CT employ diverse image priors to enhance reconstruction quality. These approaches encompass several established techniques, including Total Variation (TV) minimization⁴, dictionary learning^{5,6}, smoothed L0 norm optimization⁷, and non-local means^{8,9}. Such methodologies maintain fidelity between the reconstructed image and sparse-view projection data while

¹School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, 400065 Chongqing, China. ²School of Medical Information and Engineering, Southwest Medical University, 646000 Luzhou, China. ³Chongqing Engineering Research Center of Medical Electronics and Information Technology, Chongqing University of Posts and Telecommunications, 400065 Chongqing, China. ⁴Key Laboratory of Intelligent Computing for Advanced Manufacturing, Chongqing University of Posts and Telecommunications, 400065 Chongqing, China. ⁵School of Optoelectronic Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China. ⁶Department of Electronic and Electrical Engineering, College of Engineering Design and Physical Sciences, Brunel University London, UB8 3PH Uxbridge, UK. ^{See}email: lizy@cqupt.edu.cn simultaneously enhancing salient image features. Through the incorporation of prior knowledge regarding image characteristics, these algorithms effectively suppress noise and mitigate artifacts, thereby preserving essential anatomical structures. However, model-based iterative methods frequently encounter limitations due to their substantial computational complexity and high iteration counts. Furthermore, when the projection data becomes extremely sparse, these prior constraints alone may not be sufficient to achieve satisfactory image reconstruction quality¹⁰.

In recent years, data-driven deep learning techniques have significantly advanced sparse-view CT reconstruction by demonstrating improvements in accuracy and computational efficiency compared to traditional model-based iterative methods that require complex parameter optimization^{11,12}. Researchers have systematically organized these approaches into three main categories: image post-processing, model-based unrolling, and dual-domain learning. Early developments concentrated on image post-processing techniques, with architectures such as FBPConvNet¹³, DD-Net¹⁴, and U-Net variants¹⁵⁻²⁰ employing CNNs to refine filtered back-projection reconstructed images. Although these methods effectively reduce noise and artifacts, they face challenges in recovering fine details and occasionally misinterpret severe artifacts as genuine image features when handling heavily degraded inputs¹⁰. To address these limitations, model-based unrolling strategies emerged by integrating iterative optimization processes into deep neural frameworks. Notable implementations such as Learned Primal-Dual²¹, FISTA-Net²², DIOR²³, SOUL²⁴, DEAR²⁵, and IRON²⁶ reconstruct images through unfolded optimization steps that combine projection data consistency constraints with CNN-learned prior regularization terms²⁶. The latest advancements in dual-domain learning methodologies exploit complementary information from both sinogram and image domains²⁷, with architectures like HDNet²⁸, DuDoUFNet²⁹, Dual-AGNet³⁰, DDPTransformer³¹, MIST³², and DDDM³³ processing sinogram data prior to back-projection followed by image-domain refinement. This dual-domain framework effectively combines the distinct advantages of each domain to achieve enhanced reconstruction performance.

While deep learning methods show significant advances in sparse-view CT reconstruction, clinical adoption faces three key technical barriers. First, anatomical variations between different body regions-such as tissue composition and density distributions in thoracic versus cranial scans-limit model generalizability, as systems trained on specific anatomical domains often fail to maintain performance when applied to unseen regions²⁶. Second, these methods exhibit susceptibility to overfitting, demonstrated by substantial performance gaps between training and patient-separated test datasets³⁴. This discrepancy suggests models may prioritize memorizing training-set artifacts over learning generalizable reconstruction features. Third, real-world clinical environments introduce confounding variables including patient motion and scanner heterogeneity that challenge the robustness of purely data-driven approaches, potentially compromising reconstruction consistency.

Recent efforts to address these challenges build upon foundational work in prior-image constrained reconstruction³⁵. The integration of learning-based and model-based paradigms has seen developments through frameworks like DRONE³⁶ and DL-PICCS³⁷, which combine deep learning with compressed sensing principles. Hybrid approaches such as those proposed by Wu et al.³⁸ demonstrate improved stability through systematic integration of physical models, while self-prior enhanced methods like SPIE-DIR³⁹ leverage complementary scan information to boost reconstruction reliability.

Building upon these prior-guided methodologies, we present a Dual-domain deep Prior-guided Multiscale fusion Attention (DPMA) model for sparse-view CT reconstruction that systematically combines learning-based and model-based paradigms. The DPMA architecture implements a regularization strategy that applies constraints on the residual discrepancy between prior and target images instead of direct target image regularization. To implement this strategy, we develop a Dual-domain Multi-scale fusion Attention (DMA) reconstruction framework that generates a deep prior image to guide the iterative reconstruction process. The framework combines physical constraints from projection data with data-driven prior constraints, effectively reducing reconstruction errors and improving accuracy. The DMA framework incorporates a Multi-scale Fusion Attention (MFA) mechanism that captures features at different scales, and a Physics-Informed Consistency (PIC) module based on range-null space decomposition that ensures consistency between the generated prior image and the measured projection data.

To summarize, our contributions are as follows:

- A residual regularization strategy for DPMA that constrains the discrepancy between prior and target images rather than applying direct regularization. The prior image originates from our DMA framework, which synergizes sinogram and image domain learning through MFA mechanisms. This approach effectively combines data-driven priors with model-based optimization for enhanced reconstruction stability.
- A DMA framework that captures hierarchical features across sinogram and image domains while preserving data consistency. The framework's MFA network employs three parallel pathways to simultaneously model global context, regional dependencies, and local details. The PIC module, based on range-null space decomposition, integrates measured projection data through an optimization process that ensures structural accuracy and visual quality.
- Comprehensive evaluation of the DPMA model demonstrating improved noise suppression, artifact reduction, and detail preservation compared to existing methods.

The rest of this paper is organized as follows: Section Method presents the DPMA methodology, Section Experimental Studies and Results details experimental results, Section Discussion and Limitations discusses the model's strengths and limitations, and Section Conclusion summarizes our work.

Method

Sparse-view CT reconstruction addresses the inverse problem of estimating an image $x \in \mathbb{R}^n$ from undersampled projection measurements $s \in \mathbb{R}^m$. The forward model follows the linear relationship:

$$s = Ax,$$
 (1)

where $A \in \mathbb{R}^{m \times n}$ represents the system matrix encoding the forward projection operator. The underdetermined nature of this problem (m < n) renders it ill-posed, requiring regularization to obtain physically meaningful solutions.

Traditional reconstruction approaches employ transform-domain sparsity constraints through the regularized optimization framework:

$$\min_{\mathbf{x}} \left(\frac{1}{2} \| \mathbf{A}\mathbf{x} - \mathbf{s} \|_2^2 + \alpha \Psi(\mathbf{x}) \right), \tag{2}$$

where the data fidelity term $\frac{1}{2} ||Ax - s||_2^2$ ensures consistency with measured projections, $\Psi(x)$ denotes the regularization functional promoting desired image properties, and $\alpha > 0$ governs the trade-off between data consistency and regularization.

DPMA framework

As shown in Figure 1, the proposed DPMA model consists of two core components: a DMA prior generation framework and a prior-guided iterative reconstruction module(Figure 1a). The DMA framework generates physics-constrained prior images through its dual-domain architecture, which combines a cross-domain reconstruction network (Figure 1b) with a PIC module (Figure 1c). The prior image produced by this framework provides residual regularization during iterative reconstruction, guiding the optimization process to achieve solutions that simultaneously satisfy projection data consistency and preserve structural details through multi-scale feature constraints.

Prior image-assisted iterative module

Building upon compressed sensing theory⁴⁰, our reconstruction framework incorporates deep prior guidance to address limitations in traditional sparsity-based methods that often exhibit suboptimal performance under extreme sparse-view conditions due to insufficient sparsity representation⁴¹. Following recent developments in residual-constrained reconstruction methodologies^{36,37}, we formulate the objective function with embedded deep prior regularization:



Fig. 1. Overview of the DPMA framework. (a) Prior image-assisted iterative module, which utilizes the prior image as a residual regularization term for iterative reconstruction. (b) The dual-domain reconstruction network of the DMA prior framework for high-fidelity prior image generation. (c) Integration of the PIC module with the dual-domain network to ensure data consistency between the prior image and measured projections.

$$\min_{\mathbf{x}} \left(\frac{1}{2} \| \mathbf{A}\mathbf{x} - \mathbf{s} \|_{2}^{2} + \alpha \| D(\mathbf{x} - \mathbf{x}_{p}) \|_{1} \right)$$
(3)

where x_p represents the prior image generated by the DMA prior framework (Section DMA prior framework), and *D* denotes the discrete gradient operator implementing isotropic total variation regularization. Introducing the residual term $z = x - x_p$ and applying augmented Lagrangian framework⁴², the standard form is:

$$\mathscr{L}(\mathbf{x}, \mathbf{z}, \lambda) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{s}\|_{2}^{2} + \alpha \|D\mathbf{z}\|_{1} + \lambda^{T} (\mathbf{x} - \mathbf{x}_{p} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{x} - \mathbf{x}_{p} - \mathbf{z}\|_{2}^{2}$$
(4)

where λ is the Lagrange multiplier associated with the constraint $z = x - x_p$, and $\rho > 0$ is the penalty parameter that controls the strength of the constraint. To simplify computation, we introduce the scaled dual variable $f = \lambda/\rho$, rewriting the augmented Lagrangian as:

$$\mathscr{L}(\mathbf{x}, \mathbf{z}, \mathbf{f}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{s}\|_{2}^{2} + \alpha \|D\mathbf{z}\|_{1} + \frac{\rho}{2} \|\mathbf{x} - \mathbf{x}_{p} - \mathbf{z} + \mathbf{f}\|_{2}^{2} - \frac{\rho}{2} \|\mathbf{f}\|_{2}^{2}$$
(5)

Omitting the constant term $\frac{\rho}{2} \|f\|_2^2$ and reformulating, we derive the constrained optimization problem:

$$\min_{\mathbf{x},\mathbf{z}} \left(\frac{1}{2} \| \mathbf{A}\mathbf{x} - \mathbf{s} \|_{2}^{2} + \alpha \| D\mathbf{z} \|_{1} + \frac{\rho}{2} \| \mathbf{x} - \mathbf{x}_{p} - \mathbf{z} - \mathbf{f} \|_{2}^{2} \right)$$
(6)

In our implementation, we directly update in the iterations without explicitly computing λ . Following the Alternating Direction Method of Multipliers (ADMM) framework, we decompose the optimization into three subproblems:

$$\mathbf{x}^{k+1} = \underset{\mathbf{x}}{\arg\min} \left(\frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{s}\|_{2}^{2} + \frac{\rho}{2} \|\mathbf{x} - \mathbf{x}_{p} - \mathbf{z}^{k} - \mathbf{f}^{k}\|_{2}^{2} \right)$$
(7a)

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \left(\alpha \|D\mathbf{z}\|_{1} + \frac{\rho}{2} \|\mathbf{x}^{k+1} - \mathbf{x}_{p} - \mathbf{z} - \mathbf{f}^{k}\|_{2}^{2} \right)$$
(7b)

$$\mathbf{f}^{k+1} = \mathbf{f}^k - (\mathbf{x}^{k+1} - \mathbf{x}_p - \mathbf{z}^{k+1})$$
(7c)

The x-update minimizes (7a) through preconditioned gradient descent:

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \eta \mathbf{M}^{-1} \nabla f(\mathbf{x}^k) \tag{8}$$

where the gradient $\nabla f(\mathbf{x}^k) = \mathbf{A}^\top (\mathbf{A}\mathbf{x}^k - \mathbf{s}) + \rho(\mathbf{x}^k - \mathbf{x}_p - \mathbf{z}^k - \mathbf{f}^k)$ consists of two complementary terms. The first term $\mathbf{A}^\top (\mathbf{A}\mathbf{x}^k - \mathbf{s})$ backprojects the projection domain error into image updates, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ represents the forward projection operator and \mathbf{A}^\top its adjoint backprojection. The second term $\rho(\mathbf{x}^k - \mathbf{x}_p - \mathbf{z}^k - \mathbf{f}^k)$ enforces consistency with the prior image and residual constraints through the penalty parameter $\rho > 0$. The preconditioner $\mathbf{M} = \operatorname{diag}(\mathbf{A}^\top \mathbf{A}) + \rho \mathbf{I} \in \mathbb{R}^{n \times n}$ approximates the Hessian diagonal, balancing the update magnitude across image pixels. This matrix-free implementation⁴³ avoids explicit formation of $\mathbf{A}^\top \mathbf{A}$, using on-the-fly projection/backprojection operations with step size $\eta \in (0, 1)$ chosen to ensure convergence.

The z-update solves the TV-regularized proximal minimization problem (7b) through the Chambolle-Pock algorithm⁴⁴:

$$z^{k+1} = prox_{\alpha \| D \cdot \|_1} \left(x^{k+1} - x_p - f^k \right)$$
(9)

where $D: \mathbb{R}^n \to \mathbb{R}^{2n}$ computes image gradients in horizontal and vertical directions. The proximal operator $\operatorname{prox}_{\alpha \| D \cdot \|_1} : \mathbb{R}^n \to \mathbb{R}^n$ minimizes the sum of the TV penalty $\alpha \| D \mathbf{z} \|_1$ and a quadratic proximity term $\frac{\rho}{2} \| \mathbf{z} - (\mathbf{x}^{k+1} - \mathbf{x}_p - \mathbf{f}^k) \|_2^2$. In practice, we implement this TV-proximal step using the Chambolle-Pock algorithm (via the scikit-image library):

$$z^{k+1} = TVDenoise\left(x^{k+1} - x_p - f^k, weight = \frac{\alpha}{\rho}\right)$$
 (10)

where TVDenoise : $\mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^n$ represents the Chambolle-Pock TV denoising operator with weight parameter controlling the denoising strength.

The Lagrange multiplier updates as:

$$\mathbf{f}^{k+1} = \mathbf{f}^k + \gamma (\mathbf{x}^{k+1} - \mathbf{x}_p - \mathbf{z}^{k+1}) \tag{11}$$

where $\gamma \in (0, 2)$ is the dual step size chosen to ensure convergence of the ADMM iterations. This update enforces the constraint $\mathbf{x} = \mathbf{x}_p + \mathbf{z}$ asymptotically through dual variable adjustment.

The above iterative framework effectively combines data-driven priors with model-based optimization through residual regularization, ensuring both data consistency and structural preservation. However, the reconstruction quality heavily depends on the accuracy and reliability of the prior image x_p . This critical dependency motivates us to develop a comprehensive prior generation strategy that can produce high-fidelity prior images while maintaining physics-informed data consistency.

DMA prior framework

The proposed DMA framework generates high-quality prior images x_p for the iterative reconstruction process described above. As illustrated in Figure 1b, the framework consists of a dual-domain network and a PIC module, where the dual-domain network is composed of a Sinogram Refinement (SR) module and an Image Enhancement (IE) module. In the following sections, we will detail each of these modules.

SR module The SR module integrates a sinogram domain sub-network with a Radon Inverse Transform (RIT) block⁴⁵. Within this module, the sinogram domain sub-network works to recover fully sampled sinograms from sparse-view measurements, while the RIT block subsequently transforms these recovered sinograms into image domain representations.

For sinogram recovery, we first employ linear interpolation on sparse-view sinograms to generate intermediate sinograms s_{LI} , providing an initial estimate of the missing angular samples. However, interpolation can introduce errors, especially at angles distant from the original measurements. To address this limitation, we propose a data-driven approach that minimizes the discrepancy between the recovered and actual fully sampled sinograms. The network transformation can be expressed as:

$$\hat{\mathbf{s}} = E_{\omega}(\mathbf{s}_{LI}),\tag{12}$$

where E_{ω} denotes the MFA network with parameters ω (detailed in Section MFA network). These parameters are optimized using paired training data $\{s_{LI}^i, s_{gt}^i\}_{i=1}^N$ through empirical risk minimization. The training process utilizes a loss function measuring the mean L_1 difference between the estimated sinograms \hat{s}^i and the reference sinograms s_{gt}^i :

$$L_{S} = \frac{1}{N} \sum_{i=1}^{N} \left\| E_{\omega}(\mathbf{s}_{LI}^{i}) - \mathbf{s}_{gt}^{i} \right\|_{1}$$
(13)

where s_{gt}^i represents the ground-truth sinogram obtained from a fully sampled projection of the *i*-th training sample. A weight decay regularization term with coefficient $1e^{-5}$ is added during training to prevent overfitting.

Although E_{ω} generates estimates of fully sampled sinograms, optimization solely through the sinogramdomain loss L_S exhibits limitations in achieving accurate projection recovery, resulting in residual discrepancies between the estimated and actual full projections. These sinogram inconsistencies, when processed through RIT reconstruction to obtain intermediate images x_{LI} , manifest as reconstruction errors and introduce secondary artifacts. To address both the sinogram domain discrepancies and their subsequent image domain artifacts, we incorporate an additional image-domain Mean Squared Error (MSE) loss:

$$L_{R} = \frac{1}{N} \sum_{i=1}^{N} \left\| \mathscr{R}(E_{\omega}(s_{LI}^{i})) - x_{gt}^{i} \right\|_{2}^{2}$$
(14)

where $\mathscr{R}(\cdot)$ denotes the RIT reconstruction operator that transforms sinograms into images, and x_{gt}^i represents the ground-truth image corresponding to the *i*-th training sample.

IE module

Despite the SR module's refinement of interpolation-induced errors in the projection domain, the reconstructed images may still exhibit noise and artifacts due to the limited number of projection views. To address this, we introduce the IE module, which further enhances the images. The IE module employs the same MFA network architecture as the SR module, denoted as E_{θ} . It processes the intermediate images produced by the SR module, yielding enhanced CT images $\hat{\mathbf{x}} = E_{\theta}(\mathbf{x}_{LI})$, where $\hat{\mathbf{x}}$ represents the final output and \mathbf{x}_{LI} denotes the intermediate images from the SR module. The network parameters θ are optimized using a MSE loss function:

$$L_{I} = \frac{1}{N} \sum_{i=1}^{N} \left\| E_{\theta}(x_{LI}^{i}) - x_{gt}^{i} \right\|_{2}^{2}$$
(15)

where x_{LI}^i and represent the intermediate images of the *i*-th training sample. A weight decay regularization term with coefficient $1e^{-5}$ is added during training to prevent overfitting.

Based on the three loss terms introduced above, we train the SR and IE modules jointly to leverage their complementary capabilities. The overall loss function is defined as:

$$L_{SI} = \zeta L_S + \vartheta L_R + \xi L_I, \tag{16}$$

where L_S targets the reduction of sinogram discrepancies, L_R focuses on artifact suppression, and L_I aims to minimize noise and refine image details. The coefficients $\zeta = 0.1$, $\vartheta = 0.1$, and $\xi = 1$ balance the contributions of these three terms.

PIC module While the SR and IE modules effectively reduce noise and artifacts, they may introduce oversmoothing effects due to the lack of projection data constraints in the IE stage, potentially obscuring fine anatomical details critical for diagnosis. To preserve these essential features, we introduce the PIC module, which ensures the reconstructed images maintain both visual quality and fidelity to the measured projection data.

The PIC module leverages Range-Null Space Decomposition Theory⁴⁵ to ensure data consistency while preserving image details. According to this theory, any image x_p satisfying the data consistency constraint $Ax_p = s$ can be decomposed as:

$$\mathbf{x}_p = \mathbf{A}^{\dagger} \mathbf{s} + (\mathbf{I} - \mathbf{A}^{\dagger} \mathbf{A}) \mathbf{v} \tag{17}$$

where A^{\dagger} denotes the Moore-Penrose pseudoinverse of the system matrix, and v represents an arbitrary vector in the image space. This decomposition separates the solution into two orthogonal components: the range space term $A^{\dagger}s$ (minimum-norm solution satisfying Ax = s) and the null space term $(I - A^{\dagger}A)v$ containing anatomical features unconstrained by projection data.

In CT systems with sparse-view sampling $(m \ll n)$, the null space dimension $n - \operatorname{rank}(A)$ dominates the solution space. This necessitates careful selection of v to preserve diagnostically relevant features. The computational intractability of direct pseudoinverse calculation ($\mathcal{O}(n^3)$ complexity) and the need for anatomical guidance motivate our dual approximation strategy. First, we approximate the pseudoinverse using the Landweber iteration identity⁴⁶ $A^{\dagger} = \lim_{k \to \infty} \sum_{i=0}^{k} (I - \beta A^{\top} A)^{i} \beta A^{\top}$ with step size $\beta = 1/\tau_{\max}^{2}(A)$, where $\tau_{\max}(A)$ denotes the maximum singular value of A. Retaining only the first-order term yields the practical approximation $A^{\dagger}s \approx A^{\top}s$, valid under the quasi-unitary condition $A^{\top}A \approx cI$ common in CT systems with uniform angular sampling. Second, we substitute the arbitrary null space vector v with the enhanced image \hat{x} from the dual-domain network. This substitution introduces the approximation:

$$\mathbf{x}_p \approx \mathbf{A}^\top \mathbf{s} + (\mathbf{I} - \mathbf{A}^\top \mathbf{A})\hat{\mathbf{x}} + \epsilon_r$$
 (18)

where ϵ_r captures residual approximation errors. Algebraic rearrangement reveals this expression equivalently as:

$$\mathbf{x}_p = \hat{\mathbf{x}} - \mathbf{A}^\top (\mathbf{A}\hat{\mathbf{x}} - \mathbf{s}) + \epsilon_r \tag{19}$$

showing explicit data consistency correction through the Landweber-type term $A^{\top}(A\hat{x}-s).$

The final implementation encapsulates these approximations through the PIC network E_o , which learns to compensate residual errors while preserving valid anatomy:

$$\mathbf{x}_p = E_o \left(\hat{\mathbf{x}} - \mathbf{A}^\top (\mathbf{A} \hat{\mathbf{x}} - \mathbf{s}) \right) \tag{20}$$

Here, the network input $\hat{\mathbf{x}} - \mathbf{A}^{\top}(\mathbf{A}\hat{\mathbf{x}} - \mathbf{s})$ represents a single Landweber iteration step applied to the prior estimate, combining data consistency enforcement with learned null space refinement. Through end-to-end training, E_o develops capacity to correct approximation errors ϵ_r while maintaining fidelity to both projection data and anatomical references. The PIC network is trained to minimize:

$$L_{\rm PIC} = \frac{1}{N} \sum_{i=1}^{N} \left\| E_o \left(\hat{\mathbf{x}}^i - \mathbf{A}^{\top} (\mathbf{A} \hat{\mathbf{x}}^i - \mathbf{s}^i) \right) - \mathbf{x}_{\rm gt}^i \right\|_2^2$$
(21)

This formulation ensures the deep prior image maintains: (1) Data consistency through explicit pseudoinverse projection, (2) Anatomical fidelity via the learned null-space component, and (3) Noise-artifact suppression through the refinement network. The complete procedure of our proposed DPMA method is summarized in Algorithm 1.

Require: Sparse sinogram s, system matrix A, pretrained networks E_{ω} , E_{θ} , E_{θ} **Ensure:** Reconstructed image $\mathbf{x}^{(K)}$ 1: Phase 1: Prior Generation 2: $\mathbf{s}_{LI} \leftarrow \text{LinearInterpolation}(\mathbf{s}) \{\text{Sinogram domain interpolation}\}$ 3: $\hat{\mathbf{s}} \leftarrow E_{\boldsymbol{\omega}}(\mathbf{s}_{LI})$ {Sinogram refinement via MFA network} 4: $\mathbf{x}_{LI} \leftarrow \operatorname{RIT}(\hat{\mathbf{s}})$ {Radon inverse layer} 5: $\hat{\mathbf{x}} \leftarrow E_{\theta}(\mathbf{x}_{LI})$ {Image enhancement via MFA network} 6: $\mathbf{x}_{p} \leftarrow E_{o} \left(\hat{\mathbf{x}} - \mathbf{A}^{\top} (\mathbf{A} \hat{\mathbf{x}} - \mathbf{s}) \right)$ {PIC module processing} 7: Phase 2: Iter Optimization 8: Initialize $\mathbf{x}^{(0)} \leftarrow \mathbf{x}_{p}, \mathbf{z}^{(0)} \leftarrow \mathbf{0}, \mathbf{f}^{(0)} \leftarrow \mathbf{0}$ 9: for k = 0 to K - 1 do x-update: $10 \cdot$ Solve $(\mathbf{A}^{\top}\mathbf{A} + \boldsymbol{\rho}\mathbf{I})\mathbf{x} = \mathbf{A}^{\top}\mathbf{s} + \boldsymbol{\rho}(\mathbf{x}_{p} + \mathbf{z}^{(k)} + \mathbf{f}^{(k)})$ 11: using preconditioned conjugate gradient method⁴³ 12: 13: z-update: $\mathbf{z}^{(k+1)} \leftarrow \text{TVDenoise}\left(\mathbf{x}^{(k+1)} - \mathbf{x}_p - \mathbf{f}^{(k)}, \alpha/\rho\right)$ {Chambolle-Pock TV denoising} 14: 15: f-update: $\mathbf{\dot{f}}^{(k+1)} \leftarrow \mathbf{f}^{(k)} - \gamma(\mathbf{x}^{(k+1)} - \mathbf{x}_p - \mathbf{z}^{(k+1)})$ {Dual variable update} 16: 17: end for 18: return $\mathbf{x}^{(K)}$

Algorithm 1. DPMA Reconstruction Framework

MFA network

The MFA network is used in the SR (E_{ω}) , IE (E_{θ}) , and PIC (E_o) modules of our framework. We developed this network architecture to address the hierarchical characteristics of both sinograms and CT images, which contain structural information across global, regional, and local scales.

Conventional CNN-based approaches^{13,15–20}, while effective in capturing local features through their translation-equivariant inductive bias, suffer from rigid receptive fields due to fixed convolutional kernels that cannot adaptively aggregate spatial information. This architectural constraint limits their capacity to handle the non-uniform artifact distributions in sparse-view CT reconstruction, where artifacts propagate across multiple scales with varying intensity patterns. While transformer-based architectures^{31,32} achieve adaptive spatial aggregation through self-attention mechanisms, their quadratic computational complexity renders them impractical for high-resolution medical imaging, and more critically, they discard the convolutional inductive biases essential for preserving local structural continuity.

Non-local⁴⁷ operations, employing sparse attention patterns with dilated receptive fields, demonstrate superior capability in capturing long-range artifact correlations-a critical requirement for addressing the nonuniform artifact distributions characteristic of sparse-view CT. These mechanisms achieve computational efficiency through strategic sampling of attention heads, reducing complexity from quadratic to linear scale. However, their global interaction paradigm inherently dilutes the translation-equivariant inductive biases crucial for preserving anatomical structure continuity. Conversely, kernel-based attention⁴⁸ approaches preserve local structural fidelity through learnable convolutional priors, where different kernel bases are trained to model representative image patterns. By adaptively fusing these bases through pixel-wise coefficients, they maintain position-sensitive structural details. Yet this localized focus comes at the cost of constrained receptive fields, limiting their capacity to model the global context of radiation artifacts that often span multiple anatomical regions.

This fundamental dichotomy-where non-local methods excel at global artifact modeling but compromise local structure, while kernel-based approaches preserve anatomical details but lack global context-reveals the critical need for synergistic architectures. Effective CT reconstruction demands attention operators that simultaneously achieve: (1) efficient long-range dependency modeling through sparse non-local interactions, and (2) structure-aware local processing via adaptive kernel fusion, while maintaining the computational efficiency required for high-resolution medical imaging.

To effectively capture and synthesize multi-scale features, we propose the MFA network as depicted in Figure 2a. The network is built upon a U-Net backbone, incorporating an enhanced parallel attention MFA block. Following⁴⁸, the MFA block implements hierarchical feature integration through sequential processing stagesas, illustrated in Figure 2b. Initial layer normalization precedes a hybrid enhancement module that synergizes residual learning with channel-wise attention. The residual pathway employs depthwise separable convolutions to preserve local anatomical structures, while parallel squeeze-excitation mechanisms adaptively recalibrate channel responses through global average pooling and gated transformations. Element-wise summation combines these complementary feature streams, maintaining both high-frequency details and contextual awareness.

The block then processes features through two parallel pathways. In the global context pathway, a non-local sparse attention⁴⁷ mechanism implements angular locality-sensitive hashing with four-round bucket attention



Fig. 2. (a) Architecture of the MFA Network. (b) Detailed structure of the MFA block, incorporating non-local sparse attention, channel attention, and kernel basis attention mechanisms.

.....

for efficient long-range dependency modeling. This design enables comprehensive global context aggregation while maintaining computational efficiency for high-resolution medical images.

The local refinement pathway implements Kernel Basis Attention (KBA)⁴⁸ through a series of operations. Attention weights are generated through a combination of depthwise convolutions and SimpleGate activation, followed by projection to attention maps. These attention maps are modulated by learnable parameters and combined with a direct 1×1 convolution path. The KBA module then processes features through depthwise convolutions before applying the generated attention weights to adaptively combine kernel bases, preserving local structural patterns while enabling position-sensitive feature transformation.

Feature recombination occurs through multiplicative fusion of the global context pathway output, kerneladapted local features, and the enhanced features from the initial hybrid module. This fused representation undergoes channel reduction via 1×1 convolution before combining with the input through a residual connection modulated by learnable scaling parameters. The final stage employs a feed-forward network (FFN) with channel expansion and SimpleGate activation for nonlinear feature transformation. The FFN output combines with the intermediate features through a second residual connection, with both connections scaled by learnable parameters to maintain stable gradient flow during training. This multi-stage design enables effective capture and synthesis of features across multiple scales while preserving both global context and local anatomical details essential for high-quality CT reconstruction.

Experimental studies and results

We implemented the DPMA model using PyTorch and the Operator Discretization Library (ODL) for numerical computations. The experiments were conducted on a system with two NVIDIA Tesla V100 GPUs (32GB memory each). The training process consisted of 10 epochs for the SR and IE networks, followed by 5 epochs for the PIC network, with a batch size of 2. We used the AdamW optimizer with momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate started at 3×10^{-4} and decreased to 1×10^{-5} . For testing, we set $\rho = 0.1$, $\alpha = 0.0001$, $\gamma = 1$, and $\eta = 0.1$, with the reconstruction algorithm running for 300 iterations.

For quantitative evaluation, we established baseline comparisons against seven representative methods spanning different reconstruction paradigms: FBPConvNet (post-processing)¹³, FISTA (deep unrolling)²², DUAL (deep unrolling)²¹, FreeSeed (frequency-band-aware post-processing)⁴⁹, MVMS-RCN (dual-domain multi-sparse-view reconstruction)⁵⁰, DRONE (deep residual optimization)³⁶, and our standalone DMA prior (non-iterative deep prior). All implementations strictly followed original publications with parameter tuning on our validation set. Reconstruction quality was assessed through three established metrics: MSE, Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM)³⁶. All images were normalized to [0,1] intensity range before metric computation.

Dataset description

Simulated datasets

We evaluated DPMA on the Mayo Clinic dataset from the AAPM Low-Dose CT Grand Challenge. The dataset contains CT scans from ten patients with 2378 slices in total. We used 1943 slices from eight patients for training and 435 slices from two patients for testing. Each image has a size of 512×512 pixels. The data was split by patient to ensure independent testing. We also tested on two cases (L028 and L071) from the LDCT-and-Projection dataset⁵¹ using the model trained on AAPM without fine-tuning.

Projections were generated using Siddon's ray-driven algorithm⁵² with a fan-beam geometry. The setup used a source-to-axial distance of 590 mm and a source-to-detector distance of 1150 mm. The source trajectory had 720 views over 360 degrees with a slice thickness of 0.3 mm. For sparse-view scenarios, we generated sinograms with 64 and 32 views. To simulate realistic noise conditions, Poisson noise was added to the projection data with an incident photon number of 10^6 per ray.

Results of the experiment

Quantitative and qualitative results

Figure 3 reports the 64-view reconstruction results obtained with nine representative algorithms. FBP(Figure 3(b1)) is highly vulnerable to severe angular undersampling and consequently produces images dominated by streak artefacts and noise. Adding a learning-based post-processing stage, FBPConvNet (Figure 3(c1,c2)) suppresses most of these artefacts, yet the price is evident over-smoothing that obscures fine anatomical structures. Unfolded iterative networks, namely FISTA and DUAL (Figure 3(d1,d2) and (e1,e2)), improve visual quality by explicitly enforcing data consistency, although high-frequency details remain partially blurred. FreeSeed (Figure 3(f1,f2)) leverages frequency-band awareness to remove streaks while better preserving



Fig. 3. Reconstruction results ('1') and their corresponding difference images ('2') for the simulated AAPM dataset with 64 views: (a) reference image, (b) FBP, (c) FBPConvNet, (d) FISTA, (e) DUAL, (f) FreeSeed, (g) MVMS-RCN, (h) DRONE, (i) DMAprior, and (j) DPMA. The display window is set to [-160, 240] HU.

edges; however, the difference images reveal residual errors that most likely originate from interpolation inaccuracies misinterpreted as structural information during projection-domain refinement. MVMS-RCN (Figure 3(g1,g2)) further attenuates noise and artefacts through multi-view sparsity, yet certain high-frequency components are still missing, as evidenced by the error maps. DRONE (Figure 3(h1,h2)) demonstrates effective noise and streak suppression while maintaining texture and edge sharpness. By integrating information from both the projection and image domains and incorporating a multi-scale attention mechanism, DMAprior (Figure 3(i1,i2)) increases structural accuracy relative to pure post-processing methods, although some detail loss persists. Finally, our DPMA (Figure 3(j1,j2)), which couples the DMA prior with physics-guided iterative reconstruction, achieves the highest overall fidelity: streaks and noise are strongly suppressed, high-frequency anatomical features are accurately restored, and the resulting error maps are closest to the reference.

Figure 4 presents the reconstructions obtained from only 32 projection views, an extremely sparse-view scenario that poses a considerable challenge to all algorithms. Under this condition, FBP (Figure 4(b1)) deteriorates markedly, producing images dominated by noise and streak artefacts. FBPConvNet (Figure 4(c1,c2)) alleviates some of these artefacts, yet the enlarged ROIs reveal pronounced over-smoothing together with residual noise. The unfolded iterative schemes FISTA Figure (4(d1,d2)) and DUAL (Figure 4(e1,e2)) achieve higher visual quality than FBPConvNet, but their noise-suppression capability is constrained by the severe data sparsity, resulting in visibly blurred structures. FreeSeed (Figure 4(f1,f2)) fails to eliminate streaks in soft-tissue regions, leaving artefacts that would likely compromise diagnostic reliability. Although MVMS-RCN (Figure 4(g1,g2)) benefits from its multi-dimensional projection-recovery strategy and delivers high overall clarity, noticeable over-smoothing obscures delicate anatomical details. DRONE (Figure 4(h1,h2)) offers further improvements, yet edge blurring persists and fine structures remain partially lost. Leveraging a multi-scale feature-representation strategy, DMAprior (Figure 4(i1,i2)) succeeds in recovering the principal anatomy, yet



Fig. 4. Reconstruction results ('1') and their corresponding difference images ('2') for the simulated AAPM dataset with 32 views: (**a**) reference image, (**b**) FBP, (**c**) FBPConvNet, (**d**) FISTA, (**e**) DUAL, (**f**) FreeSeed, (**g**) MVMS-RCN, (**h**) DRONE, (**i**) DMAprior, and (**j**) DPMA. The display window is set to [-160, 240] HU.

its overall performance remains inferior to that of DRONE and MVMS-RCN, and several regions still suffer from noticeable over-smoothing. In contrast, the proposed DPMA (Figure 4(j1,j2)) combines effective noise and artefact suppression with better preservation of structural edges and subtle details. Table 3 reports the quantitative metrics for the 64-view and 32-view experiments. In both sampling scenarios, DPMA attains the highest PSNR and SSIM scores among all compared methods, indicating more accurate and perceptually faithful reconstructions across the two data-sparsity levels.

Figure 5a illustrates the 64 projection views reconstructions obtained from the LDCT-and-Projection simulated dataset when all networks are applied without retraining. FBPConvNet suppresses most streak artefacts yet leaves fine anatomical structures unresolved in the enlarged ROIs. FISTA and DUAL reveal more detail, but their outputs exhibit mild blurring, likely due to a domain gap between the training data and the current test case. FreeSeed, which processes data in both the projection and image domains, does not fully correct interpolation errors and therefore retains residual artefacts in soft-tissue regions. MVMS-RCN shows stronger cross-dataset robustness, generating generally clean images, though some structural information is still missing. DRONE benefits from its iterative architecture, achieving noise reduction while keeping many anatomical edges intact. DMAprior, however, presents noticeable artefacts comparable to those seen in FreeSeed, again reflecting insufficient handling of projection-domain interpolation errors. By coupling DMAprior with a physics-guided iterative model, DPMA produces images that align more closely with the reference than any other method in the comparison set.



Fig. 5. Reconstruction results and quantitative evaluation on the LDCT-and-Projection simulated dataset with 64 views: (**a**) coronal view of reconstructed images, (**b**) PSNR comparison, and (**c**) SSIM comparison. The display windows are [-1150, 350] HU and [-160, 240] HU, respectively.



Fig. 6. Intensity profiles along the red line for different reconstruction methods: (**a**) AAPM simulated dataset with 64 views, (**b**) AAPM simulated dataset with 32 views, and (**c**) LDCT-and-Projection simulated dataset with 64 views.

The quantitative results in Figures 5b and 5c support these visual findings: DPMA reaches the highest PSNR and SSIM values across all evaluated approaches. Consistent with these metrics, the line profiles in Figure 6 show that DPMA's attenuation curves track the reference intensities most faithfully, confirming its ability to preserve both global contrast and subtle structural variations under this challenging sparse-view, cross-domain scenario.

Noise properties analysis

Noise characteristics were assessed on the AAPM data reconstructed from 64 projection views. A square ROI of 16×16 pixels, centred at (x, y) = (220, 350) (Figure 7a), served as the analysis window. Within this patch five 4×4 sub-ROIs were positioned-one at the centre and one in each of the four cardinal directions, offset by four pixels. For each algorithm the noise map was defined as the pixel-wise difference $n(x,y) = x(x,y) - x_{gt}(x,y)$ between the reconstruction x and the reference image x_{gt} . The noise magnitude σ was obtained by computing the standard deviation inside every sub-ROI and averaging these five values, providing a measure of local noise intensity. Texture was characterised by the two-dimensional noise-power spectrum calculated on a 64×64 noise patch: the squared magnitude $|\mathscr{F}\{n\}(u,v)|^2$ of the discrete Fourier transform was normalised by its total power to yield a unit-area normalised NPS (nNPS), where (u, v) represent the spatial frequency coordinates corresponding to the spatial domain (x, y). For clearer visualization, we performed radial averaging by computing the mean nNPS value at each radial frequency $f_r = \sqrt{u^2 + v^2}$ across 50 equally spaced frequency bins, followed by Savitzky-Golay smoothing (window length 15, polynomial order 3). From this one-dimensional profile, we extracted the average frequency $f_{\rm av}$, defined as the frequency at which $\int_{0}^{f_{\text{av}}} \text{nNPS}(f) df = \frac{1}{2} \int_{0}^{f_{\text{max}}} \text{nNPS}(f) df$, and the peak frequency f_{peak} , corresponding to the maximum value of the radial nNPS curve. While σ reflects overall noise amplitude, $f_{\rm av}$ and $f_{\rm peak}$ summarise its spectral distribution-smaller values denote smoother noise, larger values indicate grainier patterns. The three metrics for all competing methods are compared in Figure 7(b-d).

The noise intensity analysis (Figure 7(b)) reveals that both DPMA and MVMS-RCN achieve the lowest variability in noise characteristics, indicating consistent noise suppression across different regions. FreeSeed demonstrates effective overall noise reduction but exhibits the highest variability, suggesting less uniform noise suppression. FBPConvNet shows moderate noise characteristics, while FISTA demonstrates higher variability than DUAL despite comparable mean noise levels. DMAprior exhibits relatively high noise variability, while



Fig. 7. Comprehensive noise analysis of reconstructions from the AAPM simulated CT dataset. (a) Noise intensity quantification (HU) with ROI selection scheme (left insets show the 16×16 primary ROI and the five 4×4 sampling regions in a cross pattern). (b) Spatial noise distribution maps across nine reconstruction methods with corresponding standard deviation (σ) values in HU. (c) Two-dimensional nNPS visualizing the frequency-domain noise characteristics for each method. (d) One-dimensional radially-averaged nNPS profiles arranged in a grid, with characteristic frequencies marked: $f_{\rm av}$ (red dashed line, frequency bisecting area under curve) and $f_{\rm peak}$ (green dashed line, frequency of maximum power), both measured in cycles/mm.

.....

DRONE maintains a balanced profile between noise magnitude and consistency. The frequency domain analysis (Figure 7(c,d)) provides complementary insights into noise texture properties. DPMA demonstrates the highest average (f_{av}) and peak (f_{peak}) frequencies among all methods, indicating better preservation of high-frequency image details while maintaining low noise magnitude. DRONE shows the second-highest frequency characteristics, particularly in peak frequency, reflecting its strength in preserving edge details. Interestingly, FreeSeed exhibits the lowest peak frequency despite moderate average frequency, suggesting effective suppression of dominant noise patterns but potentially over-smoothing fine details. MVMS-RCN maintains a well-balanced frequency profile with moderate-to-high average frequency but relatively low peak frequency. The post-processing method FBPConvNet displays nearly identical low-frequency characteristics, indicating similar noise texture properties with a tendency toward smoother results. This comprehensive analysis confirms that DPMA achieves an optimal balance between noise reduction and detail preservation compared to existing reconstruction methods.

Ablation study

To systematically investigate the contribution and efficacy of individual components within our proposed DPMA framework, we conducted comprehensive ablation experiments. These controlled analyses were performed using the 64-view reconstruction task on the AAPM simulated dataset as our experimental benchmark. By sequentially isolating and evaluating specific architectural elements, we establish empirical evidence for design choices and identify the critical factors driving performance improvements.

Effectiveness of prior image-guided iterative module

The effectiveness of our DPMA framework primarily depends on the quality of the initial prior image. To address this fundamental challenge, we developed the DMA framework, which demonstrates improved performance in generating high-fidelity prior images with enhanced structural preservation. As shown in Figure 8(a), our approach achieves better alignment with reference images compared to existing methodologies, including



Fig. 8. Ablation study on the effect of the dual-domain deep prior image-guided iterative reconstruction framework on AAPM simulated dataset with 64 views. (**a**) Comparison of different prior images generated by FBPConvNet, DDNet, FISTA, DUAL, DRONE prior, and our DMA prior. (**b**) Reconstruction results after iterative optimization with different prior images. The bottom row of each subfigure shows the zoomed ROI and error maps (difference between reconstruction and reference, color scale in HU). All images are presented with a display window of [-160, 240] HU.

Components	DRONE prior	DMA-B	DMA-G	DMA-S	DMA-I	DMA
MFA	-	UNet	GroupUNet	\checkmark	\checkmark	\checkmark
SR	-	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark
IE	-	\checkmark	\checkmark	-	\checkmark	\checkmark
PIC	-	-	-	-	-	\checkmark
PSNR	39.570	36.496	39.978	33.684	40.251	42.526
	±0.985	±2.042	±1.738	±1.225	±1.513	±1.194
SSIM	0.954	0.924	0.961	0.873	0.962	0.965
	±0.007	±0.007	±0.007	±0.010	±0.007	±0.007

Table 2. Quantitative comparison of ablation studies in DMA framework for 64 views on AAPM simulation dataset.

post-processing networks (FBPConvNet, DDNet¹⁴), iterative unfolding approaches (FISTA, DUAL), and partial implementation of deep residual frameworks (DRONE prior, which comprises only the Embedding and Refinement modules from the original DRONE architecture, without the Awareness module). The quantitative and qualitative improvements observed validate our hypothesis that dual-domain integration with physics-informed constraints provides more reliable anatomical priors for subsequent iterative refinement.

Our framework allows for the integration of various models as initial priors for iterative refinement. Figure 8(b) illustrates that the iterative optimization process enhances reconstruction quality across all models, improving the delineation of anatomical structures, organ boundaries, soft tissues, and bones. The results demonstrate the effectiveness of our prior image-guided iterative strategy. Furthermore, the reconstruction quality can be further improved when starting with higher quality initial priors, such as those generated by our DMA framework. Table 2 presents the quantitative comparison between DRONE prior and DMA prior. The results indicate that our DMA prior achieves improved performance across all metrics, validating the effectiveness of the proposed dual-domain approach and MFA attention mechanism in feature extraction and reconstruction.

Effectiveness of MFA network

To evaluate the effectiveness of our proposed MFA network, we conducted comparative experiments using different backbone architectures in our DMA framework. Specifically, we compared three variants: DMA-B using a conventional convolution-based UNet as the backbone network, DMA-G employing GroupUNet⁵³ as the backbone network, and our complete DMA framework with the MFA network. As shown in Table 2 and Figure 9(a), DMA-B shows relatively lower performance in both quantitative metrics and visual quality.



Fig. 9. Visual comparison of two ablation studies on the AAPM simulation dataset. (a) Effectiveness of MFA network: comparison of different backbone architectures from DMA-B (conventional UNet), DMA-G (GroupUNet), to our complete DMA framework with MFA. (b) Analysis of dual-domain modules: comparison of DMA-S (with SR module only), DMA-I (with SR and IE modules), and complete DMA framework (with all three modules). The display window is [-160, 240] HU.

By incorporating group convolutions, DMA-G demonstrates improved reconstruction quality with better preservation of anatomical structures. Our complete DMA framework with the MFA network further enhances the reconstruction performance, achieving improved results in both quantitative metrics and visual quality. The visual comparison in Figure 9(a) indicates that DMA produces clearer anatomical details with reduced artifacts, particularly in the zoomed regions, demonstrating the effectiveness of our proposed MFA mechanism in capturing and integrating multi-scale features for improved reconstruction.

Figure 9(b) presents a qualitative analysis of reconstructions from different architectural variants of our framework. The sinogram-only approach (DMA-S) demonstrates substantial limitations, exhibiting pronounced artifacts and reduced structural definition due to its inability to leverage image-domain refinement. The dual-domain approach without physics constraints (DMA-I) significantly improves reconstruction quality through complementary feature extraction, yet still manifests subtle artifacts in regions of complex anatomical structure. Our complete DMA framework, integrating all proposed modules, achieves superior visual quality with enhanced delineation of organ boundaries and preservation of fine tissue details. This progression in reconstruction quality clearly demonstrates the complementary nature of each architectural component. These observations validate both the effectiveness of our multi-scale fusion attention mechanism and the necessity of comprehensive dual-domain processing with physics-informed consistency constraints for optimal sparse-view CT reconstruction.

Parameter sensitivity analysis

To investigate the impact of key parameters in our DPMA framework, we conducted sensitivity analysis on the penalty parameter ρ , primal step-size parameter η , and regularization parameter α . As shown in Figure 10(a-c), the PSNR and SSIM metrics were evaluated across different parameter ranges. The penalty parameter ρ exhibits robust performance with a slight degradation as its value increases. The primal step-size parameter η demonstrates stable performance within a moderate range but shows dramatic deterioration when exceeding a certain threshold, indicating high sensitivity to large step sizes. The regularization parameter α/ρ shows consistent performance across its tested range, with 0.001 yielding the best results. Based on these experimental observations, we selected $\rho = 0.1$, $\eta = 0.15$, and $\alpha = 0.0001$ as the optimal parameter values that achieve the best balance between image quality and algorithmic stability. These parameters were used consistently across all experiments.

Analysis of loss function combinations

To validate the effectiveness of different loss components, we conducted ablation experiments with various loss combinations: $L_R + L_I$ (image domain losses), $L_S + L_I$ (sinogram and image domain losses), and $L_S + L_R + L_I$ (complete losses). As shown in Figure 10(d), the combination incorporating sinogram domain (L_S) , image domain (L_R) , and enhancement (L_I) losses achieves the best performance across MSE, PSNR, and SSIM metrics, indicating the effectiveness of each loss component.



Fig. 10. Parameter sensitivity analysis and ablation study. The top row (**a**-**c**) shows parameter sensitivity analysis of DPMA on the AAPM simulated dataset with 64 views: (**a**) the penalty parameter ρ , (**b**) the primal step-size parameter η , and (**c**) the regularization parameter α . (**d**) Validation curves of different loss combinations in terms of MSE, PSNR, and SSIM metrics, where L_S is the sinogram domain loss, L_R is the image domain loss, and L_I is the enhancement loss.

.....

Method	FBP	FBPConvNet	FISTA	DUAL	FreeSeed	MVMS-RCN	DRONE	DMAprior	DPMA
64 Views									
PSNR	26.135	35.623	39.121	40.231	42.348	42.516	43.027	41.526	43.867
SSIM	0.534	0.933	0.965	0.971	0.977	0.980	0.982	0.975	0.984
32 Views									
PSNR	23.115	32.790	33.796	34.196	37.613	38.297	38.288	37.191	38.862
SSIM	0.415	0.904	0.912	0.918	0.951	0.961	0.960	0.950	0.962

Table 1. Quantitative comparison of reconstruction methods on the AAPM simulated dataset with 64 and 32views. Best results are shown in bold.

Method	DRONE	DPMA	MVMS-RCN	FreeSeed	DMAprior
Runtime (s/slice)	66.862	23.731	2.046	1.102	0.0742
Method	DUAL	FISTA	FBPConvNet	FBP	
Runtime (s/slice)	0.0373	0.0344	0.0072	0.0013	

Table 3. Computational cost comparison of different reconstruction methods. DRONE and DPMA runtimesare with 300 iterations.

.....

Computational efficiency analysis

We analyzed the computational efficiency of different reconstruction methods, as shown in Table 3. The FBP method exhibits the highest computational efficiency but produces significant artifacts and noise in the reconstructed images. Deep learning post-processing methods like FBPConvNet demonstrate excellent computational efficiency, followed closely by iterative unfolding methods such as FISTA and DUAL. DMAprior



Fig. 11. Reconstruction results on the AAPM rebinned dataset. (**a**) FBP reconstruction from 192 views. (**b**) FBP reconstruction from 2304 full views (reference). (**c**) DPMA reconstruction from 192 views. The display window is [-160, 240] HU.

requires moderately increased computation time while achieving improved reconstruction quality. FreeSeed and MVMS-RCN represent an intermediate computational cost category, with FreeSeed being slightly more efficient than MVMS-RCN, but MVMS-RCN achieves better reconstruction quality as evidenced by the quantitative metrics. Due to their iterative optimization nature, both DRONE (300 iterations) and DPMA (300 iterations) require substantially longer computation times, with DPMA achieving approximately three times better computational efficiency compared to DRONE while maintaining superior reconstruction quality. This computational cost analysis highlights the inherent trade-off between reconstruction quality and processing speed across different methodologies.

Discussion and limitations

While our DPMA framework demonstrates promising performance in sparse-view CT reconstruction, several limitations warrant consideration, particularly concerning clinical translation and broader applicability. Firstly, similar to many supervised deep learning methods, the performance of the DMA prior generation network relies heavily on the availability of extensive, high-quality paired training data (sparse-view inputs and corresponding high-quality reference images). This requirement poses practical challenges in clinical settings due to factors like patient anatomical variability, diverse scanning protocols, and ethical constraints on acquiring fully-sampled ground-truth data, which might limit the direct applicability or necessitate domain adaptation strategies. Secondly, the generation of the deep prior image x_p involves a multi-stage training process (SR, IE, and PIC modules trained sequentially or jointly), adding complexity to the model development and deployment pipeline compared to end-to-end approaches. Furthermore, while the MFA network within the DMA framework enhances feature extraction through parallel attention mechanisms, it introduces significant computational demands, especially when processing high-resolution images (e.g., 512×512). As indicated in Table 3, the runtime for generating the DMA prior alone (DMAprior) is substantially higher than simpler post-processing methods like FBPConvNet, reflecting a trade-off between reconstruction accuracy and computational efficiency for the prior generation stage. Although the subsequent iterative optimization is efficient, the overall complexity might be a factor in resource-constrained environments.

Despite these limitations, our hybrid approach offers advantages in robustness. The integration of the physicsbased iterative optimization module (Phase 2 in Algorithm 1) with the data-driven DMA prior makes the overall DPMA framework less susceptible to imperfections in the prior image compared to purely data-driven methods. The explicit enforcement of data consistency ($||Ax - s||_2^2$) ensures that the final reconstruction adheres to the measured physics, leading to more graceful performance degradation when encountering domain shifts between training and testing data, compared to end-to-end networks that might produce physically implausible results.

To further investigate the applicability in practical scenarios, we conducted an additional experiment using rebinned data from the AAPM dataset⁵⁴. The original spiral projections were rebinned to generate fanbeam projections with 2304 views, a detector size of 736 pixels, a source-to-axis distance of 595 mm, and a source-to-detector distance of 1085 mm. Sparse-view projections were obtained by uniformly downsampling the 2304 views to 192 views. As shown in Figure 11, the DPMA method effectively reconstructs images from these rebinned sparse-view projections, significantly suppressing noise and artifacts compared to the sparse-

view FBP (Figure 11a). However, some over-smoothing and loss of fine structural details persist (Figure 11c compared to 11b), which may not be optimal for clinical diagnosis. Furthermore, these results were obtained using 192 views, a number that might still be considered high in some ultra-low-dose scenarios. Future work should focus on further improving detail preservation and exploring the framework's performance with even sparser projection data.

Conclusion

This paper proposed a DPMA model for sparse-view CT reconstruction. First, we developed a residual regularization strategy that applies constraints on the difference between the prior image and target image. Then, a multi-scale fusion attention mechanism was designed to effectively capture and integrate features across multiple scales. Besides, we incorporated a physics-informed consistency module to ensure data fidelity in the dual-domain framework. Unlike existing deep learning methods that directly learn end-to-end mappings, our DPMA leverages the dual-domain approach with deep prior guidance during the iterative optimization process, making it more robust to patient-specific variations. Experimental results on simulated datasets demonstrated the effectiveness of DPMA in terms of noise suppression, artifact reduction, and anatomical detail preservation.

Data availability

The code for the DPMA framework is publicly available at https://github.com/jia-W-w/DPKA. The experimental datasets used in this study were derived from the AAPM Low Dose CT Grand Challenge https://www.aapm.or g/grandchallenge/lowdosect/.

Received: 26 February 2025; Accepted: 12 May 2025 Published online: 15 May 2025

References

- 1. Brenner, D. J. & Hall, E. J. Computed tomography-an increasing source of radiation exposure. *New England journal of medicine* 357, 2277–2284 (2007).
- 2. Goldman, L. W. Principles of ct: radiation dose and image quality. *Journal of nuclear medicine technology* **35**, 213–225 (2007).
- Zhang, J., Chen, B., Xiong, R. & Zhang, Y. Physics-inspired compressive sensing: Beyond deep unrolling. *IEEE Signal Processing Magazine* 40, 58–72 (2023).
- Sidky, E. Y. & Pan, X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine & Biology* 53, 4777 (2008).
- 5. Yu, H. et al. Weighted adaptive non-local dictionary for low-dose ct reconstruction. Signal Processing 180, 107871 (2021).
- 6. Xu, Q. et al. Low-dose x-ray ct reconstruction via dictionary learning. *IEEE transactions on medical imaging* **31**, 1682–1697 (2012).
- Komolafe, T. E. et al. Smoothed l0-constraint dictionary learning for low-dose x-ray ct reconstruction. *IEEE Access* 8, 116961–116973 (2020).
 Wu, W., Liu, F., Zhang, Y., Wang, Q. & Yu, H. Non-local low-rank cube-based tensor factorization for spectral ct reconstruction.
- 8. Wu, W., Liu, F., Zhang, Y., Wang, Q. & Yu, H. Non-local low-rank cube-based tensor factorization for spectral ct reconstruction. IEEE transactions on medical imaging 38, 1079–1093 (2018).
- Chen, Z., Qi, H., Wu, S., Xu, Y. & Zhou, L. Few-view ct reconstruction via a novel non-local means algorithm. *Physica Medica* 32, 1276–1283 (2016).
- Xia, W., Shan, H., Wang, G. & Zhang, Y. Physics-/model-based and data-driven methods for low-dose computed tomography: A survey. *IEEE Signal Processing Magazine* 40, 89–100 (2023).
- Adler, J. & Öktem, O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems* 33, 124007 (2017).
 Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural*
- *information processing systems* 25 (2012).
 13. Jin, K. H., McCann, M. T., Froustey, E. & Unser, M. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing* 26, 4509–4522 (2017).
- 14. Zhang, Z., Liang, X., Dong, X., Xie, Y. & Cao, G. A sparse-view ct reconstruction method based on combination of densenet and deconvolution. *IEEE transactions on medical imaging* 37, 1407–1417 (2018).
- Lee, H., Lee, J., Kim, H., Cho, B. & Cho, S. Deep-neural-network-based sinogram synthesis for sparse-view ct image reconstruction. IEEE Transactions on Radiation and Plasma Medical Sciences 3, 109–119 (2018).
- Nakai, H. et al. Quantitative and qualitative evaluation of convolutional neural networks with a deeper u-net for sparse-view computed tomography reconstruction. Academic radiology 27, 563–574 (2020).
- Kang, E., Min, J. & Ye, J. C. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. *Medical physics* 44, e360–e375 (2017).
- Lee, M., Kim, H. & Kim, H.-J. Sparse-view ct reconstruction based on multi-level wavelet convolution neural network. *Physica Medica* 80, 352–362 (2020).
- Tao, X. et al. Vvbp-tensor in the fbp algorithm: its properties and application in low-dose ct reconstruction. IEEE transactions on medical imaging 39, 764–776 (2019).
- Zhang, X. et al. Artifact and detail attention generative adversarial networks for low-dose ct denoising. IEEE Transactions on Medical Imaging 40, 3901-3918 (2021).
- 21. Adler, J. & Öktem, O. Learned primal-dual reconstruction. IEEE transactions on medical imaging 37, 1322-1332 (2018).
- 22. Xiang, J., Dong, Y. & Yang, Y. Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging* **40**, 1329–1339 (2021).
- Hu, D., Zhang, Y., Liu, J., Luo, S. & Chen, Y. Dior: Deep iterative optimization-based residual-learning for limited-angle ct reconstruction. *IEEE Transactions on Medical Imaging* 41, 1778–1790 (2022).
- 24. Chen, X. et al. Soul-net: A sparse and low-rank unrolling network for spectral ct image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems* 35, 18620–18634 (2024).
- Wu, W., Guo, X., Chen, Y., Wang, S. & Chen, J. Deep embedding-attention-refinement for sparse-view ct reconstruction. *IEEE Transactions on Instrumentation and Measurement* 72, 1–11 (2022).
- 26. Pan, J. et al. Iterative residual optimization network for limited-angle tomographic reconstruction. *IEEE Transactions on Image Processing* 33, 910–925 (2024).
- 27. Shi, C., Xiao, Y. & Chen, Z. Dual-domain sparse-view ct reconstruction with transformers. Physica Medica 101, 1-7 (2022).
- 28. Hu, D. et al. Hybrid-domain neural network processing for sparse-view ct reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences* 5, 88–98 (2020).

- Zhou, B. et al. Dudoufnet: Dual-domain under-to-fully-complete progressive restoration network for simultaneous metal artifact reduction and low-dose ct reconstruction. *IEEE Transactions on Medical Imaging* 41, 3587–3599 (2022).
- Chao, L. et al. Dual-domain attention-guided convolutional neural network for low-dose cone-beam computed tomography reconstruction. *Knowledge-Based Systems* 251, 109295 (2022).
- Li, R. et al. Ddptransformer: Dual-domain with parallel transformer network for sparse view ct image reconstruction. IEEE Transactions on Computational Imaging 8, 1101–1116 (2022).
- 32. Pan, J., Zhang, H., Wu, W., Gao, Z. & Wu, W. Multi-domain integrative swin transformer network for sparse-view tomographic reconstruction. *Patterns* **3**, 100498 (2022).
- Yang, C., Sheng, D., Yang, B., Zheng, W. & Liu, C. A dual-domain diffusion model for sparse-view ct reconstruction. IEEE Signal Processing Letters 31, 1279–1283 (2024).
- 34. Zhang, C. et al. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM* 64, 107–115 (2021).
- Chen, G.-H., Tang, J. & Leng, S. Prior image constrained compressed sensing (piccs): a method to accurately reconstruct dynamic ct images from highly undersampled projection data sets. *Medical physics* 35, 660–663 (2008).
- Wu, W. et al. Drone: Dual-domain residual-based optimization network for sparse-view ct reconstruction. IEEE Transactions on Medical Imaging 40, 3002–3014 (2021).
- Zhang, C., Li, Y. & Chen, G.-H. Accurate and robust sparse-view angle ct image reconstruction using deep learning and prior image constrained compressed sensing (dl-piccs). *Medical physics* 48, 5765–5781 (2021).
- Wu, W. et al. Stabilizing deep tomographic reconstruction: Part a. hybrid framework and experimental results. Patterns 3, 100474 (2022).
- 39. Zhang, Y. et al. Spie-dir: Self-prior information enhanced deep iterative reconstruction using two complementary limited-angle scans for dect. *IEEE Transactions on Instrumentation and Measurement* **72**, 1–12 (2022).
- 40. Tsaig, Y. & Donoho, D. L. Extensions of compressed sensing. Signal processing 86, 549-571 (2006).
- 41. Zha, Z. et al. Learning nonlocal sparse and low-rank models for image compressive sensing: Nonlocal sparse and low-rank modeling. *IEEE Signal Processing Magazine* **40**, 32-44 (2023).
- Chan, S. H., Khoshabeh, R., Gibson, K. B., Gill, P. E. & Nguyen, T. Q. An augmented lagrangian method for total variation video restoration. *IEEE Transactions on Image Processing* 20, 3097–3111 (2011).
- 43. Wang, Y., Yin, W. & Zeng, J. Global convergence of admm in nonconvex nonsmooth optimization. *Journal of Scientific Computing* 78, 29–63 (2019).
- Sidky, E. Y., Jørgensen, J. H. & Pan, X. Convex optimization problem prototyping for image reconstruction in computed tomography with the chambolle-pock algorithm. *Physics in Medicine & Biology* 57, 3065 (2012).
- Genzel, M., Gühring, I., Macdonald, J. & März, M. Near-exact recovery for tomographic inverse problems via deep learning. In International Conference on Machine Learning, 7368–7381 (PMLR, 2022).
- Scherzer, O. Convergence criteria of iterative methods based on landweber iteration for solving nonlinear problems. Journal of Mathematical Analysis and Applications 194, 911–933 (1995).
- Mei, Y., Fan, Y. & Zhou, Y. Image super-resolution with non-local sparse attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3517–3526 (2021).
- 48. Zhang, Y. et al. Kbnet: Kernel basis network for image restoration. arXiv preprint arXiv:2303.02881 (2023).
- Ma, C., Li, Z., Zhang, J., Zhang, Y. spsampsps Shan, H. Freeseed: Frequency-band-aware and self-guided network for sparse-view ct reconstruction. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 250–259 (Springer Nature Switzerland, 2023).
- 50. Fan, X., Chen, K., Yi, H., Yang, Y. & Zhang, J. Mvms-rcn: A dual-domain unified ct reconstruction with multi-sparse-view and multi-scale refinement-correction. *IEEE Transactions on Computational Imaging* **10**, 1749–1762 (2024).
- Zhang, Y. et al. Dream-net: Deep residual error iterative minimization network for sparse-view ct reconstruction. *IEEE Journal of Biomedical and Health Informatics* 27, 480–491 (2023).
- 52. Siddon, R. L. Fast calculation of the exact radiological path for a three-dimensional ct array. Medical physics 12, 252-255 (1985).
- Ye, J. C., Han, Y. & Cha, E. Deep convolutional framelets: A general deep learning framework for inverse problems. SIAM Journal on Imaging Sciences 11, 991–1048 (2018).
- 54. Wagner, F. *et al.* On the benefit of dual-domain denoising in a self-supervised low-dose ct setting. In 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), 1–5 (IEEE, 2023).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. U21 A20447, No. 62471077, and No. 62171073), the Project of the Central Government in Guidance of Local Science and Technology Development (No. 2024ZYD0270), and the Southwest Medical University Natural Science Foundation (No. 2023ZD004).

Author contributions

J.W. conceived the experiment, designed the framework, and conducted the main studies. J.L. and X.J. contributed to algorithm implementation and data analysis. W.Z. assisted with mathematical formulation and optimization. L.Z. and Y.P. helped with data curation and experimental validation. H.M. provided methodological guidance and critical review. Z.L. supervised the project and acquired funding. All authors participated in manuscript writing and revision.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Z.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025