Physics-Aware Accelerated Unrolling Model for Sparse-View CT Reconstruction

Shaojie Guo[1], Yingying Fang[2], Junkang Zhang[3], Yan Wang[1]*

[1]East China Normal University, Shanghai, China
[2]Bioengineering Department and National Heart and Lung Institute, Imperial College London, London SW7 2AZ, UK
[3]Nanjing University of Posts and Telecommunications, Nanjing, China

Abstract

Deep unrolling models (DUMs) have shown great potential in sparse-view CT reconstruction by combining iterative optimization and deep learning. However, most DUMs insufficiently account for physical degradation from sparse-view imaging, leading to slow convergence and persistent artifacts. To address this, we propose PAUM, a Physics-Aware Accelerated Unrolling Model explicitly incorporating CT imaging physics into the iterative reconstruction. PAUM first introduces a Dual-Domain Physics-Aware Extrapolation (DDPE) module. By modeling dual-domain degradations, it performs row-wise extrapolation in the sinogram domain to improve missing view recovery, and pixel-wise extrapolation in the image domain to address spatially variant degradation from incomplete backprojection. This physics-aware extrapolation aligns optimization dynamics with underlying physical imaging degradation, significantly enhances structural updates, thereby accelerating convergence. Subsequently, we develop a lightweight Block-Attention Deformable Regularization Network (BDRN), leveraging deformable convolutions and block-wise attention to model spatially variant and structured artifact physical characteristics. This enables spatially adaptive regularization on extrapolated results, effectively improving reconstruction quality. Extensive experiments demonstrate PAUM achieves over 1dB improvement compared to SOTA methods, while reducing iteration count by 50%.

Code — https://github.com/DeepMed-Lab-ECNU/PAUM

Introduction

Sparse-view CT is a critical technique for reducing radiation exposure by acquiring fewer projection views during scanning (Sidky, Kao, and Pan 2006). However, reconstructing images from such severely undersampled projections remains an ill-posed problem, often leading to severe artifacts and loss of fine details in the reconstructed images.

Early deep learning-based models (Han, Yoo, and Ye 2016; Zhang et al. 2018) aimed to directly map degraded observations to high-quality CT images through learnable networks. Although effective, they treat CT reconstruction as a general natural image recovery problem, neglecting the inherent physics model of CT imaging. Therefore, they struggle to capture spatially variant degradation patterns caused by sparse-view sampling, limiting their performance.

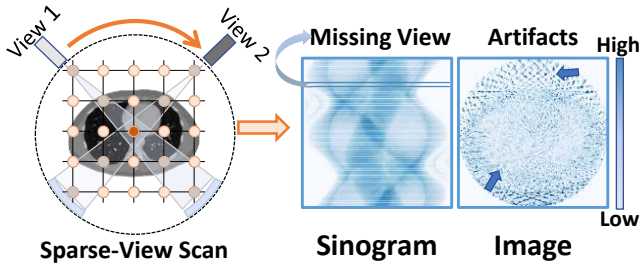


Figure 1: Visualization of degradation patterns derived from the residuals between observed sparse-view CT measurements and ground truth: missing projection views in the sinogram and spatially variant artifacts in the image domain - highlighting the need to explicitly model such physical degradations as a prior.

To better incorporate CT imaging model, Deep Unrolling Models (DUMs) (Wang et al. 2019a; Xiang, Dong, and Yang 2021; Gupta et al. 2018; Hu et al. 2022; Cheng et al. 2020; Adler and Öktem 2018) have gained popularity. These models unfold model-based iterative optimization algorithm into learnable neural networks, alternating between a model-based gradient descent step and a proximal mapping step to reconstruct sparse-view CT. In gradient descent step, the forward projection operator is used to simulate the CT acquisition process, and the reconstruction is updated by minimizing the discrepancy between its simulated and measured projections through gradient descent. In proximal mapping step, regularization networks are applied to improve the reconstruction based on learnable degradation priors.

Although existing DUMs incorporate the projection operator through gradient descent, they still treat CT reconstruction as a uniform residual minimization problem, without explicitly modeling the physical characteristics of degradation. Specifically, updates in these models are applied globally by multiplying a fixed learning rate with the gradient of the discrepancy between the reconstructed image and the measured projection. However, sparse-view acquisi-

*Corresponding author

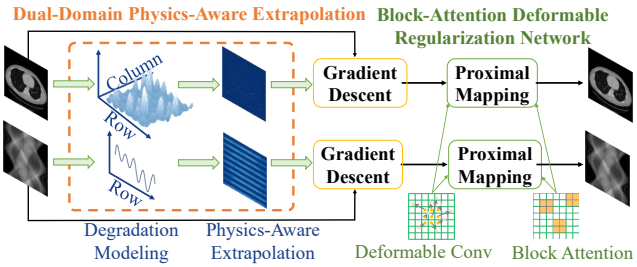


Figure 2: The proposed PAUM framework: Dual-Domain Physics-Aware Extrapolation (DDPE) for modeling physical degradation and accelerating convergence, followed by Block-Attention Deformable Regularization Network (BDRN) for spatially adaptive regularization.

tion inherently introduces spatially variant degradation patterns—such as missing projection rows in the sinogram domain and spatially variant artifacts in the image domain (see Fig. 1). These degradations introduce region-specific reconstruction challenges, with severely under-sampled structural areas demanding more accurate and adaptive updates. A globally fixed learning rate fails to accommodate such spatial variation, lacking the capacity to modulate update according to the spatial distribution of degradation. As a result, these models struggle to provide effective correction (see Fig. 4) to accurately restore the structure and details of the sparse CT during gradient descent steps, requiring more iterations to achieve acceptable reconstruction, ultimately leading to slower convergence.

Moreover, existing DUMs typically adopt generic CNN-based architectures for the proximal mapping step. However, due to limited receptive fields and fixed convolutional kernels, these networks are insufficient for capturing the spatially variant and anisotropic degradation patterns in sparse-view CT. As a result, they fail to perceive complex subtle artifacts and cannot adaptively regulate regularization strength across spatial degradation, ultimately limiting reconstruction quality.

In this work, we propose PAUM, a Physics-Aware Accelerated Unrolling Model that explicitly incorporates the imaging physics of sparse-view CT into unrolling framework (see Fig. 2). At each iteration, we introduce a Dual-Domain Physics-Aware Extrapolation(DDPE) module which models the physical distribution of degradation in dual-domain, thereby performing row-wise extrapolation in the sinogram domain to better recover missing projection views and pixel-wise extrapolation in the image domain to address spatially variant degradation caused by incomplete backprojection. This physics-aware extrapolation in DDPE aligns the update dynamics with the physics degradation, substantially improving the accuracy of dual-domain structural reconstruction and accelerating convergence.

Additionally, we propose a lightweight Block-attention Deformable Regularization Network (BDRN) specifically designed to suppress structured and spatially variant artifacts. While deformable convolutions combined with self-attention have shown effectiveness in modeling structured degradation (Zhu et al. 2020), their quadratic complexity limits efficiency in iterative reconstruction frameworks. Our BDRN addresses this by cascading deformable convolutions to capture artifact geometry from multiple directions, and employing a lightweight block-wise attention mechanism based on depthwise convolutions to spatially adaptive weighting according to the degree of region degradation. This design balances efficiency and effectiveness, significantly improving artifact suppression and reconstruction quality with minimal computational overhead.

Our contributions can be summarized as follows:

- We propose a Dual-Domain Physics-Aware Extrapolation (DDPE) module that accelerates convergence by aligning extrapolation behaviors with physical degradation patterns in both sinogram and image spaces.
- We design a Block-Attention Deformable Regularization Network (BDRN) that efficiently regulates and suppresses spatially variant degradations, achieving superior reconstruction with minimal computational cost.
- Our model achieves state-of-the-art performance on multiple sparse-view CT benchmarks, outperforming existing unrolling-based methods by over 1 dB PSNR while requiring only 50% iterations.

Related Work

Acceleration in Deep Unrolling

To alleviate the slow convergence of gradient update in DUMs, several acceleration strategies have been developed, most of which are solely based on optimization methods. U3Net (Chen, Quan, and Ji 2024) employs Nesterov momentum, Zhou et al. (Zhou et al. 2024) propose an LSTM-based recurrent momentum scheme, and MiJUN (Qin et al. 2025) adopts augmented half-quadratic splitting. QN-Mixer (Ayad, Larue, and Nguyen 2024) introduces second-order quasi-Newton updates for faster convergence, but incurs high memory cost due to the need for Hessian approximations. These methods are primarily designed for general inverse problems and fail to be tailored to the inherent degradation characteristics of reconstruction tasks, which may result in limited performance improvements. In contrast, our DDPE module introduces physics-aware extrapolation in both sinogram and image domains. By explicitly modeling dual-domain degradations based on sparse-view CT imaging physics, optimization dynamics precisely align with underlying physical degradation patterns, thereby achieving more targeted updates and faster convergence.

Deformable Convolution and Attention

Deformable convolutions (Dai et al. 2017; Wang et al. 2019b; Zhu et al. 2019) enhance spatial modeling by learning adaptive offsets, enabling flexible representation of spatially variant structures. Recent works (Zhu et al. 2020; Xia et al. 2022) further combine deformable convolutions with self-attention to fuse local flexibility and global context, improving performance on vision tasks with complex spatial dependencies. However, standard self-attention modules often introduce quadratic complexity, making them inefficient for iterative unrolling-based reconstruction frameworks.

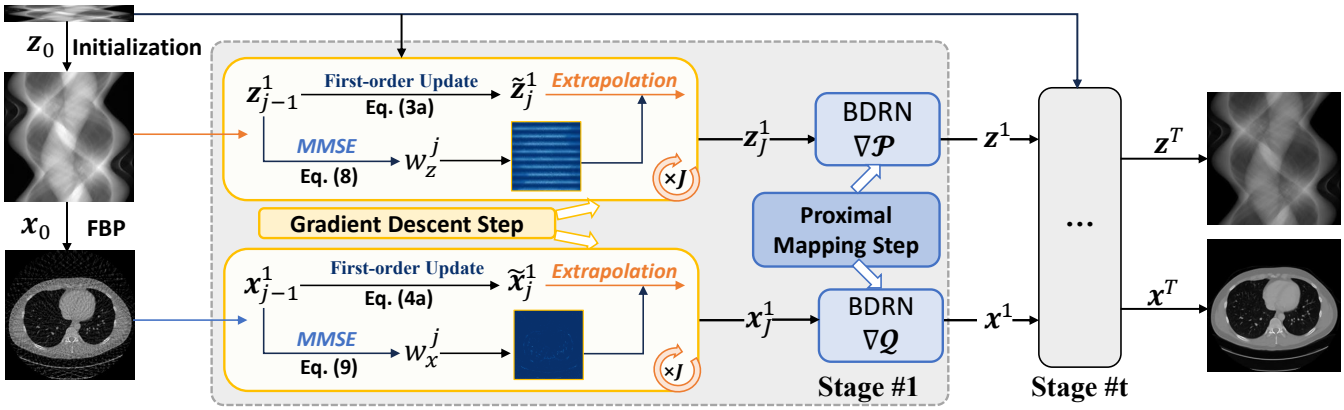


Figure 3: Overview of the proposed Physics-Aware Accelerated Unrolling Model. Gradient updates are accelerated through a physics-aware extrapolation method which models physical degradation. The extrapolation matrix is generated via Bayesian Minimum Mean Square Error (MMSE) estimation. Then, the lightweight Block-Attention Deformable Regularization Network (BDRN) is used to regularize the extrapolated results, leveraging Deformable Attention mechanisms to address spatially variant and structured artifact physical characteristics.

In sparse-view CT reconstruction, degradations vary spatially in severity across different regions. To address this, we propose a lightweight Block-attention Deformable Regularization Network (BDRN). By integrating cascaded deformable convolutions with a block-wise attention mechanism, BDRN effectively models both prominent and subtle spatially variant degradations, enabling spatially adaptive regularization at low computational cost.

Methodology

Problem Formulation

Sparse-view CT reconstruction can be formulated as a joint minimization problem in both the image and sinogram domains:

$$\min_{\boldsymbol{z},\boldsymbol{x}} \frac{1}{2}\|\mathbf{R}\boldsymbol{x}-\boldsymbol{z}\|_2^2+\frac{\lambda}{2}\|\mathbf{P}\boldsymbol{z}-\boldsymbol{z}_0\|_2^2+\mathcal{P}(\boldsymbol{x})+\mathcal{Q}(\boldsymbol{z}), \quad (1)$$

where $\boldsymbol{x}$ is the target CT image and $\boldsymbol{z}$ is its corresponding full-view projection. $\boldsymbol{z}_0$ denotes the observed sparse-view projection. $\mathbf{R}$ denotes the forward projection operator, which simulates the CT acquisition process using the Radon transform. $\mathbf{P}$ represents the sparse-view sampling operator. $\mathcal{P}(x)$ and $\mathcal{Q}(x)$ are regularization priors, and λ is a trade-off parameter.

By alternately optimizing one objective variable while keeping the other fixed, Eq. 1 can be solved as follows:

$$\begin{cases} \boldsymbol{z}^{(t)}=\min_{\boldsymbol{z}} \frac{1}{2}\|\mathbf{R}\boldsymbol{x}^{(t-1)}-\boldsymbol{z}\|_2^2+\frac{\lambda}{2}\|\mathbf{P}\boldsymbol{z}-\boldsymbol{z}_0\|_2^2+\mathcal{Q}(\boldsymbol{z}), & (2a) \\ \boldsymbol{x}^{(t)}=\min_{\boldsymbol{x}} \frac{1}{2}\|\mathbf{R}\boldsymbol{x}-\boldsymbol{z}^{(t)}\|_2^2+\mathcal{P}(\boldsymbol{x}), & (2b) \end{cases}$$

The $\boldsymbol{z}^{(0)}$ is initial full-view projection, and the initial image estimate $\boldsymbol{x}^{(0)}$ is obtained by applying the inverse Radon transform to $\boldsymbol{z}^{(0)}$.

Baseline Optimization: PGD

Previous unrolled methods typically solve Eq.2 using proximal gradient descent (PGD), where each update is decomposed into a gradient descent (GD) step followed by a proximal mapping (PM). For $\boldsymbol{z}^{(t)}$, the updates are given by:

$$\begin{cases} \tilde{\boldsymbol{z}}^{(t)}=\boldsymbol{z}^{(t-1)}-u^{(t-1)}\Big(\boldsymbol{z}^{(t-1)}-\mathbf{R}\boldsymbol{x}^{(t-1)} \\ \qquad\qquad +\lambda\mathbf{P}^{\top}(\mathbf{P}\boldsymbol{z}^{(t-1)}-\boldsymbol{z}_0)\Big), & (3a) \\ \boldsymbol{z}^{(t)}=\tilde{\boldsymbol{z}}^{(t)}-\nabla\mathcal{Q}(\boldsymbol{z}^{(t-1)}), & (3b) \end{cases}$$

where u is the globally fixed learning rate, and $\mathbf{P}^{\top}$ denotes the adjoint of the sparse-view transform operator $\mathbf{P}$. $\tilde{\boldsymbol{z}}^{(t)}$ is an auxiliary variable to decouple the gradient and proximal steps. Eq.3a performs a first-order gradient descent update on the L2 data fidelity term in Eq. 2a, while Eq.3b refines the intermediate result via the regularization prior.

Similarly, the image update $\boldsymbol{x}^{(t)}$ follows:

$$\begin{cases} \tilde{\boldsymbol{x}}^{(t)}=\boldsymbol{x}^{(t-1)}-v^{(t-1)}\mathbf{R}^{\top}(\mathbf{R}\boldsymbol{x}^{(t-1)}-\boldsymbol{z}^{(t)}), & (4a) \\ \boldsymbol{x}^{(t)}=\tilde{\boldsymbol{x}}^{(t)}-\nabla\mathcal{P}(\boldsymbol{x}^{(t)}), & (4b) \end{cases}$$

where v is the globally fixed learning rate and $\mathbf{R}^{\top}$ denotes the adjoint Radon transform.

The baseline PGD framework offers simplicity and interpretability, making it a common choice for unrolled reconstruction. But, it suffers from two limitations in CT:

1. The gradient descent steps in PGD (Eqs. 3a and 4a) treat CT reconstruction as a uniform residual minimization problem for the data fidelity term, applying a fixed learning rate across all spatial locations. However, this global update scheme fails to model the spatially variant degradation introduced by sparse-view imaging—such as missing projection rows in the sinogram domain and spatially variant artifacts in the image domain (Fig. 1). Without explicitly accommodating pattern-specific degrada-

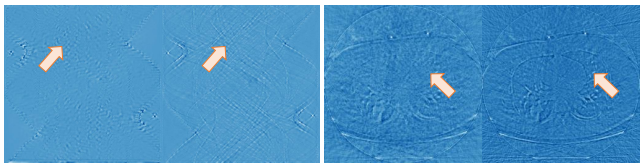


Figure 4: Visualization comparing traditional global extrapolation (GlobExt) and our extrapolated results in both sinogram and image domains. Our extrapolation better preserves the structural characteristics in both domains, effectively promoting updates in these structures.

tion, it results in inefficient updates and slow convergence.

2. The regularization terms $\nabla\mathcal{Q}(\boldsymbol{z})$ and $\nabla\mathcal{P}(\boldsymbol{x})$ are typically implemented using generic CNNs with limited receptive fields and fixed convolutional kernels, making them inadequate for capturing complex, structured degradations in sparse-view CT across both the sinogram and image domains. As a result, these regularizers struggle to encode effective priors for artifact suppression, leading to persistent artifacts and limiting reconstruction fidelity.

To address these limitations, we propose two key components: a Dual-Domain Physics-Aware Extrapolation (**DDPE**) module that models domain-specific degradation to adaptively modulate updates in two domains to accurate CT images, and a Block-Attention Deformable Regularization Network (**BDRN**) that captures structured, spatially variant artifacts for effective regularization. Fig. 3 illustrates the overall dual-domain unrolling pipeline, and Algorithm 1 further details the iterative procedure, showing how DDPE and BDRN are jointly integrated to accelerate convergence while preserving reconstruction fidelity.

Dual-Domain Physics-Aware Extrapolation

Classical Extrapolation. Extrapolation is an effective strategy in iterative optimization (Chen, Quan, and Ji 2024) for accelerating convergence. In each gradient descent step, classical extrapolation incorporates historical updates from previous iterations to estimate the next update direction, introducing momentum-like dynamics that improve optimization efficiency beyond using current gradients alone. In the context of sinogram-domain updates (Eq. 3a), this is typically applied over multiple inner iterations within each outer iteration. Letting j denote the inner step index, the extrapolated update follows:

$$\tilde{\boldsymbol{z}}_{(j)}^{(t-1)} \leftarrow \tilde{\boldsymbol{z}}_{(j)}^{(t-1)} + w\left(\tilde{\boldsymbol{z}}_{(j)}^{(t-1)} - \tilde{\boldsymbol{z}}_{(j-1)}^{(t-1)}\right), \tag{5}$$

where w is a globally fixed extrapolation factor shared across all projection rows and iterations. While global extrapolation introduces momentum to accelerate convergence, it overlooks the spatially variant degradation in sparse-view CT (see Fig. 1), where different regions suffer varying levels of reconstruction difficulty. Applying a uniform extrapolation strength across all locations fails to account for this heterogeneity, leading to suboptimal or even unstable updates(see Fig. 4).

Algorithm 1: Physics-Aware Accelerated Unrolling Model

Input:$\boldsymbol{z}^{(0)}$,$\boldsymbol{x}^{(0)}$,λ,T,J,σ

for $t = 1$ to T **do**
  // Multiple Physics-Aware Extrapolation for z
  for $j = 1$ to J **do**
    $\boldsymbol{w}_z^{(j)} \leftarrow \{\boldsymbol{z}^{(j-1)}, \boldsymbol{z}^{(j-2)}, \sigma\}$ by Eq.8
    $\boldsymbol{z}_{(j)}^{(t-1)} \leftarrow \{\boldsymbol{z}_{(j-1)}^{(t-1)}, \boldsymbol{x}^{(t-1)}, \boldsymbol{z}_0\}$ by Eq.3a
  end for
  $\boldsymbol{z}^{(t)} \leftarrow \boldsymbol{z}_{(J)}^{(t-1)}$ by Eq.3b
  // Multiple Physics-Aware Extrapolation for x
  for $j = 1$ to J **do**
    $\boldsymbol{w}_x^{(j)} \leftarrow \{\boldsymbol{x}^{(j-1)}, \boldsymbol{x}^{(j-2)}, \sigma\}$ by Eq.9
    $\boldsymbol{x}_{(j)}^{(t-1)} \leftarrow \{\boldsymbol{x}_{(j-1)}^{(t-1)}, \boldsymbol{z}^{(t)}\}$ by Eq.4a
  end for
  $\boldsymbol{x}^{(t)} \leftarrow \boldsymbol{x}_{(J)}^{(t-1)}$ by Eq.4b
end for
Output: $\boldsymbol{x}^{(T)}$, $\boldsymbol{z}^{(T)}$

Physics-Aware Extrapolation. For more accurate gradient update, we propose a Dual-Domain Physics-Aware Extrapolation (see Alg.1). Unlike conventional extrapolation schemes that rely on fixed step sizes, DDPE dynamically modulates the extrapolation strength by explicitly modeling the degradation patterns in both sinogram and image domains. Given that view-missing patterns vary across sinogram rows and backprojection-induced artifacts exhibit strong spatial non-uniformity, we model each sinogram row $\boldsymbol{z}_h$ and image pixel $\boldsymbol{x}_{h,w}$ as independent random variables.

The rationality of this assumption stems from the dominant degradation sources—row-wise view loss in the sinogram and locally structured streak artifacts in the image—are inherently local and vary across rows and spatial locations, making row-wise and pixel-wise modeling a natural and effective choice. We further assume that the extrapolated intermediate value $\boldsymbol{z}_h^{(j)}$ in the inner iterations serves as an unbiased noisy observation of an underlying clean latent variable $\boldsymbol{z}_h^*$, corrupted by independent Gaussian noise:

$$\boldsymbol{z}_h^{(j-1)} \sim \mathcal{N}(\boldsymbol{z}_h^*, \sigma_{j-1}^2 I), \quad \boldsymbol{z}_h^{(j)} \sim \mathcal{N}(\boldsymbol{z}_h^*, \sigma_j^2 I), \tag{6}$$

Here, $\boldsymbol{z}_h^*$ denotes the ideal full-view sinogram row, and $\boldsymbol{x}_{h,w}^*$ represents the ground-truth CT value at pixel (h, w). We aim to minimize the expected squared error between the next extrapolation result $\boldsymbol{z}_h^{(j+1)}$ and the latent clean signal $\boldsymbol{z}_h^*$ to find the optimal extrapolation weight $w_{z,h}^{(j+1)}$:

$$L(w_{z,h}^{(j+1)}) = \mathbb{E}\left[\left|\boldsymbol{z}_h^{(j+1)} - \boldsymbol{z}_h^*\right|^2\right], \tag{7}$$

where $w_{z,h}^{(j+1)} \in \mathbb{R}^{H\times 1}$. Using Bayesian minimum mean square error (MMSE) estimation, the optimal extrapolation

matrix can be derived as follows:

$$w_{z,h}^{(j+1)} = \frac{\sigma^2}{r_{z,h}^{(j)} + \sigma^2}, \quad r_{z,h}^{(j)} = \left\| \boldsymbol{z}_h^{(j)} - \boldsymbol{z}_h^{(j-1)} \right\|_2^2, \quad (8)$$

where σ^2 is a stable constant, and $r_{z,h}^{(j)}$ denotes the residual between successive iterations. Intuitively, our extrapolation weight $w_{z,h}^{(j+1)}$ adapts based on the magnitude of the local residual between successive iterations. Projection rows with larger residuals (typically from severely undersampled regions) are assigned smaller extrapolation weights to prevent update overshooting, enabling precise correction, while stable regions receive stronger extrapolation to accelerate convergence.

As for the image domain, we assume that the extrapolated image pixel value $\boldsymbol{x}_{h,w}^{(j)}$ at the jth iteration is an unbiased noisy observation of the latent clean signal $\boldsymbol{x}_{h,w}^{*}$, affected by spatially variant artifact. The pixel-wise extrapolation weights $\boldsymbol{w}_{x,hw}^{(j)} \in \mathbb{R}^{H \times W}$ are computed as:

$$\boldsymbol{w}_{x,hw}^{(j+1)} = \frac{\sigma^2}{r_{x,hw}^{(j)} + \sigma^2}, r_{x,hw}^{(j)} = \left\| \boldsymbol{x}_{h,w}^{(j)} - \boldsymbol{x}_{h,w}^{(j-1)} \right\|_2^2, \quad (9)$$

where σ^2 is a stabilization constant. The weight $\boldsymbol{w}_{x,hw}^{(j)}$ adaptively modulates the extrapolation strength at each pixel according to the local residual dynamics in the image.

By explicitly modeling domain-specific degradations in both sinogram and image domains, DDPE dynamically adjusts extrapolation strength to enable targeted corrections. As shown in Fig. 4, the extrapolated updates modulated by $\boldsymbol{w}_z$ and $\boldsymbol{w}_x$ offer finer and more reliable structure recovery than traditional global extrapolation, particularly in severe degraded regions. Quantitative comparisons (Tab. 3) further verify that DDPE significantly accelerates convergence and improves reconstruction stability over globally fixed step extrapolation methods.

Block-Attention Deformable Regularization

To effectively address the spatially variant and structured degradation in sparse-view CT, we propose a unified regularization network for $\mathcal{P}(\boldsymbol{x})$, which integrates two key components: cascaded deformable convolutions and a lightweight block-wise attention mechanism. This design enables precise modeling of anisotropic artifact patterns and provides an efficient regularizer for sparse-view reconstruction.

The pipeline begins with the current reconstructed image $\boldsymbol{x}$, which is first passed through a multi-scale convolutional (MSC) to extract hierarchical feature representations with varying receptive fields. These features are then passed through a series of cascaded deformable convolution (DCN) blocks, where each block adaptively samples from the input using learned offsets Δp_k and modulation scalars m_k, allowing the network to align distorted features caused by structured artifacts such as directional streaks. The output of each DCN layer is computed as:

$$\hat{\boldsymbol{F}}(p_0) = \sum_{k=1}^{K} w_k \cdot \boldsymbol{F}(p_0 + p_k + \Delta p_k) \cdot m_k, \quad (10)$$

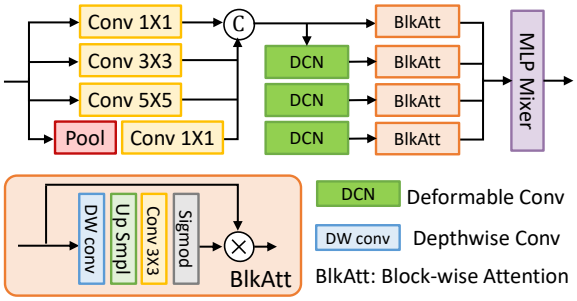


Figure 5: Overview of the proposed Block-Attention Deformable Regularization (BDRN) network.

where w_k are the kernel weights, and the independently learned offsets and masks across cascaded deformable convolution blocks allow the network to progressively model complex degradation patterns across scales and directions.

To highlight degraded regions without incurring heavy computation, we insert a block-wise attention module after each cascade DCN block. Specifically, the feature map is partitioned into non-overlapping spatial blocks $\mathcal{B}_i$ and each block is assigned a scalar attention weight $a_i \in [0, 1]$, which is applied uniformly to all pixels within that block:

$$\boldsymbol{F}_{\text{att}}(p) = a_i \cdot \boldsymbol{F}(p), \quad \forall p \in \mathcal{B}_i. \quad (11)$$

The attention weight a_i are generated via a depthwise convolution with stride equal to block size, effectively aggregating local context within each spatial block. These weights are then upsampled to match the feature resolution and applied multiplicatively to modulate the corresponding regions.

By combining cascaded deformable convolutions with lightweight block-wise attention, our regularizer adaptively aligns local distortions and emphasizes spatially degraded regions. The learned offsets and attention weights enable domain-aware correction, and the concatenated features jointly encode geometric misalignment and degradation saliency. To enhance global context, a lightweight MLP-Mixer is appended. The regularizer shares structure and parameters across unrolled stages for $\nabla\mathcal{P}(\boldsymbol{x})$, and the same design is applied to $\nabla\mathcal{Q}(\boldsymbol{z})$ for consistency and efficiency.

Experiments

Datasets. We evaluate our method on two benchmarks: the AAPM 2016 Low-Dose CT Challenge dataset (McCollough 2016) and the NBIA dataset (NIH2011). Following (Ding et al. 2024), we sample the same 500 images and their corresponding sinograms from AAPM, and 200 from NBIA, using 80% for training and 20% for testing. Each full-view sinogram contains 512 detector elements and 1024 projection angles, which are downsampled to 64 or 128 views for sparse-view experiments. Reconstruction quality is measured by PSNR and SSIM on 256×256 CT images.

Training Details. We adopt distance-driven projection/back-projection algorithms (De Man and Basu 2002, 2004) via CTLIB (Xia et al. 2021) in PyTorch. The Adam optimizer is used with learning rates of 1e-4 for the image regularizer and 6e-5 for the sinogram regularizer. Following

Model		AAPM				NBIA			
		64 Views		128 Views		64 Views		128 Views	
Type	**Name**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
-	FBP	27.18 ± 1.11	0.596 ± 9e-4	32.88 ± 1.35	0.763 ± 1e-3	25.85 ± 1.89	0.596 ± 2e-3	30.80 ± 3.33	0.748 ± 2e-3
Deep learning	FBPConvNet	35.16 ± 2.83	0.896 ± 2e-4	39.74 ± 2.24	0.958 ± 6e-5	31.60 ± 4.27	0.830 ± 1e-3	36.74 ± 1.89	0.929 ± 7e-5
	FreeSeed	38.21 ± 0.84	0.935 ± 9e-3	42.45 ± 0.75	0.964 ± 5e-5	35.94 ± 1.14	0.863 ± 2e-4	39.62 ± 1.18	0.933 ± 1e-4
	DuDoTrans	35.82 ± 0.92	0.961 ± 3e-5	40.96 ± 0.74	0.984 ± 8e-6	33.05 ± 0.80	0.937 ± 1e-4	37.41 ± 1.25	0.968 ± 7e-5
	PrideDiff	39.45 ± 1.95	0.963 ± 2e-2	46.35 ± 1.33	0.991 ± 3e-4	39.02 ± 2.01	0.962 ± 6e-4	44.35 ± 1.53	0.986 ± 1e-3
Unrolling	LAMA	44.92 ± 1.11	0.987 ± 7e-6	49.92 ± 0.66	0.995 ± 6e-7	43.31± 3.37	0.983 ± 4e-5	48.39 ± 3.64	0.993 ± 8e-6
	PAIL	45.36 ± 1.14	0.986 ± 1e-5	51.24 ± 0.41	0.994 ± 4e-6	43.11 ± 2.82	0.947 ± 1e-3	48.50 ± 5.45	0.960 ± 9e-5
	Ours	**46.47 ± 1.21**	**0.990 ± 4e-6**	**51.91 ± 0.49**	**0.996 ± 2e-7**	**44.49 ± 3.11**	**0.985 ± 3e-5**	**49.38 ± 4.40**	**0.994 ± 7e-6**

Table 1: Quantitative evaluation on the AAPM and NBIA datasets with 64 and 128 views. **Bold**: Best, Underline: Second best.

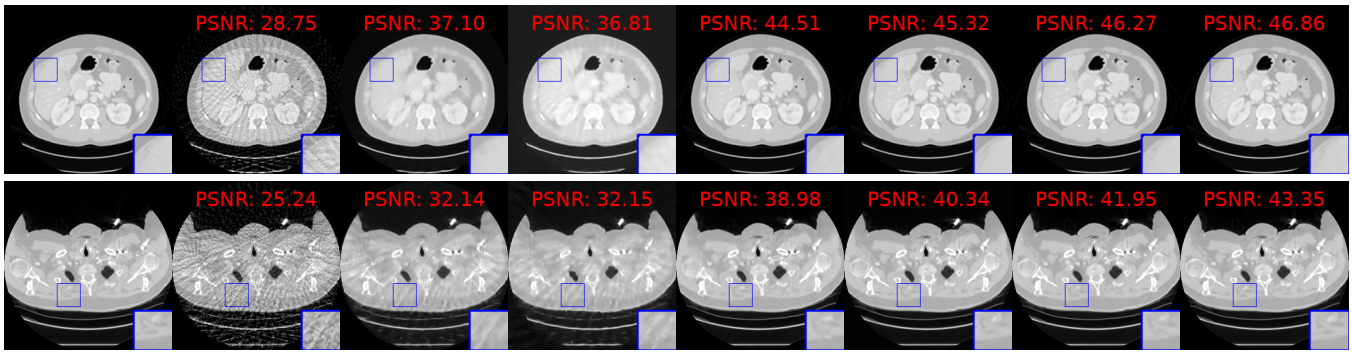


Figure 6: Visual comparison on the AAPM dataset (first row) and the NBIA dataset (second row).

LAMA (Ding et al. 2024), training starts with 3 unrolled iterations and increases by 2 per stage. All stages run for 200 epochs, except the third, which trains for 300 epochs. All hyperparameters, including those in the regularizers and extrapolation modules, are learned end-to-end during training.

Comparison with State-of-the-Art Methods

We compare our method with both deep learning-based and unrolling-based approaches. Deep learning baselines include FBPConvNet (Jin et al. 2017), DuDoTrans (Wang et al. 2022), FreeSeed (Ma et al. 2023), and PrideDiff (Lu et al. 2024), a representative diffusion-based reconstruction method. For unrolling-based counterparts, we benchmark against Learn++ (Yi et al. 2023), LAMA (Ding et al. 2024), and the recent PAIL (Zhang et al. 2025), which represents a strong latest advancement in this category. We also include Filtered Back Projection (FBP) as a conventional baseline for reference.

Quantitative Comparison. As shown in Tab. 1, our method consistently achieves the best performance across all projection settings. Notably, on the NBIA dataset with only 64 views, our model surpasses the dual-domain method LAMA by 1.18 dB and outperforms the diffusion-based PrideDiff by over 5 dB, demonstrating superior robustness in severely undersampled scenarios. This improvement highlights the effectiveness of our design, which tightly integrates physical degradation modeling with spatially adaptive regularization.

Method	PrideDiff	Learn++	LAMA	PAIL	Ours
Iters	10	50	15	10	5
Params(M)	4.22	6.22	0.46	7.54	1.32
PSNR	39.20	40.52	43.31	43.10	44.49

Table 2: Comparison of the computational efficiency.

Visual Comparison. Fig. 4 presents visual results on the AAPM and NBIA datasets under the 64-view setting. While PAIL and LAMA yield reasonable reconstructions, they struggle to preserve fine anatomical details, especially around structural boundaries. In contrast, our method achieves sharper contours and clearer textures, demonstrating superior restoration of subtle features. This highlights the strength of our model in adaptively enhancing critical regions through physics-aware modulation.

Efficiency Comparison Tab. 2 summarizes the computational efficiency of different methods. Although our inference time is not the lowest, our model requires the fewest iterations—half of SOTA methods—and uses only 1.32M parameters while achieving the highest PSNR. This demonstrates that, through physics-aware extrapolation and a powerful regularization network, our approach combines effi-

Method	NoExt	GlobExt	InvMod	Ours
PSNR	39.46	42.44	43.78	**44.49**

Table 3: Comparison of different extrapolation strategies. Our DDEP achieves the highest PSNR.

Steps	0	2	4	6	8	10	12
PSNR	39.46	40.96	41.93	43.34	44.49	44.55	44.60
SSIM	0.967	0.973	0.977	0.982	0.985	0.985	0.986
Time(s)	0.20	0.22	0.23	0.25	0.27	0.30	0.31

Table 4: Effect of inner extrapolation steps.

ciency and performance, providing high-quality reconstruction with minimal iteration cost.

Ablation Study and Analysis

We conduct ablation experiments on the NBIA dataset (64 views) to validate the contributions of our key components.

Physics-Aware Extrapolation. As shown in Tab. 3, our EEDP clearly outperforms both the baseline without extrapolation (NoExt) and the traditional global strategy (GlobExt), demonstrating the advantage of domain-specific modeling. By jointly capturing structure priors in two domains, EEDP enables more accurate and efficient reconstruction, consistently leading to superior performance.

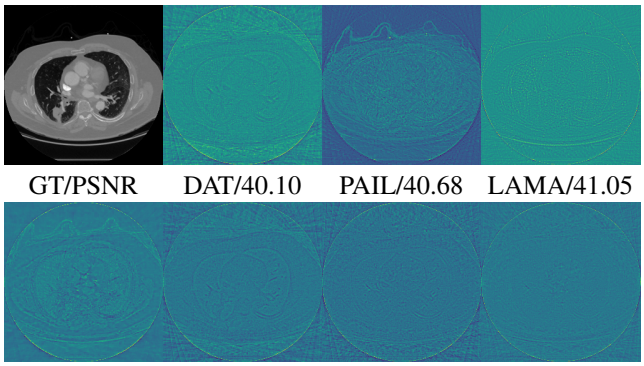
To further validate the importance of domain-specific modeling, we evaluate an inverted variant model (InvMod) that swaps the extrapolation schemes(i.e., applying row-wise extrapolation in the image domain and pixel-wise extrapolation in the sinogram domain). The significant performance drop confirms the necessity of aligning extrapolation strategies with the physical characteristics of each domain.

Additionally, Tab. 4 reports the performance and runtime under different numbers of internal extrapolation steps J. The results show that our method achieves a significant +5.03 dB improvement in reconstruction quality with just 8 internal steps, while incurring only 0.07 seconds of additional inference time. This demonstrates the efficiency and scalability of our physics-aware extrapolation strategy.

Regularization Network BDRN. We conduct a detailed ablation study to assess the contributions of each component in our Block-Attention Deformable Regularization network. We use a lightweight MLP-Mixer baseline and incrementally add components to assess their contributions. As shown in Tab. 5, adding a multi-scale convolutional (MSC) feature extractor enhances local representation, yielding 42.36 dB. Further integrating a block-wise attention (BlkAtt) mechanism with cascaded standard convolutions provides a large boost to 43.88 dB by emphasizing spatially degraded regions. Replacing standard convolutions with deformable convolutions (DCN) leads to a 0.61 dB gain, confirming the effectiveness of our cascaded Block-Attention Deformable Regularization design.

Method	Baseline	+MSC	+BlkAtt	+DCN	DAT
PSNR	40.51	42.36	43.88	**44.49**	42.66
Params(M)	**1.07**	1.08	1.44	1.32	1.19
Memory(GB)	**3.50**	5.12	8.65	6.48	47.44

Table 5: Ablation study of regularization module.



GT/PSNR DAT/40.10 PAIL/40.68 LAMA/41.05

Baseline/39.00 +MSC/41.02 +BlkAtt/42.25 +DCN/**42.79**

Figure 7: Feature maps of different regularizers. Unlike others focusing on anatomical structures, ours highlights artifacts and noise for precise suppression.

We also adopt the Deformable Attention Transformer (DAT) (Xia et al. 2022) as the regularizer to compare with our Block-Attention Deformable design. While DAT offers greater modeling flexibility by ViT, its repeated use across unrolling stages leads to high memory overhead and unstable training. In contrast, our lightweight design achieves better performance with higher efficiency.

Visual Analysis of Regularization Features. As shown in Fig. 7, general-purpose regularizers (e.g., the MLPMixer-based baseline and U-Net-based PAIL) tend to produce broad responses aligned with prominent anatomical structures, while largely ignoring subtle, spatially varying artifacts. In contrast, our regularizer exhibits focused activations specifically in artifact-affected regions. This artifact-aware behavior closely aligns with the underlying degradation patterns, enabling more precise artifact suppression and better preservation of details, ultimately enhancing reconstruction.

Conclusion

In this paper, we propose a physics-aware accelerated unrolling framework for sparse-view CT reconstruction. By explicitly modeling degradation in both the sinogram and image domains via Dual-Domain Physics-Aware Extrapolation, and integrating a Block-wise Attention Deformable Regularization Network, PAUM achieves faster convergence and higher reconstruction quality. Extensive experiments demonstrate that PAUM outperforms state-of-the-art methods in both accuracy and efficiency. However, our model still relies on multiple iterative updates. Future work will explore further reducing computational overhead to achieve reconstruction efficiency comparable to deep learning methods.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 62471182), Science and Technology Commission of Shanghai Municipality Basic Research Program (Grant No. 25JD1401300), Shanghai Rising-Star Program (Grant No. 24QA2702100), and the Science and Technology Commission of Shanghai Municipality (Grant No. 22DZ2229004)

References

Adler, J.; and Öktem, O. 2018. Learned primal-dual reconstruction. *IEEE transactions on medical imaging*, 37(6): 1322–1332.

Ayad, I.; Larue, N.; and Nguyen, M. K. 2024. QN-Mixer: A Quasi-Newton MLP-Mixer Model for Sparse-View CT Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 25317–25326.

Chen, Z.; Quan, Y.; and Ji, H. 2024. Unsupervised Deep Unrolling Networks for Phase Unwrapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 25182–25192.

Cheng, W.; Wang, Y.; Li, H.; and Duan, Y. 2020. Learned full-sampling reconstruction from incomplete data. *IEEE Transactions on Computational Imaging*, 6: 945–957.

Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; and Wei, Y. 2017. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, 764–773.

De Man, B.; and Basu, S. 2002. Distance-driven projection and backprojection. In *2002 IEEE Nuclear Science Symposium Conference Record*, volume 3, 1477–1480. IEEE.

De Man, B.; and Basu, S. 2004. Distance-driven projection and backprojection in three dimensions. *Physics in Medicine & Biology*, 49(11): 2463.

Ding, C.; Zhang, Q.; Wang, G.; Ye, X.; and Chen, Y. 2024. LAMA: Stable Dual-Domain Deep Reconstruction For Sparse-View CT. *arXiv preprint arXiv:2410.21111*.

Gupta, H.; Jin, K. H.; Nguyen, H. Q.; McCann, M. T.; and Unser, M. 2018. CNN-based projected gradient descent for consistent CT image reconstruction. *IEEE transactions on medical imaging*, 37(6): 1440–1453.

Han, Y. S.; Yoo, J.; and Ye, J. C. 2016. Deep residual learning for compressed sensing CT reconstruction via persistent homology analysis. *arXiv preprint arXiv:1611.06391*.

Hu, D.; Zhang, Y.; Liu, J.; Luo, S.; and Chen, Y. 2022. DIOR: Deep iterative optimization-based residual-learning for limited-angle CT reconstruction. *IEEE Transactions on Medical Imaging*, 41(7): 1778–1790.

Jin, K. H.; McCann, M. T.; Froustey, E.; and Unser, M. 2017. Deep convolutional neural network for inverse problems in imaging. *IEEE transactions on image processing*, 26(9): 4509–4522.

Lu, Z.; Gao, Q.; Wang, T.; Yang, Z.; Wang, Z.; Yu, H.; Chen, H.; Zhou, J.; Shan, H.; and Zhang, Y. 2024. PrideDiff: Physics-regularized generalized diffusion model for CT reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*.

Ma, C.; Li, Z.; Zhang, J.; Zhang, Y.; and Shan, H. 2023. FreeSeed: Frequency-band-aware and self-guided network for sparse-view CT reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 250–259. Springer.

McCollough, C. 2016. TU-FG-207A-04: overview of the low dose CT grand challenge. *Medical physics*, 43(6Part35): 3759–3760.

NIH2011. 2011. National Institutes of Health. https://imaging.nci.nih.gov/ncia/login.jsf. Accessed 28 February 2011.

Qin, M.; Feng, Y.; Wu, Z.; Zhang, Y.; and Yuan, X. 2025. Detail Matters: Mamba-Inspired Joint Unfolding Network for Snapshot Spectral Compressive Imaging. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 6594–6602.

Sidky, E. Y.; Kao, C.-M.; and Pan, X. 2006. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-ray Science and Technology*, 14(2): 119–139.

Wang, C.; Shang, K.; Zhang, H.; Li, Q.; and Zhou, S. K. 2022. DuDoTrans: dual-domain transformer for sparse-view CT reconstruction. In *International Workshop on Machine Learning for Medical Image Reconstruction*, 84–94. Springer.

Wang, J.; Zeng, L.; Wang, C.; and Guo, Y. 2019a. ADMM-based deep reconstruction for limited-angle CT. *Physics in Medicine & Biology*, 64(11): 115011.

Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; and Lau, R. W. 2019b. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12270–12279.

Xia, W.; Lu, Z.; Huang, Y.; Shi, Z.; Liu, Y.; Chen, H.; Chen, Y.; Zhou, J.; and Zhang, Y. 2021. MAGIC: Manifold and graph integrative convolutional network for low-dose CT reconstruction. *IEEE transactions on medical imaging*, 40(12): 3459–3472.

Xia, Z.; Pan, X.; Song, S.; Li, L. E.; and Huang, G. 2022. Vision transformer with deformable attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4794–4803.

Xiang, J.; Dong, Y.; and Yang, Y. 2021. FISTA-Net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging. *IEEE Transactions on Medical Imaging*, 40(5): 1329–1339.

Yi, Z.; Hu, C.; Wenjun, X.; Yang, C.; Baodong, L.; Yan, L.; Huaiqiang, S.; and Jiliu, Z. 2023. LEARN++: Recurrent dual-domain reconstruction network for compressed sensing CT. *IEEE TRPMS*, 7(132-142): 2–3.

Zhang, J.; Li, Z.; Pan, J.; Wang, S.; and Wu, W. 2025. Trustworthy Limited Data CT Reconstruction Using Progressive Artifact Image Learning. *IEEE Transactions on Image Processing*.

Zhang, Z.; Liang, X.; Dong, X.; Xie, Y.; and Cao, G. 2018. A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE transactions on medical imaging*, 37(6): 1407–1417.

Zhou, Q.; Qian, J.; Tang, J.; and Li, J. 2024. Deep unrolling networks with recurrent momentum acceleration for nonlinear inverse problems. *Inverse Problems*, 40(5): 055014.

Zhu, X.; Hu, H.; Lin, S.; and Dai, J. 2019. Deformable convnets v2: More deformable, better results. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9308–9316.

Zhu, X.; Su, W.; Lu, L.; Li, B.; Wang, X.; and Dai, J. 2020. Deformable detr: Deformable transformers for end-to-end object detection. *arXiv preprint arXiv:2010.04159*.