PTF-CT: Polar-Aware Temporal-Frequential Iterative Reconstruction for Sparse-View CT

Supplementary Material

This supplementary document provides additional implementation details, extended ablation studies, frequency-domain analyses, visual comparisons, and further discussions to complement the main paper.

This supplementary document is organized as follows: Sec. A introduces our task and measurement formulation in detail ; Sec. B presents implementation details; Sec. C shows the pseudo-code of IDM and TFM; Sec. D discusses limitations; Sec. E shows extended visual results; Sec. F summarizes the notation.

A. Task and Measurement Formulation

A.1. Sparse-view CT Reconstruction

Sparse-view computed tomography (SVCT) aims to recover a volumetric image $x \in \mathbb{R}^{H\times W}$ from a severely undersampled set of X-ray projections. Instead of the standard 360° scan with hundreds of uniformly sampled views, SVCT uses only $n_v \ll 360$ projections (e.g., 24–48 views), dramatically reducing radiation dose but significantly increasing the ill-posedness of the inverse problem.

The forward acquisition process follows the discrete Radon transform:

$$y = Ax + n, \tag{1}$$

where A is the projection operator determined by scanner geometry, $y \in \mathbb{R}^{n_v\times n_d}$ is the measured sinogram, n denotes mixed Poisson–Gaussian noise, and n_v, n_d denote the number of views and detector bins, respectively. In sparse-view regimes, A becomes severely rank-deficient, leading to heavy streak artifacts and unstable reconstruction when directly inverted.

A.2. Sinogram Characteristics Under Sparse Views

Figure 1 (supplementary) illustrates sinograms at different view counts. Full-view sinograms exhibit smooth angular continuity, whereas sparse-view sinograms contain large angular gaps, producing strong frequency aliasing when backprojected. These discontinuities directly manifest as long-range streaks in the reconstructed images, motivating the need for both geometry-aware priors and frequency-progressive refinement (as addressed in our PTF-CT design).

A.3. Filtered Backprojection (FBP) Baseline

The classical reconstruction approach applies filtered backprojection (FBP):

$$x_{\text{FBP}} = A^\top(h * y), \tag{2}$$

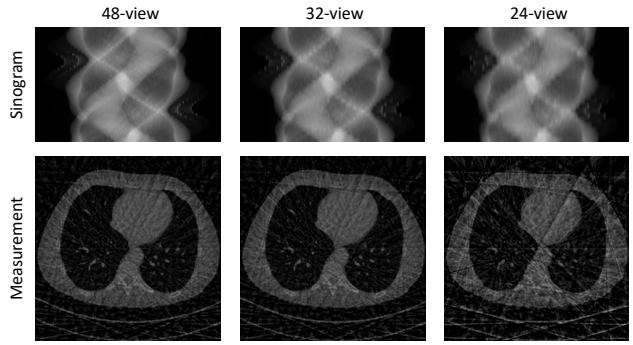

Figure 1. **Visualization of CT reconstruction mechanism.** The first row shows sinograms under different views; The second row shows measurement (FBP reconstruction) under different views; (left: **48-view**, middle: **32-view**, right: **24-view**).

where h is the Ram-Lak (or equivalent) frequency-domain filter and $A^\top$ denotes backprojection. Under dense-view settings, FBP provides fast and reliable reconstructions. However, in sparse-view CT, the angular gaps create:

- pronounced streak artifacts along projection directions,
- loss of medium- and high-frequency structures,
- strong sensitivity to measurement noise.

As a result, FBP serves as a useful but severely degraded baseline in the sparse-view regime, motivating data-driven restoration or iterative reconstruction pipelines. In our work, FBP or $A^\top y$ acts as the initialization for iterative refinement, while PTF-CT progressively corrects physics-consistent errors using IDM and TFM.

B. Implementation Details

B.1. Training Configuration

We train all models with AdamW optimizer (β_1=0.9, β_2=0.999), batch size 1, and cosine annealing schedule with warmup. The initial learning rate is 2×10^{-4} and the total training epochs are fixed at 50 for all models.

Hardware. Each experiment is conducted on a single NVIDIA RTX A5000 GPU (24GB).

B.2. Network Architecture Details

Restormer backbone. Each refinement block uses a 4-scale U-shaped Restormer:

$$\{L_1, L_2, L_3, L_4\} = \{1, 1, 2, 2\}$$

The other hyperparameters all follows Restormer.

IDM module. The map $W(\rho)$ is pre-computed once using Eq. (7) of the main paper, normalized to $(0, 1]$, broadcasted across channels, and multiplied with Inception-v2 transformed features.

TFM module. Each TFM block contains: - DCT $\rightarrow$ masking $\rightarrow$ IDCT - two lightweight convolutions: 1×1, 3×3 - temporal masks for $t = 1..4$, with increasing passband radius.

C. Pseudo-code of IDM and TFM

This section provides explicit pseudo-code for the proposed Information Density Module (IDM) and Temporal-Frequential Module (TFM), complementing the architectural descriptions in the main paper. Both modules are lightweight, deterministic, and plug into each iteration of the refinement network.

C.1. Information Density Module (IDM)

The IDM modulates intermediate features using an inverse information-density map derived from Radon geometry. Given a feature map x and a precomputed normalized inverse-density map $W(\rho)$, IDM applies an Inception-v2 block followed by density-aware modulation and a residual update:

$$x' \;=\; x + (f_\theta(x) \odot W)\,,$$

where $W \in (0, 1]^{H\times W}$ is broadcast across channels. The pseudo-code is given below.

C.2. Temporal-Frequential Module (TFM)

TFM progressively injects frequency components across iterations and U-shaped scales. For iteration t and scale s, TFM applies:

1) DCT transform:

$U_t^s \;=\; \mathrm{DCT}(F_t^s)$

2) multiplication with an iteration-/scale-specific Gaussian mask M_t^s: $\widetilde{U}_t^s \;=\; U_t^s \odot M_t^s$

3) lightweight convolutional refinement and 4) IDCT back to spatial domain:

$\widehat{F}_t^s \;=\; \mathrm{IDCT}\Big(\,\mathrm{Conv}_{3\times3}\big(\mathrm{Conv}_{1\times1}(\widetilde{U}_t^s)\big) + \widetilde{U}_t^s\Big)$

5) residual reinjection:

$F_{t+1}^s \;=\; F_t^s \;+\; \mathrm{Conv}_{1\times1}\Big(\widehat{F}_t^s\Big)$

D. Limitations

Although PTF-CT achieves strong performance across sparsity and noise regimes, several limitations should be acknowledged.

Algorithm 1: Information Density Module (IDM)

Input: Feature map $x \in \mathbb{R}^{H\times W\times C}$;
Precomputed inverse-density map $W \in (0, 1]^{H\times W}$
Output: Modulated feature map x'

1. Precompute inverse-density map
Fixed, non-learnable geometry prior
$W \leftarrow \mathrm{Normalize}\big(1/I(\rho)\big)$

2. Inception-v2 multi-branch feature transform
Four-branch Inception-v2 block:
(a) 1×1 conv:
$z_1 \leftarrow \mathrm{Conv}_{1\times1}(x)$
(b) $1 \times 1 \rightarrow 3 \times 3$ conv:
$z_2 \leftarrow \mathrm{Conv}_{3\times3}(\mathrm{Conv}_{1\times1}(x))$
(c) Average pooling $\rightarrow 1 \times 1$ conv:
$z_3 \leftarrow \mathrm{Conv}_{1\times1}(\mathrm{AvgPool}_{3\times3}(x))$
(d) Identity projection:
$z_4 \leftarrow \mathrm{Conv}_{1\times1}(x)$
Concatenate multi-scale features:
$z \leftarrow \mathrm{Concat}(z_1, z_2, z_3, z_4)$
Final fusion
$z \leftarrow \mathrm{Conv}_{1\times1}(z)$

3. Density-aware modulation
Broadcast W to all C channels
$z' \leftarrow z \odot W$

4. Residual update
$x' \leftarrow x + z'$

return x'

D.1. Fundamental Ambiguity Under Extreme Sparsity

Sparse-view CT inherently constrains the dimensionality of the measurement space. When the number of projection views becomes too small, the acquired data may lack sufficient information to uniquely determine the underlying image, even in noiseless settings. Deep learning approaches—including ours—resolve this ambiguity through learned anatomical priors. However, if the training distribution does not adequately cover the variability of real patients, the reconstructed output may be a plausible solution rather than the correct one. This limitation is shared by all reconstruction methods, including classical MBIR techniques and even full-view CT, which can still be non-unique under noise or out-of-manifold structures.

D.2. Challenges in Clinical Adoption

Although PTF-CT significantly enhances reconstruction quality for reduced-view protocols, translating such advances into clinical workflows is nontrivial. Radiologists and technicians may be hesitant to revise acquisition proto-

Algorithm 2: Temporal-Frequential Module (TFM) at iteration t and scale s

Input: Feature map $F_t^s \in \mathbb{R}^{H\times W\times C}$;
Mask bank $\{M_\tau^s\}_{\tau=1}^{T}$ with $M_\tau^s \in [0,1]^{H\times W}$.
Output: Updated feature map F_{t+1}^s.

Step 1: Select temporal–frequential mask
$M \leftarrow M_t^s$ `// Mask for current iteration` t `at scale` s

Step 2: DCT transform
$U \leftarrow \mathrm{DCT}(F_t^s)$

Step 3: Apply progressive frequency mask
$\widetilde{U} \leftarrow U \odot M$ `//` M `emphasizes a low→high band as` t `grows`

Step 4: Frequency-domain refinement
$V \leftarrow \mathrm{Conv}_{1\times1}(\widetilde{U})$
$V \leftarrow \mathrm{Conv}_{3\times3}(V) + \widetilde{U}$ `// Residual in frequency domain`

Step 5: IDCT back to spatial domain
$\widehat{F} \leftarrow \mathrm{IDCT}(V)$

Step 6: Residual reinjection in spatial domain
$F_{t+1}^s \leftarrow F_t^s + \mathrm{Conv}_{1\times1}(\widehat{F})$

return F_{t+1}^s

cols due to concerns about robustness, interpretability, compatibility with existing diagnostic tools, and regulatory approval processes. Therefore, despite algorithmic progress, widespread adoption of aggressive sparse-view CT may require additional clinical studies, calibration procedures, and user acceptance evaluations.

D.3. Dose-Fidelity Trade-off Remains Task-Dependent

Sparse-view acquisition reduces radiation exposure and shortens scan time, enabling faster examinations and reducing patient motion blur. However, extremely low view counts may compromise high-frequency anatomical fidelity beyond what any prior—learning-based or handcrafted—can reliably recover. Thus, balancing radiation reduction and diagnostic precision remains an intrinsically task-dependent trade-off. PTF-CT improves reconstruction robustness within this spectrum but does not eliminate the underlying tension between safety and fidelity.

E. Extended Visual Comparisons

We provide an extended set of qualitative results, including a total of 18 representative cases arranged in a top–down anatomical order of the human body. Specifically, the visual comparisons contain:

- 9 cases from the AAPM dataset, and
- 3 cases from the DeepLesion dataset (out-of-distribution).

As illustrated in Fig. 2, each case includes reconstructions generated by the following methods: FBP, FBPConvNet, DudoTrans Learned PD, LEARN, DiffMBIR, Regformer, QN-Mixer, along with the ground truth (GT).

F. Notation Summary

Table 1 summarizes the key notations used throughout the paper.

Table 1. Lookup table of notations and hyperparameters used in PTF-CT.

Notation	Domain of Definition	Value(s)	Description
x	$\mathbb{R}^{H\times W}$	256×256	CT image to reconstruct
y	$\mathbb{R}^{n_v \times n_d}$	–	Sparse-view sinogram
n_v	$\mathbb{N}^+$	$\{24, 32, 48\}$	Number of projection views
n_d	$\mathbb{N}^+$	368	Number of detector bins
A	$\mathbb{R}^{(n_v n_d)\times(HW)}$	–	Radon forward operator
$A^\top$	$\mathbb{R}^{(HW)\times(n_v n_d)}$	–	Adjoint Radon operator
x_0	$\mathbb{R}^{H\times W}$	$FBP(y)$	Initial measurement (FBP reconstruction)
x_k	$\mathbb{R}^{H\times W}$	–	Reconstruction at iteration k
z_k	$\mathbb{R}^{H\times W}$	–	Physics-consistent update after DC step
x_{k+1}	$\mathbb{R}^{H\times W}$	–	Updated reconstruction
T	$\mathbb{N}^+$	4	Number of unrolled iterations
α_k	$\mathbb{R}$	learnable	Step size of data-consistency update
Φ_θ	–	Restormer U-shape	Refinement operator main frame
		Information Density Module (IDM)	
$I(\rho)$	$\mathbb{R}$	-	Radial information density function (Eq. 7)
$W(\rho)$	$H \times W$	$(0, 1]$	Normalized inverse-density map (Eq. 8)
$f_\theta(x)$	$\mathbb{R}^{C\times H\times W}$	–	Inception-v2 feature transform
$\mathrm{IDM}(x)$	$\mathbb{R}^{C\times H\times W}$	–	Geometry-aware modulation feature
		Temporal-Frequential Module (TFM)	
F_t^s	$\mathbb{R}^{C_s\times H_s\times W_s}$	–	Features at iteration t, scale s
U_t^s	$\mathbb{R}^{C_s\times H_s\times W_s}$	–	DCT-transformed features (Eq. 11)
M_t^s	$H_s \times W_s$	$(0, 1]$	Gaussian radial mask at iteration t
$\widetilde{U}_t^s$	$\mathbb{R}^{C_s\times H_s\times W_s}$	–	Masked DCT spectrum (Eq. 12)
$F_t'^s$	$\mathbb{R}^{C_s\times H_s\times W_s}$	–	Refined spectrum after Conv + IDCT (Eq. 13)
F_{t+1}^s	$\mathbb{R}^{C_s\times H_s\times W_s}$	–	Residual frequency update (Eq. 14)
s	$\{0, 1, 2, 3\}$	–	U-shaped spatial scale index
		Training / Data Settings	
N_0	–	0	No-noise condition
N_1	–	5% Gaussian + 10^6 Poisson	Low-noise condition
N_2	–	5% Gaussian + 5×10^5 Poisson	High-noise condition
Optimizer	–	AdamW	$\beta_1 = 0.9,\ \beta_2 = 0.999$
LR	–	2×10^{-4}	Initial learning rate
Epochs	–	50	Training epochs
Backbone depth	–	$\{1, 1, 2, 2\}$	Restormer block configuration
Loss	–	ℓ_1	Training objective (Eq. 16)

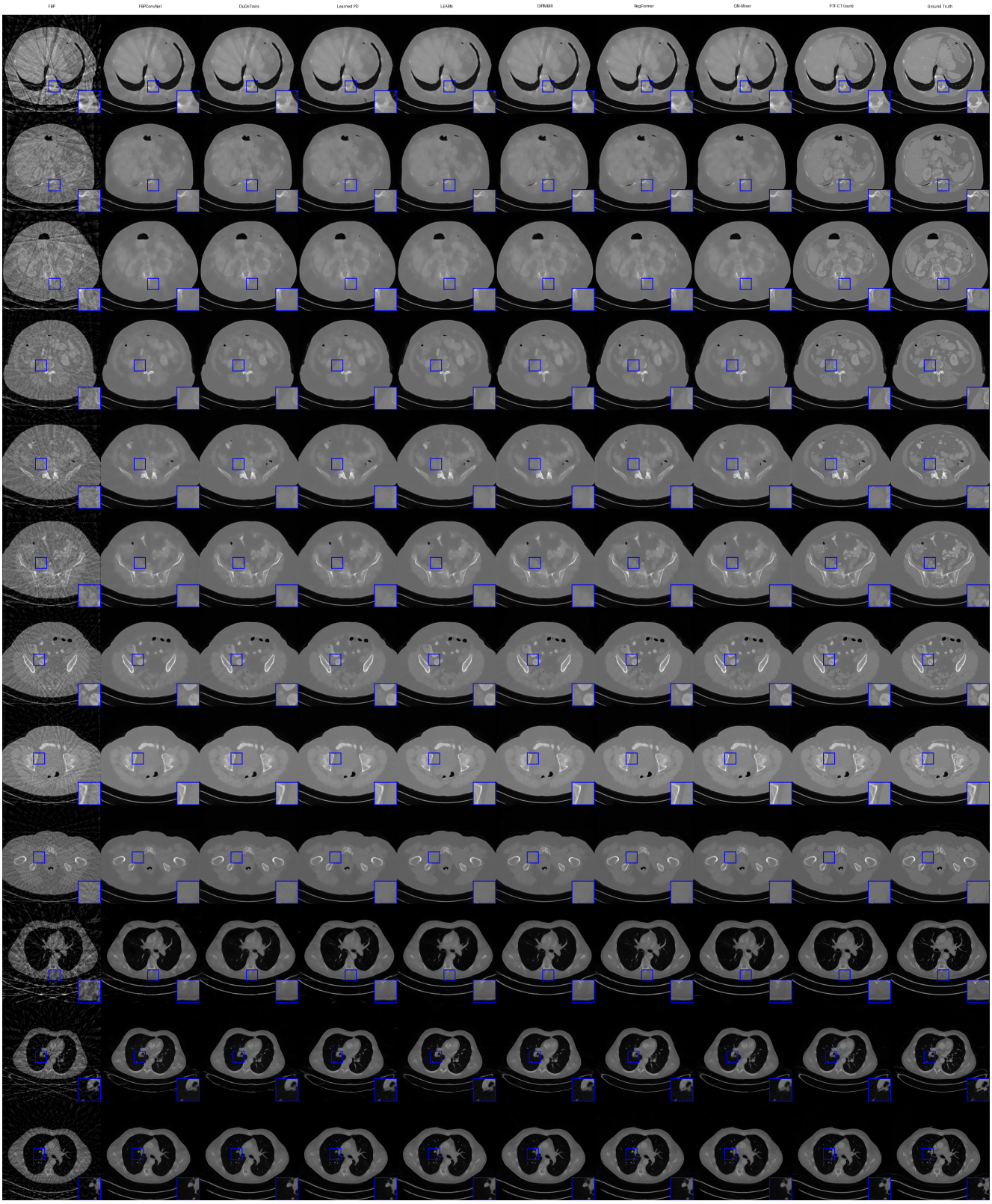

Figure 2. **Extended visual comparisons.** The first 9 rows are evaluated on the AAPM test set and last 3 rows on the DeepLesion dataset, under different sparse-view & noise regimes: (Row 1-3,10) **24-view**, (Row 4-6,11) **32-view**, (Row 7-9,12) **48-view**, (Row 1,4,7,10) **No noise**, (Row 2,5,8,11) **Low noise**, (Row 3,6,9,12) **High noise**,