TCCT: Trajectory-Conditioned Non-Circular Orbits CBCT Reconstruction with Fourier Neural Operator

Anonymized Authors

Anonymized Affiliations
email@anonymized.com

Abstract. Robotic cone-beam computed tomography (CBCT) enables flexible, non-circular source trajectories, offering potential reductions in metal and cone-beam artifacts compared with standard circular orbits. Accurate reconstruction for such trajectories remains challenging: classical filtered backprojection (FBP) fails to account for view-dependent redundancy, while iterative methods are computationally expensive. To address this challenge, we present a trajectory-conditioned non-circular orbits CBCT reconstruction (TCCT) framework, which employs a geometry-aware spectral operator to infer redundancy weights from the acquisition trajectory. Concretely, a Fourier neural operator learns the continuous mapping from per-view geometric descriptors to weights in the Radon domain, followed by a lightweight CNN that refines them spatially. During training, the predicted weights are integrated into a differentiable shift-variant FBP pipeline and end-to-end optimized using reconstruction-driven supervision on simulated volumes, without requiring ground-truth redundancy weights. Experimental results demonstrate high-quality reconstructions while achieving an order-of-magnitude speedup compared with iterative methods. TCCT lays the foundation for general non-circular trajectory CBCT reconstruction and enables future end-to-end dual-domain optimization for noise suppression, artifact reduction, and trajectory optimization. Open source code will be released upon acceptance.

Keywords: CBCT reconstruction · Non-circular trajectory · Known operator learning · Fourier neural operator.

1 Introduction

Robotic cone-beam computed tomography (CBCT) has enabled flexible non-circular source trajectories, which can improve angular coverage and reduce artifacts [**?**] compared to traditional circular orbits. However, this acquisition flexibility also introduces a reconstruction bottleneck, as conventional analytic reconstruction pipelines such as filtered backprojection (FBP) [**?**] are designed around circular-orbit assumptions and fail to model the view-dependent redundancy in non-circular CBCT geometries. Iterative reconstruction methods such

as algebraic iterative reconstruction [?] and model-based iterative reconstruction [?] can in principle handle arbitrary geometries but are computationally expensive and unsuitable for real-time applications. To address non-circular trajectory reconstruction, Defrise and Clack [?] introduced a shift-variant FBP (SV-FBP) algorithm, which is mathematically exact, but the orbit-dependent filtering is difficult to design for complex trajectories.

Recent deep learning approaches offer alternative reconstruction strategies. End-to-End approaches such as AUTOMAP [?] directly map projection data to reconstructed volumes, requiring large training datasets and extensive parameters. Other strategies, including implicit neural representations [?], neural radiance fields (NeRF) [?], or Gaussian-splatting methods [?], optimize per-object or per-scan representations and are therefore limited in their ability to generalize across subjects or acquisition trajectories.

Known-operator learning [?] offers a hybrid paradigm, combining analytic operators with learnable components, drastically reducing parameter counts while retaining physical interpretability. Building on this idea, Ye et al. [?,?] mapped Defrise and Clack's SV-FBP pipeline into a neural network, allowing data-driven estimation of trajectory-specific weighting parameters. However, this approach requires training for each new CT-trajectory, which is a key limitation for deployment in robotic CBCT, where trajectories may vary between procedures, patient setup constraints, or task-driven orbit selection.

Our approach follows the known-operator principle by fixing the differentiable SV-FBP operator chain and restricting learning to the redundancy-weight operator. Concretely, we treat redundancy weights as trajectory-conditioned functions and use an operator-learning model [?,?] to predict them from the source trajectory. We then plug these weights into a differentiable SV-FBP pipeline and train the resulting system with a reconstruction-driven objective on simulated data, yielding a single model that can be applied across trajectories within the considered orbit family without retraining. Our main contributions are as follows:

1. A trajectory-conditioned spectral neural operator (TC-SNO) that models view- and detector-dependent redundancy weighting as a continuous function of trajectory parameters.
2. Integration of TC-SNO into a differentiable SV-FBP pipeline, enabling end-to-end training using reconstruction-driven supervision on simulated phantoms, eliminating the need for ground-truth redundancy weights.
3. Extensive evaluation on clinical CT volumes with simulated CBCT projections, demonstrating robust reconstruction quality across multiple sinusoidal trajectories without trajectory-specific retraining.

2 Method

2.1 TC-SNO: Trajectory-Conditioned Spectral Neural Operator

To model trajectory-dependent redundancy weights, we propose the TC-SNO module, which maps a sequence of per-view geometry parameters to view- and

detector-dependent redundancy weights $W \in \mathbb{R}^{N\times N_\phi \times N_s}$, where N is the number of projections and (ϕ, s) denote the angle and distance in the discretized Radon domain $N_\phi \times N_s$. Following the cone-beam geometry representation in [**?**], each projection view is encoded by a 12-dimensional pose vector $\mathbf{g}_n = [\mathbf{s}_n; \mathbf{d}_n; \mathbf{u}_n; \mathbf{v}_n] \in \mathbb{R}^{12}$, where n indexes the projection views, $\mathbf{s}_n, \mathbf{d}_n \in \mathbb{R}^3$ denote normalized source and detector center directions, and $\mathbf{u}_n, \mathbf{v}_n \in \mathbb{R}^3$ define the detector coordinate. Stacking all views yields $\mathbf{G} \in \mathbb{R}^{N\times 12}$, which can be interpreted as a discrete sampling of the continuous acquisition trajectory.

The design of TC-SNO is guided by key properties of trajectory-dependent redundancy weights reported in [**?**]. First, the redundancy weighting process can be interpreted as a continuous mapping of the acquisition trajectory, motivating operator learning to model this function-to-function relationship. Second, the weights are highly compressible [**?**], suggesting the existence of a compact latent representation. Third, non-smooth weighting across views or detector coordinates can manifest as reconstruction artifacts, motivating explicit smoothness-inducing modeling along the view dimension and spatial refinement on the detector plane. To reflect these properties, TC-SNO combines global spectral modeling across views with local spatial refinement on the detector plane, as well as consists of four components as illustrated in Fig. 1(a).

Trajectory feature extraction: The geometry sequence $\mathbf{G}$ is first lifted into a higher-dimensional feature space $X^{(0)} \in \mathbb{R}^{N\times C}$ using pointwise convolutions, where C denotes the feature dimension after lifting. This operation embeds per-view geometric information independently without modeling inter-view dependencies, providing a flexible representation for subsequent global modeling.

Spectral operator layers: Global inter-view correlations are modeled using stacked Fourier layers to efficiently aggregate orbit-wide context. For layer l, we compute

$$X^{(l+1)} = \sigma\Big(\mathcal{F}^{-1}\big(\mathcal{R} \odot \mathcal{F}(X^{(l)})\big) + W^{(l)}X^{(l)}\Big), \tag{1}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the 1D Fourier transform along the view dimension, $\mathcal{R}$ are learnable complex spectral weights, $W^{(l)}$ denotes a pointwise linear mapping, $\sigma(\cdot)$ is a GELU nonlinearity. Truncating high-frequency modes enforces smoothness along the trajectory, consistent with smooth angular evolution in continuous CBCT acquisition. The residual connection improves optimization stability and preserves per-view information during spectral mixing.

Latent mapping: After the spectral layers, the features are projected to a latent representation using pointwise (1×1) convolutions and layer normalization. This stage increases feature dimensionality while stabilizing per-view feature scale, producing a compact latent sequence suitable for decoding redundancy weights.

Spatial decoding: Each per-view latent vector is expanded by a fully connected layer and reshaped into a coarse $h \times w$ grid, which serves as a low-resolution detector-plane template. This grid is then upsampled by a lightweight CNN using bilinear upsampling and convolutional refinement to produce full-resolution $W \in \mathbb{R}^{N_\phi \times N_s}$. Batch normalization and leaky ReLU activations fur-

ther improve training stability and allow the decoder to capture detector-dependent variations in redundancy weights.

Overall, TC-SNO couples trajectory-aware spectral mixing with a lightweight spatial decoder to produce smooth, geometry-dependent redundancy weights. Because these weights are not uniquely defined, we do not supervise them directly and instead learn them end-to-end through their impact on the differentiable SV-FBP reconstruction objective, which we describe next.

2.2 SV-FBP: Differentiable Shift-Variant Filtered Backprojection

The SV-FBP algorithm [**?**] provides a mathematically exact and computationally efficient reconstruction framework for general CBCT trajectories, consisting of shift-variant filtering followed by 3D cone-beam backprojection A_{3d}^T. As shown in Fig. 1(b), the projections are first transformed to the Radon domain via the Radon transform A_{2d}. A redundancy weight W_{red} is then applied in the Radon domain through element-wise multiplication. The weighted data are subsequently backprojected using the 2D backprojection operator A_{2d}^T, yielding filtered projections, which are finally reconstructed into a volume via A_{3d}^T.

Among these components, only the redundancy weight depends on the acquisition trajectory. In practice, computing it for complex trajectories is difficult or even intractable. Ye et al. [**?**] therefore reformulated SV-FBP using known-operator learning, expressing the reconstruction pipeline as a differentiable model and explicitly treating the redundancy weight as a learnable layer, enabling end-to-end data-driven estimation.

This hybrid design preserves physical interpretability while allowing the redundancy weighting function to adapt to complex acquisition geometries. However, existing approaches require retraining for each trajectory, which limits deployment in robotic CBCT with procedure-dependent or task-driven orbits. We therefore predict W_{red} directly from trajectory parameters using TC-SNO, enabling a single trained model to generalize across trajectories.

2.3 TCCT: Trajectory-Conditioned CBCT Reconstruction

To enable trajectory-generalizable CBCT reconstruction without direct supervision of redundancy weight, we propose the TCCT framework, which integrates the TC-SNO module into the differentiable SV-FBP pipeline (Fig. 1). Specifically, given a simulated ground-truth volume f_{gt} and trajectory parameters $\mathbf{G}$, the reconstruction can be expressed compactly as

$$\hat{f} = \text{SV-FBP}(\mathbf{p}, \text{TC-SNO}(\mathbf{G})), \mathbf{p} = A_{3d}(f_{\text{gt}}, \mathbf{G}), \tag{2}$$

where p represents simulated data generated by the cone-beam forward projector A_{3d} based on the trajectory $\mathbf{G}$, and TC-SNO predicts view- and detector-dependent redundancy weights from the trajectory parameters $\mathbf{G}$, which are incorporated into the SV-FBP pipeline to produce the reconstructed volume $\hat{f}$.

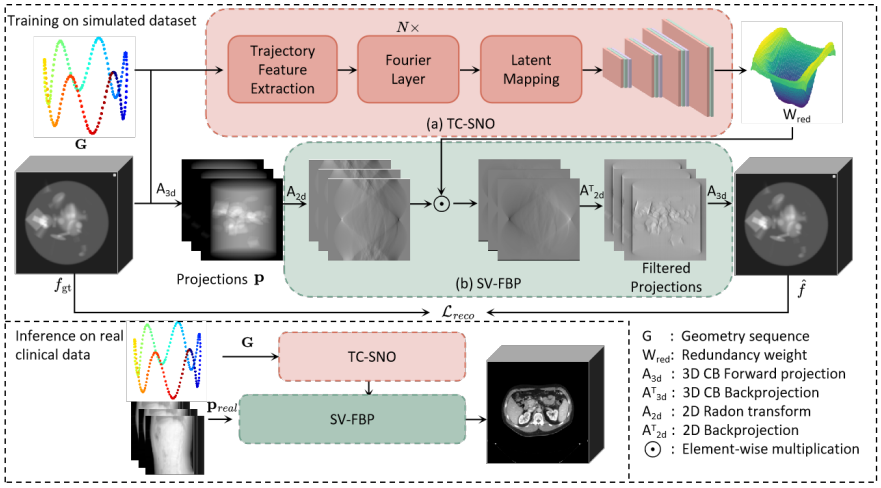

Fig. 1: Overview of the proposed TCCT framework. During the training phase, the method operates on synthetically generated projection data, whereas during inference it reconstructs volumetric representations directly from measured projections.

Because redundancy weights are not uniquely defined, TC-SNO is trained indirectly through minimizing the discrepancy between the reconstructed volume and the ground-truth:

$$\mathcal{L}_{reco} = \|\hat{f} - f_{\text{gt}}\|_2^2 + \lambda\big(1 - \text{SSIM}(\hat{f}, f_{\text{gt}})\big), \tag{3}$$

where $\|\cdot\|_2^2$ denotes the voxel-wise ℓ_2 loss, $\text{SSIM}(\cdot)$ is the structural similarity index (SSIM) computed between reconstructed and reference volumes, and λ balances fidelity and structural consistency. This physics-based supervision encourages the network to learn redundancy weights that are optimal for reconstruction quality rather than matching a predefined analytical expression.

At inference time, reconstruction requires only measured projections p_{real} and their associated trajectory parameters $\mathbf{G}_{\text{real}}$:

$$\hat{f}_{\text{real}} = \text{SV-FBP}\big(\text{p}_{\text{real}}, \text{TC-SNO}(\mathbf{G}_{\text{real}})\big). \tag{4}$$

This trajectory-conditioned formulation enables a single reconstruction model to generalize across diverse acquisition geometries while preserving the physical interpretability of the SV-FBP pipeline.

3 Experiments and Results

We evaluate the proposed TCCT framework using simulated data for training and real clinical volumes for testing. All experiments are conducted using geometry parameters of a clinical C-arm system (Artis zeego, Siemens Healthineers).

Table 1: Quantitative evaluation of reconstruction performance across sinusoidal trajectories. Values represent mean ± standard deviation across 20 test volumes.

Metric	Method	Frequency (sinusoidal trajectories)								
		1	2	3	4	5	6	7	8	9
MSE ↓	ART(50) [?]	0.1161 ±0.0161	0.1131 ±0.0159	0.1152 ±0.0161	0.1130 ±0.0181	0.1142 ±0.0170	0.1146 ±0.0162	0.1183 ±0.0170	0.1148 ±0.0158	0.1177 ±0.0181
	SV-FBP [?]	**0.0892** ±0.0101	**0.0886** ±0.0113	0.0878 ±0.0104	**0.0877** ±0.0113	0.0954 ±0.0107	**0.0903** ±0.0113	0.0915 ±0.0106	0.0936 ±0.0111	0.0938 ±0.0107
	Ours*	0.0922 ±0.0086	0.0899 ±0.0111	**0.0847** ±0.0094	0.0885 ±0.0114	**0.0874** ±0.0091	0.0910 ±0.0113	**0.0856** ±0.0096	**0.0933** ±0.0108	**0.0884** ±0.0099
PSNR(dB) ↑	ART(50) [?]	29.23 ±0.54	29.47 ±0.66	29.30 ±0.68	29.50 ±0.78	29.39 ±0.72	29.35 ±0.68	29.08 ±0.77	29.33 ±0.57	29.14 ±0.79
	SV-FBP [?]	**31.50** ±0.38	**31.57** ±0.45	31.64 ±0.41	**31.66** ±0.47	**30.91** ±0.48	**31.40** ±0.46	31.27 ±0.42	31.08 ±0.44	31.06 ±0.43
	Ours*	31.19 ±0.40	31.43 ±0.43	**31.94** ±0.42	31.58 ±0.47	31.66 ±0.42	31.33 ±0.44	**31.85** ±0.45	**31.10** ±0.42	**31.57** ±0.46
SSIM ↑	ART(50) [?]	0.8756 ±0.0312	0.8722 ±0.0271	0.8650 ±0.0357	0.8721 ±0.0299	0.8670 ±0.0336	0.8679 ±0.0236	0.8536 ±0.0334	**0.8667** ±0.0204	**0.8578** ±0.0322
	SV-FBP [?]	**0.8825** ±0.0165	**0.8846** ±0.0162	**0.8819** ±0.0160	**0.8817** ±0.0166	**0.8794** ±0.0182	**0.8758** ±0.0172	**0.8660** ±0.0172	0.8647 ±0.0180	0.8552 ±0.0178
	Ours*	0.8559 ±0.0191	0.8808 ±0.0157	0.8720 ±0.0180	0.8742 ±0.0176	0.8652 ±0.0182	0.8675 ±0.0178	0.8587 ±0.0186	0.8577 ±0.0186	0.8431 ±0.0193

*Our method handles multiple trajectories; SV-FBP will require separate models.

The source-to-detector distance is 1200 mm and the source-to-isocenter distance is 750 mm. The detector operates in 4×4 binning mode with a resolution of 620×480 pixels and an isotropic pixel size of 0.616 mm. Each scan consists of 200 projections, and reconstructed volumes have a size of $256 \times 512 \times 512$ with 0.25 mm isotropic voxels.

For trajectory parameterization, we adopt a family of sinusoidal source trajectories parameterized by frequency. The gantry tilt angle is defined as $\phi_\omega = \phi_{\max} \cos(f\,\theta_\omega)$, where θ_ω denotes the gantry rotation angle, ϕ_ω is the corresponding tilt angle, $\phi_{\max}$ is the maximum tilt amplitude, and f controls the sinusoidal frequency, which is uniformly sampled from 1 to 9 with a step size of 0.1.

The proposed framework is implemented in PyTorch 2.1.1 [?]. Differentiable projection and backprojection operators are implemented using PyroNN framework [?]. All analytic operators in the SV-FBP pipeline remain fixed, while TC-SNO is optimized via backpropagation. The weighting parameter in the loss function is set to $\lambda = 0.005$, balancing voxel-wise fidelity and structural similarity. Training is performed using the AdamW optimizer with a learning rate of 1×10^{-3} on an NVIDIA A40 GPU.

3.1 Dataset Preparation

Simulated training data: We generate data from 30 procedurally created 3D phantoms composed of randomly placed geometric primitives with varying shapes, sizes, and orientations, using 24 for training and 6 for validation. Projection data are generated with a cone-beam forward projector under randomly

sampled sinusoidal trajectories. This ensures the network learns trajectory-dependent redundancy weighting independent of anatomical content.

Clinical testing data: We evaluate the trained model on the Pancreatic-CT-CBCT-SEG dataset [?]. Projections are generated from 20 volumes using the same imaging geometry and sinusoidal trajectories as in the simulation setup, enabling assessment of generalization to realistic anatomical structures.

3.2 Performance Comparison

We compare the TCCT framework with two representative baselines: differentiable SV-FBP and iterative algebraic reconstruction technique (ART). All methods are evaluated using mean squared error (MSE), peak signal-to-noise ratio (PSNR), and SSIM. Table 1 shows that TCCT achieves reconstruction quality comparable to differentiable SV-FBP across all sinusoidal frequencies, with relative differences below 2.1% for MSE, 0.56% for PSNR, and 0.72% for SSIM. Unlike differentiable SV-FBP, which requires trajectory-specific training, TCCT generalizes across trajectories without retraining, providing greater flexibility in practical applications. In contrast, the iterative ART method with 50 iterations exhibits consistently higher MSE and lower PSNR and SSIM. ART is also computationally expensive, requiring approximately 30 s per volume, whereas TCCT completes reconstruction in only 2 s, demonstrating a substantial efficiency advantage. Representative central slices are shown in Fig. 2. Consistent with the quantitative results, TCCT preserves structural details across the four selected frequencies, illustrating its stable performance under varying trajectory settings.

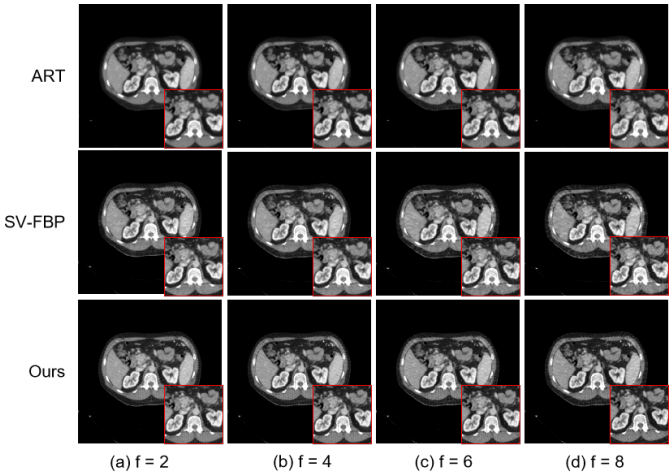

Fig. 2: Reconstructed results under sinusoidal trajectories with frequencies ($f = 2, 4, 6, 8$). Rows correspond to different reconstruction methods. Visualization parameters: window width = 400, window level = 60.

Table 2: Ablation study: reconstruction performance across multiple sinusoidal trajectories. Values show mean ± std over 20 test volumes.

Method	MSE ↓	PSNR ↑	SSIM ↑	Params (M)
SV-FBP baseline	0.1117 ± 0.0109	29.52 ± 0.85	0.7953 ± 0.0303	215.9
Transformer+CNN	0.1154 ± 0.0127	29.26 ± 0.78	0.7903 ± 0.0312	149.8
Resnet1D+CNN	0.1187 ± 0.0120	29.00 ± 0.79	0.7815 ± 0.0318	146.9
MLP+CNN	0.1212 ± 0.0137	28.83 ± 0.87	0.7750 ± 0.0277	157.6
TC-SNO (ours)	**0.0926 ± 0.0105**	**31.17 ± 0.43**	**0.8660 ± 0.0199**	**143.6**

To further evaluate robustness to unseen trajectory parameters, we randomly sampled 20 additional sinusoidal frequencies within the training range while excluding the discrete frequencies used during training. TCCT achieved MSE = 0.0867, PSNR = 31.74 dB, and SSIM = 0.8654, consistent with Table 1. This result indicates stable interpolation across continuous trajectory variations rather than overfitting to discrete training frequencies.

3.3 Ablation Study

We evaluate the impact of architectural choices on multi-trajectory reconstruction using 20 test volumes with randomly sampled sinusoidal orbits. As shown in Table 2, the differentiable SV-FBP baseline is trained across trajectories, but it converges to a trajectory-agnostic compromise solution that fails to capture orbit-specific geometry variations. Furthermore, using a Transformer [**?**], ResNet1D [**?**], or MLP backbones results in unstable training behavior and degraded reconstruction accuracy across all metrics. In contrast, the proposed Fourier-layer design achieves the best MSE, PSNR, and SSIM with the smallest model size, highlighting the effectiveness of ill-posed function-to-function mapping for trajectory-dependent redundancy weighting.

4 Discussion and Conclusion

We present TCCT, a trajectory-conditioned differentiable reconstruction framework that enables trajectory-aware reconstruction within a family of sinusoidal orbits. The method integrates a spectral neural operator with CNN into a differentiable SV-FBP pipeline and is trained using reconstruction-driven supervision, enabling a single model to handle multiple trajectories instead of training separate differentiable SV-FBP models [**?**] for each trajectory. Experimental results demonstrate that, compared with traditional iterative reconstruction [**?**], the proposed method achieves substantially faster reconstruction. Moreover, the results indicate stable performance on intermediate sinusoidal frequencies not used during training. This suggests that the network generalizes not only across object variations in the data domain but also to unseen trajectories within the training range. Although this does not constitute out-of-distribution generalization, it demonstrates robust interpolation in the continuous trajectory space, implying

that the network learns a smooth mapping from acquisition geometry to redundancy weighting rather than overfitting to discrete trajectories. Future work will investigate extrapolation beyond the training range, extension to more complex trajectories, more comprehensive ablation studies of network components, and validation on real-world datasets.

References