Resolution-Agnostic Neural Operators for Multi-Rate Sparse-View CT

Aujasvit Datta[1,3†] Jiayun Wang[1,2†] Asad Aali[4] Anima Anandkumar[1]
[1]Caltech [2]Georgia Tech [3]IIT Kanpur [4]Stanford University
aujasvitd22@iitk.ac.in, {peterw, anima}@caltech.edu, asadaali@stanford.edu
[†]Equal contribution

Abstract. Sparse-view Computed Tomography (CT) reconstructs images from a limited number of X-ray projections to reduce radiation and scanning time, which is an ill-posed inverse problem. Existing methods achieve high-fidelity reconstructions but overfit to a fixed acquisition setup, failing to generalize well across sampling rates. For example, convolutional neural networks (CNNs) use the same kernels across resolutions, leading to artifacts when data resolution changes. This is a critical limitation in clinical practice, where acquisition sampling settings vary across organs and diagnostic protocols.

We propose Computed Tomography neural Operator (CTO), the first neural operator (NO) framework for CT reconstruction. CTO extends learning from fixed discretized grids to continuous function space, enabling a single model to generalize across measurement sampling rates without retraining. We also propose new NO architectural designs for CT: (i) a dual-domain NO architecture in both sinogram and image spaces, capturing complementary spatial–frequency information, and (ii) rotation-equivariant DIScrete–COntinuous convolutions (DISCO) that exploit the rotational structure inherent in tomographic acquisition. Empirically, CTO outperform CNNs ($>$ 3.4dB PSNR gain) and other baselines in multi-resolution settings across multiple CT datasets. Compared to state-of-the-art diffusion methods, CTO has 500$\times$ faster inference with an average 3dB gain. CTO further demonstrates strong out-of-distribution robustness, maintaining gains under cross-dataset transfer and noisy sinogram conditions. Ablation studies also validate each design choice. CTO establishes neural operators as a principled and practical paradigm for flexible, discretization-agnostic CT reconstruction. Our code is available at https://github.com/neuraloperator/sparse_ct.

1 Introduction

Computed Tomography (CT) is a widely used imaging modality that reconstructs cross-sectional images of internal anatomy from X-ray projections acquired at multiple angles. The measured projection data (sinogram) represent line integrals of the object's linear attenuation coefficient along different ray paths, mathematically modeled (for parallel-beam) by the Radon transform [22, 40]. CT image reconstruction aims to recover the underlying attenuation

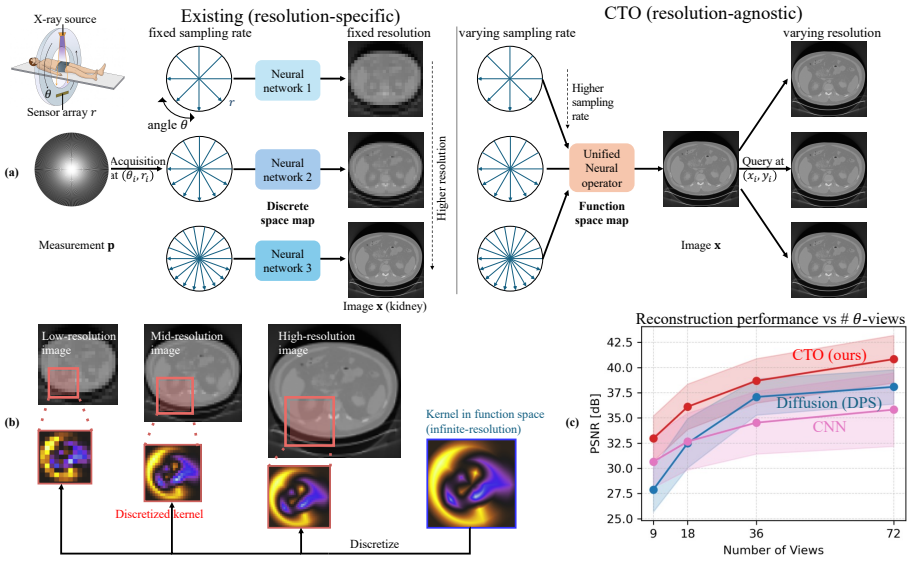

Fig. 1: **(a)** We propose **CTO**, a unified **CT** reconstruction framework based on resolution-agnostic neural **O**perator parameterized in function space. CTO reconstructs CT images across different measurement sampling rates and resolutions, whereas existing methods [6, 49] are resolution-dependent and need to be trained separately for different sampling rates due to resolution-specific backbones. **(b)** CTO's basic building block is DISCO (discrete-continuous convolution) [38], a convolution layer defined in function space which works for images at different resolutions. Specifically, DISCO learns convolution kernels in the infinite-dimensional function space and can be discretized to a specific resolution based on the input data resolution to ensure a consistent receptive field across resolutions. **(c)** CTO consistently outperforms the CNN baseline [3] across different sampling rates for kidney CT [21]. They both follow an unrolled network design with similar model size; CTO uses resolution-agnostic DISCO while CNN uses regular convolutions.

map from these projections, forming an inverse problem—namely, inversion of the Radon transform. Under ideal conditions with fully sampled data and sufficient angular coverage, this inverse transform yields accurate reconstructions. However, in sparse-view CT (SVCT), where the number of projection angles is significantly reduced and uniformly subsampled to lower radiation dose or shorten scan time [20], the inverse problem becomes ill-posed. In such cases, direct inversion leads to severe noisy artifacts, and additional regularization or learning-based methods are required to achieve high-quality reconstructions.

Classical CT reconstruction relies on filtered back projection (FBP) [43], which is efficient but produces severe artifacts under sparse sampling [1]. Other classical compressed sensing methods [9,18,44] improve quality by enforcing data consistency and sparsity priors, but at high computational cost. Deep learning has since transformed SVCT reconstruction: CNNs [12, 13] and transformer-based architectures [60] substantially outperform classical methods. Unrolled model-based networks [2, 3, 11] bridge the gap between classical and learned ap-

proaches by unfolding iterative optimization into trainable network layers, embedding data consistency directly into the architecture. However, most of these approaches operate solely in the image domain, treating reconstruction as a post-processing step after FBP and discarding the rich information present in the raw sinogram. Dual-domain methods [31,49,52,61] address this by jointly performing sinogram completion and image refinement with improved performance.

Yet, existing learning methods are tied to a specific discretization setting with a fixed input measurement sampling rate—a model trained for one measurement sampling rate does not generalize well to another [25,32] (See Table 2a). A common solution to avoid sampling-rate-overfitting (e.g., [42]) is to train several models, each for a different sampling rate. However, maintaining distinct models for different sparsity regimes limits scalability, and the approach does not generalize across sampling rates/image resolutions. This is a critical limitation in clinical settings, where acquisition settings like measurement sampling rates and target resolution constantly change across organs, protocols, and diagnostic purposes. *Thus, this paper adopts a setting where a unified model is used across acquisition sampling rates.* This enables flexible CT reconstruction under varying conditions, thus broadening its clinical applicability.

Our approach. We propose CTO, a unified CT reconstruction framework based on resolution-agnostic Neural Operators (NOs) [27] that learns mappings between infinite-dimensional function spaces, enabling a single model to work and generalize across measurement sampling rates and image resolutions without retraining. To the best of our knowledge, CTO is the first application of neural operators to sparse-view CT.

CTO uses unrolled networks with two U-shaped DISCO [25,34] blocks that process the measurements and images: (NO_s) in the *sinogram domain* and (NO_i) in the *image domain*. The sinogram operator (NO_s) is defined in polar coordinates (r, θ) — the natural space of CT measurements — and jointly learns in both *spatial and frequency spaces* of sinograms, inspired by FBP [43]. The image operator (NO_i), defined on the 2D Cartesian grid, further refines spatial features and improves reconstruction quality. Both operators are built on Discrete–Continuous (DISCO) convolutions [25, 34], which parametrize kernels in the continuous function space and discretize them to match different input or output resolution—making the entire framework inherently resolution-agnostic. DISCO has been proven to be discretization-convergent [8,34]: the approximation error is bounded across resolutions and converges to zero as resolution increases.

This design yields two structural properties. First, CTO is resolution- and sampling-agnostic: the function-space parametrization allows consistent performance across different data resolutions without architectural changes. Second, learning in the sinogram domain's polar coordinate system makes CTO rotation-equivariant—a rotation in the measurement coordinates induces an equivalent rotation in the reconstructed image—improving robustness to varying acquisition angles. Empirically, replacing CNNs with the proposed neural operators yields an average gain of over 3.4 dB PSNR across sinogram subsampling rates. CTO outperforms state-of-the-art diffusion methods [16] by over 3 dB while be-

ing at least 500× faster at inference. CTO also outperforms all SOTA baselines in SVCT reconstruction and zero-shot super-resolution reconstruction.

Our main contributions are: **1)** We propose CTO, a unified end-to-end CT reconstruction framework based on neural operators that generalizes across measurement sampling rates and image resolutions. To the best of our knowledge, we are the first to use Neural Operators for sparse CT reconstruction. **2)** We introduce a joint spatial-frequency sinogram-domain operator which respects the CT rotational geometry and acquisition physics. **3)** By learning in the function space, CTO is inherently rotation-equivariant and resolution-agnostic via DISCO operators in both sinogram and image domains, achieving consistent reconstructions across varying conditions. In summary, our central novelty is a *unified function-space formulation for sparse-view CT*: instead of learning separate discrete models for each sampling rate, we learn a single continuous operator from which different discretizations can be sampled, enabling *multi-rate reconstruction and cross-discretization consistency without retraining*.

The remainder of the paper is organized as follows. Sec. 2 reviews related work on accelerated CT reconstruction and neural operators. Sec. 3 establishes the theoretical framework for the problem, formulating CT reconstruction in function space and introducing DISCO convolutions. Sec. 4 details CTO's design, including its dual-domain operators and rotation-equivariant convolution. Sec. 5 presents experimental results, and Sec. 6 concludes.

2 Related works

Accelerated CT Reconstruction. Methods broadly fall into (i) *physics-based iterative* and (ii) *learning-based* approaches. Classical compressed-sensing methods impose handcrafted priors, especially total variation (TV) and prior-image constrained CS (PICCS), to stabilize ill-posed sparse-view or limited-angle reconstruction. They reduce artifacts, but rely on slow multi-iteration solvers and degrade under extreme undersampling [10, 33, 58], highlighting the *speed–fidelity trade-off* of optimization-based CT. Learning-based methods instead learn data-driven priors for faster inference [6, 23, 36]. Early image-domain CNNs exploit hierarchical feature extraction and nonlinear representation capacity to outperform classical methods in sparse-view reconstruction [12, 13], while architectures such as TransCT [60] introduce dual-path transformers to refine high-frequency details through piecewise reconstruction. However, these methods treat reconstruction as post-hoc denoising of an initial FBP image, discarding rich sinogram information and often producing artifacts that violate physical consistency [57, 59]. Model-based unrolled networks such as Learned Primal-Dual (LPD) [2], RegFormer [56], LEARN [11], and QN-Mixer [6] address this by integrating forward and adjoint operators for improved data consistency, but they remain tied to fixed sampling patterns, limiting adaptability. Dual-domain networks [28, 31, 47, 49, 50, 52, 55, 61, 62] jointly perform sinogram completion and image refinement, addressing aliasing artifacts at their source and showing greater robustness. Yet across these families, models remain bound to specific discretization settings: networks trained for one measurement sampling rate or output

resolution generalize poorly to others [25, 32]. Shao et al. [42] broaden coverage by synergistically training separate models for ultra-sparse and sparse regimes, but maintaining distinct networks for each regime limits scalability and does not generalize well across sampling rates. Fan et al. [19] train a single dual-domain unfolding model across multiple sparse-view rates, but remain CNN-based and tied to fixed discretizations, unable to generalize across image resolutions. Generative diffusion models [14, 32, 45, 63] achieve SOTA fidelity in SVCT by alternating data-consistency and denoising steps, but their high inference cost limits practical use. Across all families, the main challenges remain [15]: (1) discretization binding to specific sampling rates and resolutions and (2) inference-time scalability. *Our method addresses these challenges by introducing a NO formulation that learns in function space, enabling cross-resolution generalization and rotation-equivariant reconstruction without retraining.*

Neural Operators for Computational Imaging. For resolution-agnostic image reconstruction, diffusion models have shown empirically stable performance across data resolutions in accelerated CT reconstruction [32, 63]. Yet, they typically incur higher inference times and require substantial hyperparameter tuning to achieve optimal results. Further, diffusion models are not inherently discretization-agnostic. In contrast, a neural operator (NO) [7, 27] is inherently resolution-agnostic as it learns a map between infinite-dimensional function spaces, enabling evaluation at arbitrary resolutions and converging toward the target operator as the grid resolution increases. NOs were first used for accelerating the numerical solutions to partial differential equations (PDEs) [27, 51, 53]. More recently, NOs have been used for resolution-agnostic computational imaging tasks in various domains, e.g., MRI [25], neuro-imaging [48], and ultrasound imaging [53, 54]. NO architectures are tailored to each application. For instance, the Fourier NO (FNO) [30] employs global convolutions and has achieved robust discretization-agnostic performance across diverse tasks [7]. Other NOs [29, 34] use locally-supported kernels to better capture localized structures, which is beneficial in applications involving localized dynamics such as turbulent fluid modeling [29]. Additionally, NOs with local integral formulations are more amenable to parallel computation and can offer improved efficiency compared to architectures that rely on global operations. *Our CT reconstruction framework leverages neural operators with local integral kernels, making it inherently agnostic to both sinogram sampling rates and output image resolution.*

3 CT Reconstruction in Function Space

CTO combines unrolled networks and function space learning to allow physical awareness (of the forward operator) and resolution-agnosticism (Fig. 2). Among multiple architectural designs [29, 30] available for neural operators, we choose discrete-continuous convolution (DISCO), especially for image reconstruction due to its similarity to convolutional layers. We finally present a U-shaped DISCO block, the basic building block enabling multi-scale receptive field.

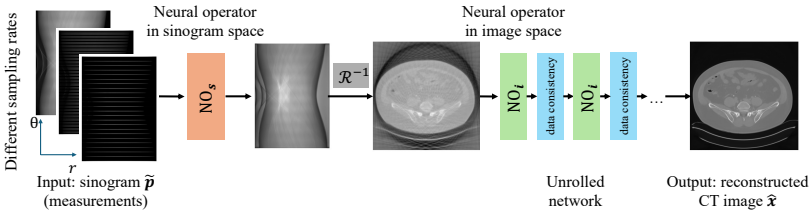

Fig. 2: CTO architecture overview. It follows an unrolled network design (Sec. 3.1). The input sensory signal sinogram is first fed to neural operator NO_s defined in sinogram space, allowing CTO to be resolution-agnostic to different sampling rates for sinograms. Then, we use unrolled networks with multiple cascades mimicking a classical iterative algorithm, with each cascade consisting of image-space NO_i and a data consistency (physical update) term. They are both defined in function space and make CTO discretization-agnostic to sampling rates and image resolutions. $\mathcal{R}^{-1}$ refers to the inverse Radon transform that transforms a sinogram to an image.

3.1 CT Reconstruction with Unrolled Networks

In CT, cross-sectional images are reconstructed from a series of X-ray measurements acquired at multiple angles around an object (Fig. 1a). As the beams propagate through the object, their intensity is attenuated according to the material along each path, and detectors record this attenuation as a one-dimensional projection. Collecting projections over many angles produces a two-dimensional sinogram, which compactly represents the measurement data.

Mathematically, the sinogram can be described as a function with polar coordinates $\mathbf{p}(\theta, r)$, where θ is the projection angle and r is the detector position. The sinogram's relationship with the underlying image x can be written as

$$\mathbf{p}(\theta, r) := \mathcal{R}(\mathbf{x}) + \epsilon \tag{1}$$

where $\mathcal{R}$ is the Radon transform and ϵ is the measurement noise. For a specific machine, sensor position r is fixed, but the sampling of θ varies via the programming of the scanning speed. To accelerate acquisition and reduce radiation, fewer projections are taken and measurements are subsampled as $\tilde{\mathbf{p}} = M\mathbf{p}$, where M is an angle sampling operator, which can be implemented as a binary mask that selects a subset of projection angles. Recovering the image $\mathbf{x}$ from $\tilde{\mathbf{p}}$ becomes an ill-posed inverse problem. Classical compressed sensing methods solve the inverse problem and reconstruct the underlying image $\hat{\mathbf{x}}$ via optimization

$$\hat{\mathbf{x}} = \text{argmin}_{\mathbf{x}} \frac{1}{2} \|\mathcal{A}(\mathbf{x}) - \tilde{\mathbf{p}}\|_2^2 + \lambda \Psi(\mathbf{x}) \tag{2}$$

where $\mathcal{A}(\cdot) := M\mathcal{R}$ is the linear forward operator, $\Psi(\mathbf{x})$ is a regularization term and λ is the weight of the regularization term. The optimization can be solved with gradient descent. Recently, unrolled networks [2, 3, 11, 56] approximate the iterative solver by cascading learned updates. We adopt [3], a popular choice that uses multiple cascades of deep learning architectures such as CNNs, where a specific cascade t mimics a gradient descent step from $\mathbf{x}^t$ to $\mathbf{x}^{t+1}$:

$$\mathbf{x}^{t+1} \leftarrow \mathbf{x}^t - \eta^t \mathcal{A}^*(\mathcal{A}(\mathbf{x}^t) - \tilde{\mathbf{p}}) + \lambda^t \, \text{CNN}^t(\mathbf{x}^t) \tag{3}$$

where η^t, λ^t are weights of the data-consistency term and deep-learning-based regularization term. $\mathcal{A}^*$ is the hermitian of the forward operator $\mathcal{A}$.

Reconstruction in Function Space. Both sinograms (measurements) and images are *functions*: $\mathbf{p} : \mathbb{S}^1 \times \mathbb{R} \to \mathbb{R}$ and $\mathbf{x} : \Omega \subset \mathbb{R}^2 \to \mathbb{R}$. A pixel grid is merely a discretization (sampling) of pixel value $\mathbf{x}$ over coordinates (x, y); higher image resolution corresponds to denser (x, y) sampling. Likewise, acquisition with more view angles corresponds to denser sampling of θ for $\mathbf{p}$.

While Eqn. (2) naturally accommodates different samplings via $\mathcal{A}$, CNN-based unrolled networks (Eqn. (3)) optimize in a *discrete subspace* tied to a fixed grid and often overfit to that resolution. In contrast, neural operators define the learned map directly between *function spaces*, enabling a single model to act consistently across discretizations of both $\mathbf{p}$ and $\mathbf{x}$. We therefore propose to replace the discrete CNN prior with an image-space neural operator NO_i:

$$\mathbf{x}^{t+1} \leftarrow \mathbf{x}^t - \eta^t \mathcal{A}^*\big(\mathcal{A}(\mathbf{x}^t) - \tilde{\mathbf{p}}\big) + \lambda^t \, \mathrm{NO}_i^t(\mathbf{x}^t), \tag{4}$$

so the learned update is resolution-agnostic: the same model applies to different data discretizations (Fig. 1a), i.e., different angle samplings for $\mathbf{p}$ and to different spatial resolutions for $\mathbf{x}$. We choose a neural operator design specifically adapted for computational imaging as follows.

3.2 DISCO (Discrete-Continuous) Convolutions: Basic Unit for Function Space Learning

The fundamental building block of the proposed neural operator is the discrete-continuous convolution (DISCO) neural operator [34], which mimics the convolution layer but optimizes in the function space. It also provides a principled way to embed local inductive biases into neural operators while remaining resolution-agnostic. Classical CNNs achieve locality through finite-support convolutional kernels. However, a CNN layer converges to pointwise linear mappings as input resolution increases, thus losing its interpretation as local integral operators in the infinite-resolution limit [34]. This limits their ability to generalize across varying resolutions and downsampling patterns. DISCO addresses this by defining the kernel as a continuous function in the function space and discretizing it only at the resolution of the input. For a kernel κ and input g over a compact domain $D \subset \mathbb{R}^d$, the continuous convolution is given by $(\kappa \star g)(v) = \int_D \kappa(u - v)\, g(u)\, du$ and is approximated by

$$(\kappa \star g)(v_i) \approx \sum_{j=1}^{m} \kappa(u_j - v_i)\, g(u_j)\, q_j. \tag{5}$$

The kernel is parameterized as $\kappa = \sum_{\ell=1}^{L} \theta_\ell \kappa_\ell$ with learnable coefficients θ_ℓ and fixed basis functions κ_ℓ, where κ_ℓ is a piecewise-linear basis following [25] (visualized in Fig. 1b). Because the basis is defined in the function space, the convolution scales with the domain rather than the grid, converging to a local integral operator as resolution increases. This property makes DISCO an effective

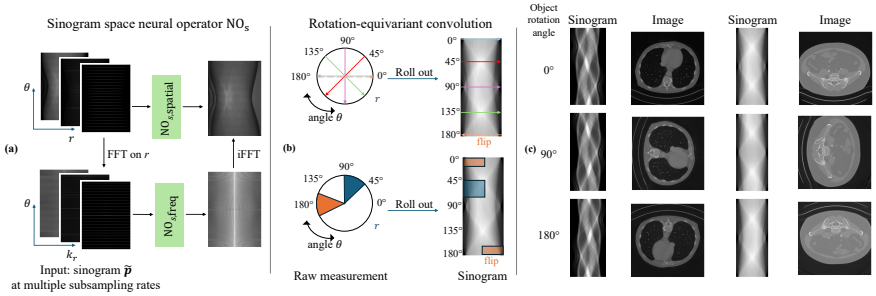

Fig. 3: (a) **Sinogram-space neural operator** NO_s**.** The subsampled sinogram $\tilde{\mathbf{p}}$ is processed by two complementary operators: $\mathrm{NO}_{s,\mathrm{spatial}}$ in the spatial domain and $\mathrm{NO}_{s,\mathrm{freq}}$ in the frequency domain via a 1D Fourier transform along the detector axis r. Their outputs are combined and passed to the image-space stage. (b) **Rotation-equivariant learning.** Object rotation corresponds to a shift along the angular axis θ in the sinogram. Circular padding along θ preserves this rotational structure during learning. Note the angle flips at 180° (See Supp. Sec. A). (c) **Visualization of rotational equivariance.** We rotate the object by 0°, 90°, and 180° and show the corresponding sinograms (left) and CTO reconstructions (right). Reconstructions rotate consistently with the inputs, confirming stable performance across orientations.

backbone for both NO_s and NO_i, enabling consistent feature learning across resolutions and sampling conditions.

UDNO (U-Shaped DISCO Block). UDNO [25] is a core architectural unit. Inspired by U-Net [41], it progressively downsamples to aggregate coarse contextual information and upsamples to recover fine spatial detail. U-shaped networks [41] have shown strong performance across a range of vision tasks [17,39,41] due to their ability to learn multi-scale features. UDNO builds on this design but replaces standard convolutions with DISCO layers, combining multi-scale feature learning with the resolution-agnostic properties of NOs.

4 CTO Design Details

Framework Overview. The proposed CTO is an end-to-end model following the unrolled network architecture [1](Fig. 2). Specifically, the input sinogram $\mathbf{p}$ is first fed to a sinogram space neural operator NO_s for function space learning that unifies different sampling rates for θ. The intermediate feature map is then fed to the unrolled network for multi-cascade learning, where the image-space neural operator NO_i and the physical update/data-consistency term denoise and reconstruct the final images (Eqn. (4)).

DISCO in NO_i is defined in 2D Cartesian coordinates, introducing a local inductive bias in the image space that captures fine details across discretizations. In NO_s, we extend DISCO from Cartesian to polar coordinates (r, θ), allowing

[1] We use a cascaded dual-domain design because it is necessary for strong state-of-the-art CT performance, but our main contribution lies in replacing discretization-specific operators with a single function-space operator that supports multiple view samplings and image resolutions within one model.

the kernels to provide localized support in the sinogram domain. Due to Radon transform duality, localized support corresponds to direction-selective global support in the image space, thus allowing NO_s to learn global structural information. The two components jointly model complementary information about the image, resulting in improved structural consistency and sharper detail.

4.1 Sinogram-Space Operator NO_s

Sinogram-Space NO. We introduce a sinogram-space neural operator NO_s that acts on the function space of $\mathbf{p}(\theta, r)$, with detector position r and view angle θ treated as the two sinogram axes. DISCO convolutions are applied in this (θ, r) coordinate system—without projection or reparameterization to (x, y)—so the model learns directly in the intrinsic measurement domain, avoiding projection errors. Extending DISCO from 2D Cartesian to polar coordinates requires honoring the periodic boundary condition in θ; we therefore use circular padding along θ (and standard padding along r), yielding θ-shift equivariance (rotational equivariance) and consistent handling of arbitrary angular offsets. As a neural *operator*, NO_s maps between function spaces rather than fixed grids, so the same model applies across discretizations (e.g., different detector bins).

We adapt DISCO, originally defined in the image space, to the sinogram space with a novel dual-branch design. The NO_s module comprises two parallel UDNO branches (Fig. 3a): $\mathrm{NO}_{s,\mathrm{spatial}}$ and $\mathrm{NO}_{s,\mathrm{freq}}$. $\mathrm{NO}_{s,\mathrm{spatial}}$ operates on $\mathbf{p}(\theta, r)$ to capture local, geometry-aware correlations, while $\mathrm{NO}_{s,\mathrm{freq}}$ operates on the sinogram's 2D frequency representation to model long-range angular–radial dependencies. The two outputs are averaged to form the final sinogram-space estimation, which is then forwarded to the downstream reconstruction stage. The design is inspired by FBP [43], where the additional r-frequency processing of the sinogram greatly reduces reconstruction blur.

Rotation-Equivariant Convolution. Parallel-beam CT induces two symmetries on the sinogram $\mathbf{p}(\theta, r)$: (i) rotation of the image by ϕ is a *shift in view angle* of the sinogram, $(\mathbf{p} \circ R_\phi)(\theta, r) = \mathbf{p}(\theta - \phi, r)$, and (ii) $\mathbf{p}(\theta + \pi, r) = \mathbf{p}(\theta, -r)$ (the π-periodicity-with-flip identity). We design NO_s to be rotation-equivariant (see Supp. A). For implementation, NO_s adopts a circular padding along the projection angle (θ) axis of the sinogram to account for the inherent periodicity of projection data. Projections at $\theta = 0$ and $\theta = 180^\circ$ correspond to parallel beam paths and are therefore identical up to a flip along the detector r axis. More generally, the projection at $\theta = 180^\circ + \alpha$ is equivalent to the projection at $\theta = \alpha$ after a flip along the detector r axis. To preserve this physical continuity, we use a modified circular padding scheme that first flips the boundary data along the detector r axis and then wraps it around the projection angle θ axis.

We visualize the equivariant design in Fig. 3b. This design choice is significant for two reasons. 1) it provides physically consistent boundary conditions in NO_s, preventing discontinuities at $\theta = 0$ and $\theta = 180^\circ$ and the artifacts they can cause. 2) It enhances rotational robustness by aligning the network's inductive biases with the geometry of the sinogram domain. Since DISCO operators in NO_s act over the (r, θ) axes, which we treat as rectangular axes for NO learning, rotational

transformations of the image x manifest as translations along these axes, which can be naturally modeled by the translational equivariance of DISCO. As this periodicity arises only along the angular dimension, we apply flipped circular padding along θ and standard reflective padding along r.

UDNO for Spatial Sinogram. $\mathrm{NO}_{s,\mathrm{spatial}}$ is a UDNO with circular padding, which learns in the spatial domain of the sinogram. Mathematically, its output $\hat{\mathbf{p}}_{\mathrm{spatial}}$ can be written as $\hat{\mathbf{p}}_{\mathrm{spatial}} = \mathrm{NO}_{s,\mathrm{spatial}}(\tilde{\mathbf{p}})$.

UDNO for Sinogram's Frequency Component. To learn in the frequency domain of the sinogram, we first apply a one-dimensional Fourier transform along the detector (r) axis of the sinogram. The transformed representation is then processed by the circular-padded UDNO $\mathrm{NO}_{s,\mathrm{freq}}$, after which an inverse Fourier transform is applied along the same axis to return the output to the spatial domain. The overall output $\hat{\mathbf{p}}_{\mathrm{freq}}$ can be written as $\hat{\mathbf{p}}_{\mathrm{freq}} = \mathcal{F}_r^{-1}(\mathrm{NO}_{s,\mathrm{freq}}(\mathcal{F}_r(\tilde{\mathbf{p}}(r,\theta))))$.

This design is motivated by two complementary considerations. First, the Fourier slice theorem provides a theoretical foundation for learning in this domain. The theorem states that the one-dimensional Fourier transform of the projection along r satisfies

$$\mathcal{F}_r\{\mathbf{p}(\theta, r)\}(\omega) = \hat{f}(\omega\cos\theta, \omega\sin\theta), \tag{6}$$

where $\hat{f}(\xi_x, \xi_y)$ is the two-dimensional Fourier transform of the object. This result implies that the Fourier transform of the sinogram encodes the object's frequency spectrum in polar coordinates. Consequently, localized operations in this domain correspond to global modifications in the reconstructed image, allowing $\mathrm{NO}_{s,\mathrm{freq}}$ to learn high-level global features about the underlying image.

Second, classical reconstruction algorithms such as FBP incorporate an explicit prior on the frequency domain of the sinogram by applying fixed frequency-domain filters (such as the ramp filter) to the one-dimensional Fourier transform of the sinogram along the detector (r) axis, suppressing low-frequency components before backprojection. Our approach can be interpreted as a data-driven generalization of this idea: rather than imposing a hand-crafted prior, $\mathrm{NO}_{s,\mathrm{freq}}$ learns a frequency-domain prior directly from data. This learned prior captures how the sinogram should be completed under different subsampling conditions, thus reducing image artifacts and enabling the model to generalize better.

Final Output. The final output of NO_s is given by: $\mathrm{NO}_s(\tilde{\mathbf{p}}) = \frac{\hat{\mathbf{p}}_{\mathrm{spatial}} + \hat{\mathbf{p}}_{\mathrm{freq}}}{2}$, which is then passed downstream for processing in the image domain.

4.2 Image-Space Operator NO_i

NO_i is implemented as a UDNO (Sec. 3.2) that operates directly in the spatial domain of the image space. DISCO allows NO_i to train and infer at different image resolutions. NO_i refines and denoises the intermediate reconstruction obtained from the sinogram-space operator by capturing high-frequency local information and finer structural details that may not be fully recovered in earlier stages. This ensures final reconstructions preserve multi-scale features, including large-scale structural consistency and small-scale details.

Table 1: Sparse-view CT reconstruction results for AAPM dataset [37]. Lower is better for RMSE; higher is better for PSNR/SSIM. For all learning methods, a single model is trained on multi-view sinograms with $N_v = 9, 18, 36, 72$ views together. For fairness, all methods are trained using the same amount of data – this differs from the resolution-specific models trained in their original paper since multi-resolution training generally reduces overfitting (Table 2a). All methods are reported on the entire test set. 9-view test results are in Supp. C.6 due to space limit.

Category	Method	18-view			36-view			72-view		
		RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑	RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑	RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑
Learning-free	FBP [43]	632.58 ± 42.17	13.14 ± 1.63	33.87 ± 6.32	616.66 ± 41.07	13.36 ± 1.63	36.07 ± 5.89	584.72 ± 38.87	13.82 ± 1.63	40.35 ± 5.66
	SART [4]	161.91 ± 5.93	24.97 ± 1.46	59.51 ± 4.53	145.15 ± 2.59	25.91 ± 1.41	69.09 ± 3.37	133.24 ± 1.81	26.66 ± 1.40	76.85 ± 2.06
Diffusion	DPS [16]	149.97 ± 17.34	25.68 ± 1.77	60.08 ± 8.28	100.50 ± 7.33	29.13 ± 1.51	76.71 ± 4.02	75.19 ± 4.71	31.64 ± 1.43	82.61 ± 2.94
	ALD [24]	148.13 ± 14.63	25.78 ± 1.71	63.42 ± 7.19	110.19 ± 8.78	28.33 ± 1.46	75.50 ± 3.92	85.17 ± 5.35	30.56 ± 1.41	80.47 ± 3.12
Single pass	DuDoTrans [49]	154.64 ± 6.18	25.36 ± 1.50	69.52 ± 4.32	92.17 ± 3.59	29.88 ± 1.62	78.21 ± 3.59	69.28 ± 4.99	32.35 ± 1.57	81.49 ± 3.38
	GloReDi [28]	96.89 ± 10.39	29.47 ± 1.70	78.52 ± 3.69	74.42 ± 6.24	31.74 ± 1.64	83.59 ± 3.05	64.63 ± 4.34	32.96 ± 1.55	85.78 ± 2.60
	MDPRNet [42]	151.78 ± 17.64	25.38 ± 1.82	73.05 ± 4.53	116.85 ± 11.28	27.91 ± 1.79	81.60 ± 3.85	74.02 ± 8.93	30.87 ± 1.55	87.18 ± 3.64
Unrolled Networks	Learned PD [2]	155.14 ± 13.54	25.35 ± 1.61	50.21 ± 5.62	126.36 ± 9.79	27.13 ± 1.60	57.7 ± 5.82	91.07 ± 5.70	29.96 ± 1.53	71.88 ± 4.59
	LEARN [11]	153.39 ± 22.22	25.51 ± 1.94	73.42 ± 3.90	121.31 ± 16.15	27.53 ± 1.81	77.64 ± 3.41	92.49 ± 12.91	29.90 ± 1.59	81.37 ± 2.72
	RegFormer [56]	150.57 ± 21.34	25.67 ± 1.91	72.95 ± 4.02	114.93 ± 15.18	28.00 ± 1.82	77.87 ± 3.56	86.83 ± 10.46	30.43 ± 1.70	81.55 ± 3.01
	Unrolled CNN [3]	100.65 ± 11.36	29.14 ± 1.84	78.21 ± 5.47	74.44 ± 8.32	31.76 ± 1.80	82.01 ± 4.95	62.38 ± 11.16	33.35 ± 1.81	84.49 ± 5.27
	CTO (ours)	**75.97 ± 7.32**	**31.57 ± 1.73**	**81.62 ± 4.02**	**50.79 ± 4.27**	**35.06 ± 1.64**	**87.14 ± 3.12**	**36.64 ± 2.44**	**37.88 ± 1.56**	**91.55 ± 1.81**

5 Experimental Results

We study a setting in which a single model operates across varying sinogram sampling rates (Fig. 1a-Right). Using sampling rate as an example, we observe that existing approaches [6, 11] train separate models for each rate. Such rate-specific training overfits to the corresponding configuration (Table 2a, row 1-2). In contrast, training on multiple-sampling-rate data mitigates this issue and improves generalization (Table 2a, row 3-5). For fairness, all methods and baselines are trained using the same amount of multi-rate data. Baseline method hyperparameters are tuned to ensure the best multi-rate performance.

5.1 Dataset and Setup

Datasets: We consider two CT datasets in the paper:

- **AAPM Low-Dose Abdominal CT Dataset** [37] consists of 5,936 axial slices with patient ID. Following the split in [28,49], we select 90% of patients for training and the rest (526 slices) for testing.
- **Kidney CT Scan Dataset** [21] is from 2019 Kidney and Kidney Tumor Segmentation Challenge (C4KC-KiTS). For each CT volume, we select the central 50% of slices and a uniform 10% sample of peripheral slices to ensure coverage across anatomical variability, resulting in 33,000 training slices from 170 patients. For evaluation, we use 10,000 slices from 40 held-out patients.

Main-text quantitative tables focus on AAPM [37]; Kidney [21] visualization and results are discussed in main text and further expanded in Supp. C.1.

CTO Implementation. CTO follows unrolled networks [3] with 3 cascades. Each cascade consists of an image-space operator (NO_i) followed by a data

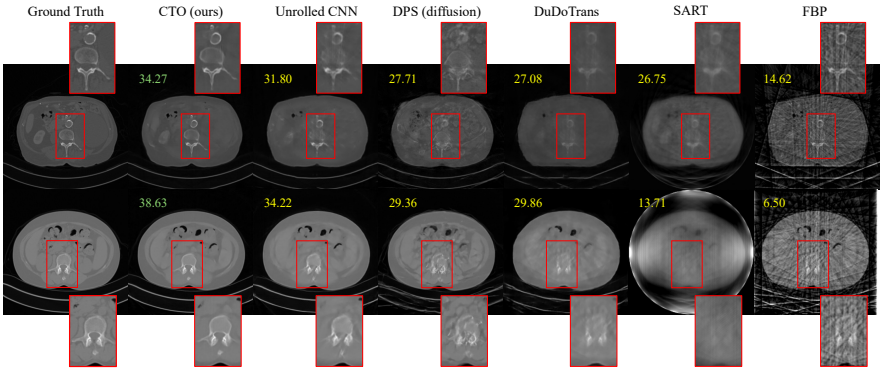

Fig. 4: Reconstruction on 18-view sampling rate for Abdomen Low-Dose CT dataset [37] (top row) and Kidney dataset [21] (bottom row). CTO outperforms baselines for PSNR (upper left) and reconstruction quality.

consistency module. All UDNOs (Sec. 3.2) share the same architectural configuration: DISCO parameterized with a piecewise-linear kernel basis [25] composed of one isotropic basis and 5 anisotropic basis rings, each containing 7 basis functions (hyperparameter-tuned, CNN baseline matches the model size). Each UDNO uses one input and one output channel, 32 hidden channels, and four encoder–decoder (pooling) levels. The DISCO kernels use a radius cutoff of 0.02 and are defined over an input domain of size 256×256. Models are trained using mean squared error loss with the Adam optimizer and a learning rate of $1e-3$.

Baseline: Unrolled Networks. We compare with existing SOTA Model-based Unrolled Networks like Learned Primal-Dual [2], RegFormer [56], and LEARN [11]. We also compare with an Unrolled CNN model which uses the same architecture as CTO, with DISCOs replaced by regular convolutions. Unrolled CNN can be considered an ablation study that compares neural operator (function space) and neural network (discrete space) architecture. However, Unrolled CNN has slightly different number of parameters than CTO due to a different kernel configuration.

Baseline: Single-pass methods. We compare with single-pass methods like DuDoTrans [49], GloReDi [28] and MDPRNet [42]. Details of baseline implementations, sinogram sampling settings, hardware and training are in Supp. Sec. B.

Evaluation Protocols. CT images represent linear attenuation coefficients that are typically interpreted in *Hounsfield Units* (HU), where water is defined as 0 HU and air as -1000 HU. Following [14], we report RMSE after converting reconstructed images to HU to provide a clinically meaningful measure of reconstruction error. We also report PSNR and SSIM in the original reconstruction space (before HU conversion) to assess structural and perceptual fidelity independent of scanner calibration. Thus, RMSE (HU) reflects quantitative clinical accuracy, while PSNR and SSIM capture image similarity.

5.2 Reconstruction with Different Sampling Rates

Multi-Sparse-View Rate Results. We train and evaluate reconstruction performance at 9-, 18-, 36-, and 72-view acquisition settings, where fewer projections

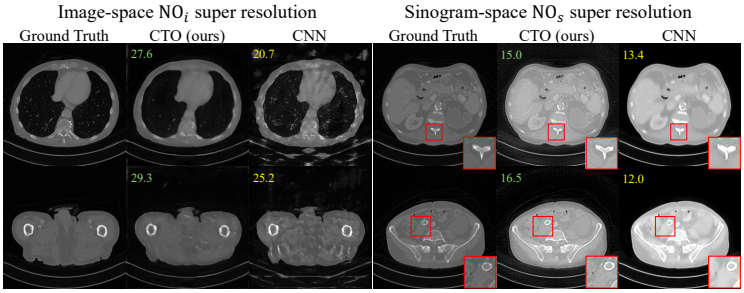

Fig. 5: Zero-shot super resolution results. CTO outperforms CNN baseline for both NO_i and NO_s super resolution.

correspond to lower radiation at the cost of more ill-posed reconstruction. For all methods, a single model is trained across all sampling rates. In each training batch, a sparsity rate is randomly and uniformly chosen, ensuring that all models train on equal amounts of data for each sparsity rate. Results for the AAPM dataset [37] are shown in Table 1, and for Kidney dataset in Fig. 1c. Full results for the C4KC-KiTS Kidney dataset [21] are reported in Supp. C.1.

CTO outperforms learning-free methods [4, 43] and deep learning methods, including both diffusion methods [16, 24] and architectures like CNNs [3] and transformers [49]. On the kidney dataset [21], on average across different numbers of views, we perform 1) 4 dB PSNR better than unrolled CNN baseline; 2) 3 dB PSNR better than diffusion-DPS [16]; 3) 5 dB better than transformer-DuDoTrans [49]. On the abdomen dataset [37], we perform 1) 3 dB PSNR better than unrolled CNN baseline; 2) 6 dB better than DPS [16]; 3) 6 dB better than DuDoTrans [49] on average. Visualizations are in Fig. 4. Note that some baseline models show different performance from what was reported in their original papers. This is expected since they train multiple models for each rate.

5.3 Zero-Shot Super-Resolution Results

Super-Resolution on NO_i. We evaluate the zero-shot super-resolution capability of NO in the image domain. All models—CTO, the Unrolled CNN baseline [3], RegFormer [56], and LEARN [11]—are trained to produce 256×256 reconstructions from 18-view sinograms. At inference, we test generalization to a higher resolution *without any fine-tuning*: the sinogram-space operator NO_s is kept fixed, and the intermediate 256×256 output is bilinearly upsampled to 512×512 before image-space processing. Reconstructions are evaluated against fully sampled ground truth images that have been bilinearly interpolated to 512×512. When the resolution doubles, CNN's fixed-size kernels cover half the relative receptive field, whereas a NO's kernel, defined as a continuous function, maintains the same relative coverage. Consequently, CTO achieves a 3 dB PSNR gain over the CNN baseline, and 4.8 dB gain over LEARN [11]. As shown in Fig. 5, the CNN baseline exhibits substantial blurring and aliasing at the higher resolution, while CTO preserves much sharper detail.

Table 2: (a) Left: Baseline methods overfit to the seen sampling rate, performing substantially worse on other rates (row 1-2). Training on all rates together (row 3-5) mitigates this problem. We adopt multi-rate co-training for all methods for the all-view setting (row 3-5). **(b) Right:** Inference and tuning time of methods tested on NVIDIA A100. CTO is $> 500\times$ faster than diffusion [24].

Method/training rate	Test rate		
	18-view	**36-view**	**72-view**
LEARN-72 view	4.16	22.40	32.32
RegFormer-72 view	5.04	24.63	35.11
LEARN-all view	25.51	27.53	29.90
RegFormer-all view	25.67	28.0	30.43
CTO-all view	**31.57**	**35.06**	**37.88**

Category	**Method**	**Inference Time (s)**	**Tuning Required**
Learning-free	SART [4]	0.417	✓
Diffusion	DPS [16]	51.58	✓
	ALD [24]	32.72	✓
Single-pass	RegFormer [56]	0.085	
	CTO (ours)	**0.065**	✗

Super-Resolution on NO_s. We evaluate zero-shot super-resolution in the sinogram domain by training all models—CTO, the Unrolled CNN baseline [3], RegFormer [56], and LEARN [11]—at one sinogram sampling setting and testing at another. Models are trained on sinograms with 72 views and 672 detectors, then evaluated on two increased-sinogram-size settings: 1) increased number of acquisition views: sinograms with 144 views and 672 detectors (CTO outperforms Unrolled CNN by 3.5 dB PSNR, LEARN by 3.2 dB PSNR, RegFormer by 2.8 dB PSNR); 2) further increased number of detectors: sinograms with 144 views and 1344 detectors (CTO outperforms Unrolled CNN by 3.7 dB PSNR, LEARN by 2.7 dB PSNR, RegFormer by 2.4 dB PSNR.). As shown in Fig. 5, CTO preserves fine anatomical structures with fewer artifacts compared to the baselines.

5.4 Analysis and Ablation Study

Rotational Equivariance. As discussed in Sec. 4.1, circular padding enforces the periodic boundary conditions inherent to the sinogram's angular domain, enabling the network to better exploit rotational equivariance. Supp. Table 5 shows that this design leads to a 0.83 dB PSNR gain under 24-view subsampling (33 dB with circular padding, 32.17 dB without circular padding). Qualitative results (Fig. 3c) further demonstrate that the reconstruction rotates as the object rotates, confirming that the model captures true rotational structure. Our equivariant design gives a notable improvement in PSNR, indicating that handling angular periodicity explicitly enhances reconstruction quality.

Ablation of NO_s and NO_i. Averaged over view counts, removing NO_i and NO_s individually leads to PSNR drops of 3.86 dB and 4.82 dB, respectively. Within NO_s, removing the spatial branch causes a 3.1 dB drop and the frequency branch a 2.7 dB drop. Further NO_s design ablations are in Supp. Sec. C.3.

DISCO Kernel Hyperparameter Sensitivity. We perform a Bayesian search over the kernel basis, kernel radius, number of basis rings, and bases per ring for DISCO in CTO. The results are reported in Supp. Sec. C.4.

Model Inference and Tuning Time. We compare the model development and inference time of our end-to-end NO with other methods (Table 2b): Diffusion models are at least $500\times$ slower in inference and require pattern-specific hyperparameter tuning for each sampling rate, since each rate corresponds to a

different measurement distribution and the same model cannot be applied directly across rates. Classical learning-free methods like SART [4] also require hyperparameter tuning for specific **p** sampling rates during optimization.
Out of Distribution (OOD) Generalization. We perform several experiments to test OOD generalization of CTO. Results are in Supp. Sec. C.5.
LIDC-IDRI Lung CT Dataset. We also evaluate Unrolled CNN and CTO on the LIDC-IDRI lung CT dataset [5] in Supp. Sec. C.9.
Complexity analysis. Comparisons with other NO architectures are in Supp. Sec. D.1, and with the Unrolled CNN baseline in Supp. Sec. D.2.

6 Summary

We present CTO, a unified neural operator-based framework for sparse-view CT reconstruction that generalizes across different sampling rates of sensory data, overcoming the limitations of existing deep learning methods, which are tied to a fixed acquisition sampling rate. Specifically, we adopt the Neural Operator, a deep learning framework that learns maps between infinite-dimensional function spaces and thus is agnostic to different data resolutions. By combining rotation-equivariant discrete-continuous convolutions (DISCO) (a specific neural operator design that closely mimics convolutions but uses function-space kernels) with joint spatial–frequency learning in the sinogram domain, CTO achieves robust, resolution-agnostic reconstructions and consistently outperforms state-of-the-art methods across varying downsampling regimes, enabling a scalable and flexible alternative to traditional CT reconstruction techniques. Future work could extend CTO to more clinically relevant datasets, 3D CT and other computational imaging tasks, conduct uncertainty analysis on CTO, and evaluate its clinical applicability across broader scanner protocols.

Acknowledgment

This work is supported in part by ONR (MURI grant N000142312654 and N000142012786). A.D. is supported in part by the Undergraduate Research Fellowships (SURF) at Caltech. J.W. is supported in part by the Pritzker AI+Science initiative and Schmidt Sciences. A.A. is supported in part by Bren endowed chair and the AI2050 senior fellow program at Schmidt Sciences.

References

1. Abbas, S., Lee, T., Shin, S., Lee, R., Cho, S.: Effects of sparse sampling schemes on image quality in low-dose ct. Medical physics **40**(11), 111915 (2013). https://doi.org/10.1118/1.4825096
2. Adler, J., Öktem, O.: Learned primal-dual reconstruction. IEEE transactions on medical imaging **37**(6), 1322–1332 (2018). https://doi.org/10.1109/TMI.2018.2799231

3. Aggarwal, H.K., Mani, M.P., Jacob, M.: Modl: Model-based deep learning architecture for inverse problems. IEEE Transactions on Medical Imaging **38**(2), 394–405 (Feb 2019). https://doi.org/10.1109/tmi.2018.2865356, http://dx.doi.org/10.1109/TMI.2018.2865356
4. Andersen, A.H., Kak, A.C.: Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm. Ultrasonic imaging **6**(1), 81–94 (1984). https://doi.org/10.1177/016173468400600107
5. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al.: The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. Medical physics **38**(2), 915–931 (2011). https://doi.org/10.1118/1.3528204
6. Ayad, I., Larue, N., Nguyen, M.K.: 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. pp. 25317–25326 (2024). https://doi.org/10.1109/CVPR52733.2024.02392
7. Azizzadenesheli, K., Kovachki, N., Li, Z., Liu-Schiaffini, M., Kossaifi, J., Anandkumar, A.: Neural operators for accelerating scientific simulations and design. Nature Reviews Physics pp. 1–9 (2024). https://doi.org/10.1038/s42254-024-00712-5
8. Berner, J., Liu-Schiaffini, M., Kossaifi, J., Duruisseaux, V., Bonev, B., Azizzadenesheli, K., Anandkumar, A.: Principled approaches for extending neural architectures to function spaces for operator learning. arXiv preprint arXiv:2506.10973 (2025). https://doi.org/10.48550/arXiv.2506.10973
9. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on information theory **52**(2), 489–509 (2006). https://doi.org/10.1109/TIT.2005.862083
10. 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**(2), 660–663 (2008). https://doi.org/10.1118/1.2836423
11. Chen, H., Zhang, Y., Chen, Y., Zhang, J., Zhang, W., Sun, H., Lv, Y., Liao, P., Zhou, J., Wang, G.: Learn: Learned experts' assessment-based reconstruction network for sparse-data ct. IEEE transactions on medical imaging **37**(6), 1333–1347 (2018). https://doi.org/10.1109/TMI.2018.2805692
12. Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Zhou, J., Wang, G.: Low-dose ct with a residual encoder-decoder convolutional neural network. IEEE transactions on medical imaging **36**(12), 2524–2535 (2017). https://doi.org/10.1109/TMI.2017.2715284
13. Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., Wang, G.: Low-dose ct denoising with convolutional neural network. In: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017). pp. 143–146. IEEE (2017). https://doi.org/10.1109/ISBI.2017.7950488
14. Chen, J., Lin, Y., Qin, Y., Wang, H., Li, X.: Cross-view generalized diffusion model for sparse-view ct reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 140–150. Springer (2025). https://doi.org/10.48550/arXiv.2508.10313
15. Cheng, S., Chen, Y., Yang, B., Liu, C., Zhang, L., Yin, L., Zheng, W.: Deep learning for sparse-view ct reconstruction: A survey. Available at SSRN 5366340 (2025)

16. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: The Eleventh International Conference on Learning Representations (2023). https://doi.org/10.48550/arXiv.2209.14687
17. Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M.: Diffusion models in vision: A survey. IEEE transactions on pattern analysis and machine intelligence **45**(9), 10850–10869 (2023). https://doi.org/10.1109/TPAMI.2023.3261988
18. Donoho, D.L.: Compressed sensing. IEEE Transactions on information theory **52**(4), 1289–1306 (2006). https://doi.org/10.1109/TIT.2006.871582
19. Fan, X., Chen, K., Yi, H., Yang, Y., Zhang, J.: MVMS-RCN: A dual-domain unfolding CT reconstruction with multi-sparse-view and multi-scale refinement-correction. IEEE Transactions on Computational Imaging (2024). https://doi.org/10.1109/TCI.2024.3489040
20. Han, Y., Ye, J.C.: Framing u-net via deep convolutional framelets: Application to sparse-view ct. IEEE transactions on medical imaging **37**(6), 1418–1429 (2018). https://doi.org/10.1109/TMI.2018.2823768
21. Heller, N., Sathianathen, N., Kalapara, A., Walczak, E., Moore, K., Kaluzniak, H., Rosenberg, J., Blake, P., Rengel, Z., Oestreich, M., et al.: The kits19 challenge data: 300 kidney tumor cases with clinical context, ct semantic segmentations, and surgical outcomes. arXiv preprint arXiv:1904.00445 (2019). https://doi.org/10.48550/arXiv.1904.00445
22. Hermena, S., Young, M.: Ct-scan image production procedures. StatPearls (2023)
23. Hu, D., Zhang, Y., Quan, G., Xiang, J., Coatrieux, G., Luo, S., Coatrieux, J.L., Ji, X., Han, H., Chen, Y.: Cross: Cross-domain residual-optimization-based structure strengthening reconstruction for limited-angle ct. IEEE Transactions on Radiation and Plasma Medical Sciences **7**(5), 521–531 (2023). https://doi.org/10.1109/TRPMS.2023.3242662
24. Jalal, A., Arvinte, M., Daras, G., Price, E., Dimakis, A.G., Tamir, J.: Robust compressed sensing mri with deep generative priors. Advances in neural information processing systems **34**, 14938–14954 (2021). https://doi.org/10.48550/arXiv.2108.01368
25. Jatyani, A.S., Wang, J., Chandrashekar, A., Wu, Z., Liu-Schiaffini, M., Tolooshams, B., Anandkumar, A.: A unified model for compressed sensing mri across undersampling patterns. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 26004–26013 (2025). https://doi.org/10.48550/arXiv.2410.16290
26. Karras, T., Aittala, M., Aila, T., Laine, S.: Elucidating the design space of diffusion-based generative models. Advances in neural information processing systems **35**, 26565–26577 (2022). https://doi.org/10.48550/arXiv.2206.00364
27. Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., Anandkumar, A.: Neural operator: Learning maps between function spaces with applications to pdes. Journal of Machine Learning Research **24**(89), 1–97 (2023)
28. Li, Z., Ma, C., Chen, J., Zhang, J., Shan, H.: Learning to distill global representation for sparse-view ct. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 21196–21207 (2023). https://doi.org/10.48550/arXiv.2308.08463
29. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A.: Neural operator: Graph kernel network for partial differential equations. arXiv preprint arXiv:2003.03485 (2020). https://doi.org/10.48550/arXiv.2003.03485

30. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A.: Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv:2010.08895 (2021). https://doi.org/10.48550/arXiv.2010.08895
31. Lin, W.A., Liao, H., Peng, C., Sun, X., Zhang, J., Luo, J., Chellappa, R., Zhou, S.K.: Dudonet: Dual domain network for ct metal artifact reduction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10512–10521 (2019). https://doi.org/10.1109/CVPR.2019.01076
32. Liu, J., Anirudh, R., Thiagarajan, J.J., He, S., Mohan, K.A., Kamilov, U.S., Kim, H.: Dolce: A model-based probabilistic diffusion framework for limited-angle ct reconstruction. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 10498–10508 (2023). https://doi.org/10.1109/ICCV51070.2023.00963
33. Liu, Y., Liang, Z., Ma, J., Lu, H., Wang, K., Zhang, H., Moore, W.: Total variation-stokes strategy for sparse-view x-ray ct image reconstruction. IEEE transactions on medical imaging **33**(3), 749–763 (2013). https://doi.org/10.1109/TMI.2013.2295738
34. Liu-Schiaffini, M., Berner, J., Bonev, B., Kurth, T., Azizzadenesheli, K., Anandkumar, A.: Neural operators with localized integral and differential kernels. arXiv preprint arXiv:2402.16845 (2024). https://doi.org/10.48550/arXiv.2402.16845
35. Lu, L., Jin, P., Karniadakis, G.E.: Deeponet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. arXiv preprint arXiv:1910.03193 (2019). https://doi.org/10.48550/arXiv.1910.03193
36. Ma, C., Li, Z., Zhang, J., Zhang, Y., Shan, H.: Freeseed: Frequency-band-aware and self-guided network for sparse-view ct reconstruction. In: International conference on medical image computing and computer-assisted intervention. pp. 250–259. Springer (2023). https://doi.org/10.1007/978-3-031-43999-5_24
37. McCollough, C.: Overview of the low dose ct grand challenge. Medical physics **43**(6Part35), 3759–3760 (2016). https://doi.org/10.1118/1.4957556
38. Ocampo, J., Price, M.A., McEwen, J.D.: Scalable and equivariant spherical cnns by discrete-continuous (disco) convolutions. arXiv preprint arXiv:2209.13603 (2022). https://doi.org/10.48550/arXiv.2209.13603
39. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4195–4205 (2023). https://doi.org/10.1109/ICCV51070.2023.00387
40. Radon, J.: 1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. Classic papers in modern diagnostic radiology **5**(21), 124 (2005)
41. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015). https://doi.org/10.1007/978-3-319-24574-4_28
42. Shao, J., Chen, H., Li, Q., Huang, X., Shu, J., Liu, L., Dou, S., Wang, H.: Multi-stage dual-domain progressive network with synergistic training for sparse-view ct reconstruction. Neural Networks p. 108221 (2025). https://doi.org/10.1016/j.neunet.2025.108221
43. Shepp, L.A., Logan, B.F.: The fourier reconstruction of a head section. IEEE Transactions on nuclear science **21**(3), 21–43 (1974). https://doi.org/10.1109/TNS.1974.6499235

44. Sidky, E.Y., Pan, X.: Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. Physics in Medicine & Biology **53**(17), 4777 (2008). https://doi.org/10.1088/0031-9155/53/17/021
45. Song, B., Hu, J., Luo, Z., Fessler, J., Shen, L.: Diffusionblend: Learning 3d image prior through position-aware diffusion score blending for 3d computed tomography reconstruction. Advances in Neural Information Processing Systems **37**, 89584–89611 (2024). https://doi.org/10.48550/arXiv.2406.10211
46. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. arXiv preprint arXiv:2011.13456 (2020). https://doi.org/10.48550/arXiv.2011.13456
47. Sun, C., Salimi, Y., Angeliki, N., Boudabbous, S., Zaidi, H.: An efficient dual-domain deep learning network for sparse-view ct reconstruction. Computer methods and programs in biomedicine **256**, 108376 (2024). https://doi.org/10.1016/j.cmpb.2024.108376
48. Tolooshams, B., Lin, L., Callier, T., Wang, J., Pal, S., Chandrashekar, A., Rabut, C., Li, Z., Blagden, C., Norman, S.L., et al.: Vars-fusi: Variable sampling for fast and efficient functional ultrasound imaging using neural operators. bioRxiv pp. 2025–04 (2025). https://doi.org/10.1101/2025.04.16.649237
49. Wang, C., Shang, K., Zhang, H., Li, Q., Zhou, S.K.: Dudotrans: dual-domain transformer for sparse-view ct reconstruction. In: International workshop on machine learning for medical image reconstruction. pp. 84–94. Springer (2022). https://doi.org/10.1007/978-3-031-17247-2_9
50. Wang, C., Zhang, H., Li, Q., Shang, K., Lyu, Y., Dong, B., Zhou, S.K.: Improving generalizability in limited-angle ct reconstruction with sinogram extrapolation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 86–96. Springer (2021). https://doi.org/10.1007/978-3-030-87231-1_9
51. Wang, C., Berner, J., Li, Z., Zhou, D., Wang, J., Bae, J., Anandkumar, A.: Coarse graining with neural operators for simulating chaotic systems (2025). https://doi.org/10.48550/arXiv.2408.05177, https://arxiv.org/abs/2408.05177
52. Wang, H., Li, Y., Zhang, H., Chen, J., Ma, K., Meng, D., Zheng, Y.: Indudonet: An interpretable dual domain network for ct metal artifact reduction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 107–118. Springer (2021). https://doi.org/10.1007/978-3-030-87231-1_11
53. Wang, J., Aborahama, Y., Khokhar, A., Zhang, Y., Wang, C., Sastry, K., Berner, J., Luo, Y., Bonev, B., Li, Z., Azizzadenesheli, K., Wang, L.V., Anandkumar, A.: Accelerating 3d photoacoustic computed tomography with end-to-end physics-aware neural operators (2025). https://doi.org/10.48550/arXiv.2509.09894, https://arxiv.org/abs/2509.09894
54. Wang, J., Ostras, O., Sode, M., Tolooshams, B., Li, Z., Azizzadenesheli, K., Pinton, G.F., Anandkumar, A.: Ultrasound lung aeration map via physics-aware neural operators. ArXiv pp. arXiv–2501 (2025). https://doi.org/10.48550/arXiv.2501.01157
55. Wu, J., Lin, J., Jiang, X., Zheng, W., Zhong, L., Pang, Y., Meng, H., Li, Z.: Dual-domain deep prior guided sparse-view ct reconstruction with multi-scale fusion attention. Scientific Reports **15**(1), 16894 (2025). https://doi.org/10.1038/s41598-025-02133-5
56. Xia, W., Yang, Z., Zhou, Q., Lu, Z., Wang, Z., Zhang, Y.: A transformer-based iterative reconstruction model for sparse-view ct reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 790–800. Springer (2022). https://doi.org/10.1007/978-3-031-16446-0_75

57. Ye, D.H., Buzzard, G.T., Ruby, M., Bouman, C.A.: Deep back projection for sparse-view ct reconstruction. In: 2018 ieee global conference on signal and information processing (globalsip). pp. 1–5. IEEE (2018). https://doi.org/10.1109/GlobalSIP.2018.8646669
58. Yu, Z., Wen, X., Yang, Y.: Reconstruction of sparse-view x-ray computed tomography based on adaptive total variation minimization. Micromachines **14**(12), 2245 (2023). https://doi.org/10.3390/mi14122245
59. 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**(10), 5765–5781 (2021). https://doi.org/10.1002/mp.15183
60. Zhang, Z., Yu, L., Liang, X., Zhao, W., Xing, L.: Transct: dual-path transformer for low dose computed tomography. In: International conference on medical image computing and computer-assisted intervention. pp. 55–64. Springer (2021). https://doi.org/10.1007/978-3-030-87231-1_6
61. Zhou, B., Chen, X., Xie, H., Zhou, S.K., Duncan, J.S., Liu, C.: 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**(12), 3587–3599 (2022). https://doi.org/10.1109/TMI.2022.3189759
62. Zhou, B., Chen, X., Zhou, S.K., Duncan, J.S., Liu, C.: Dudodr-net: Dual-domain data consistent recurrent network for simultaneous sparse view and metal artifact reduction in computed tomography. Medical Image Analysis **75**, 102289 (2022). https://doi.org/10.1016/j.media.2021.102289
63. Zhou, C., Sun, Y., Liang, J., Liu, J., Liu, Q.: Sparse-view ct image reconstruction using conditional embedding fusion diffusion model. Neurocomputing p. 131748 (2025). https://doi.org/10.1016/j.neucom.2025.131748

Resolution-Agnostic Neural Operators for Multi-Rate Sparse-View CT

– Supplementary Material –

Aujasvit Datta[1,3†] Jiayun Wang[1,2†] Asad Aali[4] Anima Anandkumar[1]

[1]Caltech [2]Georgia Tech [3]IIT Kanpur [4]Stanford University

aujasvitd22@iitk.ac.in, {peterw,anima}@caltech.edu, asadaali@stanford.edu

[†]Equal contribution

In this supplementary material, we provide details omitted in the main text, including:

- Sec. A: more mathematical details of the proposed rotational equivariant DISCO. (Sec. 4.1 "Sinogram-Space Operator NO_s" of the main paper.)
- Sec. B: additional details on experimental setup, specifically sinogram sampling setting, and implementation details of baseline methods (Sec. 5.2 "Reconstruction with Different Sampling Rates" of the main paper.)
- Sec. C: more experimental results and visualizations of the method, including numerical results/tables that cannot fit in the main text due to page limit, as well as additional reconstruction visualizations of multiple methods under different sampling rates. (Sec. 5 "Experimental Results" of the main paper.)
- Sec. D: complexity analysis of the proposed method to existing neural architectures, specifically other neural operator architectures and CNNs. (Sec. 5.4 "Analysis and Ablation Study" of the main paper.)

A Rotational Equivariant DISCO

We describe the rotational equivariant DISCO [34, 38] design details below.

Rotation–equivariant convolution in (r, θ). Parallel-beam CT induces two key symmetries on the sinogram $\mathbf{p}(\theta, r)$: (i) a rotation of the image by ϕ is a *shift in view angle* of the sinogram,

$$(\mathbf{p} \circ R_\phi)(\theta, r) \;=\; \mathbf{p}(\theta - \phi, r), \tag{7}$$

and (ii) the π-periodicity-with-flip identity

$$\mathbf{p}(\theta + \pi, r) \;=\; \mathbf{p}(\theta, -r). \tag{8}$$

Define the group action of in-plane rotations T_ϕ on sinogram functions by

$$(T_\phi f)(\theta, r) := f(\theta - \phi, r), \tag{9}$$

with the identification

$$f(\theta + \pi, r) \equiv f(\theta, -r). \tag{10}$$

Let $\star$ denote a stationary convolution on (θ, r) (implemented by DISCO with quadrature),

$$(f \star \psi)(\theta, r) \;=\; \int_{\mathbb{R}} \int_{0}^{\pi} f(\theta - \tau,\, r - \rho)\, \psi(\tau, \rho)\, d\tau\, d\rho, \tag{11}$$

where the domain obeys Eqn. (8). Then $\star$ is *SO(2)-equivariant* with respect to the action Eqn. (9):

$$\big((T_\phi f) \star \psi\big)(\theta, r) = \int\!\!\int f(\theta - \phi - \tau,\, r - \rho)\, \psi(\tau, \rho)\, d\tau\, d\rho \tag{12}$$

$$= (f \star \psi)(\theta - \phi, r) \tag{13}$$

$$= \big(T_\phi (f \star \psi)\big)(\theta, r). \tag{14}$$

In the discrete DISCO implementation, (12) holds up to the usual quadrature error.

Padding to realize the symmetry. Eqns. (7)–(12) require boundary conditions consistent with Eqn. (8). We therefore use *flipped circular padding* along θ: when indices cross the $\theta = 0/\pi$ boundary, we wrap the data while flipping r,

$$\mathbf{p}(\theta \pm \pi, r) \;\mapsto\; \mathbf{p}(\theta, -r), \tag{15}$$

and apply standard reflective padding along r. This enforces physical continuity at $\theta = 0$ and $\theta = \pi$, prevents seam artifacts, and makes the DISCO convolutions in NO_s intrinsically rotation-robust: rotations of x become θ-translations of p (Eqn. (7)), which are naturally handled by the translational equivariance of Eqn. (11).

B Implementation Details

B.1 Sinogram Sampling Settings

For both datasets, we generate ground-truth sinograms using 720 projection angles. For the C4KC-KiTS dataset, we use a parallel-beam geometry with a 256×256 image resolution and a detector size of 300. Projection angles are uniformly sampled over $[0, \pi)$, which is standard for parallel-beam CT. For the AAPM Low-Dose CT dataset, we use a fan-beam geometry with a 256×256 resolution and 672 detector elements, with a source-to-detector distance of 1075. Projection angles are uniformly sampled over $[0, 2\pi)$ to match the acquisition geometry of the dataset. In both datasets, we train and evaluate reconstruction performance at 9-view, 18-view, 36-view, and 72-view acquisition settings by subsampling the corresponding number of projection angles from the 720-view ground-truth sinograms.

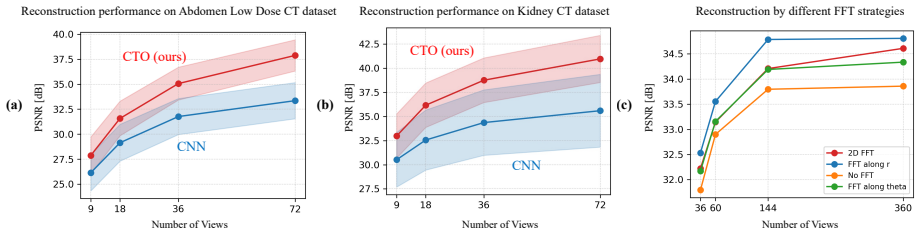

Fig. 6: (a) Reconstruction performance on the AAPM abdomen Low-Dose CT dataset [37] (entire test set). **(b)** Reconstruction performance on the Kidney CT dataset [21] (entire test set). In both datasets, CTO consistently outperforms the CNN baseline across all sampling rates. **(c)** Comparison of reconstruction performance under four frequency-space configurations: 1) 2D FFT: 2D Fourier transform across both axes; 2) FFT along r: 1D Fourier transform along the detector (r) axis; 3) No FFT: no Fourier transform (spatial-domain learning only). 4) FFT along θ: 1D Fourier transform along the angular (θ) axis; Introducing frequency-space processing yields consistent improvements over the spatial-only baseline, with detector-axis transforms providing the largest gains, in line with the analysis in Sec. 4.1.

B.2 Baseline Setup

For fair comparisons, CTO and baseline methods all use the same amount and configurations of training data.

1. **Learning-free iterative methods.** We include two classical reconstruction baselines: Filtered Back-Projection (FBP) [43] and the Simultaneous Algebraic Reconstruction Technique (SART) [4]. FBP reconstruction applies the inverse Radon transform configured as in Sec. B.1 on the subsampled sinograms after passing them through a ramp frequency filter. For SART, we use the public `scikit-image iradon_sart` operator. For each test image, we perform one initial SART sweep followed by four additional refinement iterations (five total). The relaxation parameter is left at the default internal setting of `iradon_sart`. Following standard conventions, all SART updates are clipped to the valid attenuation range [0, 0.549] to avoid divergence.
2. **DuDoTrans**. DuDoTrans [49] is a dual-domain sparse-view CT reconstruction framework that jointly enhances sinograms and images. A Sinogram Restoration Transformer (SRT) models long-range dependencies in the projection domain to restore informative sinograms, which are converted to an intermediate image through a differentiable consistency layer and fused with an initial FBP reconstruction. A Residual Image Reconstruction Module then refines the fused representation to produce the final output. We use the official implementation, training for 100 epochs with Adam (initial learning rate 1e−4) and a step decay (learning rate multiplied by 0.1 after epoch 20).
3. **GloReDi**. GloReDi [28] is an image-domain framework for sparse-view CT that distills global representations (GloRe) from intermediate-view recon-

structions to improve sparse-view quality. The teacher network (intermediate) guides the student (sparse) through EMA-updated weights. We use the official implementation, training for 100 epochs with Adam ($\alpha = 1e-4$).

4. **MDPRNet**. MDPRNet [42] is a multi-stage dual-domain progressive reconstruction network for sparse-view CT. A U-shaped subnetwork first restores the sparse sinogram, which is converted to an intermediate image via a differentiable FBP layer. A second U-shaped subnetwork refines this image, and a final single-scale feature subnetwork (SSFNet) operates at the original resolution to preserve fine spatial details. Cross-stage Feature Adapters (CFA), comprising a Learnable Global Attention Gate and a Supervised Attention Module, regulate information flow between stages. The architecture further employs a Multi-view Synergistic Training Strategy (MSTS) that groups sparse-view settings into ultra-sparse and sparse-view categories, enabling a single model to handle multiple view counts. We use the official implementation of the method.
5. **Learned Primal-Dual (LPD)**. LPD [2] is a learned iterative reconstruction scheme that unrolls the primal-dual hybrid gradient (PDHG) algorithm, replacing the proximal operators with convolutional neural networks. Each unrolled iteration maintains extended primal and dual variables (N_{primal}=N_{dual}=5 channels) and applies a three-layer residual CNN (with 3×3 convolutions, 32 hidden channels, and PReLU activations) as the learned proximal in both primal and dual spaces. The forward operator and its adjoint are evaluated once per iteration, connecting the two domains. The algorithm works directly from raw projection data with zero-initialization, requiring no initial FBP reconstruction. We use the official implementation with I=10 unrolled iterations (60 total convolutional layers, $\sim 2.4 \times 10^5$ parameters), trained for 10^5 batches using Adam (initial learning rate 1e−3 with cosine annealing, β_2=0.99) and an MSE loss with gradient norm clipping set to 1.
6. **LEARN**. LEARN [11] is an unrolled iterative reconstruction network that casts the fields-of-experts (FoE) regularized optimization into a deep feed-forward architecture. Each iteration block comprises a data-fidelity gradient step $\lambda^t A^\top (Ax^t - y)$ and a three-layer CNN acting as the learned regularizer, connected via a residual shortcut. All regularization terms and balancing parameters are iteration-dependent and learned end-to-end from projection–image pairs. The network uses 48 filters of size 5×5 in the first two convolutional layers and a single output filter per iteration, with the FBP result as initialization. We use the official implementation with N_t=30 unrolled iterations (90 total layers), trained with the Adam optimizer (initial learning rate 1e−4, decayed to 1e−5) using an accumulated MSE loss.
7. **RegFormer**. RegFormer [56] is a transformer-based unrolled iterative reconstruction network that learns both local and nonlocal regularization for sparse-view CT. The gradient descent iteration is unrolled into alternating local and nonlocal iteration blocks: local blocks use three CNN layers with 5×5 kernels for artifact suppression, while nonlocal blocks employ two successive Swin Transformer blocks to capture long-range dependencies for structure preservation. An FBP operator replaces the standard back-projection in the

data-fidelity step to accelerate convergence within the fixed iteration budget. An iteration transmission (IT) module passes intermediate feature maps between consecutive blocks via skip connections, enabling deeper feature extraction across iterations. We use the official implementation with N_t=18 iteration blocks and C=96 feature channels, trained for 200 epochs with AdamW (initial learning rate 1e−4, decayed to 0).

8. **Unrolled networks with CNNs.** We use an unrolled U-Net [41] baseline composed of 3 image space cascades, where each cascade consists of a U-Net refinement module followed by a data consistency update. This baseline operates entirely in the image space and does not apply any learned processing in the sinogram domain. The U-Net uses a standard encoder–decoder configuration with 32 hidden channels and four pooling levels. The model is trained under the same loss function and optimization settings as our method to ensure a fair comparison.
9. **Diffusion models**. Diffusion probabilistic models (DPMs) have been shown as robust methods for solving ill-posed inverse problems. In such inverse problems, we aim to recover x_0 from corrupted measurements y, requiring sampling from the posterior $p(x_0|y)$. This stochastic process is modeled by a stochastic differential equation (SDE) [26, 46]: $dx = -2\dot{\sigma}(t)(\mathbb{E}[x_0|x_t, y] - x_t)dt + g(t)dw$. Because computing likelihood $p(y|x_t)$ is computationally intractable, we use approximations, such as Diffusion Posterior Sampling (DPS) [16] and Annealed Langevin Dynamics (ALD) [24], that leverage DPMs as priors for accelerated reconstruction and approximate $p(y|x_t)$ as $p(y|x_0 = \mathbb{E}[x_0|x_t])$. For the diffusion baseline, we train a DPM following the EDM [26] loss formulation. The model is trained for $2,000$ iterations, employing a UNet-style architecture comprising 65 million trainable parameters, where the learning rate, noise schedule, and training were set following [26]. For inference, we utilize a $1,000$-step noise schedule (data-consistency weighting parameter $\gamma = 5e - 3$), and use the trained DPM to perform reconstruction across both approximation methods: (i) DPS [16] and (ii) ALD [24].

B.3 Hardware and Training.

The training of our model and baselines is with a batch size of 8 across 4 A100 (40GB) GPUs.

C Additional Results and Visualizations

C.1 Kidney CT Dataset Results

Subset results. Due to the long inference time and computation of diffusion methods, we report all methods (including diffusion ones) on a 500-slice subset.
Full-set results. We also provide results of all methods excluding diffusion ones on the entire test set of 10,000 images for the Kidney CT dataset [21] in Table 4. We plot the numerical results in Fig. 6b.
We also plot results on AAPM dataset in Fig. 6a.

Table 3: Sparse-view CT reconstruction results (including diffusion models) for smaller test set of 500 images from the kidney CT dataset [21]. Results on entire test set of 10,000 images are provided in Table 4. Lower is better for RMSE; higher is better for PSNR/SSIM.

Category	Method	18-view			36-view			72-view		
		RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑	RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑	RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑
Learning-free	FBP [43]	1541.70 ± 190.59	6.74 ± 1.45	5.00 ± 3.00	1502.42 ± 185.69	6.96 ± 1.45	8.78 ± 2.81	1423.67 ± 175.85	7.43 ± 1.45	16.54 ± 2.48
	SART [4]	678.26 ± 140.11	14.13 ± 3.09	48.93 ± 6.60	678.86 ± 144.47	14.16 ± 3.27	50.93 ± 7.61	675.58 ± 145.43	14.22 ± 3.35	51.00 ± 8.84
Diffusion	DPS [16]	81.67 ± 21.72	32.50 ± 2.45	88.03 ± 4.10	47.25 ± 8.73	37.08 ± 1.80	94.38 ± 1.63	41.95 ± 6.99	38.08 ± 1.70	95.34 ± 1.53
	ALD [24]	92.26 ± 16.09	31.25 ± 1.72	85.58 ± 3.15	53.21 ± 8.97	36.02 ± 1.63	93.27 ± 1.80	43.28 ± 6.70	37.79 ± 1.59	95.27 ± 1.53
Single pass	DuDoTrans [49]	111.86 ± 25.87	29.62 ± 2.07	84.64 ± 5.19	64.69 ± 16.01	34.42 ± 2.18	90.94 ± 3.85	48.03 ± 13.06	37.04 ± 2.37	94.49 ± 3.19
	GloReDi [28]	108.36 ± 84.92	30.76 ± 3.86	85.28 ± 7.42	94.30 ± 86.62	32.19 ± 4.09	87.76 ± 7.11	84.60 ± 88.21	33.37 ± 4.32	90.01 ± 7.15
	MDPRNet [42]	119.52 ± 27.41	27.57 ± 2.43	84.97 ± 6.44	82.76 ± 20.17	33.42 ± 1.96	87.47 ± 6.41	65.43 ± 6.79	34.67 ± 2.64	89.27 ± 4.46
Unrolled Networks	Learned PD [2]	140.94 ± 20.47	27.52 ± 1.58	71.66 ± 6.40	106.43 ± 15.78	29.96 ± 1.60	77.51 ± 4.94	76.17 ± 13.49	32.90 ± 1.77	86.49 ± 3.45
	LEARN [11]	122.01 ± 25.14	28.86 ± 1.83	83.25 ± 4.22	86.27 ± 17.76	31.87 ± 1.85	88.33 ± 3.25	62.22 ± 14.49	34.75 ± 2.02	92.52 ± 2.45
	RegFormer [56]	119.03 ± 84.37	29.86 ± 3.82	84.47 ± 7.24	88.62 ± 87.77	32.94 ± 4.38	89.08 ± 6.46	69.71 ± 59.65	35.68 ± 4.90	92.56 ± 6.01
	Unrolled CNN [3]	82.52 ± 41.09	32.66 ± 2.82	89.54 ± 3.33	67.67 ± 38.30	34.53 ± 3.11	91.92 ± 3.12	60.65 ± 43.93	35.82 ± 3.64	93.84 ± 2.95
	CTO (ours)	**53.50 ± 13.55**	**36.11 ± 2.25**	**92.56 ± 2.50**	**39.67 ± 9.59**	**38.68 ± 2.20**	**94.85 ± 1.85**	**31.07 ± 8.11**	**40.83 ± 2.36**	**96.42 ± 1.55**

Table 4: Sparse-view CT reconstruction results for Kidney CT dataset [21] for the larger test set of 10,000 images. Diffusion models are omitted due to high computational costs. Lower is better for RMSE; higher is better for PSNR/SSIM.

Category	Method	18-view			36-view			72-view		
		RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑	RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑	RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑
Learning-free	FBP [43]	1539.09 ± 198.54	6.83 ± 1.89	5.20 ± 3.46	1499.88 ± 193.44	7.05 ± 1.89	8.97 ± 3.27	1421.26 ± 183.22	7.52 ± 1.89	16.72 ± 2.94
	SART [4]	675.36 ± 142.01	14.24 ± 3.30	49.08 ± 7.84	675.91 ± 146.27	14.27 ± 3.47	51.08 ± 7.94	672.58 ± 147.23	14.33 ± 3.54	51.19 ± 9.16
Single pass	DuDoTrans [49]	112.77 ± 27.57	29.63 ± 2.14	84.61 ± 5.32	65.34 ± 17.20	34.42 ± 2.25	90.88 ± 3.99	48.52 ± 14.22	37.05 ± 2.40	94.41 ± 3.26
	GloReDi [28]	111.95 ± 92.06	30.68 ± 4.04	85.08 ± 7.81	97.92 ± 93.81	32.10 ± 4.28	87.57 ± 7.51	88.50 ± 95.46	33.24 ± 4.52	89.74 ± 7.54
	MDPRNet [42]	117.46 ± 25.57	27.80 ± 2.49	84.98 ± 8.28	81.43 ± 19.73	33.37 ± 2.04	88.16 ± 6.14	66.13 ± 6.29	35.04 ± 2.78	89.42 ± 3.97
Unrolled Networks	Learned PD [2]	141.94 ± 21.43	27.53 ± 1.69	71.66 ± 6.60	107.34 ± 16.87	29.97 ± 1.68	77.47 ± 5.17	76.71 ± 14.24	32.92 ± 1.83	86.45 ± 3.59
	LEARN [11]	123.22 ± 26.74	28.85 ± 1.91	83.22 ± 4.35	87.28 ± 20.80	31.87 ± 1.93	88.30 ± 3.36	63.15 ± 18.87	34.75 ± 2.04	92.48 ± 2.50
	RegFormer [56]	123.05 ± 91.37	29.76 ± 3.97	84.29 ± 7.45	92.43 ± 85.41	32.84 ± 4.57	88.88 ± 6.77	73.55 ± 68.62	35.55 ± 5.14	92.34 ± 6.35
	Unrolled CNN [3]	82.51 ± 48.28	32.56 ± 3.11	89.48 ± 3.50	70.74 ± 45.33	34.37 ± 3.39	91.82 ± 3.35	63.13 ± 45.85	35.60 ± 3.77	93.71 ± 3.26
	CTO (ours)	**53.60 ± 13.81**	**36.17 ± 2.30**	**92.63 ± 2.55**	**39.71 ± 10.05**	**38.76 ± 2.31**	**94.93 ± 1.88**	**30.85 ± 8.08**	**40.97 ± 2.44**	**96.49 ± 1.50**

C.2 Rotational Equivariance Results

We perform an ablation study to evaluate the impact of circular padding on reconstruction quality. As discussed in Sec. 4.1, circular padding enforces the periodic boundary conditions inherent to the sinogram's angular domain and enables the network to better exploit rotational equivariance. Table 5 shows the reconstruction performance with and without circular padding under a 24-view subsampling setting. Incorporating circular padding leads to a notable improvement in PSNR, indicating that handling angular periodicity explicitly enhances the overall reconstruction quality.

Table 5: Ablation study demonstrating the effect of the proposed rotational equivariant DISCO (via circular padding) on reconstruction quality.

Method	PSNR on 24-view subsampled reconstruction (dB)
CTO	**33.00**
CTO (no Rot. Eqv.)	32.17

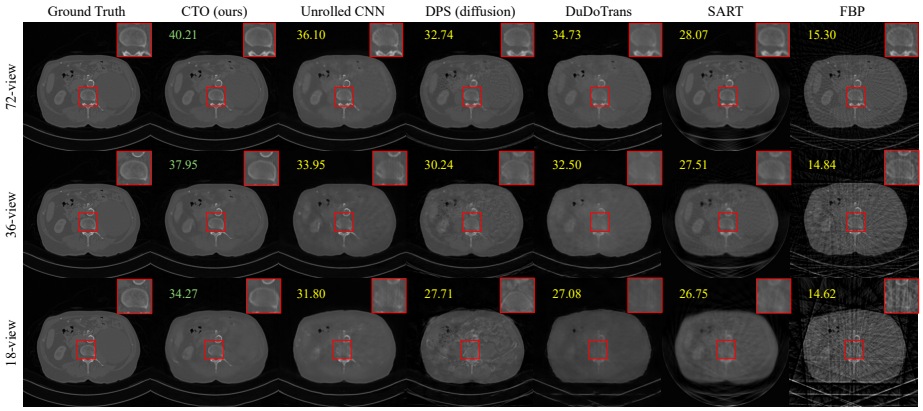

Fig. 7: Reconstruction on multiple sampling rates for Abdomen Low-Dose CT dataset [37]. CTO outperforms baseline methods for PSNR (upper left) and reconstruction quality across all sinogram sampling rates.

C.3 Sinogram Space NO_s Design Variants

To determine the most effective configuration for frequency-space learning, we evaluate different Fourier transform strategies for the sinogram prior to operator learning by $\mathrm{NO}_{s,\mathrm{freq}}$. Specifically, we consider four variants: 1) **Detector-axis transform:** one-dimensional Fourier transform applied along the detector (r) axis. 2) **Angle-axis transform:** one-dimensional Fourier transform applied along the angular (θ) axis. 3) **2D transform:** two-dimensional Fourier transform applied along both detector and angular axes. 4) **Baseline (no transform):** no frequency-space learning is applied, and the network operates only in the spatial domain of the sinogram and image.
The comparative results of these configurations are summarized in Supplementary-Fig. 6. We observe that incorporating frequency-space learning in any form leads to a consistent boost in reconstruction quality compared to purely spatial-domain learning (at least 1 dB PSNR gain). Among the variants, applying the Fourier transform along the detector axis provides the largest performance gain (2.7 dB PSNR gain), consistent with the theoretical advantages of this strategy as detailed in Sec. 4.1.

C.4 DISCO Hyperparameter Analysis

Hyperparameter optimization was performed using Bayesian search over six configuration parameters: basis type (piecewise linear, Morlet, Zernike), number of basis functions (`num_basis_func`), number of bases (`num_basis`), and radial cutoff (`radius_cutoff`). Parameter importance was estimated using a random forest regressor trained on the swept configurations with validation PSNR as the target metric, with feature importances serving as a proxy for predictive

Table 6: Hyperparameter importance and correlation with validation PSNR. Importance scores were computed by training a random forest regressor with the swept hyperparameters as input features and validation PSNR as the target, reporting each feature's importance as a measure of its predictive relevance. Correlation values report the Pearson linear correlation between each hyperparameter and validation PSNR, ranging from -1 to 1. Importance captures nonlinear and interaction effects between hyperparameters, whereas correlation isolates linear associations of individual parameters.

Config Parameter	Importance	Correlation
`basis_type`: piecewise linear	0.429	0.468
`num_basis_func`	0.278	0.032
`num_basis`	0.145	-0.176
`radius_cutoff`	0.139	0.214
`basis_type`: Morlet	0.008	-0.412
`basis_type`: Zernike	0.002	-0.196

relevance. Linear association between each parameter and validation PSNR was additionally quantified via Pearson correlation.

As summarized in Table 6, `basis_type.piecewise_linear` emerged as the most important predictor of reconstruction quality, followed by `num_basis_func` and `num_basis`. While Morlet and Zernike basis types showed negligible importance scores, both exhibited moderate negative correlation with validation PSNR, indicating that their selection is actively detrimental rather than merely uninformative. These findings guided our final model configuration, with piecewise linear basis functions selected as the default.

C.5 Out-of-Distribution Generalization

Generalization on noisy sinogram data. We evaluate our zero-noise trained CTO and unrolled CNN on sinogram data with added Poisson noise ($I_0 = 10^6$). On average over different subsampling rates (9, 18, 36, 72-view), CTO outperforms unrolled CNN by 3.06 dB PSNR. This shows that CTO generalizes better to noisy conditions observed in real clinical settings.

Generalization across different datasets. To evaluate the ability of our model to generalize across different datasets, we test the CTO and unrolled CNN models trained on C4KC-KiTS dataset on the AAPM dataset. We observe that CTO performs substantially better than unrolled CNN with an average gain of 9.8 dB PSNR across different subsampling rates.

C.6 9-View AAPM Test Results

We provide test results for 9-view SVCT reconstruction on the AAPM dataset [37] in Table 7.

Table 7: 9-view CT reconstruction results for AAPM dataset [37]. Lower is better for RMSE; higher is better for PSNR/SSIM. For all learning methods, a single model is trained on multi-view sinograms with $N_v = 9, 18, 36, 72$ views together. For fairness, all methods are trained using the same amount of data – this differs from the resolution-specific models trained in their original paper since multi-resolution training generally reduces overfitting (Table 2a). Test results for the rest of the rates are reported in main text Table 1.

Category	**Method**	**9-view**		
		RMSE (HU)↓	PSNR↑	SSIM (x10^{-2}) ↑
Learning-free	FBP [43]	640.64 ± 42.73	13.03 ± 1.63	33.29 ± 6.09
	SART [4]	681.87 ± 130.27	14.07 ± 2.88	45.31 ± 6.46
Single pass	DuDoTrans [49]	350.61 ± 18.22	18.26 ± 1.60	48.64 ± 4.91
	GloReDi [28]	159.11 ± 86.01	27.12 ± 3.45	78.30 ± 8.07
	MDPRNet [42]	346.89 ± 22.17	20.81 ± 1.67	64.57 ± 6.40
Unrolled Networks	Learned PD [2]	203.95 ± 21.92	22.99 ± 1.73	39.87 ± 5.11
	LEARN [11]	205.77 ± 25.54	22.94 ± 1.73	61.49 ± 4.29
	RegFormer [56]	196.114 ± 23.68	23.35 ± 1.79	65.52 ± 4.84
	Unrolled CNN [3]	142.45 ± 20.73	26.14 ± 1.82	68.99 ± 8.37
	CTO (ours)	**116.69 ± 15.86**	**27.87 ± 1.85**	**74.35 ± 4.98**

C.7 Multi-Rate Visualization

Fig. 7 shows reconstructions on the AAPM abdomen low-dose CT dataset [37] across multiple sampling rates. CTO preserves reconstruction quality even under heavy undersampling with noticeably fewer artifacts compared to other methods.

C.8 Single Rate training and inference

To further analyze the performance of our model, we also train/test per rate as an ablation. This differs from our default setting, where a single model is trained jointly across all sampling rates and must generalize across them at test time; here, instead, a separate model is trained and evaluated separately on each fixed rate (18, 36, or 72 views), which is the default setting most other methods are designed and evaluated on. Under this setting, CTO still outperforms all baselines at every view count (Table 8), and the margin widens further under our default multi-rate regime.

C.9 LIDC-IDRI Lung CT Dataset Results

We also compare CTO and Unrolled CNN on the LIDC-IDRI dataset [5] under the same multi-rate training protocol as the main paper (a single model trained jointly across 18-, 36-, and 72-view rates). LIDC-IDRI is a publicly available lung CT benchmark of 1,018 thoracic scans from 1,010 patients. Results are in Table 9. CTO outperforms Unrolled CNN across all rates, with gains of +1.54 dB, +2.47 dB, and +3.29 dB PSNR at 18-, 36-, and 72-view respectively.

D Complexity Analysis

D.1 Computational Complexity of Neural Operators

We analyze the per-layer computational cost of four neural operators on an $N \times N$ grid with c hidden channels. FNO [30] retains $k_{\max}$ Fourier modes per spatial axis for its spectral convolution. DISCO [34] uses a compactly supported kernel with stencil size s (the number of spatial neighbors within the support radius). DeepONet [35] parameterizes its branch and trunk MLPs with hidden width d, basis rank p, and depth L. LocalNO [34] augments FNO with two parallel local branches—a differential kernel and a DISCO convolution—summed pointwise per layer. For a fair comparison, we set $c = 64$, $N = 256$, $k_{\max} = 256$ (no spectral truncation, giving FNO maximum expressivity), $s = 9$ (a 3×3 local stencil), $d = 128$, $p = 64$, and $L = 4$.

Why FNO is insufficient for imaging tasks. Although FNO is the cheapest operator ($1\times$), it is equivalent to a global convolution in the spatial domain: every output pixel depends equally on all input pixels. As demonstrated in [34], this global mixing severely degrades performance on tasks that require spatially localized processing, such as image reconstruction, where preserving edges, textures, and fine anatomical structures is critical. Global operators over-smooth these local features, producing blurry reconstructions. Image reconstruction fundamentally requires a *local integral operator*—one whose kernel has compact support so that each output is determined by a local neighborhood of the input. FNO lacks this property, making it unsuitable despite its low cost.

DISCO is the most efficient neural local operator. For tasks requiring local support, we compare DISCO against two alternatives. DeepONet [35] lacks explicit local spatial structure and costs $65\times$ more than FNO, making it impractical for high-resolution imaging. LocalNO [34], the current state-of-the-art local neural operator, combines FNO with parallel differential and DISCO branches, achieving strong accuracy but at $9\times$ FNO's cost. Since DISCO is already the local component within LocalNO, it represents the minimal unit for introducing

Table 8: Single-rate vs. multi-rate training comparison (PSNR↑, dB). Multi-rate is the setting we study: a single model is trained jointly across all sampling rates and must generalize across them at test time, whereas single-rate trains and tests a separate model per rate. CTO outperforms all baselines under both protocols, and the margin of improvement widens in the Multi-rate training setting. The results for multi-rate training are same as those in Table 1

	Single rate training			**Multi rate training**		
Method (PSNR↑)	**18-view**	**36-view**	**72-view**	**18-view**	**36-view**	**72-view**
LEARN	27.97	30.76	32.32	25.51	27.53	29.90
RegFormer	29.47	32.18	35.11	25.67	28.00	30.43
Unrolled CNN	32.35	35.63	38.35	29.14	31.76	33.35
CTO (ours)	**33.94**	**36.26**	**38.83**	**31.57**	**35.06**	**37.88**

Table 9: Sparse-view CT reconstruction on the LIDC-IDRI lung CT dataset [5]. Higher is better for PSNR / SSIM.

Method	18-view (PSNR / SSIM)	36-view (PSNR / SSIM)	72-view (PSNR / SSIM)
Unrolled CNN	33.89 / 0.889	35.11 / 0.914	36.38 / 0.941
CTO	**35.43** / **0.901**	**37.58** / **0.932**	**39.67** / **0.960**

local inductive bias—at just 4× FNO's cost, which is 2.25× cheaper than LocalNO. Furthermore, as shown in Sec. D.2, DISCO achieves the same inference FLOPs as standard CNNs, since its continuous kernel is pre-discretized at test time and reduces to a regular convolution. This makes CTO competitive in efficiency with CNN-based methods while retaining the resolution-agnostic benefits of neural operators.

D.2 Run Time Comparison with CNNs

Since CTO employs only a single neural operator layer in each of NO_s and NO_i, its computational cost closely matches that of a CNN with equivalent depth. A neural operator layer with DISCO convolution has the same asymptotic FLOP count as a standard convolutional layer with identical channel dimensions, as both scale as $\mathcal{O}(C_{\text{in}} \cdot C_{\text{out}} \cdot H \cdot W)$. The only additional cost arises during training from repeated kernel sampling; at inference, the discretized filter is pre-computed and fixed, making DISCO and CNN identical in runtime.

Training. During training, the continuous integration kernel must be sampled onto the discretized grid at each forward pass. At 256×256 resolution, a single DISCO layer requires 4.68 MFLOPs compared to 3.28 MFLOPs for the equivalent CNN layer—an increase of 42%. At 512×512, the corresponding figures are 63.45 MFLOPs vs 52.63 MFLOPs, a 20.5% increase. The relative overhead diminishes as spatial dimensions grow, since the kernel sampling cost is fixed while the convolution cost scales with the spatial size.

Inference. During inference, the discretized filter is pre-computed and fixed for the given input resolution—the continuous kernel does not need to be sampled again. The operation then reduces to a standard convolution, making DISCO and CNN identical in inference runtime. In Table 11, we compare GFLOPs/forward pass and parameter count across the Unrolled CNN baseline, a CTO variant with

Table 10: Per-layer computational complexity of neural operators at channel $c_{\text{in}} = c_{\text{out}} = c = 64$, image feature map size $N = 256$.

Operator	Complexity per layer	FLOPs	Relative
FNO [30]	$\mathcal{O}(c^2 N^2 + c\, N^2 \log N)$	6.0×10^8	1×
DISCO [34]	$\mathcal{O}(c^2\, s\, N^2)$	2.4×10^9	4×
LocalNO [34]	$\mathcal{O}(c^2\, s\, N^2 + c^2 N^2 + c\, N^2 \log N)$	5.4×10^9	9×
DeepONet [35]	$\mathcal{O}\big(N^2(c\, d + L\, d^2 + d\, p\, c)\big)$	3.9×10^{10}	65×

Table 11: Computational cost vs. accuracy. We compare GFLOPs per forward pass, parameter count, and 18-view PSNR for the Unrolled CNN baseline, a CTO variant using the identical architecture as the CNN ablation (*Identical no. of param ablation*), and our full CTO model. At matched configuration, CTO and CNN have identical FLOPs and parameter count, confirming that the continuous parameterization introduces no inference-time overhead. Our full model instead tunes its kernel configuration for optimal performance, incurring a modest 1.3× FLOPs and 1.1× parameter overhead while improving PSNR by 2.43 dB over the baseline.

Method	GFLOPS/forward pass	No. of parameters	18-view PSNR
Unrolled CNN	36.33 (1×)	23.1 M (1×)	29.14
CTO (Identical no. of param ablation)	36.33 (1×)	23.1M (1×)	31.11
CTO	47.53 (1.3×)	25.8 M (1.1×)	31.57

the identical architecture to the CNN ablation, and our original CTO model. At matched configuration, CTO and CNN have exactly the same FLOPs and parameter count, confirming that the continuous parameterization introduces no inference-time overhead. The full CTO model, whose kernel configuration is instead tuned for optimal performance, incurs a modest 1.3× FLOPs and 1.1× parameter overhead relative to the baseline, while improving 18-view PSNR from 29.14 to 31.57 dB.