NOIR: NEURAL OPERATOR MAPPING FOR IMPLICIT REPRESENTATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper presents NOIR, a framework that reframes core medical imaging tasks as operator learning between continuous function spaces, challenging the prevailing paradigm of discrete grid-based deep learning. Instead of operating on fixed pixel or voxel grids, NOIR embeds discrete medical signals into shared Implicit Neural Representations and learns a Neural Operator that maps between their latent modulations, enabling resolution-independent function-to-function transformations. We evaluate NOIR across multiple 2D and 3D downstream tasks, including segmentation, shape completion, image-to-image translation, and image synthesis, on five datasets and compare its performance against convolutional, attention-based, and neural-operator baselines. NOIR achieves competitive performance at native resolution while exhibiting strong robustness to unseen discretizations, further providing empirical validation of the ϵ-ReNO Bartolucci et al. (2023) property of neural operators through its ability to generalize effectively across resolutions and discretizations not encountered during training.

1 INTRODUCTION

Medical image computing is a specialized subfield of computer vision that aims to computationally interpret medical images acquired from modalities such as X-ray, CT, MRI, and ultrasound Xiang (2025), thereby supporting many clinical workflows like diagnosis, treatment planning and guidance. Core tasks in this domain include segmentation Gao et al. (2025), which delineates anatomical structures; registration Darzi & Bocklitz (2024), which aligns images across time, modalities, or subjects; shape completion Friedrich et al. (2023), which estimates missing or occluded anatomical structures from partial medical observations, and image-to-image translations Chen et al. (2025), which map images between different domains, aiming for synthesis or quality enhancement.

With recent advances in deep learning, these tasks are now largely solved, or at least extensively exploited, using data-driven approaches Liu et al. (2021); El hadramy et al. (2023). Deep neural networks have become the dominant tools in medical image computing, achieving state-of-the-art performance across a wide range of benchmarks Suganyadevi et al. (2022). Despite their success, these methods share a common and often implicit assumption: medical images are represented and processed as discrete grid-based signals. Indeed, in the prevailing deep learning framework, medical images are represented as pixels or voxels arranged on regular lattices and fed directly into neural networks Suganyadevi et al. (2022). Despite its effectiveness, this representation raises a fundamental question about its appropriateness for medical image computing tasks. This question is particularly relevant given that **medical images originate from inherently continuous physical domains** but are frequently resampled, interpolated, and transformed during clinical workflows, making discrete grid-based neural network models sensitive to voxel spacing, interpolation, and resampling artifacts. A continuous formulation would enable resolution-independent representations, improving stability under geometric transformations and across heterogeneous acquisition protocols.

In this paper, we challenge the prevailing deep learning paradigm in medical image computing that learns over discrete image grids and introduce **NOIR**: **N**eural **O**perator mapping for **I**mplicit **R**epresentations, a general framework that reframes medical image computing downstream tasks as a continuous operator learning problem. NOIR is not intended to outperform every task-specific, heavily tuned architecture at its native resolution; rather, it is a unifying formulation, applicable

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

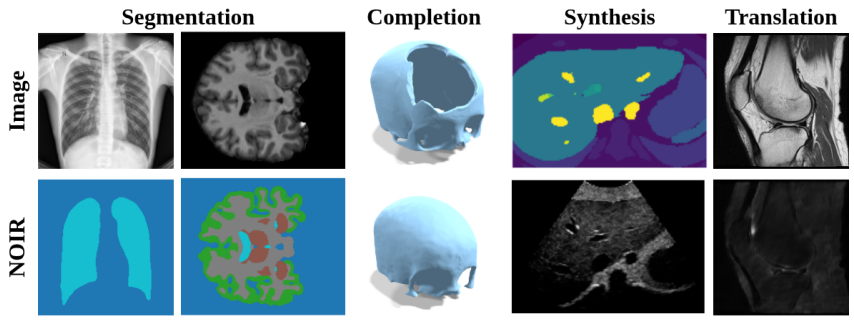


Figure 1: **Example of downstream tasks solved by NOIR.** By framing a given task as operator learning between function spaces, NOIR tackles supervised 2D and 3D medical image analysis problems and is inherently robust to resolution changes.

across multiple downstream tasks, whose distinguishing property is a controlled, bounded degradation under discretization changes that grid-based and spectral models do not offer.

Contributions. We (i) challenge the standard grid-based deep learning framework in medical image analysis by questioning the choice of image representation, (ii) formulate downstream medical image computing tasks as Neural Operator (NO) learning problems mapping Implicit Neural Representations (INRs), rather than as discrete grid-based prediction tasks, (iii) demonstrate the applicability of NOIR to core medical imaging tasks on several datasets in both 2D and 3D (Figure 1) and validate it against grid-based, attention-based, and neural-operator baselines, and (iv) empirically validate that NOIR fulfills the fundamental criteria of NOs Kovachki et al. (2021) and approximates a Representation equivalent NO (ε-ReNO) Bartolucci et al. (2023).

2 Related work

Representations in medical image analysis. In medical image computing, state-of-the-art approaches predominantly rely on convolutional and attention-based architectures: segmentation on U-Net variants Ronneberger et al. (2015); Fabian et al. (2018); Chen et al. (2021); Hu et al. (2021); Oktay et al. (2018) and Vision Transformers Dosovitskiy et al. (2020); image-to-image translation on generative and diffusion models Cai et al. (2023); Baldini et al. (2023); Ho et al. (2020); Friedrich et al. (2024b); Durrer et al. (2025); and deformable registration on dense displacement or velocity fields Balakrishnan et al. (2018); Chen et al. (2021); Tian et al. (2023). These methods process medical signals as discrete grids of pixels or voxels. Several alternative methods avoid learning from fixed discretizations. In MRI, downstream tasks can be performed directly in k-space, without explicit image reconstruction Rempe et al. (2024a); Schlemper et al. (2017); Li et al. (2024a;b); Rempe et al. (2024b); Friedrich et al. (2024a), and similar signal-domain learning has been explored on raw acquisition data in CT and ultrasound Zhu et al. (2018); Han & Ye (2018); Khan et al. (2020). Functional map representations have likewise been used for 3D-3D registration, mapping functions between shapes for smooth, resolution-independent alternatives to voxel-based registration Litany et al. (2017); Li & Li (2025). Together, these approaches highlight viable continuous-domain alternatives to grid-based representations, providing a natural context for INRs.

Modeling Images with Implicit Neural Representations. INRs Sitzmann et al. (2020b) model images, shapes, and physical fields as continuous functions via coordinate-based networks, enabling resolution-independent representations of data Genova et al. (2019); Molaei et al. (2023); Sitzmann et al. (2020b). Subsequent work improved their expressiveness and stability through sinusoidal activations Sitzmann et al. (2020b), Fourier features Tancik et al. (2020), and multiresolution hash encodings Müller et al. (2022), enabling state-of-the-art neural rendering Mildenhall et al. (2020), 3D reconstruction Fan & Musialski (2024), and signal compression Dupont et al. (2022b). Because standard INRs are trained per instance, recent work improves generalizability through dataset-level learning: hypernetworks that condition an INR on a signal representation Ha et al. (2016); Sitzmann et al. (2020a), auto-decoder formulations that learn a shared INR with instance-specific latent codes Park et al. (2019), and meta-learning approaches that fit new instances with only a few gradient

steps Dupont et al. (2022b); Friedrich et al. (2025); Kazerouni et al. (2025). Despite these advances, existing INR-based methods primarily represent individual signals Vyas et al. (2025); Hadramy et al. (2026) rather than the task-specific mappings between them, a gap that motivates NOs for learning mappings between continuous functions.

Neural Operators. NOs learn mappings between infinite-dimensional function spaces by operating directly on functions rather than specific discretizations Kovachki et al. (2021), and have been extensively applied to PDEs and physics-based problems through architectures such as DeepONet Lu et al. (2019), FNO Li et al. (2020b), CNO Raonić et al., and GNO Li et al. (2020a). FNO, in particular, relies on spectral convolutions in Fourier space and is therefore restricted to regular grids and complex or irregular geometries; CORAL Serrano et al. (2023) addresses this by leveraging INRs within an operator-learning framework for resolution-invariant mappings. These NO methods have been extensively validated on canonical PDEs (Burgers', Darcy flow, Navier-Stokes, elasticity), where they consistently surpass classical networks. However, **NOs remain sparsely explored in medical image computing, limited to FNO-based methods** such as FNO3DSeg Wong et al. (2023) and FNOReg Drozdov & Sorokin (2025), which inherit FNO's grid restriction and aliasing errors on high-frequency features Serrano et al. (2023).

Our motivation. These three directions expose complementary gaps. Grid-based architectures couple predictions to a fixed discretization, making them brittle to the resampling that pervades clinical pipelines. INR-based methods decouple signals from their discretization but have so far represented individual signals rather than task-specific mappings between them, e.g., from an image to its segmentation. NOs target function-to-function mappings with resolution-invariance guarantees, but **existing medical applications are confined to grid-restricted, aliasing-prone FNO variants.** NOIR closes this gap by using INRs to obtain a discretization-free representation of each signal, and casting the downstream task itself as a mapping between the resulting function spaces, inheriting NO resolution-invariance guarantees while remaining applicable to arbitrary medical imaging tasks.

3 METHODS

3.1 PROBLEM FORMULATION

Let $\mathcal{X} \subset \mathbb{R}^d$ denote a continuous spatial domain. We consider **supervised learning** of mappings between continuous functions defined on $\mathcal{X}$. Given a dataset $\mathcal{D} = \{(f_i^d, g_i^d)\}_{i=1}^N$, each input–output pair consists of discontinuous representations: $f_i^d : \mathcal{X} \to \mathbb{R}^{c_{\text{in}}}$ and $g_i^d : \mathcal{X} \to \mathbb{R}^{c_{\text{out}}}$, where f_i^d represents an input discrete signal and g_i^d denotes a target discrete signal associated with a downstream task. We first seek continuous representations, denoted $\{f_i, g_i\}$, of any given sample $\{f_i^d, g_i^d\}$ of the dataset $\mathcal{D}$, embedding them into suitable continuous function spaces on $\mathcal{X}$, denoted $\mathcal{F}$ and $\mathcal{G}$ respectively. Subsequently, we aim to learn an operator: $\mathcal{T} : \mathcal{F} \to \mathcal{G}$, that maps between these function spaces such that for any given $i < N$, $\mathcal{L}(\mathcal{T}(f_i), g_i) \approx 0$, where $\mathcal{L}$ is a criterion related to the downstream task. No assumptions are made regarding the semantic nature of g_i or the specific form of $\mathcal{L}$, allowing the formulation to remain task-agnostic.

Our proposed framework, NOIR, addresses this problem by learning mappings between function spaces (Fig. 2). NOIR contains two main blocks, described below: **first**, the continuous function representation for discrete medical signals, constructing implicit representations of the sampled input (blue) and output (red) functions; **second**, the NO mapping that learns to transform between the resulting function spaces.

3.2 CONTINUOUS FUNCTION REPRESENTATION

Given a discrete signal $f_i^d \in \mathcal{D}$, we seek a continuous function $f_i \in \mathcal{F}$ such that $f_{i|\mathcal{X}} \approx f_i^d$. Following Dupont et al. (2022a); Friedrich et al. (2025); Serrano et al. (2023), we model f_i with an INR $\phi_i(x; \theta, \gamma_i)$, where θ are dataset-level shared parameters and γ_i are signal-specific parameters predicted from a latent $z_i \in \mathbb{R}^{p_{\text{in}}}$ via a hypernetwork $\gamma_i = M_\psi(z_i)$, rather than being optimized directly. These parameters modulate the shared weights θ, adapting a dataset-level representation to each sample: θ captures common content (background, anatomy, acquisition characteristics), while γ_i encodes signal-specific deviations such as anatomical variation or pathology. The resulting continuous approximation is $f_i(x) = \phi_i\big(x; \theta, M_\psi(z_i)\big) \approx f_i^d(x)$. Training follows a meta-learning

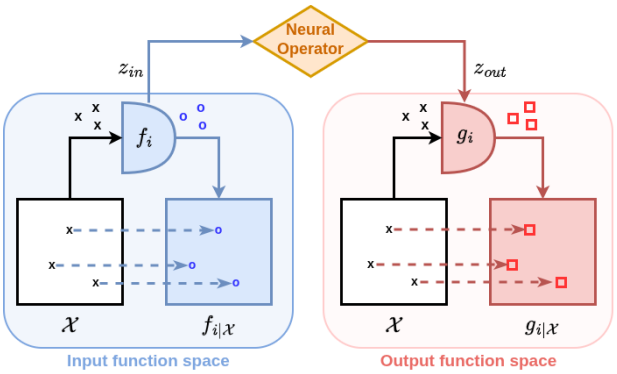


Figure 2: **Architecture of NOIR**. Discrete input (f_i^d) and output (g_i^d) signals are embedded as continuous functions ($f_{i|\mathcal{X}}$ and $g_{i|\mathcal{X}}$) using INRs with shared dataset-level parameters and signal-specific modulations. A neural operator maps the modulations of input (z_{in}) and output (z_{out}) INRs, resulting in a continuous function-to-function mapping applicable to multiple downstream tasks.

scheme (Algorithm 1 in the Appendix A.2): the shared parameters (θ, ψ) are optimized in an outer loop across all training signals, while for each signal f_i^d an inner loop optimizes a latent z_i (initialized to zero) through K gradient steps minimizing a similarity loss $\mathcal{L}_i$ between $\phi(x; \theta, M_\psi(z_i))$ and f_i^d, before a single outer gradient step updates (θ, ψ). At test time (Algorithm 2 in the Appendix A.2), (θ, ψ) are frozen and only z^* is optimized through the same K-step inner loop, yielding $\phi(x; \theta, M_\psi(z^*))$ without retraining the shared parameters.

3.3 Neural Operator Mapping

Following Section 3.1, we operate on a dataset $\mathcal{D} = \{(f_i^d, g_i^d)\}_{i=1}^N$ of paired discrete signals. We apply the continuous function approximation above to both signals of each pair, obtaining latent representations $z_{\text{in}}^i, z_{\text{out}}^i$ encoding f_i^d, g_i^d (via Algorithm 2 in the Appendix A.2), and learn a NO $\hat{z}_{\text{out}} = \mathcal{T}(z_{\text{in}})$, where $\mathcal{T}$ is any data-driven parametric model. We instantiate $\mathcal{T}$ as an MLP with residual connections, sufficient to model this transformation, supervised by $\mathcal{L}_{\text{NO}} = \frac{1}{N}\sum_{i=1}^N \left\|\mathcal{T}(z_{\text{in}}^i) - z_{\text{out}}^i\right\|_2^2$.

3.4 NOIR is an ϵ-ReNO

Intuitively, a NO should give the same answer regardless of the grid used to discretize its input: segmenting an image resampled to 64^2 or to 200^2 should recover the same underlying continuous prediction, up to unavoidable information loss. When this is not the case, the discrepancy is called *aliasing error*, and it is what causes grid-based and spectral models to degrade when evaluated away from their training resolution Kovachki et al. (2021). Following the framework of Bartolucci et al. (2023), we define a *Representation equivalent Neural Operator* (ReNO) as a neural operator $(\mathcal{T}, \mathcal{T}^d)$ for which the aliasing error $\varepsilon(\mathcal{T}, \mathcal{T}^d, \Psi, \Phi) = 0$ for all admissible frame sequence pairs (Ψ, Φ), i.e., for all discretizations, where Ψ, Φ are frame sequences for $\mathcal{F}, \mathcal{G}$ with synthesis operators T_Ψ, T_Φ (mapping discrete coefficients to continuous functions) and analysis operators $T_\Psi^\dagger, T_\Phi^\dagger$ (the reverse). The aliasing error operator measures the discrepancy between $\mathcal{T}$ and its discretization $\mathcal{T}^d$ after synthesis and analysis:

$$\varepsilon(\mathcal{T}, \mathcal{T}^d, \Psi, \Phi) = \mathcal{T} - T_\Phi \circ \mathcal{T}^d \circ T_\Psi^\dagger, \tag{1}$$

and when $\|\varepsilon(\mathcal{T}, \mathcal{T}^d)\| \leq \epsilon$ for a small constant ϵ, the operator is an ϵ-ReNO: predictions made at two different discretizations then provably differ by at most a bounded amount. In practice, this property matters because acquisition protocols in medical imaging vary across scanners, sites, and clinical constraints, so a model trained at one resolution routinely encounters inputs at another at deployment time. An ϵ-ReNO guarantees that the discrepancy between predictions at two discretizations stays bounded.

Empirical evidence that NOIR is an ϵ-ReNO. As one of the main contributions of NOIR is its resolution-independent formulation, we first empirically validate its ϵ-ReNO Bartolucci et al. (2023)

property before evaluating downstream tasks. We train NOIR on a full-resolution dataset of X-ray images Jaeger et al. (2014) and evaluate the resulting model across five input discretizations, 32^2, 64^2, 128^2, 160^2, and 200^2. For each test sample and resolution r, we optimize the input latent $z_{\text{in}}^{(r)}$ from the observed samples, predict $\hat{z}_{\text{out}}^{(r)} = \mathcal{T}^d(z_{\text{in}}^{(r)})$, and reconstruct the output at its native resolution.

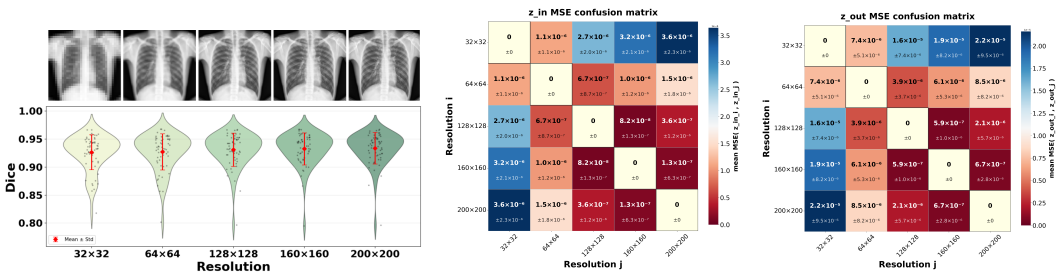
Formally, when $\|\varepsilon(\mathcal{T}, \mathcal{T}^d)\| \leq \epsilon$, the representation-equivalence error between two discrete operators $\mathcal{T}^d$ and $\mathcal{T}'^d$ associated with different frame sequences satisfies

$$\|\tau(\mathcal{T}^d, \mathcal{T}'^d)\| \leq 2\epsilon\sqrt{B_\Psi}\sqrt{A_\Phi}, \tag{2}$$

where A_Φ and B_Ψ are the corresponding frame bounds and

$$\tau(\mathcal{T}^d, \mathcal{T}'^d) = \mathcal{T}^d - T_\Phi^\dagger \circ T_{\Phi'} \circ \mathcal{T}'^d \circ T_{\Psi'}^\dagger \circ T_\Psi. \tag{3}$$

Thus, ϵ bounds the discrepancy between discrete operators associated with different discretizations, providing the theoretical basis for resolution-independent predictions. In NOIR, the frame sequences Ψ and Φ correspond to spatial discretizations $\mathcal{X}^{(r)} \subset \mathcal{X}$ of the input and output signals. The synthesis operators T_Ψ and T_Φ map latent modulations to continuous functions through the input and output INRs, while $T_\Psi^\dagger$ and $T_\Phi^\dagger$ recover the corresponding latent representations through the inner optimization procedure. The continuous operator $\mathcal{T} : \mathcal{F} \to \mathcal{G}$ is therefore approximated by its discrete latent counterpart $\mathcal{T}^d : \mathbb{R}^{p_{\text{in}}} \to \mathbb{R}^{p_{\text{out}}}$ through the pipeline $T_\Phi \circ \mathcal{T}^d \circ T_\Psi^\dagger$. In this setting, $\epsilon_{z_{\text{in}}}$ and $\epsilon_{z_{\text{out}}}$ quantify the maximum pairwise discrepancy between the input and output latent representations across discretizations, respectively, while ϵ_{seg} measures the corresponding discrepancy at the reconstructed segmentation level. Empirically, the pairwise MSE heatmaps in Fig. 3 show that both optimized input and predicted output modulations remain highly consistent across the five discretizations. Specifically, Fig. 3b reports pairwise MSE between the optimized input modulations $z_{\text{in}}^{(r)}$, while Fig. 3c reports the corresponding MSE between the predicted output modulations $\hat{z}_{\text{out}}^{(r)}$. Across all resolution pairs and test samples, we obtain $\epsilon_{z_{\text{in}}} = 4.0 \times 10^{-6}$ and $\epsilon_{z_{\text{out}}} = 2.2 \times 10^{-5}$. At the segmentation level, the maximum MSE between reconstructed predictions and ground truth is $\epsilon_{\text{seg}} = 1.8 \times 10^{-2}$. The resulting Dice distributions remain stable across resolutions, with mean Dice around 0.93–0.94 (Fig. 3a). These small ϵ values provide empirical evidence that NOIR preserves the underlying continuous operator across different discretizations.



(a) Resolution robustness of NOIR. Dice distributions across discretizations show stable performance.

(b) Pairwise MSE between optimized input modulation vectors $z_{\text{in}}^{(r)}$ across resolutions.

(c) Pairwise MSE between predicted output modulation vectors $\hat{z}_{\text{out}}^{(r)}$.

Figure 3: **Resolution robustness and latent consistency across discretizations.** (a): Dice distributions across discretizations show stable segmentation performance. (b): pairwise MSE between optimized input modulation vectors obtained at resolutions 32×32, 64×64, 128×128, 160×160, and 200×200. (c): pairwise MSE between predicted output modulation vectors produced by the NO across the same discretizations. The low off-diagonal discrepancies indicate consistency across different discretizations.

Table 1: Performance (DSC ↑) and inference time (s ↓) across segmentation tasks.

Dataset	Res	U-Net	ViT	AttU-Net	FNO	CNO	DiffFNO	Transolver-3	**NOIR**
Shenzhen	10%	0.75	0.91	0.01	0.92	0.72	0.86	0.88	**0.93**
	25%	0.92	0.93	0.74	**0.94**	0.92	0.92	0.93	**0.94**
	100%	**0.95**	0.94	**0.95**	0.94	**0.95**	0.94	0.93	0.94
	Time (ms)	3.1	3.1	4.7	2.9	288	24	190	830
OASIS	10%	0.40	0.66	0.40	0.65	0.53	0.48	0.66	**0.85**
	25%	0.79	0.85	0.76	0.84	**0.86**	0.84	0.83	0.85
	100%	**0.95**	0.90	**0.95**	0.91	**0.95**	0.85	0.94	0.85
	Time (ms)	2.7	5.8	3.6	2.6	261	21	132	240

4 EXPERIMENTS

4.1 DATASETS AND IMPLEMENTATION DETAILS

To demonstrate the effectiveness of NOIR, we evaluated our method on the following medical image computing tasks: segmentation, shape completion, image translation, and synthesis. Experiments were conducted using four public datasets: Shenzhen Jaeger et al. (2014), OASIS Marcus et al. (2007), SkullBreak Kodym et al. (2021), fastMRI Knoll et al. (2020), along with an in-house phantom-acquired ultrasound dataset. Each dataset was divided into training, validation, and test splits, and all reported results were obtained on held-out test sets after model selection on validation performance. Segmentation and anatomy completion were evaluated with the Dice Similarity Coefficient (DSC); translation and synthesis with Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). For each task, we first trained separate INRs for the input and output domains, then trained the NO to map between them. Details about reproducibility are in Appendix A.1.

Baseline methods: As a general, task-agnostic framework, NOIR is evaluated against baselines representing distinct architectural paradigms: U-Net Ronneberger et al. (2015), ViT Dosovitskiy et al. (2020), Attention U-Net (AttU-Net) Oktay et al. (2018), FNO Li et al. (2020b), and DDPM Ho et al. (2020). Since FNO is, to our knowledge, the only neural operator previously applied to medical imaging, we additionally adapt three other operators: CNO Raonić et al., Transolver-3 Zhou et al. (2026), and DiffFNO Liu & Tang (2025). All baselines use vanilla architectures and identical datasets and splits (details in Appendix A.1), enabling a direct comparison of native- and unseen-resolution performance.

4.2 NOIR FOR IMAGE SEGMENTATION

We present two representative settings: binary segmentation on Shenzhen chest X-rays Jaeger et al. (2014) and multi-label segmentation on OASIS brain MRI Marcus et al. (2007). The input INR shares the same architecture in both cases, with modulation vectors of size 64 (Shenzhen) and 2048 (OASIS); see Appendix A.1 for architecture details. Table 1 reports segmentation performance at 10%, 25%, and 100% (full) resolution, and Figure 4 shows qualitative results.

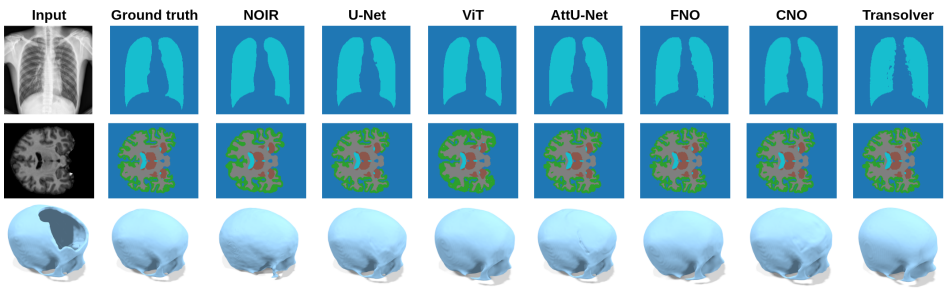


Figure 4: Qualitative results for segmentation and anatomy completion. Rows correspond to Shenzhen binary segmentation, OASIS multi-label segmentation, and SkullBreak anatomy completion, respectively. A test sample is shown with the input, ground truth, NOIR and comparison methods.

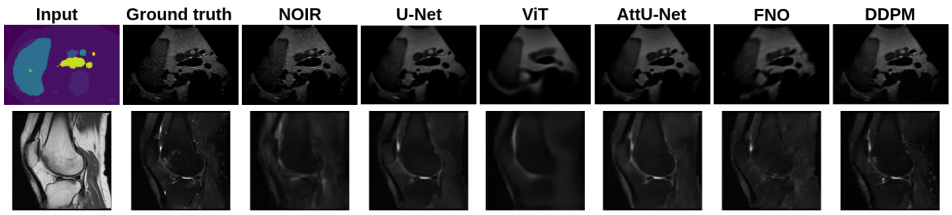


Figure 5: Qualitative results for image synthesis (top row) and image translation (bottom row). A randomly selected test sample is shown, including the input image, ground truth, NOIR results, and results from comparison methods.

4.3 NOIR FOR SHAPE COMPLETION

We evaluated NOIR on SkullBreak Kodym et al. (2021) (512^3 volumes of defective and corresponding complete skulls, split as provided by Kodym et al. (2021)). The input INR is trained on defective volumes and the output INR on complete volumes, both using latent codes of size 64; the NO maps these with a three-layer, 128-unit fully connected network (See Appendix A.1). Table 2 reports quantitative results, and Figure 4 (third row) shows qualitative completions.

Table 2: Performance (DSC ↑) and inference time (s ↓) on anatomy completion.

Dataset	Res	U-Net	ViT	AttU-Net	FNO	CNO	Transolver-3	**NOIR**
SkullBreak	10%	0.71	**0.86**	0.52	0.74	0.67	<u>0.78</u>	**0.86**
	25%	<u>0.90</u>	**0.91**	0.84	0.87	0.87	0.79	0.86
	100%	<u>0.96</u>	0.91	<u>0.96</u>	0.90	**0.98**	0.79	0.87
	Time (s)	<u>0.21</u>	0.11	<u>0.24</u>	0.12	0.51	5.03	0.35

4.4 NOIR FOR IMAGE TRANSLATION AND SYNTHESIS

To demonstrate NOIR's ability to bridge different modalities, we consider two tasks. For translation, fastMRI Knoll et al. (2020) provides paired, registered knee volumes in Proton Density (PD) and T2; we extract 11 central 2D slices per volume from 533 patients (5863 PD/T2 pairs). For synthesis, we use an in-house phantom-acquired ultrasound dataset of 333 paired samples, with tracked-probe pose enabling ground-truth alignment between ultrasound images and CT-derived label maps. Both datasets are split at the patient level (0.8/0.1/0.1). Separate INRs are trained per domain, with modulation size 4096 (fastMRI) and 64 (ultrasound); input/output INRs model PD/T2 for fastMRI and label maps/ultrasound images for synthesis. Table 3 reports quantitative PSNR and SSIM results comparing NOIR with baseline methods, while Figure 5 presents qualitative comparisons.

4.5 COMPUTATIONAL COST

Tables 1 and 2 report the average inference time for NOIR and all baselines; for NOIR this includes the inner-loop latent optimization (once per new sample) plus the operator forward pass. NOIR is one to three orders of magnitude slower than grid-based feed-forward models, since it requires test-time latent optimization (10 inner-loop steps), the price of discretization robustness. Overhead scales with latent size: fastMRI ($dim(z)$=4096) is slowest at 1.58s, while $dim(z)$=64 configurations stay under 1s; this remains compatible with offline or near-real-time clinical use (6–9× faster than DDPM's ∼5s).

4.6 DISENTANGLING INR FIDELITY FROM OPERATOR MAPPING

NOIR combines two components: an output INR that determines how faithfully a signal can be represented, and a neural operator (NO) that maps the input latent representation to the output latent representation. This decomposition raises a central question: which component ultimately limits task performance? In particular, does the latent bottleneck that enables resolution-independent representations also impose a ceiling on full-resolution fidelity? To disentangle these effects, we measure

Table 3: Image-to-image translation performance for domain adaptation and ultrasound synthesis. PSNR and SSIM are reported across different resolutions, along with inference time per sample (ms), averaged over the test set.

Task	Method	Res = 25%		Res = 50%		Res = 100%		t (ms) ↓
		PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	PSNR ↑	SSIM ↑	
US Syn.	U-Net	17.51	0.84	17.45	0.84	17.54	0.89	3.0
	ViT	28.22	0.78	28.31	0.79	28.36	0.79	4.0
	AttU-Net	27.76	**0.87**	28.29	**0.89**	28.55	**0.90**	4.7
	FNO	27.37	0.73	27.94	0.75	28.15	0.76	1.6
	DDPM	26.35	0.65	26.51	0.65	26.91	0.64	270
	NOIR	**30.94**	0.86	**30.94**	0.86	**31.83**	0.87	590
FastMRI	U-Net	21.68	0.60	23.49	0.68	24.55	0.72	1.9
	ViT	22.34	0.62	22.54	0.63	22.61	0.63	2.7
	AttU-Net	21.75	0.60	23.26	0.67	24.46	0.72	2.8
	FNO	22.61	0.58	22.74	0.60	22.82	0.60	1.6
	DDPM	22.66	**0.69**	**25.59**	**0.76**	**25.45**	**0.74**	5270
	NOIR	**22.87**	0.61	22.87	0.61	22.87	0.61	1580

the *fidelity ceiling* imposed by the output INR alone. For each task, we fit the output INR directly to the ground-truth target, bypassing the neural operator, and evaluate the resulting reconstruction using the same task-specific metric. This provides an upper bound on the performance achievable by NOIR with a given latent representation: regardless of the operator, the final prediction must be decoded through the output INR and therefore cannot exceed the fidelity of the representation it can express. Table 4 compares this INR ceiling with NOIR's actual full-resolution performance reported in Tables 1 and 3.

Table 4: Fidelity ceiling of the output INR (fit directly to ground truth, bypassing the operator) versus NOIR's actual end-task performance at full resolution.

	Shenzhen (DSC)	OASIS (DSC)	SkullBreak (DSC)	fastMRI (PSNR)	Ultrasound (PSNR)
Output INR ceiling	0.98	0.91	0.90	29.09	33.48
NOIR (task)	0.94	0.85	0.87	22.87	31.83

The gap to the INR ceiling depends strongly on latent dimensionality. For Shenzhen, OASIS, and SkullBreak, NOIR remains within 2–4 DSC points of the ceiling, indicating that for compact latents ($dim(z) \leq 2048$), the operator recovers most of the available representational capacity and INR fidelity is the primary limitation. In contrast, fastMRI shows a larger gap (29.09 vs. 22.87 PSNR) at $dim(z) = 4096$, where mapping between high-dimensional latent spaces from only a few thousand training pairs makes the operator a more challenging regression problem. We directly investigate whether this gap can be attributed to the choice of operator architecture in the ablation study (Table 5). Although NOIR instantiates $\mathcal{T}$ as a residual MLP, this choice is not intrinsic to the framework: replacing $\mathcal{T}$ leaves T_Ψ, T_Φ, $T_\Psi^\dagger$, and $T_\Phi^\dagger$ unchanged and therefore preserves the underlying function-space mapping. Across classical models, a Transformer, and a conditioned DDPM, however, more expressive operators do not consistently close the gap to the INR ceiling. This suggests that the remaining limitation is primarily associated with the latent dimensionality and data available for learning the latent-to-latent mapping, rather than the specific architecture of $\mathcal{T}$. Thus, latent size should be kept as small as the task's intrinsic complexity permits, while NOIR remains compatible with alternative operators when greater mapping complexity or data scale warrants them.

Table 5: Ablation study on NO model selection (DSC ↑)

Dataset	Lin. Reg.	SVR	Rdm. For.	Transformer	Cond. DDPM	**MLP**
Shenzhen ($dim(z) = 64$)	0.93	**0.94**	0.93	**0.94**	0.87	**0.94**
OASIS ($dim(z) = 2048$)	0.80	0.79	0.80	0.84	0.74	**0.85**

5 DISCUSSION

We emphasize NOIR's positioning before interpreting the numbers: it is a general framework unifying diverse medical imaging tasks under a single continuous formulation, evaluated against distinct architectural paradigms (convolutional, attention-based, spectral, diffusion, and operators), rather than a task-specific architecture optimized to beat the state of the art on any single benchmark. With that scope in mind, across segmentation, shape completion, translation, and synthesis, NOIR consistently performs comparably to or better than established architectures, particularly under resolution shift, while remaining competitive with the strongest baseline at native resolution in most settings. In **segmentation**, NOIR matches the strongest baselines at full resolution while preserving this performance across sub-training resolutions; FNO and, to a lesser extent, DiffFNO are the only other relatively stable baselines, though with a small consistent aliasing error (a pattern that holds on OASIS too: the CNO, DiffFNO, and Transolver-3 baselines all show the same low-resolution degradation as the older grid-based methods, dropping 0.56, 0.37, and 0.48 Dice points, respectively). For **shape completion**, NOIR is competitive at full resolution (3–4% Dice below FNO/ViT) and best at the lowest resolution. It also clearly outperforms Transolver-3, which tends to regress toward the mean shape of the dataset rather than adapting to the individual input. While NOIR produces plausible completions, it struggles to recover fine surface details already present in the input skull (Figure 4); adding skip connections that pass input information directly to the decoder could help preserve these details. For **image synthesis**, NOIR's point-wise INR reconstruction avoids the over-smoothing of convolutional/attention baselines, giving it the best PSNR consistently across resolutions (Figure 5). **Image translation** is the hardest task: NOIR is best at the lowest resolution and comparable to ViT at high resolution, though, as Section 4.6 shows, this is also where it sits furthest from its own fidelity ceiling, and where DDPM's iterative refinement gives it a clear edge in absolute quality. Section 3.4 empirically validates a desirable property of NOs: **NOIR is an ϵ-ReNO**, which FNO and the other operators evaluated here do not satisfy. NOIR's modular, task-agnostic design further means the separately trained input INR, output INR, and NO can, in theory, be reused across downstream tasks, adapting only the NO and one of the INRs.

Limitations. NOIR's stability under resolution change reflects a latent bottleneck that also caps its best-case fidelity. Section 4.6's ceiling analysis says: partially. The bottleneck is not binding on segmentation/completion (compact latents), but does dominate on fastMRI translation (large latents), precisely where NOIR is weakest. Resolution robustness is thus real, not free, trading off against native-resolution fidelity when the latent must be large. We see NOIR, currently, as complementing rather than replacing task-tuned models where diagnosis hinges on native-resolution detail; its advantage is largest under heterogeneous acquisition (e.g., point-of-care ultrasound, multi-site cohorts). These observations highlight the limitations of the current framework. The fidelity of NOIR ultimately depends on the INRs' ability to represent the underlying continuous functions, while increasing latent dimensionality improves representational capacity but makes the operator mapping harder to learn (Section 4.6). NOIR therefore requires the latent size and operator capacity to be chosen jointly for each task. In addition, the current framework relies on supervised paired data, limiting its applicability to tasks such as registration where ground-truth fields are typically unavailable; unsupervised and weakly supervised extensions are important directions for future work.

6 CONCLUSION

We introduced NOIR, a framework to challenge mainstream grid-based deep learning approaches not only conceptually but also empirically. By combining INRs with operator learning, the framework provides competitive performance while introducing additional advantages in resolution robustness, task flexibility, and architectural simplicity, which we have shown to hold against a broader set of recent operator baselines and under an explicit analysis of where this robustness trades off against native-resolution fidelity. Future work will investigate the trade-off between modulation size and training stability of the NO, as larger latent representations improve signal fidelity but increase the complexity of the operator learning problem, and will explore architectural or training strategies (e.g., progressive latent growth, regularized high-dimensional operators) that could raise native-resolution fidelity without sacrificing resolution robustness. Furthermore, we will focus on extending NOIR to unsupervised learning of downstream medical imaging tasks, in particular, deformable image registration.

AI USE STATEMENT

We used generative AI tools to assist with rewording and reformulating parts of the manuscript to improve clarity and readability. All AI-generated content was reviewed and verified by the authors, who take full responsibility for the final manuscript.

REPRODUCIBILITY STATEMENT

Full training configurations (architectures, learning rates, optimizers, number of sampled points, dataset splits) for the INRs and neural operators used in every experiment are provided in Appendix and in the accompanying supplementary material, together with additional runtime measurements and INR reconstruction diagnostics that support the analysis in Sections 4.6 and 4.5. Source code and pretrained weights will be released upon acceptance.

ETHICS STATEMENT

The ultrasound dataset used for the synthesis task was acquired on a tissue-mimicking phantom with a tracked probe under an institutional research protocol, not on human subjects; we refer to it as phantom-acquired throughout to avoid overstating clinical realism. All other datasets (Shenzhen, OASIS-4, SkullBreak, fastMRI) are publicly available and de-identified. This work does not involve any new collection of patient data.

REFERENCES

Guha Balakrishnan, Amy Zhao, Mert R Sabuncu, John Guttag, and Adrian V Dalca. VoxelMorph: A learning framework for deformable medical image registration. September 2018.

Giulia Baldini, Melanie Schmidt, Charlotte Zäske, and Liliana L Caldeira. MRI scan synthesis methods based on clustering and Pix2Pix. December 2023.

Francesca Bartolucci, Emmanuel de Bézenac, Bogdan Raonić, Roberto Molinaro, Siddhartha Mishra, and Rima Alaifari. Representation equivalent neural operators: A framework for alias-free operator learning. May 2023.

Yingchao Cai, Mengxiao Li, Shiqi Liu, and Changhao Zhou. CycleGAN-based image translation from MRI to CT scans. *J. Phys. Conf. Ser.*, 2646(1):012016, December 2023.

Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. TransUNet: Transformers make strong encoders for medical image segmentation. February 2021.

Junxin Chen, Zhiheng Ye, Renlong Zhang, Hao Li, Bo Fang, Li-Bo Zhang, and Wei Wang. Medical image translation with deep learning: Advances, datasets and perspectives. *Med. Image Anal.*, 103(103605):103605, July 2025.

Fatemehzahra Darzi and Thomas Bocklitz. A review of medical image registration for different modalities. *Bioengineering (Basel)*, 11(8):786, August 2024.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. October 2020.

Nikita A Drozdov and Dmitry V Sorokin. FNOReg: Resolution-robust medical image registration method based on fourier neural operator. In *Lecture Notes in Computer Science*, Lecture Notes in Computer Science, pp. 163–177. Springer Nature Switzerland, Cham, 2025.

Emilien Dupont, Hyunjik Kim, S M Ali Eslami, Danilo Rezende, and Dan Rosenbaum. From data to functa: Your data point is a function and you can treat it like one. January 2022a.

Emilien Dupont, Hrushikesh Loya, Milad Alizadeh, Adam Goliński, Yee Whye Teh, and Arnaud Doucet. COIN++: Neural compression across modalities. January 2022b.

Alicia Durrer, Florentin Bieder, Paul Friedrich, Bjoern Menze, Philippe C Cattin, and Florian Kofler. fastWDM3D: Fast and accurate 3D healthy tissue inpainting. July 2025.

Sidaty El hadramy, Juan Verde, Nicolas Padoy, and Stéphane Cotin. Intraoperative CT augmentation for needle-based liver interventions. In *Lecture Notes in Computer Science*, Lecture Notes in Computer Science, pp. 291–301. Springer Nature Switzerland, Cham, 2023.

Isensee Fabian, Petersen Jens, Klein Andre, Zimmerer David, F Jaeger Paul, Kohl Simon, Wasserthal Jakob, Koehler Gregor, Norajitra Tobias, Wirkert Sebastian, and H Maier-Hein Klaus. NnU-Net: Self-adapting framework for U-Net-based medical image segmentation. September 2018.

Shen Fan and Przemyslaw Musialski. Optimizing 3D geometry reconstruction from implicit neural representations. October 2024.

Paul Friedrich, Julia Wolleb, Florentin Bieder, Florian M Thieringer, and Philippe C Cattin. Point cloud diffusion models for automatic implant generation. In *International conference on medical image computing and computer-assisted intervention*, pp. 112–122. Springer, 2023.

Paul Friedrich, Alicia Durrer, Julia Wolleb, and Philippe C Cattin. cwdm: Conditional wavelet diffusion models for cross-modality 3d medical image synthesis. *arXiv preprint arXiv:2411.17203*, 2024a.

Paul Friedrich, Julia Wolleb, Florentin Bieder, Alicia Durrer, and Philippe C Cattin. WDM: 3D wavelet diffusion models for high-resolution medical image synthesis. February 2024b.

Paul Friedrich, Florentin Bieder, Julian McGinnis, Julia Wolleb, Daniel Rueckert, and Philippe C Cattin. Medfuncta: A unified framework for learning efficient medical neural fields. *arXiv preprint arXiv:2502.14401*, 2025.

Yuxiao Gao, Yang Jiang, Yanhong Peng, Fujiang Yuan, Xinyue Zhang, and Jianfeng Wang. Medical image segmentation: A comprehensive review of deep learning-based methods. *Tomography*, 11 (5):52, April 2025.

Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T Freeman, and Thomas Funkhouser. Learning shape templates with structured implicit functions. April 2019.

David Ha, Andrew Dai, and Quoc V Le. HyperNetworks. September 2016.

Sidaty El Hadramy, Oumeymah Cherkaoui, and Philippe C Cattin. cidir: Conditioned implicit neural representation for regularized deformable image registration. In *Lecture Notes in Computer Science*, Lecture Notes in Computer Science, pp. 87–96. Springer Nature Switzerland, Cham, 2026.

Yoseob Han and Jong Chul Ye. Framing u-net via deep convolutional framelets: Application to sparse-view ct. *IEEE transactions on medical imaging*, 37(6):1418–1429, 2018.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. June 2020.

Cao Hu, Wang Yueyue, Chen Joy, Jiang Dongsheng, Zhang Xiaopeng, Tian Qi, and Wang Manning. Swin-Unet: Unet-like pure transformer for medical image segmentation. May 2021.

Stefan Jaeger, Sema Candemir, Sameer Antani, Yì-Xiáng J Wáng, Pu-Xuan Lu, and George Thoma. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quant. Imaging Med. Surg.*, 4(6):475–477, December 2014.

Amirhossein Kazerouni, Soroush Mehraban, Michael Brudno, and Babak Taati. LIFT: Latent implicit functions for task- and data-agnostic encoding. March 2025.

Shujaat Khan, Jaeyoung Huh, and Jong Chul Ye. Adaptive and compressive beamforming using deep learning for medical ultrasound. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 67(8):1558–1572, 2020.

Florian Knoll, Jure Zbontar, Anuroop Sriram, Matthew J Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzalv, Adriana Romero, Michael Rabbat, Pascal Vincent, James Pinkerton, Duo Wang, Nafissa Yakubova, Erich Owens, C Lawrence Zitnick, Michael P Recht, Daniel K Sodickson, and Yvonne W Lui. FastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. *Radiol. Artif. Intell.*, 2 (1):e190007, January 2020.

Oldřich Kodym, Jianning Li, Antonio Pepe, Christina Gsaxner, Sasank Chilamkurthy, Jan Egger, and Michal Španěl. SkullBreak / SkullFix - dataset for automatic cranial implant design and a benchmark for volumetric shape learning tasks. *Data Brief*, 35(106902):106902, April 2021.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces. August 2021.

Ruochen Li, Jiazhen Pan, Youxiang Zhu, Juncheng Ni, and Daniel Rueckert. Classification, regression and segmentation directly from k-space in cardiac mri. In *International Workshop on Machine Learning in Medical Imaging*, pp. 31–41. Springer, 2024a.

Siyu Li and Minqi Li. Fm-pot: non-rigid point cloud registration based on functional maps and partial optimal transport fusion. In *Fifth International Conference on Optical Imaging and Image Processing (ICOIP 2025)*, volume 13688, pp. 389–396. SPIE, 2025.

Yikang Li, Yulong Qi, Zhanli Hu, Ke Zhang, Sen Jia, Lei Zhang, Wenjing Xu, Shuai Shen, Yì Xiáng J Wáng, Zongyang Li, et al. A novel automatic segmentation method directly based on magnetic resonance imaging k-space data for auxiliary diagnosis of glioma. *Quantitative Imaging in Medicine and Surgery*, 14(2):2008–2020, 2024b.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations. March 2020a.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. October 2020b.

Or Litany, Tal Remez, Emanuele Rodola, Alex Bronstein, and Michael Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE international conference on computer vision*, pp. 5659–5667, 2017.

Xiaoqing Liu, Kunlun Gao, Bo Liu, Chengwei Pan, Kongming Liang, Lifeng Yan, Jiechao Ma, Fujin He, Shu Zhang, Siyuan Pan, and Yizhou Yu. Advances in deep learning-based medical image analysis. *Health Data Sci.*, 2021:8786793, May 2021.

Xiaoyi Liu and Hao Tang. DiffFNO: Diffusion fourier neural operator. `https://arxiv.org/abs/2411.09911`, 2025. CVPR 2025.

Lu Lu, Pengzhan Jin, and George Em Karniadakis. DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. October 2019.

Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.*, 19(9):1498–1507, September 2007.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. March 2020.

Amirali Molaei, Amirhossein Aminimehr, Armin Tavakoli, Amirhossein Kazerouni, Bobby Azad, Reza Azad, and Dorit Merhof. Implicit neural representation in medical imaging: A comparative survey. In *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 2373–2383. IEEE, October 2023.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. January 2022.

Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, Ben Glocker, and Daniel Rueckert. Attention U-Net: Learning where to look for the pancreas. 2018.

Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. March 2019.

Bogdan Raonić, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, and Emmanuel De Bézenac. Convolutional neural operators for robust and accurate learning of PDEs. https://proceedings.neurips.cc/paper_files/paper/2023/file/f3c1951b34f7f55ffaecada7fde6bd5a-Paper-Conference.pdf. Accessed: 2026-2-1.

Moritz Rempe, Florian Mentzel, Kelsey L Pomykala, Johannes Haubold, Felix Nensa, Kevin Kroeninger, Jan Egger, and Jens Kleesiek. k-strip: A novel segmentation algorithm in k-space for the application of skull stripping. *Comput. Methods Programs Biomed.*, 243(107912):107912, January 2024a.

Moritz Rempe, Florian Mentzel, Kelsey L Pomykala, Johannes Haubold, Felix Nensa, Kevin Kroeninger, Jan Egger, and Jens Kleesiek. k-strip: A novel segmentation algorithm in k-space for the application of skull stripping. *Computer Methods and Programs in Biomedicine*, 243:107912, 2024b.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. May 2015.

Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging*, 37(2):491–503, 2017.

Louis Serrano, Lise Le Boudec, Armand Kassaï Koupaï, Thomas X Wang, Yuan Yin, Jean-Noël Vittaut, and Patrick Gallinari. Operator learning with neural fields: Tackling PDEs on general geometries. June 2023.

Vincent Sitzmann, Eric R Chan, Richard Tucker, Noah Snavely, and Gordon Wetzstein. MetaSDF: Meta-learning signed distance functions. June 2020a.

Vincent Sitzmann, Julien N P Martel, Alexander W Bergman, David B Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. 2020b.

S Suganyadevi, V Seethalakshmi, and K Balasamy. A review on deep learning in medical image analysis. *Int. J. Multimed. Inf. Retr.*, 11(1):19–38, 2022.

Matthew Tancik, Pratul P Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. June 2020.

Lin Tian, Hastings Greer, François-Xavier Vialard, Roland Kwitt, Raúl San José Estépar, Richard Jarrett Rushmore, Nikolaos Makris, Sylvain Bouix, and Marc Niethammer. Gradicon: Approximate diffeomorphisms via gradient inverse consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18084–18094, 2023.

Kushal Vyas, Ashok Veeraraghavan, and Guha Balakrishnan. Fit pixels, get labels: Meta-learned implicit networks for image segmentation. October 2025.

Ken C L Wong, Wang Hongzhi, and Syeda-Mahmood Tanveer. FNOSeg3D: Resolution-robust 3D image segmentation with fourier neural operator. October 2023.

Bangyin Xiang. A comprehensive review on applications of computer vision in medical imaging. In Chunyi Chen and Tao Lei (eds.), *Fifth International Conference on Computer Vision and Pattern Analysis (ICCPA 2025)*, pp. 11. SPIE, August 2025.

Hang Zhou, Haixu Wu, Haonan Shangguan, Yuezhou Ma, Huikun Weng, Jianmin Wang, and Mingsheng Long. Transolver-3: Scaling up transformer solvers to industrial-scale geometries. *arXiv preprint arXiv:2602.04940*, 2026.

Bo Zhu, Jeremiah Z. Liu, Stephen F. Cauley, Bruce R. Rosen, and Mikhail S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555:487–492, 2018. doi: 10.1038/nature25988.

A Appendix

A.1 Reproducibility Details

To ensure reproducibility of our experiments, we summarize all architectures and training hyperparameters used across datasets in the tables below. Table 6 reports the training configuration for the input implicit neural representations (INRs), Table 7 reports the configuration for the output INRs, and Table 8 reports the configuration for the neural operator (NO) that maps between the input and output INRs. For clarity, we adopt the following abbreviations for column names: **In** denotes the input size, **Out** the output size, **#HL** the number of hidden layers, **HS** the hidden layer size, **HAct** the hidden layer activation function, **LAct** the activation function of the last layer, and **Lat** the size of the latent representation. Additionally, **#HHL** and **HHS** denote the number of hidden layers and the size of hidden layers in the hypernetwork, respectively. **LRL** denotes the learning rate for the latent representation, **LR** the learning rate, **Opt** the optimizer, **WD** the weight decay, and **Dropout** the dropout rate. For the neural operator tables, **HD** refers to the hidden dimension of each layer, **Act** to the activation function, and **Res** indicates whether residual connections are used. The datasets used, their image sizes, splits, and the number of sampled points are summarized in Table 9.

Table 6: Training configuration for input implicit neural representations (INRs).

Dataset	In	Out	#HL	HS	HAct	LAct	Lat	#HHL	HHS	LRL	LR	Opt
Shenzhen	2	1	6	256	`Sine`	`Sigmoid`	64	1	64	10^{-2}	5×10^{-6}	`AdamW`
OASIS-4	2	1	6	256	`Sine`	`Sigmoid`	2048	1	64	10^{-2}	5×10^{-6}	`AdamW`
SkullBreak	3	1	10	256	`Sine`	`Softmax`	64	1	64	10^{-2}	5×10^{-6}	`AdamW`
fastMRI	2	1	10	256	`Sine`	`Sigmoid`	4096	1	64	10^{-2}	5×10^{-6}	`AdamW`
Ultrasound	2	1	10	256	`Sine`	`Sigmoid`	64	1	64	10^{-2}	5×10^{-6}	`AdamW`

Table 7: Training configuration for output implicit neural representations (INRs).

Dataset	In	Out	#HL	HS	HAct	LAct	Lat	#HHL	HHS	LRL	LR	Opt
Shenzhen	2	2	6	256	`Sine`	`Softmax`	64	1	64	10^{-2}	5×10^{-6}	`AdamW`
OASIS-4	2	5	10	256	`Sine`	`Softmax`	2048	1	64	10^{-2}	5×10^{-6}	`AdamW`
SkullBreak	3	2	6	256	`Sine`	`Softmax`	64	1	64	10^{-2}	5×10^{-6}	`AdamW`
fastMRI	2	1	10	256	`Sine`	`Sigmoid`	4096	1	64	10^{-2}	5×10^{-6}	`AdamW`
Ultrasound	2	1	10	256	`Sine`	`Sigmoid`	64	1	64	10^{-2}	5×10^{-6}	`AdamW`

A.2 Learning Implicit Representations

Algorithms 1 and 2 illustrate the process of continuous function approximation, which is performed by training and test-time optimizing an INR as explained in Section 3.2. All INR layers used sine activations except the final layer (sigmoid for the input INR, softmax for the output INR); the output INR differs across tasks only in this final layer's dimensionality. All experiments ran on a single NVIDIA A100 GPU (40 GB). INRs were trained with inner-/outer-loop learning rates of 1×10^{-2}/5×10^{-6}, batch size 1, and 10,000 sampled points per iteration, with 5 inner-loop steps during training and 10 at test time; NOs were trained separately with a learning rate of 5×10^{-6}. All models used early stopping (patience 50, max. 1000 epochs). While the main paper focuses on the performance of NOIR on each dataset and its corresponding task, an important component of the

Table 8: Training configuration for the neural operator (NO) mapping input to output INRs.

	Shenzhen	OASIS-4	SkullBreak	fastMRI	Ultrasound
In	64	2048	64	4096	64
Out	64	2048	64	4096	64
#HL	1	3	1	1	3
HD	128	2048	2048	2048	512
Act	`SiLU`	`SiLU`	`SiLU`	`SiLU`	`ReLU`
Res	`True`	`True`	`True`	`True`	`True`
LR	10^{-4}	10^{-4}	10^{-4}	5×10^{-5}	7×10^{-4}
Opt	`AdamW`	`AdamW`	`AdamW`	`AdamW`	`AdamW`
WD	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-5}
Dropout	0	0.2	0	0.1	0

Table 9: Dataset information including training, validation, and test splits, image sizes, and number of points/slices used for each experiment.

Dataset	Image Size	Train	Val	Test	Num. Points
Shenzhen	200×200	0.8	0.1	0.1	40000
OASIS-4	160×192	0.7	0.15	0.15	30720
SkullBreak	$512 \times 512 \times 512$	provided	provided	provided	40000
fastMRI	128×128	0.8	0.1	0.1	10000
Ultrasound (in-house)	64×64	0.8	0.1	0.1	1820

framework is ensuring that both the input and output INRs accurately learn the underlying signals. In particular, the latent representations optimized for each new sample should enable the INR to provide a faithful functional approximation of the data. To verify this property, we reconstruct the signals directly from the trained INRs and compare the reconstructed images with the corresponding ground truth samples. Table 10 reports the reconstruction quality across datasets in terms of PSNR and SSIM for both the input and output INRs, highlighting that the learned representations preserve the relevant information required to accurately model the signals.

Input: dataset $\{f_i^d\}_{i=1}^N$, θ, M_ψ
  while not converged **do**
    Sample minibatch $\mathcal{B} \subset \{1, \dots, N\}$
    $z_i \leftarrow 0,\ \forall i \in \mathcal{B}$
    for each $i \in \mathcal{B}$ **do**
      for $k = 1, \dots, K$ **do**
        $z_i \leftarrow z_i - \alpha \nabla_{z_i} \mathcal{L}_i$
      end for
      $(\theta, \psi) \leftarrow (\theta, \psi) - \beta \nabla_{(\theta,\psi)} \sum_{i \in \mathcal{B}} \mathcal{L}_i$
    end for
  end while
Return: trained θ, ψ

Algorithm 1: Training

Input: new signal f^*, trained (θ, ψ)
$z^* \leftarrow 0$
  for $k = 1, \dots, K$ **do**
    $z^* \leftarrow z^* - \alpha \nabla_{z^*} \mathcal{L}(f^*, \phi_{\theta,\psi}(z^*))$
  end for
Return: $\phi\big(.; \theta, M_\psi(z^*)\big)$

Algorithm 2: Test-ime optimization

In the following figures, we present qualitative reconstruction results obtained from the output INR, which is responsible for modeling the signals used in the downstream tasks. These visualizations allow us to assess how accurately the INR approximates the target signals by reconstructing them directly from the learned latent representations. Specifically, we show reconstruction examples for the Shenzhen dataset in Figure 6, the OASIS-4 dataset in Figure 7, the fastMRI dataset in Figure 8, and the Ultrasound dataset in Figure 9. In each case, we compare the reconstructed signals obtained from the output INR with the corresponding ground truth samples.

Table 10: Reconstruction performance of the input and output INRs across datasets.

	Input INR			Output INR		
Dataset	PSNR ↑	SSIM ↑	DSC ↑	PSNR ↑	SSIM ↑	DSC ↑
Shenzhen	22.72	0.66	–	–	–	0.98
OASIS-4	24.76	0.79	–	–	–	0.91
SkullBreak	–	–	0.87	–	–	0.90
fastMRI	22.26	0.65	–	29.09	0.71	–
Ultrasound	33.77	0.94	–	33.48	0.70	–



Figure 6: Reconstruction results from the output INR for the Shenzhen dataset. Each example shows the ground truth image and the reconstruction produced by the output INR.

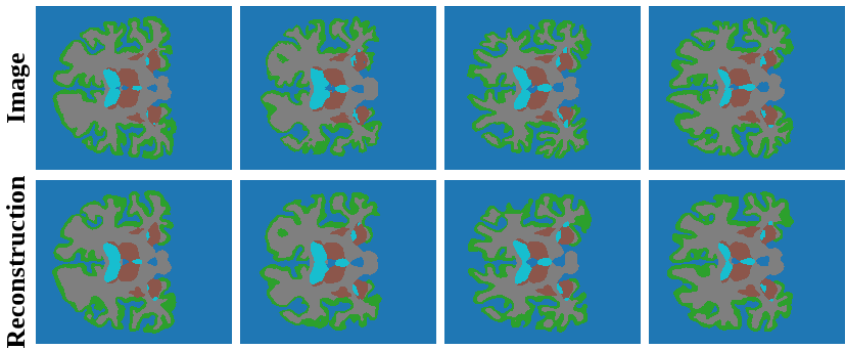


Figure 7: Reconstruction results from the output INR for the OASIS-4 dataset. Ground truth segmentations are compared with the corresponding INR reconstructions.

B NOIR IS AN ϵ-RENO

Figure 10 shows qualitative example demonstrating consistent segmentation predictions across multiple input resolutions. We show a an example from the Shenzhen test set evaluated at different resolutions, illustrating that the NOIR produces consistent segmentation predictions despite changes in input resolution.

C BASELINE METHODS

We provide additional implementation details for the baseline models used in our experiments. All models are implemented in both two-dimensional and three-dimensional variants depending on the dimensionality of the input signals. Inputs are resized to a fixed spatial resolution and normalized to zero mean and unit variance before being provided to the networks.

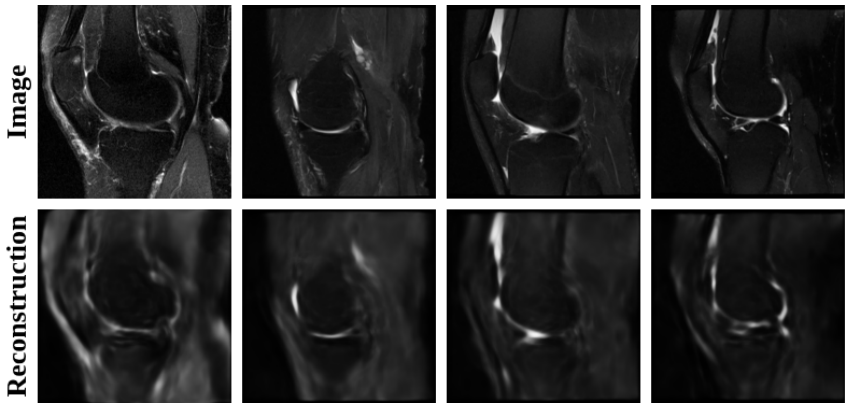


Figure 8: Reconstruction results from the output INR for the fastMRI dataset. The reconstructed images are compared with the ground truth MRI slices.

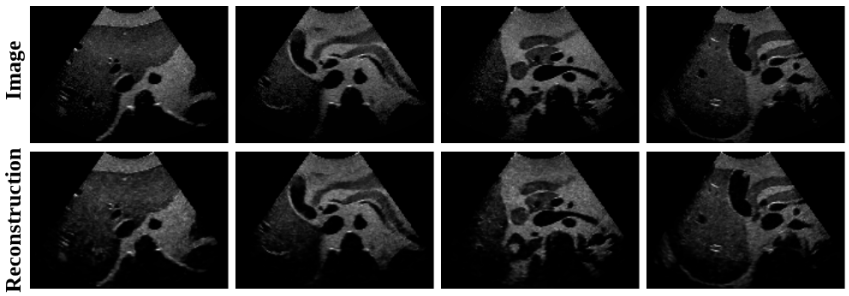


Figure 9: Reconstruction results from the output INR for the ultrasound dataset, showing the ability of the INR to approximate the ultrasound signal from the learned latent representation.

The training objective depends on the type of output signal. For segmentation and shape completion tasks, models are trained using a combination of cross-entropy loss and Dice loss with equal weighting. For image synthesis and image translation tasks, models are trained using a mean squared error (MSE) loss between the predicted output and the ground truth signal. In all experiments, optimization is performed using the `AdamW` optimizer with a learning rate of 10^{-4} and weight decay of 10^{-4}.

U-Net. As a convolutional baseline, we implemented a standard U-Net architecture Ronneberger et al. (2015) following the classical encoder–decoder design with skip connections between corresponding resolution levels. Each encoder block consists of two convolutional layers with kernel size 3×3 (or $3 \times 3 \times 3$ in the 3D case), followed by batch normalization and `ReLU` activations. Downsampling is performed using max-pooling layers. The decoder mirrors the encoder using transposed convolutions (or trilinear upsampling in the 3D case) followed by convolutional blocks. Skip connections concatenate encoder and decoder feature maps at matching resolutions. The network starts with 32 feature channels which are doubled at each downsampling stage, resulting in feature sizes of $32, 64, 128, 256$, and 512 in the deepest layer. In the 3D version used for volumetric reconstruction tasks, the architecture follows the same structure but uses volumetric convolutions and trilinear upsampling, with a base feature size of 16 channels to control memory usage. A final 1×1 (or $1 \times 1 \times 1$) convolution produces the output prediction.

Vision Transformer (ViT). As a transformer-based baseline, we implement a Vision Transformer architecture Dosovitskiy et al. (2020). The input image or volume is first divided into non-overlapping patches using a strided convolutional embedding layer. Each patch is projected to a token embedding space and augmented with learnable positional encodings. The sequence of tokens is then processed by a stack of transformer encoder blocks composed of multi-head self-attention layers and feed-forward multilayer perceptrons with `GELU` activations, using a pre-layer

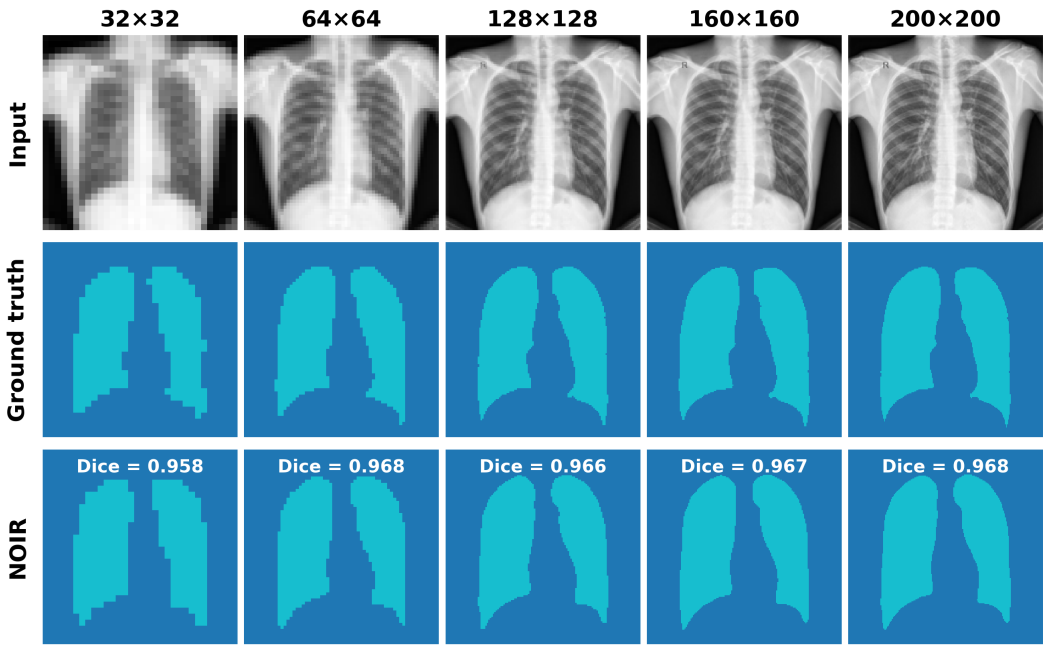


Figure 10: Qualitative example demonstrating consistent segmentation predictions across different input resolutions.

normalization configuration. In the 2D case, the model uses an embedding dimension of 192, 8 transformer layers, and 3 attention heads. In the 3D variant, volumetric patches are extracted and processed using the same transformer architecture. To obtain dense predictions, the token sequence is reshaped into a spatial feature grid and passed through a hierarchical convolutional decoder with progressive upsampling stages. Skip projections from intermediate transformer layers are used to guide the decoder. The decoder uses feature dimensions of $128, 64, 32$, and 16 channels before a final convolutional layer produces the output signal.

Attention U-Net. We implemented an Attention U-Net architecture that extends the standard U-Net by introducing attention gates in the skip connections between encoder and decoder stages Oktay et al. (2018). The encoder and decoder structures follow the same convolutional design as the U-Net baseline. At each resolution level, the feature maps from the encoder are passed through an additive attention mechanism before being concatenated with the decoder features. These attention gates learn spatial attention coefficients that highlight task-relevant regions and suppress irrelevant activations, allowing the network to focus on informative structures in the input. The model uses the same feature scaling strategy as the U-Net baseline, starting from 32 channels in the 2D case and 16 channels in the 3D case, doubling the number of channels at each downsampling stage. In the volumetric implementation, all convolutional operations are replaced with their three-dimensional counterparts.

Fourier Neural Operator (FNO). As an operator-learning baseline, we implement a Fourier Neural Operator architecture Li et al. (2020b). The model follows a lifting–operator–projection design. First, the input signal is lifted to a higher-dimensional latent representation using a 1×1 (or $1\times1\times1$) convolution. The lifted representation is then processed by a stack of Fourier operator blocks that combine spectral and spatial transformations. In each block, the feature maps are transformed to the Fourier domain, where a truncated set of low-frequency modes is retained. A learnable complex weight tensor is applied to these modes before transforming the representation back to the spatial domain using the inverse Fourier transform. In parallel, a residual pointwise convolution operates in the spatial domain to capture local interactions. The outputs of the spectral and spatial paths are combined and passed through a nonlinear activation. In the 2D configuration, we retain 16×16 Fourier modes and use four operator layers with hidden feature dimension 64. In the 3D variant,

volumetric Fourier transforms are applied and $12 \times 12 \times 12$ spectral modes are retained to maintain computational efficiency. The final representation is projected to the desired output space using a small convolutional head.

Denoising Diffusion Probabilistic Model (DDPM). We implement a conditional Denoising Diffusion Probabilistic Model following Ho et al. (2020) for image-to-image translation. The model learns to iteratively denoise a Gaussian-corrupted target image, conditioned on a source image. During training, the target image is progressively noised according to a forward diffusion process defined by a linear variance schedule over $T = 1000$ timesteps. At each training step, a random timestep t is uniformly sampled, the target is noised to that level, and the noised target is channel-wise concatenated with the conditioning image to form a two-channel input. The model is trained to predict the added noise (epsilon parameterization) using a mean squared error loss. The denoising network is a U-Net architecture with sinusoidal timestep embeddings injected via Feature-wise Linear Modulation like scale-and-shift conditioning in each residual block. The encoder–decoder structure uses a channel multiplier schedule of $(1, 2, 4, 8)$ starting from 64 base channels, with two residual blocks per resolution level. Multi-head self-attention with four heads is applied at resolutions $32 \times 32, 16 \times 16$, and 8×8, and additionally at the bottleneck. Down- and upsampling are performed using residual blocks with learned strided or transposed convolutions. The encoder and decoder stages are connected via concatenative skip connections. At inference, sampling starts from pure Gaussian noise and iteratively denoises over 1000 steps using the learned reverse process, concatenating the conditioning image at each step.

Transolver-3. We implement Transolver-3 Zhou et al. (2026), a physics-attention transformer for PDEs on general geometries that treats the domain as an unordered set of coordinate-feature points. Each block applies a slice-based attention: points are aggregated into G learnable "slice tokens" via softmax-weighted assignment, attended over at a cost independent of the point count N, then de-sliced back to point-level features, wrapped in pre-norm residual connections with an MLP. This coordinate-native design needs no architectural change for irregular meshes or non-square image resolutions. **For 2D image** data we simply treat each pixel as a token carrying its normalized 2D coordinates and intensity (e.g. 160×192 OASIS-4 MRI slices, 200×200 Shenzhen X-rays). We use $L = 5$ blocks, hidden width 256, 8 heads, $G = 32$ slice tokens, MLP ratio 1, and `GELU` activations, for approximately 1.8M trainable parameters. **For 3D data** ($256 \times 256 \times 256$ SkullBreak), we sample $N = 32{,}768$ points directly on the anatomy and decode the slices onto the target 3D grid. We use $L = 6$ blocks, hidden width 128, 8 heads, $G = 128$ slice tokens, MLP ratio 1, and `GELU` activations (≈ 1.18M parameters).

CNO. We implement the Convolutional Neural Operator (CNO) Raonić et al. following a U-Net design, with a hierarchical encoder-decoder structure of three downsampling and three upsampling stages connected via concatenative skip connections. In each CNO block, convolutions are combined with an antialiased activation function that interpolates features to double the resolution before applying `LeakyReLU` and downsampling back to the target size; residual blocks with two convolutional layers are applied at each resolution level, with a residual bottleneck block at the deepest layer. Feature channels progressively scale as $8, 16, 32$, and 64. For 2D data, we extend the original square-only formulation to track height and width independently, allowing arbitrary rectangular inputs (e.g. 160×192 MRI slices), using 3×3 convolutions with batch normalization and a lifting/projection block, for approximately 515k trainable parameters. For 3D data, the same design is applied directly on volumes with $3 \times 3 \times 3$ convolutions and instance normalization, resulting in approximately 610k trainable parameters.

DiffFNO. We adapted the deterministic core of DiffFNO as a dual-branch architecture: an EDSR-style convolutional encoder feeds a Weighted Fourier Neural Operator (WFNO) branch, which applies learnable spectral filters with Mode Rebalancing $w(\xi) = 1 + \gamma|\xi|^{\alpha}$ to reweight Fourier modes, in parallel with an Attention Neural Operator (AttnNO) branch using $O(N)$ Galerkin-type linear attention over flattened spatial tokens; a Gated Fusion Mechanism combines the two branches via a learned per-pixel gate before a final projection head. For 2D data, this uses 3×3 convolutions and a 64-channel encoder, giving approximately 8.8M trainable parameters.