DISENTANGLED LEARNING IMPROVES IMPLICIT NEURAL REPRESENTATIONS FOR MEDICAL RECONSTRUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Implicit neural representations (INRs) have emerged as a powerful paradigm for medical imaging via physics-informed unsupervised learning. Classical INRs optimize an entire network from scratch for each subject, leading to inefficient training and suboptimal imaging quality. Recent initialization-based approaches attempt to inject population priors into pre-trained networks, yet they rely on high-quality images and update the prior-bearing parameters during fine-tuning. We present DisINR, a novel INR framework that explicitly disentangles shared and subject-specific representations. DisINR introduces a shared encoder–decoder pair and subject-specific encoders, whose features are jointly decoded for image reconstruction. By integrating differentiable forward models, it pre-trains the shared modules directly from limited raw measurements, removing the need for pre-acquired high-quality images. During test-time adaptation, only the subject-specific encoder is optimized, while the shared pair remains frozen, effectively preserving learned priors. Extensive evaluations on three representative medical imaging tasks show that DisINR significantly outperforms state-of-the-art INRs in both reconstruction accuracy and efficiency. The code is available at: https://anonymous.4open.science/r/DisINR-34E2

1 INTRODUCTION

Medical imaging is a cornerstone of modern clinical practice, enabling the detailed visualization of internal anatomical structures of the human body (Harisinghani et al., 2019; Rubin, 2014). Inverse problems in medical imaging aim to recover anatomical images from raw measurements (*e.g.*, projections in CT or k-space data in MRI). However, due to factors such as undersampling, these problems are inherently ill-posed with multiple suboptimal solutions (Huang et al., 2024; Wang et al., 2023), calling for more advanced reconstruction algorithms.

Recently, implicit neural representations (INRs) have shown great potential in solving medical inverse problems through physics-informed unsupervised learning (Molaei et al., 2023; Luo et al., 2025). INRs model each image as a continuous function parameterized by a coordinate-based neural network. With differentiable forward models (*e.g.*, Fourier transform for MRI), the network can be optimized without using external training data. Owing to the intrinsic bias of neural networks toward structured image patterns (Rahaman et al., 2019; Mildenhall et al., 2021), INRs can resolve high-quality images in an unsupervised way. Yet, existing INR approaches (Wu et al., 2023b; Zha et al., 2022; Feng et al., 2023; Cai et al., 2024; Wu et al., 2025) typically optimize an entire network from scratch for each subject, which leads to low efficiency and limited reconstruction quality.

A promising direction is to learn an effective network initialization from pre-acquired high-quality images. Early studies (Tancik et al., 2021; Lee et al., 2021; Chen & Wang, 2022) employ meta-learning, which performs well but requires a large number of pre-training images. More recently, Vyas et al. (2024) proposed STRAINER, which learns transferable representations via a shared encoder and subject-specific decoders. This approach enables INRs to achieve a strong initialization even from just a few images. However, both types of initialization techniques face two key challenges: ***1) Dependence on high-quality images.*** Such diagnosis-quality images are often difficult to obtain in medical scenarios, particularly for rare diseases; ***2) Preserving population priors.*** Case-specific fine-

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.

tuning updates the prior-bearing parameters, potentially altering the learned population representation. These limitations reduce practical applicability and potential performance improvements. Moreover, existing initialization-based INR techniques are mainly developed for natural RGB images, which may limit their effectiveness on medical volumetric data.

In this work, we propose DisINR, a novel INR framework that explicitly disentangles shared and subject-specific representations. Specifically, DisINR introduces a shared encoder–decoder pair along with subject-specific encoders, whose features are jointly decoded to reconstruct images. By integrating differentiable forward models, DisINR can pre-train the encoder-decoder pair shared across diverse subjects directly from a limited set of raw measurements. This can effectively alleviate the reliance on high-quality images, significantly improving the model's applicability in a wide range of clinical scenarios. During test-time adaptation, we freeze the pre-trained pair, while optimizing only a subject-specific encoder for a given unseen test sample. This design preserves the shared representation while allocating separate capacity for case-specific adaptation, unlike full-network initialization techniques (Tancik et al., 2021; Vyas et al., 2024). Therefore, our method can effectively incorporate learned population priors into INR optimization, enabling faster and higher-quality reconstructions. We evaluate DisINR on three representative medical imaging tasks, including 3D volume fitting, undersampled MRI, and sparse-view CT. Extensive experiments show DisINR substantially outperforms state-of-the-art INRs in both accuracy and efficiency.

2 Background

Medical Inverse Problems The data acquisition process of a medical imaging system can generally be formulated as below: $\mathbf{y} = \boldsymbol{A}\mathbf{x} + \boldsymbol{\epsilon}$, where $\mathbf{y} \in \mathbb{R}^m$ denotes measured data, $\mathbf{x} \in \mathbb{R}^n$ is the desired image, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the system matrix (*i.e.*, forward model), and $\boldsymbol{\epsilon} \in \mathbb{R}^m$ represents system noise. Note that the formulations are presented in the real domain for simplicity, but they also apply to complex-valued data, such as *k*-space measurements in MRI.

Medical inverse problems aim to reconstruct unknown images $\mathbf{x}$ from measurements $\mathbf{y}$. Due to various factors like undersampling (*i.e.*, $m \ll n$), these inverse problems are often highly ill-posed, where multiple suboptimal solutions exist. Conventional model-based algorithms (Beister et al., 2012; Fessler, 2010; Thibault et al., 2007) formulate it as the following optimization problem:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|\boldsymbol{A}\mathbf{x} - \mathbf{y}\|_1 + \lambda \cdot \mathcal{R}(\mathbf{x}), \tag{1}$$

where $\mathbf{x}^*$ denotes the optimal solution, $\mathcal{R}$ is an explicit regularizer (*e.g.*, total variation (Rudin et al., 1992) for image smoothness), and λ is a hyperparameter controlling the contribution of the regularizer. The use of the regularizer can effectively constrain the solution space, thus enabling improved image reconstructions. However, such handcrafted regularizers often fail to capture the complex distribution of medical images, thereby limiting the resulting image quality.

INR for Medical Reconstruction As a signal representation way based on neural networks, implicit neural representations (INRs) have shown great potential in solving medical reconstruction, such as undersampled MRI (Shen et al., 2022; Huang et al., 2023; Feng et al., 2023; Liu et al., 2025; Feng et al., 2025; Wu et al., 2025; Xu et al., 2025; Li et al., 2025) and CT (Sun et al., 2021; Zang et al., 2021; Shen et al., 2022; Zha et al., 2022; Wu et al., 2023b;a; Cai et al., 2024; Du et al., 2024). Technically, INR models an image as a continuous function of coordinates: $\mathbf{c} \in \Omega \rightarrow \mathbf{x}$, and uses a coordinate-based neural network $\boldsymbol{h}_\theta$ to learn this mapping. Formally, INR solves the optimization problem:

$$\theta^* = \arg\min_{\theta} \|\boldsymbol{A}\mathbf{x} - \mathbf{y}\|_1, \quad \text{s.t.} \quad \mathbf{x} = \boldsymbol{h}_\theta(\mathbf{c}), \tag{2}$$

where θ^* denotes the optimized network parameters. Under this formulation, recovering the image $\mathbf{x}$ is transformed into optimizing the parameters θ. Therefore, the desired image can be resolved as $\mathbf{x}^* = \boldsymbol{h}_{\theta^*}(\mathbf{c})$ in an unsupervised manner. The key insight is that neural networks possess an inherent and generalizable inductive bias toward continuous image structures (Rahaman et al., 2019; Mildenhall et al., 2021), which serves as an implicit prior enabling high-quality reconstructions.

Advances for INR Initialization As shown in Eq. (2), classical INRs for medical reconstruction optimize an independent network from random initialization for each subject. To enhance reconstruction efficiency and accuracy, recent works explore INR initialization strategies that generally fall

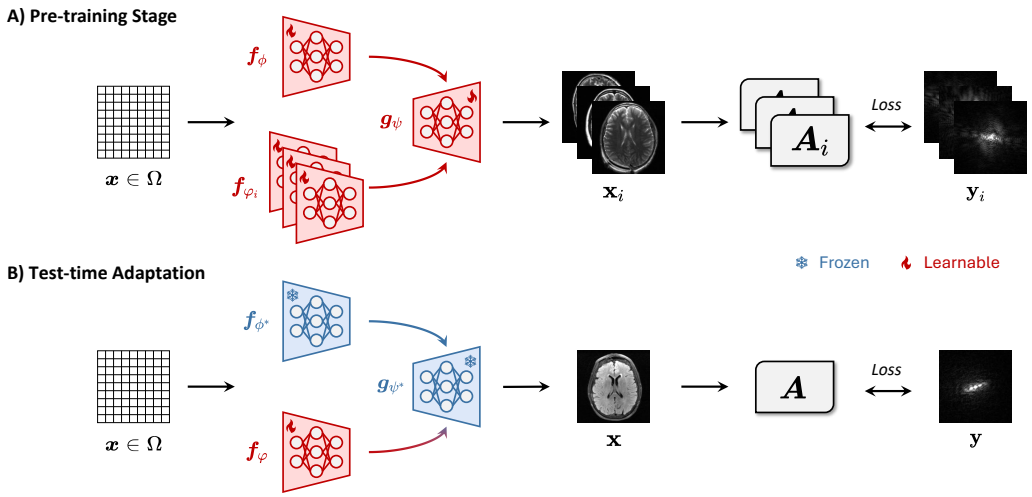

Figure 1: Overview of the proposed DisINR, which consists of a shared encoder–decoder pair $(\boldsymbol{f}_\phi, \boldsymbol{g}_\psi)$ and multiple subject-specific encoders $\{\boldsymbol{f}_{\varphi_i}\}_{i=1}^N$. **A) Pre-training Stage:** All modules in DisINR are jointly optimized from a few raw measurements $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$ by incorporating differentiable forward models $\{\boldsymbol{A}_i\}_{i=1}^N$. **B) Test-time Adaptation:** Given a new measurement $\mathbf{y} \notin \mathbf{Y}$ and the corresponding forward model $\boldsymbol{A}$, the shared pair $(\boldsymbol{f}_{\phi^*}, \boldsymbol{g}_{\psi^*})$ is frozen while only the subject-specific encoder $\boldsymbol{f}_\varphi$ is optimized. This effectively injects learned population priors into the INR optimization process, enabling faster and higher-quality reconstruction $\mathbf{x}$.

into two categories: 1) Meta-learning (Sitzmann et al., 2020a; Tancik et al., 2021; Lee et al., 2021; Dupont et al., 2022; Chen & Wang, 2022; Kim et al., 2023; Friedrich et al., 2026). For example, Tancik et al. (2021) showed that MAML- and Reptile-based meta-learning can significantly accelerate INR optimization and boost reconstruction quality across diverse tasks; 2) Transferable feature learning (Vyas et al., 2024; 2025; Rangarajan et al., 2025; Stolt-Ansó et al., 2023). Vyas et al. (2024) adopts a shared encoder and individual decoders, enabling feature transfer from a few high-quality images. However, these approaches typically initialize INRs from high-quality images and update the prior-bearing parameters during adaptation, coupling prior preservation with case-specific fitting.

3 Proposed Method

Overview Given a set $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$ consisting of a limited number of undersampled measurements (*e.g.*, projections in CT or *k*-space data in MRI), our goal is to reconstruct the corresponding high-quality image $\mathbf{x}$ from a new measurement $\mathbf{y} \notin \mathbf{Y}$ in a fully unsupervised way. This inverse problem involves two key challenges: 1) How to extract rich population priors directly from the set $\mathbf{Y}$? 2) How to inject the learned population priors into the model optimization?

To achieve this, we propose DisINR, a new INR framework that explicitly disentangles shared and subject-specific representations. Unlike conventional INR methods (Feng et al., 2023; Cai et al., 2024) for medical reconstructions, which optimize an independent network $\boldsymbol{h}_\theta$ from scratch for each subject, as shown in Eq (2), our DisINR inherits a pre-trained encoder–decoder pair $(\boldsymbol{f}_{\phi^*}, \boldsymbol{g}_{\psi^*})$, while learning only a subject-specific encoder $\boldsymbol{f}_\varphi$ for the target. Formally, given a target $\mathbf{x}$, DisINR represents it as follows:

$$\mathbf{x} = \boldsymbol{g}_{\psi^*}(\boldsymbol{f}_{\phi^*}(\mathbf{c}) \odot \boldsymbol{f}_\varphi(\mathbf{c})), \tag{3}$$

where $\odot$ denotes the concatenation operator, ϕ^* and ψ^* represent the parameters of the pre-trained pair shared across diverse subjects, and φ are the parameters of the encoder specialized for the target $\mathbf{x}$. Here, disentanglement refers to the factorization of shared and subject-specific learning roles, rather than a guarantee of statistically independent or uniquely identifiable semantic factors.

Pre-training Stage Suppose the measurement set $\mathbf{Y}$ and the corresponding forward model set $\boldsymbol{A} = \{\boldsymbol{A}_i\}_{i=1}^N$, we first seek to embed the population priors inherent in the set $\mathbf{Y}$ into the shared

Algorithm 1 Pre-training Stage

Input: Measurement set $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$, forward model set $\{\boldsymbol{A}_i\}_{i=1}^N$, coordinates $\mathbf{c} \in \Omega$, learning rate β
Output: Parameters ϕ^*, ψ^*, $\{\varphi_i^*\}_{i=1}^N$
Initialize $\phi, \psi, \{\varphi_i\}_{i=1}^N$.
for iteration $t = 1$ **to** T **do**
  for subject $i = 1$ **to** N **do**
    $\mathcal{L}_i \leftarrow \|\boldsymbol{A}_i \boldsymbol{g}_\psi(\boldsymbol{f}_\phi(\mathbf{c}) \odot \boldsymbol{f}_{\varphi_i}(\mathbf{c})) - \mathbf{y}_i\|_1$
  end for
  $\varphi_i \leftarrow \varphi_i - \beta \nabla_{\varphi_i} \mathcal{L}_i,\ \forall i$
  $(\phi, \psi) \leftarrow (\phi, \psi) - \beta \nabla_{\phi,\psi} \sum_{i=1}^N \mathcal{L}_i$
end for
return ϕ^*, ψ^*, $\{\varphi_i^*\}_{i=1}^N$

Algorithm 2 Test-time Adaptation

Input: New measurement $\mathbf{y} \notin \mathbf{Y}$, forward model $\boldsymbol{A}$, coordinates $\mathbf{c} \in \Omega$, learning rate β, and encoder-decoder pair $(\boldsymbol{f}_{\phi^*}, \boldsymbol{g}_{\psi^*})$
Output: Reconstruction $\mathbf{x}^*$
Initialize the new encoder parameters φ.
for iteration $t = 1$ **to** T **do**
  $\mathcal{L} \leftarrow \|\boldsymbol{A} \boldsymbol{g}_{\psi^*}(\boldsymbol{f}_{\phi^*}(\mathbf{c}) \odot \boldsymbol{f}_\varphi(\mathbf{c})) - \mathbf{y}\|_1$
  $\varphi \leftarrow \varphi - \beta \nabla_\varphi \mathcal{L}$
end for
$\mathbf{x}^* \leftarrow \boldsymbol{g}_{\psi^*}(\boldsymbol{f}_{\phi^*}(\mathbf{c}) \odot \boldsymbol{f}_{\varphi^*}(\mathbf{c}))$
return $\mathbf{x}^*$

encoder-decoder pair of DisINR. Fig. 1**A** illustrates this pre-training stage. For each sample $\mathbf{y}_i \in \mathbf{Y}$, we first assign a specific encoder $\boldsymbol{f}_{\varphi_i}$. Both the shared encoder $\boldsymbol{f}_\phi$ and the subject-specific encoder $\boldsymbol{f}_{\varphi_i}$ take the spatial coordinates at imaging space $\mathbf{c} \in \Omega$ as inputs, generating shared and subject-specific representations $\boldsymbol{f}_\phi(\mathbf{c})$ and $\boldsymbol{f}_{\varphi_i}(\mathbf{c})$, respectively. The decoder $\boldsymbol{g}_\psi$ then maps the concatenated features to the reconstructed image as $\mathbf{x}_i = \boldsymbol{g}_\psi(\boldsymbol{f}_\phi(\mathbf{c}) \odot \boldsymbol{f}_{\varphi_i}(\mathbf{c}))$. Next, the predicted image $\mathbf{x}_i$ is projected to the measurement domain through the differentiable forward operator $\boldsymbol{A}_i$. Finally, we jointly optimize the three networks by minimizing the prediction error in the measurement domain. Mathematically, this process can be expressed as follows:

$$\phi^*, \psi^*, \{\varphi_i^*\}_{i=1}^N = \underset{\phi,\psi,\{\varphi_i\}_{i=1}^N}{\arg\min} \sum_{i=1}^N \|\boldsymbol{A}_i \mathbf{x}_i - \mathbf{y}_i\|_1, \quad \text{s.t.} \quad \mathbf{x}_i = \boldsymbol{g}_\psi(\boldsymbol{f}_\phi(\mathbf{c}) \odot \boldsymbol{f}_{\varphi_i}(\mathbf{c})), \tag{4}$$

where ϕ^*, ψ^*, and $\{\varphi_i^*\}_{i=1}^N$ denote learned parameters of the shared pair and the subject-specific encoders. The detailed procedure of the pre-training stage is presented in Algorithm 1.

Test-time Adaptation After the model pre-training, the shared encoder $\boldsymbol{f}_{\phi^*}$ and decoder $\boldsymbol{g}_{\psi^*}$ capture rich representations that generalize across diverse subjects. During test-time adaptation, our goal is to recover the corresponding high-quality image $\mathbf{x}^*$ from a new measurement $\mathbf{y} \notin \mathbf{Y}$, without using any external data. As shown in Fig. 1**B**, we freeze the shared encoder-decoder pair $(\boldsymbol{f}_{\phi^*}, \boldsymbol{g}_{\psi^*})$, and optimize only a new subject-specific encoder $\boldsymbol{f}_\varphi$ from scratch, by incorporating the differentiable forward model $\boldsymbol{A}$. Formally, we solve the following optimization problem:

$$\varphi^* = \underset{\varphi}{\arg\min} \|\boldsymbol{A}\mathbf{x} - \mathbf{y}\|_1, \quad \text{s.t.} \quad \mathbf{x} = \boldsymbol{g}_{\psi^*}(\boldsymbol{f}_{\phi^*}(\mathbf{c}) \odot \boldsymbol{f}_\varphi(\mathbf{c})), \tag{5}$$

where φ^* represents the learned parameters of the subject-specific encoder. Hence, the reconstruction is given by $\mathbf{x}^* = \boldsymbol{g}_{\psi^*}(\boldsymbol{f}_{\phi^*}(\mathbf{c}) \odot \boldsymbol{f}_{\varphi^*}(\mathbf{c}))$. The procedure of the test-time adaptation is presented in Algorithm 2. Freezing prevents updates to the shared parameters and their coordinate-dependent features, while the individual encoder remains trainable. We evaluate the relationship between shared-feature drift and reconstruction quality in Sec. 4.3.

Network Architecture Our DisINR framework supports different INR backbones through its shared and subject-specific encoder design. Owing to its efficient multi-resolution hash encoding, neural graphics primitive (NGP) (Müller et al., 2022) has become one of the leading INR architectures for various medical imaging tasks, such as sparse-view CT (Zha et al., 2022; Cai et al., 2024; Wu et al., 2023b) and undersampled MRI (Feng et al., 2023; 2025).

In this study, we thus use NGP as the backbone of DisINR. Both the shared encoder $\boldsymbol{f}_\phi$ and the subject-specific encoder $\boldsymbol{f}_{\varphi_i}$ share the same design, consisting of a multi-resolution hash encoding module followed by an MLP with two hidden layers. The shared decoder $\boldsymbol{g}_\psi$ also uses two hidden layers and takes as input the concatenated features from the two encoders. Reference implementation settings and task-specific evaluation protocols are provided in Appendix A.1.

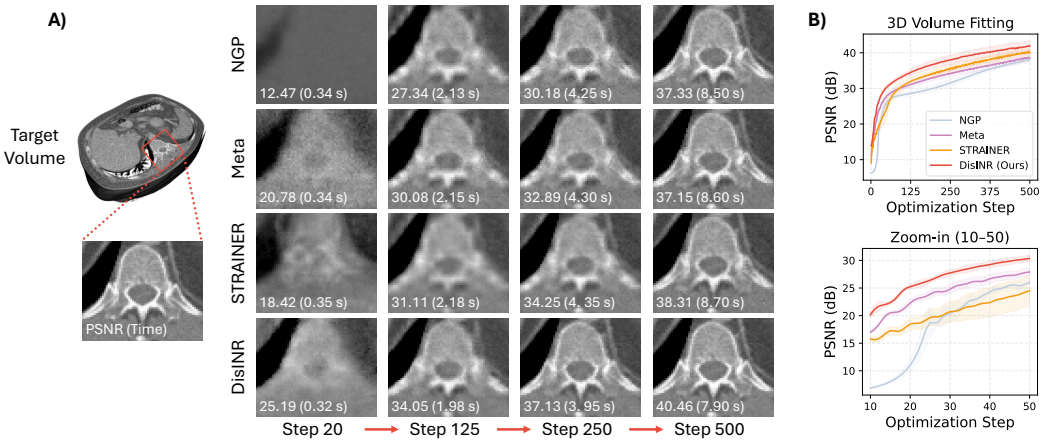

Figure 2: Qualitative result (**A**) on a representative case and optimization curves (**B**) over all cases of three SOTA INR baselines and DisINR for 3D CT volume fitting from the AAPM dataset (McCollough et al., 2017).

4 Experiments

In this section, we evaluate the effectiveness and generalization of DisINR on three representative medical reconstruction tasks, including *3D Volume Fitting*, *Undersampled MRI*, and *Sparse-view CT*. We also conduct extensive ablation studies to analyze the impact of key components in DisINR. ***More details and additional results are provided in Appendix A.***

4.1 Experimental Settings

Dataset Our experiments involve multiple datasets across three tasks:

- *3D Volume Fitting.* We employ 9 human body CT volumes from the AAPM dataset (McCollough et al., 2017), each with a size of 256×256×100. The volumes are first thresholded using a Hounsfield Unit (HU) window of [-800, 400] and then normalized to the range [0, 1]. Finally, we split these volumes into two subsets: 6 volumes for pre-training and 3 volumes for test.
- *Undersampled MRI.* The fastMRI dataset (Knoll et al., 2020) is one of the most commonly used public datasets for MRI reconstruction, containing large-scale multi-contrast *k*-space data. Here, we use 150 multi-coil 2D brain *k*-space samples of size 256×256 from the fastMRI dataset (Knoll et al., 2020), including 100 T2w and 50 FLAIR scans. We simulate a 1D uniform Cartesian pattern with acceleration factors (AF) of 6 and 8. The dataset is divided into three subsets: 50 T2w samples for pre-training, 50 T2w samples for in-domain test, and 50 FLAIR samples for out-of-domain test.
- *Sparse-view CT.* The DeepLesion (Yan et al., 2018) and LIDC (Armato III et al., 2011) are two widely used public CT datasets, containing diverse human body scans collected from multiple clinical institutions. Specifically, we extract 100 and 50 slices of size 256×256 from the two datasets, respectively. Each slice is thresholded using a HU window of [-800, 400] and normalized to the range [0, 1]. We simulate 2D fan-beam projections using `CIL` (Jørgensen et al., 2021; Papoutsellis et al., 2021), with 90 views for pre-training and 60/90 views for testing. The dataset is split into three parts: 50 cases from DeepLesion for pre-training, 50 cases from DeepLesion for in-domain test, and 50 cases from LIDC for out-of-domain test.

Baselines & Metrics We compare DisINR with several representative baselines across three tasks, including both general-purpose INR methods and task-specific reconstruction approaches. ***The general-purpose INR baselines*** include NGP (Müller et al., 2022), a grid-based INR architecture; Meta (Tancik et al., 2021), a meta-learning-based initialization framework; and STRAINER (Vyas et al., 2024), an INR model for learning transferable representations. ***For undersampled MRI***, we include Zero-Filling (ZF), a classical reconstruction method, and IMJENSE (Feng et al., 2023), a state-of-the-art INR-based model for parallel MRI. Since IMJENSE is originally built upon SIREN (Sitzmann et al., 2020b), we reproduce it using the NGP backbone (Müller et al., 2022) for consistency.

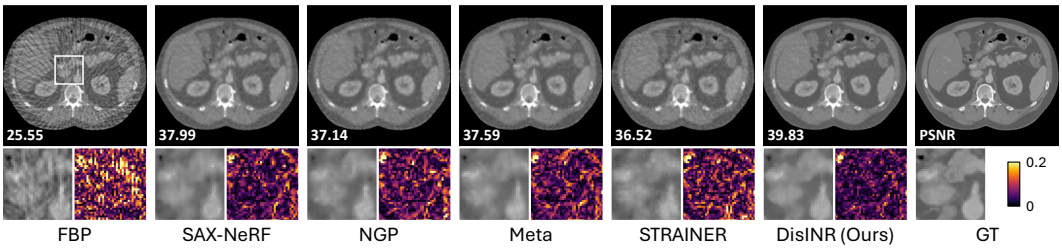

Figure 3: Qualitative results of five baselines and DisINR for sparse-view CT with 60 projection views on a representative sample of the DeepLesion dataset (Yan et al., 2018).

Table 2: Quantitative results of five baselines and DisINR for sparse-view CT on the DeepLesion (Yan et al., 2018) and LIDC (Armato III et al., 2011) datasets.

Method	DeepLesion (in-domain)				LIDC (out-of-domain)			
	60 Views		90 Views		60 Views		90 Views	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FBP	23.98±1.25	0.362±0.039	26.70±1.25	0.478±0.043	24.47±2.03	0.403±0.062	27.40±2.39	0.524±0.076
SAX-NeRF	37.71±1.60	0.965±0.010	38.92±1.49	0.975±0.007	37.34±3.00	0.948±0.036	38.13±3.02	0.956±0.033
NGP	36.89±1.42	0.957±0.010	38.09±1.34	0.969±0.007	36.50±2.73	0.940±0.036	37.50±2.74	0.951±0.033
Meta	36.96±1.35	0.958±0.010	38.08±1.34	0.969±0.007	36.61±2.72	0.941±0.036	37.52±2.69	0.951±0.032
STRAINER	36.17±1.46	0.945±0.014	37.05±1.40	0.956±0.011	35.35±2.77	0.920±0.043	36.17±2.76	0.932±0.039
DisINR	**39.24**±1.65	**0.976**±0.008	**40.30**±1.54	**0.982**±0.006	**38.02**±2.97	**0.954**±0.034	**38.85**±2.98	**0.962**±0.031

For sparse-view CT, we include FBP, a standard analytical reconstruction algorithm, and SAX-NeRF (Cai et al., 2024), a recent INR-based method based on a line Transformer. We use the official implementation of SAX-NeRF in our experiments. For quantitative evaluation, we report peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) on all three tasks.

4.2 Main Results

3D Volume Fitting Fig. 2**A** provides qualitative results. DisINR reconstructs clear anatomical structures with sharp tissue boundaries as early as step 125, whereas other methods still produce blurry results. Fig. 2**B** shows the optimization curves. DisINR converges significantly faster, reaching over 30 dB PSNR within 50 steps, while other methods remain around 25 dB. After full convergence (500 steps), it outperforms the second-best method STRAINER (Vyas et al., 2024) by more than 2 dB, indicating that the learned shared representations provide a strong prior for both fast convergence and accurate reconstruction.

Table 1: Comparison of learnable parameters and running time between three SOTA INR baselines and DisINR for 3D volume fitting on the AAPM dataset (McCollough et al., 2017).

Method	Pre-training		Test	
	#Param.	Time	#Param.	Time
NGP	–	–	23.40 M	8.5 s
Meta	23.40 M	558.3 s	23.40 M	8.6 s
STRAINER	23.63 M	**124.9** s	23.63 M	8.7 s
DisINR	**21.71 M**	186.5 s	**10.83 M**	**7.9** s

In terms of efficiency, as summarized in Table 1, our DisINR maintains a compact model size (21.71 M parameters during pre-training) and reduces the number of learnable parameters to 10.83 M during test-time adaptation, since only the subject-specific encoder is optimized. This design also leads to the fastest inference speed (7.9 s). Instead, these baselines require updating the entire network at test time, resulting in higher computational cost and slower adaptation.

Sparse-View CT Table 2 shows the quantitative results. We observe that DisINR achieves the highest PSNR and SSIM in all settings, outperforming the second-best baseline by 0.68–1.53 dB in

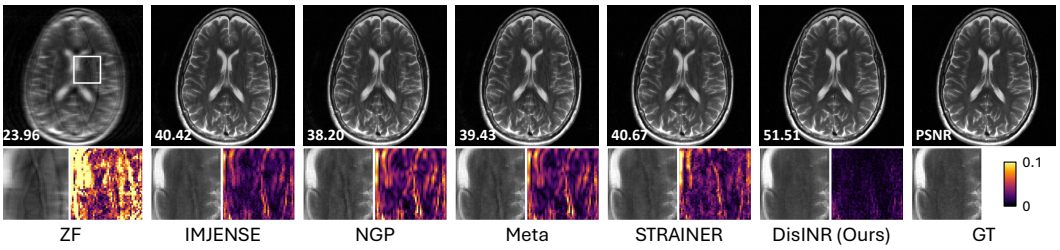

Figure 4: Qualitative comparison of five baselines and DisINR for undersampled MRI with a Cartesian pattern of AF = 6× on a representative sample of the fastMRI-T2w dataset (Knoll et al., 2020).

PSNR. We also provide the qualitative comparison in Fig. 3. From the visualization, our DisINR recovers the sharper anatomical details while reducing streak artifacts.

To assess generalization to unseen pathology and anatomy, models with population priors are pre-trained on DeepLesion and evaluated on unseen COVID-19 CT (Shakouri et al., 2021); DisINR achieves the best results in Table 9 and Fig. 9. We further evaluate KiTS23 kidney-tumor slices and HNSCC head CT. DisINR improves both global reconstruction and tumor-region accuracy in these stress tests (Appendix A.6). These observations support transfer in the evaluated cases, rather than universal robustness to population shifts.

Undersampled MRI Fig. 4 shows that DisINR suppresses undersampling artifacts while preserving anatomical details. Quantitative results (Table 7, Appendix A.2) show the best performance across all evaluated AFs and domains. At AF = 6×, DisINR achieves 48.42 dB on in-domain fastMRI-T2w, exceeding STRAINER (Vyas et al., 2024) by over 8 dB, and 45.64 dB on out-of-domain fastMRI-FLAIR, surpassing competing methods by 3–5 dB despite the contrast shift.

We further evaluate robustness under different sampling trajectories. Models are pre-trained on fastMRI-T2w with Cartesian sampling and directly tested on radial (AF = 10×) and Poisson (AF = 20×) sampling without re-pretraining the shared modules; case-specific adaptation is still performed. As shown in Table 8 and Fig. 8 of Appendix, DisINR consistently achieves the best performance, outperforming STRAINER by +3.34 dB and +2.07 dB in PSNR, respectively. This study further demonstrates the superiority of DisINR.

Reconstruction Runtime We additionally measure wall-clock reconstruction time for 6× Cartesian MRI on fastMRI-T2w and 60-view SVCT on DeepLesion. Under the evaluated budgets, DisINR achieves higher PSNR with shorter reconstruction time than the baselines, despite its additional shared branch. Appendix A.5 reports the complete quality–time comparison and hardware setup.

4.3 Discussion

Visualization and Analysis of Disentangled Learning The strong performance gains of DisINR are primarily driven by its ability to learn disentangled representations. To better understand this property, we conduct experiments on the fastMRI-T2w dataset (Knoll et al., 2020) and examine the shared and subject-specific representations during pre-training. Specifically, we apply PCA to the features $\boldsymbol{f}_{\phi}(\mathbf{c})$ and $\boldsymbol{f}_{\varphi_i}(\mathbf{c})$, and visualize the resulting principal components.

The visualizations are shown in Fig. 5. We make the following key observations: 1) the shared representation $\boldsymbol{f}_{\phi}(\mathbf{c})$ captures smooth, population-level information; 2) the shared representation is not simply the mean image, but instead encodes higher-level semantic structures; and 3) the individual features $\boldsymbol{f}_{\varphi_i}$ capture subject-specific details, such as ***the hemorrhage regions***, as indicated by the yellow arrow. This study demonstrates that DisINR can effectively extract and disentangle population-level representations, offering empirical evidence of its robustness and strong generalization capability.

Table 3 extends this analysis to all pre-training cases in MRI, SVCT, and 3D fitting. Starting from identical initialization, the shared and individual branches develop nonzero PCA-map distances across all three tasks, supporting learned representation separation. Appendix A.7 provides further

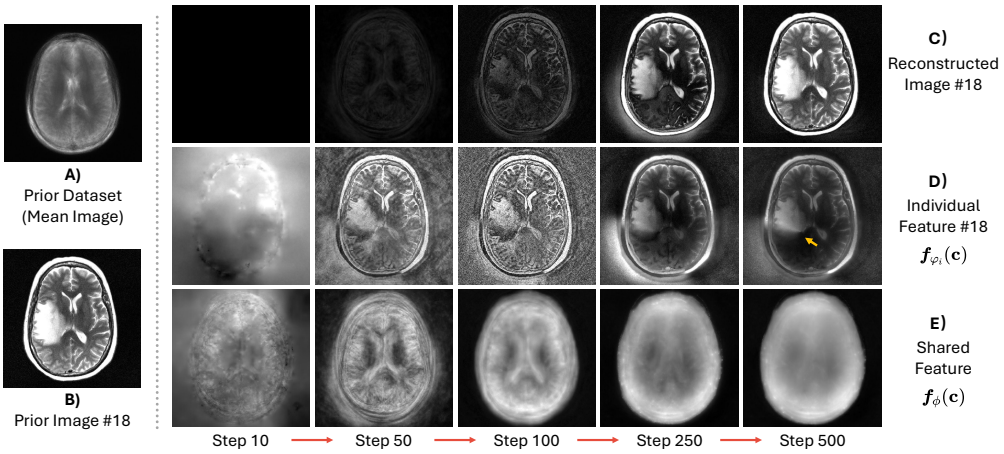

Figure 5: Visualization of DisINR pretraining on the fastMRI-T2w dataset (Knoll et al., 2020): **A)** Prior dataset, mean image of all prior samples. **B)** Prior image #18. **C)** Reconstructed image #18 by DisINR. **D)** Individual feature extracted by the subject-specific encoder of DisINR for sample #18. **E)** Shared feature extracted by the shared encoder of DisINR from the prior dataset.

Table 3: Shared/case-specific feature separation during pre-training on fastMRI-T2w, DeepLesion, and AAPM (Knoll et al., 2020; Yan et al., 2018; McCollough et al., 2017). Relative L_2 and cosine distances compare the two encoders' sign-aligned PCA feature maps, evaluated on the same coordinate grid at each epoch.

Task	**PCA-map distance**	**Epoch**				
		0	**10**	**250**	**500**	**1,000**
MRI	Relative L_2	≈ 0	0.561±0.123	0.612±0.090	0.636±0.077	0.656±0.074
	Cosine	≈ 0	0.091±0.025	0.168±0.056	0.154±0.048	0.164±0.049
SVCT	Relative L_2	≈ 0	0.741±0.155	0.621±0.057	0.755±0.112	1.127±0.195
	Cosine	≈ 0	0.129±0.035	0.195±0.033	0.197±0.044	0.204±0.028
3D fitting	Relative L_2	≈ 0	0.272±0.092	0.502±0.042	0.477±0.039	0.498±0.040
	Cosine	≈ 0	0.038±0.026	0.122±0.017	0.121±0.022	0.131±0.024

details and a complementary feature-drift experiment comparing frozen and unfrozen shared modules during test-time adaptation.

Effect of Weight Freezing in DisINR To study the effect of weight freezing on model performance, we perform an ablation study using the DeepLesion dataset (Yan et al., 2018). As shown in Table 4, we make three observations: 1) Freezing the shared components in DisINR slightly improves performance (+1.45 dB in PSNR). 2) Regardless of freezing, DisINR consistently outperforms existing SOTA INR-based methods (NGP (Müller et al., 2022) and STRAINER (Vyas et al., 2024)). 3) Freezing the shared components in STRAINER (Vyas et al., 2024) significantly reduces performance (about -8 dB in PSNR). Fig. 6 shows the qualitative results, where DisINR (w/ Frozen) consistently recovers the best reconstructions.

A separate matched experiment directly measures shared-representation drift during SVCT-60 adaptation. Unfreezing reduces PSNR from 39.581 to 38.092 dB and produces a feature relative L_2 distance of 0.673, versus zero with freezing (Table 13). Thus, retaining the shared representation is beneficial in this setting. Drift and target-case performance do not, however, measure retention on earlier cases and should not be equated with proof that catastrophic forgetting is eliminated.

Effect of Network Backbone We implement naive INR, STRAINER (Vyas et al., 2024), and DisINR with NeRF (Mildenhall et al., 2021), SIREN (Sitzmann et al., 2020b), and NGP (Müller et al., 2022), using comparable test-time trainable parameter counts within each MRI backbone

Table 4: Quantitative results of two baselines and DisINR ablating frozen components for sparse-view CT with 60 projection views on the DeepLesion dataset (Yan et al., 2018).

Method	Frozen Components	PSNR	SSIM
NGP (Naive INR)	N/A	37.30±1.20	0.960±0.009
STRAINER (w/o Frozen)	N/A	36.65±1.24	0.950±0.012
STRAINER (w/ Frozen)	Shared Encoder	28.40±2.34	0.832±0.057
DisINR (w/o Frozen)	N/A	38.18±1.12	0.967±0.007
DisINR (w/ Frozen)	Shared Encoder-Decoder Pair	**39.63**±1.39	**0.977**±0.006

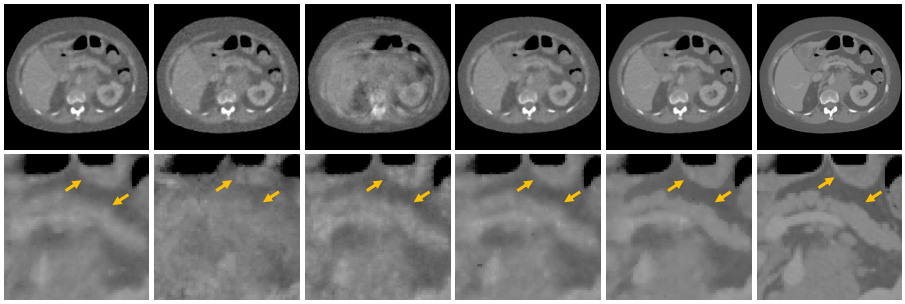

Figure 6: Qualitative results of two baselines and DisINR ablating frozen components for sparse-view CT with 60 projection views on a representative sample of the DeepLesion dataset (Yan et al., 2018).

Table 5: Quantitative results of naive INR, STRAINER (Vyas et al., 2024), and DisINR using different backbones for undersampled MRI with a Cartesian pattern of AF = 6× on the fastMRI-T2w dataset (Knoll et al., 2020).

Method	NeRF		SIREN		NGP	
	PSNR	#Param.	PSNR	#Param.	PSNR	#Param.
Naive INR	37.03±2.23	0.34M	35.73±1.15	0.46M	38.73±5.56	4.92M
STRAINER	36.78±1.61	0.34M	35.76±2.54	0.43M	41.75±2.33	4.97M
DisINR	**40.16**±1.96	**0.31M**	**38.26**±1.04	0.44M	**49.61**±3.82	4.94M

(Table 5). DisINR performs best in these MRI comparisons. We additionally evaluate 3D fitting and SVCT-60: DisINR exceeds naive INR in all six task–backbone combinations and ranks first in four, while STRAINER is better in the other two (Appendix A.8). The framework therefore supports multiple backbones, with task-dependent performance gains.

5 Conclusion

In this work, we present DisINR, a new framework to incorporate population priors into INR learning for medical reconstruction. Using physics-informed unsupervised learning, DisINR effectively pre-trains a shared encoder–decoder pair to capture rich representations directly from a few raw measurements. By freezing the learned encoder–decoder pair, a subject-specific encoder is then optimized for each new subject. Evaluations on three classic medical imaging tasks show DisINR achieves SOTA performance compared to existing INR techniques in both accuracy and efficiency.

Limitations A key limitation of this work is the limited exploration of scalability with respect to the size of the pre-training dataset. Our preliminary study with 5, 10, and 50 pre-training cases demonstrates strong performance with small pre-training sets (Appendix A.9); more extensive large-scale validation is left for future work.

AI USE STATEMENT

Generative AI tools assisted with manuscript editing, table formatting, and figure adjustments. The authors reviewed all AI-assisted content and take responsibility for the final manuscript.

REFERENCES

Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, 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.

Marcel Beister, Daniel Kolditz, and Willi A Kalender. Iterative reconstruction methods in x-ray ct. *Physica medica*, 28(2):94–108, 2012.

Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view x-ray 3d reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11174–11183, 2024.

Yinbo Chen and Xiaolong Wang. Transformers as meta-learners for implicit neural representations. In *European Conference on Computer Vision*, pp. 170–187. Springer, 2022.

Chenhe Du, Xiyue Lin, Qing Wu, Xuanyu Tian, Ying Su, Zhe Luo, Rui Zheng, Yang Chen, Hongjiang Wei, S Kevin Zhou, et al. Dper: Diffusion prior driven neural representation for limited angle and sparse view ct reconstruction. *arXiv preprint arXiv:2404.17890*, 2024.

Emilien Dupont, Hyunjik Kim, SM Ali Eslami, Danilo Jimenez Rezende, and Dan Rosenbaum. From data to functa: Your data point is a function and you can treat it like one. In *International Conference on Machine Learning*, pp. 5694–5725. PMLR, 2022.

Jie Feng, Ruimin Feng, Qing Wu, Xin Shen, Lixuan Chen, Xin Li, Li Feng, Jingjia Chen, Zhiyong Zhang, Chunlei Liu, et al. Spatiotemporal implicit neural representation for unsupervised dynamic mri reconstruction. *IEEE Transactions on Medical Imaging*, 2025.

Ruimin Feng, Qing Wu, Jie Feng, Huajun She, Chunlei Liu, Yuyao Zhang, and Hongjiang Wei. Imjense: scan-specific implicit representation for joint coil sensitivity and image estimation in parallel mri. *IEEE Transactions on Medical Imaging*, 43(4):1539–1553, 2023.

Jeffrey A Fessler. Model-based image reconstruction for mri. *IEEE signal processing magazine*, 27(4):81–89, 2010.

Paul Friedrich, Florentin Bieder, Julian McGinnis, Julia Wolleb, Daniel Rueckert, and Philippe C. Cattin. MedFuncta: A unified framework for learning efficient medical neural fields. In Yuankai Huo, Mingchen Gao, Chang-Fu Kuo, Yueming Jin, and Ruining Deng (eds.), *Proceedings of The 9th International Conference on Medical Imaging with Deep Learning*, volume 315 of *Proceedings of Machine Learning Research*, pp. 56–87. PMLR, 08–10 Jul 2026. URL https://proceedings.mlr.press/v315/friedrich26a.html.

A. Grossberg, H. Elhalawani, A. Mohamed, S. Mulder, B. Williams, A. L. White, J. Zafereo, A. J. Wong, J. E. Berends, S. AboHashem, J. M. Aymard, A. Kanwar, S. Perni, C. D. Rock, S. Chamchod, M. Kantor, T. Browne, K. Hutcheson, G. B. Gunn, S. J. Frank, D. I. Rosenthal, A. S. Garden, C. D. Fuller, and M.D. Anderson Cancer Center Head and Neck Quantitative Imaging Working Group. HNSCC [dataset]. The Cancer Imaging Archive, 2020. URL https://doi.org/10.7937/k9/tcia.2020.a8sh-7363.

Aaron J Grossberg, Abdallah SR Mohamed, Hesham Elhalawani, William C Bennett, Kirk E Smith, Tracy S Nolan, Bowman Williams, Sasikarn Chamchod, Jolien Heukelom, Michael E Kantor, et al. Imaging and clinical data archive for head and neck squamous cell carcinoma patients treated with radiotherapy. *Scientific data*, 5(1):180173, 2018. doi: 10.1038/sdata.2018.173.

Mukesh G Harisinghani, Aileen O'Shea, and Ralph Weissleder. Advances in clinical mri technology. *Science Translational Medicine*, 11(523):eaba2591, 2019.

Nicholas Heller, Fabian Isensee, Resha Tejpau, Andrew Wood, Nikolaos Papanikolopoulos, and Christopher Weight. 2023 kidney and kidney tumor segmentation challenge. Zenodo, 2023. URL https://doi.org/10.5281/zenodo.7840134.

Jiahao Huang, Yinzhe Wu, Fanwen Wang, Yingying Fang, Yang Nan, Cagan Alkan, Daniel Abraham, Congyu Liao, Lei Xu, Zhifan Gao, et al. Data-and physics-driven deep learning based reconstruction for fast mri: Fundamentals and methodologies. *IEEE Reviews in Biomedical Engineering*, 2024.

Wenqi Huang, Hongwei Bran Li, Jiazhen Pan, Gastao Cruz, Daniel Rueckert, and Kerstin Hammernik. Neural implicit k-space for binning-free non-cartesian cardiac mr imaging. In *International Conference on Information Processing in Medical Imaging*, pp. 548–560. Springer, 2023.

Jakob S Jørgensen, Evelina Ametova, Genoveva Burca, Gemma Fardell, Evangelos Papoutsellis, Edoardo Pasca, Kris Thielemans, Martin Turner, Ryan Warr, William RB Lionheart, et al. Core imaging library-part i: a versatile python framework for tomographic imaging. *Philosophical Transactions of the Royal Society A*, 379(2204):20200192, 2021.

Chiheon Kim, Doyup Lee, Saehoon Kim, Minsu Cho, and Wook-Shin Han. Generalizable implicit neural representations via instance pattern composers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11808–11817, 2023.

Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Florian Knoll, Jure Zbontar, Anuroop Sriram, Matthew J Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J Geras, Joe Katsnelson, Hersh Chandarana, et al. fastmri: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning. *Radiology: Artificial Intelligence*, 2(1):e190007, 2020.

Jaeho Lee, Jihoon Tack, Namhoon Lee, and Jinwoo Shin. Meta-learning sparse implicit neural representations. *Advances in Neural Information Processing Systems*, 34:11769–11780, 2021.

Yunxiang Li, Jie Deng, and You Zhang. Universal mapping and patient-specific prior implicit neural representation for enhanced high-resolution mri in mri-guided radiotherapy. *Medical physics*, 52(7):e17863, 2025.

Yuanyuan Liu, Jinwen Xie, Jianhao Wu, Zhuo-Xu Cui, Qingyong Zhu, Jing Cheng, Haifeng Wang, Zhen Song, Dong Liang, and Yanjie Zhu. Physics-guided self-supervised implicit neural representation for accelerated $T_{1\rho}$ mapping. *IEEE Transactions on Biomedical Engineering*, pp. 1–14, 2025. doi: 10.1109/TBME.2025.3618476.

Yisi Luo, Xile Zhao, and Deyu Meng. Continuous representation methods, theories, and applications: An overview and perspectives. *arXiv preprint arXiv:2505.15222*, 2025.

Cynthia H McCollough, Adam C Bartley, Rickey E Carter, Baiyu Chen, Tammy A Drees, Phillip Edwards, David R Holmes III, Alice E Huang, Farhana Khan, Shuai Leng, et al. Low-dose ct for the detection and classification of metastatic liver lesions: results of the 2016 low dose ct grand challenge. *Medical physics*, 44(10):e339–e352, 2017.

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. *Communications of the ACM*, 65(1):99–106, 2021.

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 *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2381–2391, 2023.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022.

Evangelos Papoutsellis, Evelina Ametova, Claire Delplancke, Gemma Fardell, Jakob S Jørgensen, Edoardo Pasca, Martin Turner, Ryan Warr, William RB Lionheart, and Philip J Withers. Core imaging library-part ii: multichannel reconstruction for dynamic and spectral tomography. *Philosophical Transactions of the Royal Society A*, 379(2204):20200193, 2021.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International conference on machine learning*, pp. 5301–5310. PMLR, 2019.

Vikram Rangarajan, Shishira Maiya, Max Ehrlich, and Abhinav Shrivastava. Siedd: Shared-implicit encoder with discrete decoders. *arXiv preprint arXiv:2506.23382*, 2025.

Geoffrey D Rubin. Computed tomography: revolutionizing the practice of medicine for 40 years. *Radiology*, 273(2S):S45–S74, 2014.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

Shokouh Shakouri, Mohammad Amin Bakhshali, Parvaneh Layegh, Behzad Kiani, Farid Masoumi, Saeedeh Ataei Nakhaei, and Sayyed Mostafa Mostafavi. Covid19-ct-dataset: an open-access chest ct image repository of 1000+ patients with confirmed covid-19 diagnosis. *BMC research notes*, 14(1):178, 2021.

Liyue Shen, John Pauly, and Lei Xing. Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):770–782, 2022.

Vincent Sitzmann, Eric Chan, Richard Tucker, Noah Snavely, and Gordon Wetzstein. Metasdf: Meta-learning signed distance functions. *Advances in Neural Information Processing Systems*, 33:10136–10147, 2020a.

Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in neural information processing systems*, 33:7462–7473, 2020b.

Nil Stolt-Ansó, Julian McGinnis, Jiazhen Pan, Kerstin Hammernik, and Daniel Rueckert. Nisf: Neural implicit segmentation functions. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 734–744. Springer, 2023.

Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov. Coil: Coordinate-based internal learning for tomographic imaging. *IEEE Transactions on Computational Imaging*, 7:1400–1412, 2021.

Matthew Tancik, Ben Mildenhall, Terrance Wang, Divi Schmidt, Pratul P Srinivasan, Jonathan T Barron, and Ren Ng. Learned initializations for optimizing coordinate-based neural representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2846–2855, 2021.

Jean-Baptiste Thibault, Ken D Sauer, Charles A Bouman, and Jiang Hsieh. A three-dimensional statistical approach to improved image quality for multislice helical ct. *Medical physics*, 34(11):4526–4544, 2007.

Kushal Vyas, Ahmed I Humayun, Aniket Dashpute, Richard G Baraniuk, Ashok Veeraraghavan, and Guha Balakrishnan. Learning transferable features for implicit neural representations. *Advances in Neural Information Processing Systems*, 37:42268–42291, 2024.

Kushal Vyas, Ashok Veeraraghavan, and Guha Balakrishnan. Fit pixels, get labels: Meta-learned implicit networks for image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 194–203. Springer, 2025.

Tao Wang, Wenjun Xia, Jingfeng Lu, and Yi Zhang. A review of deep learning ct reconstruction from incomplete projection data. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 8(2):138–152, 2023.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612, 2004. doi: 10.1109/TIP.2003.819861.

Qing Wu, Lixuan Chen, Ce Wang, Hongjiang Wei, S Kevin Zhou, Jingyi Yu, and Yuyao Zhang. Unsupervised polychromatic neural representation for ct metal artifact reduction. *Advances in Neural Information Processing Systems*, 36:69605–69624, 2023a.

Qing Wu, Ruimin Feng, Hongjiang Wei, Jingyi Yu, and Yuyao Zhang. Self-supervised coordinate projection network for sparse-view computed tomography. *IEEE Transactions on Computational Imaging*, 9:517–529, 2023b.

Qing Wu, Chenhe Du, Xuanyu Tian, Jingyi Yu, Yuyao Zhang, and Hongjiang Wei. Moner: Motion correction in undersampled radial MRI with unsupervised neural representation. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=OdnqG1fYpo.

Di Xu, Hengjie Liu, Xin Miao, Daniel O'Connor, Jessica E Scholey, Wensha Yang, Mary Feng, Michael Ohliger, Hui Lin, Dan Ruan, et al. Accelerated patient-specific non-cartesian mri reconstruction using implicit neural representations. *International Journal of Radiation Oncology* Biology* Physics*, 2025.

Ke Yan, Xiaosong Wang, Le Lu, Ling Zhang, Adam P Harrison, Mohammadhadi Bagheri, and Ronald M Summers. Deep lesion graphs in the wild: relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 9261–9270, 2018.

Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1960–1970, 2021.

Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: neural attenuation fields for sparse-view cbct reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 442–452. Springer, 2022.

A Appendix

A.1 Experimental Details

Data Pre-processing In our experiments, we include two classical medical imaging tasks: undersampled MRI and sparse-view CT. Here, we detail the data pre-processing.

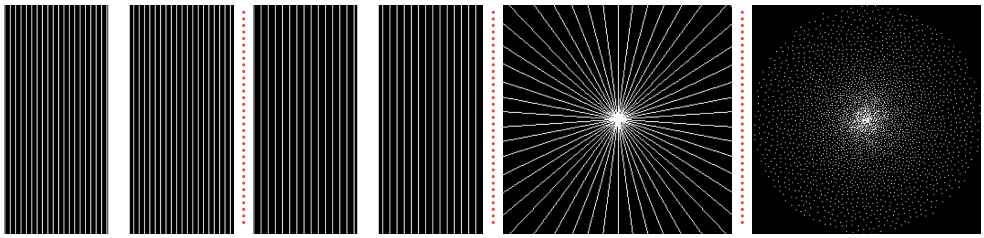

Figure 7: Four types of sampling patterns used in the undersampled MRI task.

- *Undersampled MRI.* Given multi-coil 2D brain *k*-space data of size 256×256 from the fastMRI dataset (Knoll et al., 2020), as shown in Fig. 7, we simulate 1D Cartesian sampling patterns with acceleration factors of 6 and 8, where the size of the auto-calibration region (ACS) is set to 24. Coil-sensitivity maps computed from the raw fully sampled *k*-space data are used for parallel MRI.
- *Sparse-view CT.* For raw 2D slices from the DeepLesion (Yan et al., 2018) and LIDC (Armato III et al., 2011) datasets, we generate sparse-view CT projections by using the CIL toolbox (Jørgensen et al., 2021; Papoutsellis et al., 2021) to simulate a 2D fan-beam geometry. The detailed acquisition parameters are provided in Table 6.

Table 6: Acquisition parameters of 2D fan-beam geometry used in the sparse-view CT task.

Parameters	**Values**
Type of geometry	2D fan-beam
Image Size	256×256
Voxel Size (mm^2)	1×1
View Range (°)	[0, 360)
Number of Detectors	500
Detector Spacing (mm)	2
Number of Projection Views	60/90
Distance from Source to Center (mm)	300
Distance from Center to Detector (mm)	300

Quantitative Metrics In our evaluation, we use peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) (Wang et al., 2004), two widely used visual metrics, to quantitatively assess the model performance. Specifically, for undersampled MRI reconstruction, we calculate these two metrics using the normalized amplitude map, as MRI images are complex-valued. For CT reconstruction (including 3D volume fitting and sparse-view CT), we directly compute them using the reconstructed CT images without applying additional pre-processing.

Implementations of DisINR DisINR can be implemented with different INR backbones. Our main MRI and SVCT experiments use NGP (Müller et al., 2022): each encoder combines hash encoding with a two-hidden-layer, width-128 MLP and outputs 128 features. The shared decoder takes the concatenated 256-dimensional vector and uses two width-128 hidden layers followed by a linear output. Both tasks use LeakyReLU activations and hash settings $L = 10$, $T = 2^{18}$, $F = 8$, $N_{\min} = 2$, and per-level scale $b = 2$. These settings apply to the NGP backbone, rather than defining a common parameter budget across all backbones.

We use Adam (Kingma, 2014), an initial learning rate of 10^{-3}, and multiplicative decay by 0.5. MRI uses 4,000 pre-training and 4,000 adaptation steps, decaying every 1,000 steps. SVCT uses 16,000 pre-training steps with decay every 4,000 steps, and 4,000 adaptation steps with decay every 1,000 steps. The CT preparation notebook generates 90-view pre-training projections and 60/90-view test projections. MRI adaptation additionally uses total variation with weight 10^{-3}. Runtime comparisons and cross-backbone experiments use the separate optimization budgets described in their respective sections.

A.2 Results on Cartesian-Undersampled MRI

Table 7 presents the quantitative results for Cartesian-undersampled MRI.

Table 7: Quantitative results of five baselines and DisINR for undersampled MRI on the fastMRI-T2w and fastMRI-FLAIR datasets (Knoll et al., 2020).

Method	fastMRI-T2w (in-domain)				fastMRI-FLAIR (out-of-domain)			
	AF = 6×		AF = 8×		AF = 6×		AF = 8×	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ZF	23.47±1.51	0.681±0.036	23.12±1.43	0.660±0.036	21.24±3.57	0.684±0.045	21.06±3.54	0.655±0.048
IMJENSE	38.76±5.28	0.961±0.041	29.30±2.84	0.848±0.049	41.61±6.72	0.979±0.022	29.29±5.04	0.901±0.055
NGP	36.95±4.74	0.947±0.043	28.38±2.49	0.818±0.047	41.36±7.59	0.977±0.024	28.87±5.10	0.891±0.059
Meta	37.42±4.78	0.951±0.042	28.56±2.57	0.824±0.049	42.20±8.06	0.980±0.023	29.48±5.46	0.900±0.059
STRAINER	40.22±3.78	0.972±0.032	31.36±2.99	0.892±0.041	37.28±5.17	0.969±0.015	28.17±4.90	0.886±0.046
DisINR	**48.42**±5.18	**0.992**±0.021	**35.46**±3.96	**0.952**±0.036	**45.64**±4.79	**0.990**±0.009	**33.55**±4.82	**0.946**±0.026

A.3 Results on Unseen MRI Sampling Patterns

Table 8 and Fig. 8 present results with unseen MRI sampling patterns.

Table 8: Quantitative results of five baselines and DisINR for undersampled MRI with unseen radial (AF = 10×) and Poisson patterns (AF = 20×) on the fastMRI-T2w dataset (Knoll et al., 2020).

Method	Radial (AF = 10×)		Poisson (AF = 20×)	
	PSNR	SSIM	PSNR	SSIM
ZF	18.84±1.95	0.520±0.056	16.57±1.53	0.402±0.043
IMJENSE	34.51±3.42	0.968±0.010	34.78±3.92	0.951±0.014
NGP	31.94±3.50	0.926±0.024	30.99±3.38	0.843±0.055
Meta	32.01±3.52	0.924±0.026	30.97±3.58	0.841±0.063
STRAINER	33.88±3.20	0.960±0.011	34.88±3.71	0.955±0.015
DisINR	**37.22**±2.79	**0.976**±0.006	**36.95**±3.88	**0.964**±0.011

A.4 Results on Unseen Abnormal Anatomical Structures

Table 9 and Fig. 9 present results on the unseen COVID-19 dataset.

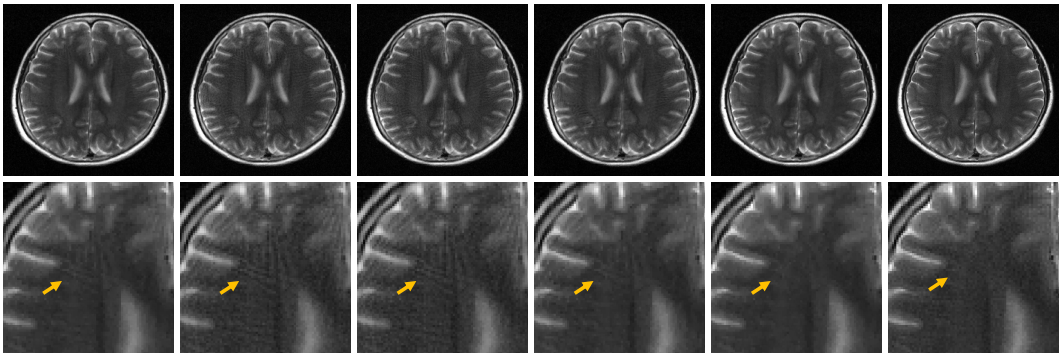

Figure 8: Qualitative results of four baselines and DisINR for undersampled MRI with a radial pattern of AF = 10× on a representative sample of the fastMRI-T2w dataset (Knoll et al., 2020).

Table 9: Quantitative results of five baselines and DisINR for sparse-view CT with 60 and 90 projection views on the unseen COVID-19 dataset (Shakouri et al., 2021) containing abnormal anatomical structures.

Method	60 Views		90 Views	
	PSNR	SSIM	PSNR	SSIM
FBP	18.95±0.90	0.280±0.011	21.87±0.90	0.358±0.017
SAX-NeRF	32.09±1.39	0.926±0.003	33.02±1.36	0.940±0.003
NGP	31.50±1.30	0.916±0.004	32.45±0.97	0.933±0.005
Meta	31.55±1.23	0.916±0.004	32.43±1.03	0.932±0.005
STRAINER	30.68±1.27	0.899±0.006	31.45±1.11	0.916±0.003
DisINR	**32.73**±1.43	**0.935**±0.005	**33.72**±1.18	**0.948**±0.005

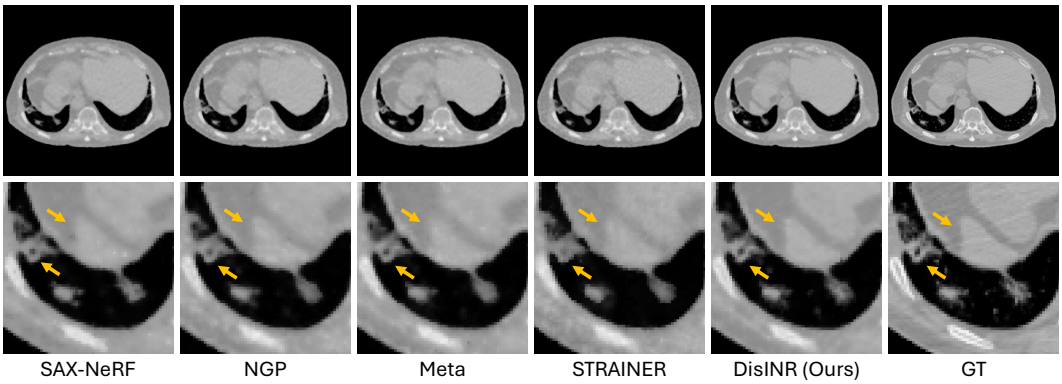

Figure 9: Qualitative results of four baselines and DisINR for sparse-view CT with 60 projection views on a representative sample of the unseen COVID-19 dataset (Shakouri et al., 2021).

A.5 Reconstruction Runtime

Setup We evaluate 6× Cartesian-accelerated MRI on fastMRI-T2w (Knoll et al., 2020) and 60-view SVCT on DeepLesion (Yan et al., 2018). All methods run on the same workstation with one NVIDIA GeForce RTX 5060 Ti GPU (16 GB) and 64 GB of system memory, without multi-GPU parallelism. Table 10 reports reconstruction time and the corresponding PSNR at the stated iteration budget. These are test-time reconstruction benchmarks, not measurements of the cost of population pre-training.

Table 10: Reconstruction quality and wall-clock time for 6× Cartesian MRI on fastMRI-T2w (Knoll et al., 2020) and 60-view SVCT on DeepLesion (Yan et al., 2018). Lower time and higher PSNR are better; iteration budgets are shown explicitly.

Task / dataset	Method	Iterations	Time (s)	PSNR (dB)
MRI 6× fastMRI-T2w	NGP	8,000	52.99	38.04
	MetaNGP	8,000	52.66	38.43
	STRAINER	8,000	67.53	43.57
	IMJENSE-NGP	8,000	56.64	40.73
	DisINR	4,000	**40.80**	**50.41**
SVCT-60 DeepLesion	NGP	8,000	57.90	39.96
	MetaNGP	8,000	59.27	40.07
	STRAINER	8,000	61.68	39.20
	SAX-NeRF	4,000	54.63	39.26
	DisINR	4,000	**32.70**	**41.49**

Quality–time comparison DisINR reconstructs the MRI example in 40.80 s, compared with 52.66–67.53 s for the baselines, while achieving higher PSNR. For SVCT, the corresponding times are 32.70 s and 54.63–61.68 s. The added shared branch therefore does not prevent a reduction in total reconstruction time under these budgets. The comparison does not imply lower cost per iteration, identical iteration counts, or lower total cost including pre-training.

A.6 Pathological and Anatomical OOD Evaluation

KiTS23 kidney-tumor reconstruction To complement the COVID-19 experiment, we evaluate five fixed tumor-containing slices from a KiTS23 CT volume (Heller et al., 2023) under SVCT-60. KiTS23 is excluded from pre-training; all population-based methods retain their DeepLesion-pre-trained models. In addition to image-level PSNR and SSIM, we report mean absolute error (MAE, in HU) within the tumor and along its boundary. These local metrics assess whether a global improvement also extends to the unseen lesion.

Table 11: Reconstruction on five KiTS23 tumor-containing slices (Heller et al., 2023) under SVCT-60 with DeepLesion pre-training. Values are slice-wise mean ± SD, not statistics over five independent patients. Tumor and boundary MAE are measured in HU.

Method	PSNR (dB) ↑	SSIM ↑	Tumor MAE ↓	Boundary MAE ↓
NGP	33.47±0.06	0.9180±0.0035	24.94±4.86	33.81±10.38
MetaNGP	33.43±0.13	0.9171±0.0025	25.11±5.98	34.79±12.45
STRAINER	32.56±0.11	0.8979±0.0028	27.85±5.73	33.98±10.26
SAX-NeRF	33.21±0.50	0.9208±0.0082	25.41±4.09	32.73±7.20
DisINR	**35.07**±0.28	**0.9459**±0.0010	**19.45**±2.08	**23.16**±5.20

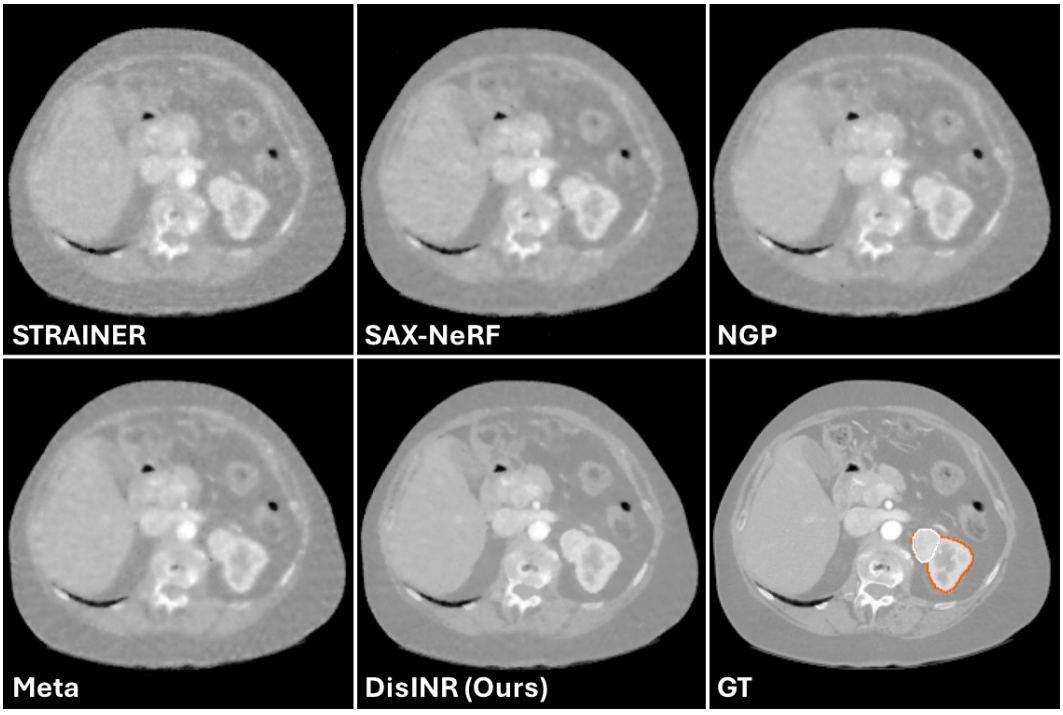

Figure 10: Qualitative comparison of four baselines and DisINR for SVCT-60 on an unseen KiTS23 slice (Heller et al., 2023). Orange and white contours in the reference image indicate kidney and tumor boundaries, respectively.

As shown in Table 11, DisINR has the highest PSNR and SSIM and the lowest tumor and boundary MAE. Its advantage therefore extends beyond global image quality to the tumor region. Fig. 10 provides the corresponding qualitative comparison.

HNSCC anatomical stress test We additionally evaluate five fixed adjacent axial head slices from an HNSCC CT volume (Grossberg et al., 2020; 2018) excluded from pre-training, retaining the DeepLesion-pre-trained models. This tests transfer to a markedly different anatomical region; it is not a tumor-specific evaluation of these head slices.

Table 12: Reconstruction on five adjacent HNSCC head CT slices (Grossberg et al., 2020) under SVCT-60 with DeepLesion pre-training. Values are within-volume slice-wise mean ± SD.

Method	PSNR (dB) ↑	SSIM ↑
NGP	31.33±0.51	0.9510±0.0028
MetaNGP	31.35±0.55	0.9531±0.0022
STRAINER	29.99±0.50	0.9211±0.0036
SAX-NeRF	31.76±0.89	0.9580±0.0091
DisINR	**33.26**±0.61	**0.9736**±0.0015

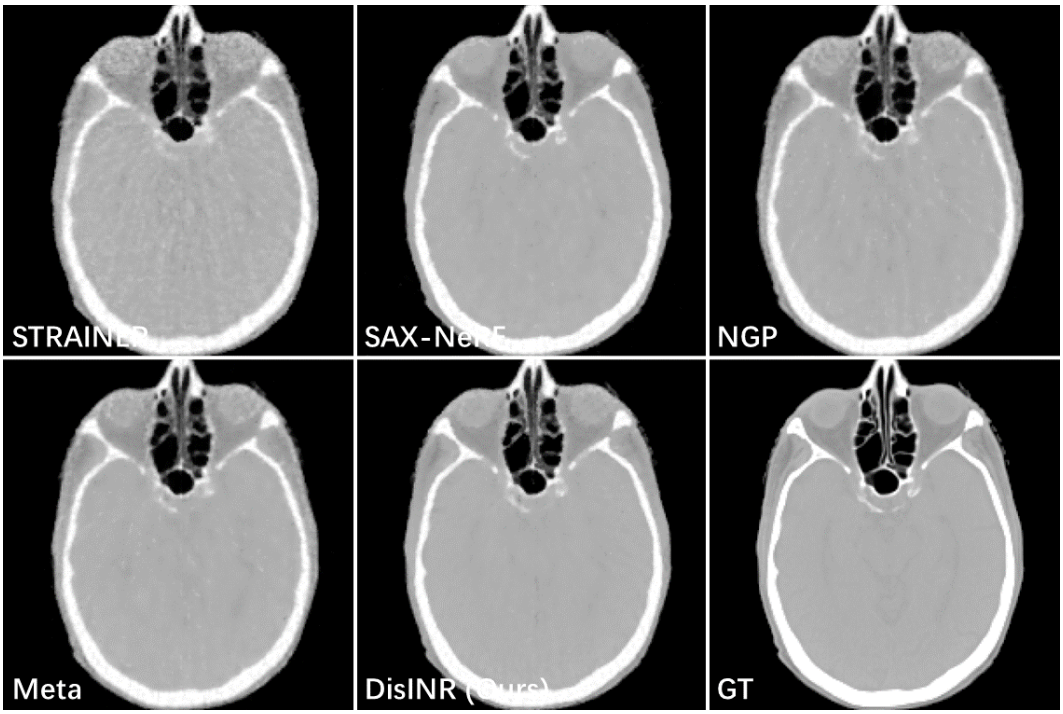

Figure 11: Qualitative comparison of four baselines and DisINR for SVCT-60 on an unseen HNSCC head CT slice (Grossberg et al., 2020). All population-based methods use DeepLesion pre-training.

DisINR exceeds NGP by 1.93 dB and SAX-NeRF by 1.50 dB (Table 12). MetaNGP provides little improvement over NGP, whereas STRAINER performs worse. Together with the MRI sampling-pattern and COVID-19 experiments, these results broaden the tested distribution shifts. Nevertheless, these small stress tests do not establish population-wide robustness or the absence of reconstruction bias under arbitrary pathology or scanner changes.

A.7 Shared and Case-Specific Representation Analysis

Separation during population pre-training We track the shared encoder and every case-specific encoder during pre-training for MRI (50 cases), SVCT (50 cases), and 3D volume fitting (six volumes). At each checkpoint, both branches are evaluated on the same fixed coordinate grid. Table 3 reports relative L_2 and cosine distances between their sign-aligned principal feature maps, summarized as mean $\pm$ SD across the pre-training cases in each task. These are PCA-map distances between branches, not the adaptation-drift metrics used below.

The branches use identical architectures and parameter initialization, so their initial feature maps coincide. Nonzero distances emerge during training in all three tasks. This is consistent with their different optimization roles: the shared encoder receives gradients across cases, whereas each individual encoder is updated only from its associated measurements. The distances describe learned representational differences; larger distance is not itself a reconstruction objective or evidence of statistical independence.

Shared-representation drift during adaptation On DeepLesion SVCT-60, we compare frozen and jointly fine-tuned shared encoder–decoder pairs using identical pre-trained weights, case-specific initialization, random seed, and 4,000 adaptation steps. Table 13 measures shared-feature and parameter changes relative to the pre-adaptation state, alongside target reconstruction quality.

Table 13: Reconstruction and shared-representation drift during frozen/unfrozen adaptation on DeepLesion (Yan et al., 2018) under SVCT-60 (mean $\pm$ SD). Feature and parameter distances compare the adapted state with its pre-adaptation reference.

Shared pair	PSNR (dB)	SSIM	Feature relative L_2	Feature cosine distance	Parameter relative L_2
Frozen	39.581$\pm$1.643	0.9770$\pm$0.0071	0.000$\pm$0.000	0.000$\pm$0.000	0.000$\pm$0.000
Unfrozen	38.092$\pm$1.303	0.9665$\pm$0.0076	0.673$\pm$0.123	0.268$\pm$0.107	0.491$\pm$0.051

Unfreezing is associated with a 1.489 dB PSNR reduction and nonzero shared-feature drift. Zero drift for the frozen branch follows from the construction. This comparison supports the benefit of preserving the shared representation in the tested setting, but does not constitute a direct old-task retention test.

A.8 Cross-Backbone Evaluation

MRI backbone comparison Fig. 12 complements the quantitative MRI backbone comparison in Table 5 with representative NeRF- and SIREN-based reconstructions.

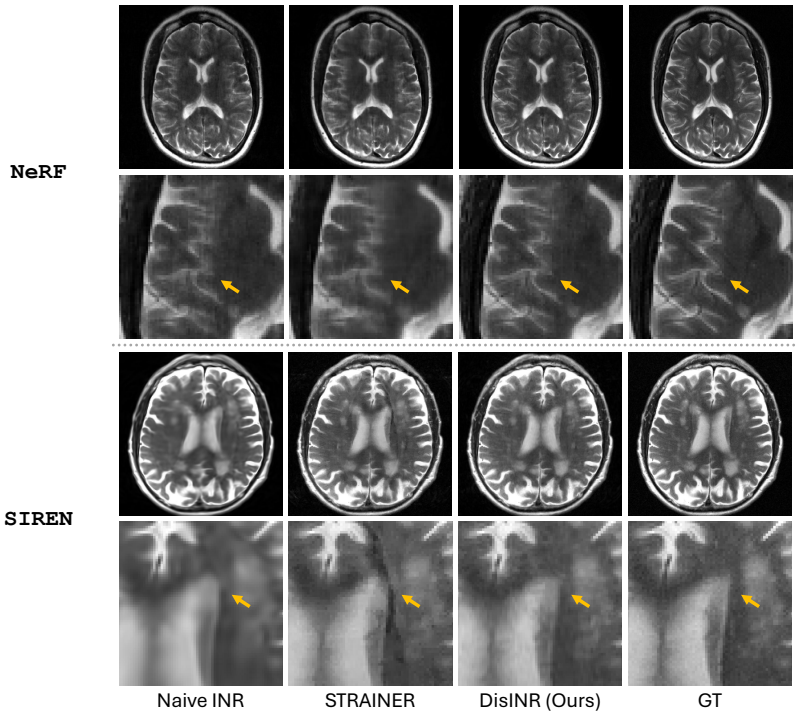

Figure 12: Qualitative results of two baselines and DisINR with different INR architectures (NeRF and SIREN) for undersampled MRI with a Cartesian pattern of AF = 6× on a representative sample of the fastMRI-T2w dataset (Knoll et al., 2020).

Cross-task comparison We extend the MRI backbone study in Table 5 to AAPM 3D volume fitting and DeepLesion SVCT-60. Naive INR, STRAINER, and DisINR use NeRF, SIREN, or NGP backbones. Within each task–backbone group, the data split, preprocessing, and test-time optimization budget are shared across methods. Table 14 reports PSNR as mean ± population SD. These controlled comparisons are distinct from the original default-backbone experiments.

Table 14: Cross-backbone reconstruction results on AAPM 3D volume fitting and DeepLesion SVCT-60 (McCollough et al., 2017; Yan et al., 2018). PSNR (dB) is reported as mean ± population SD; bold denotes the best method within each task–backbone group.

Task	Backbone	Naive INR	STRAINER	DisINR
3D volume fitting	NeRF	28.98±0.97	**31.21±1.48**	31.12±1.44
	SIREN	31.38±1.33	30.02±0.45	**31.90±1.66**
	NGP	37.39±0.95	37.67±0.37	**38.75±0.72**
SVCT-60	NeRF	31.13±0.99	33.47±1.12	**34.70±1.62**
	SIREN	31.56±1.25	**34.16±1.01**	32.31±1.08
	NGP	37.36±1.17	36.46±1.22	**39.54±1.40**

DisINR improves on naive INR in every combination and achieves the highest PSNR in four of six settings. STRAINER performs better for NeRF-based volume fitting and SIREN-based SVCT. Compatibility across these architectures therefore does not imply uniform superiority. Figs. 13 and 14 provide representative reconstructions.

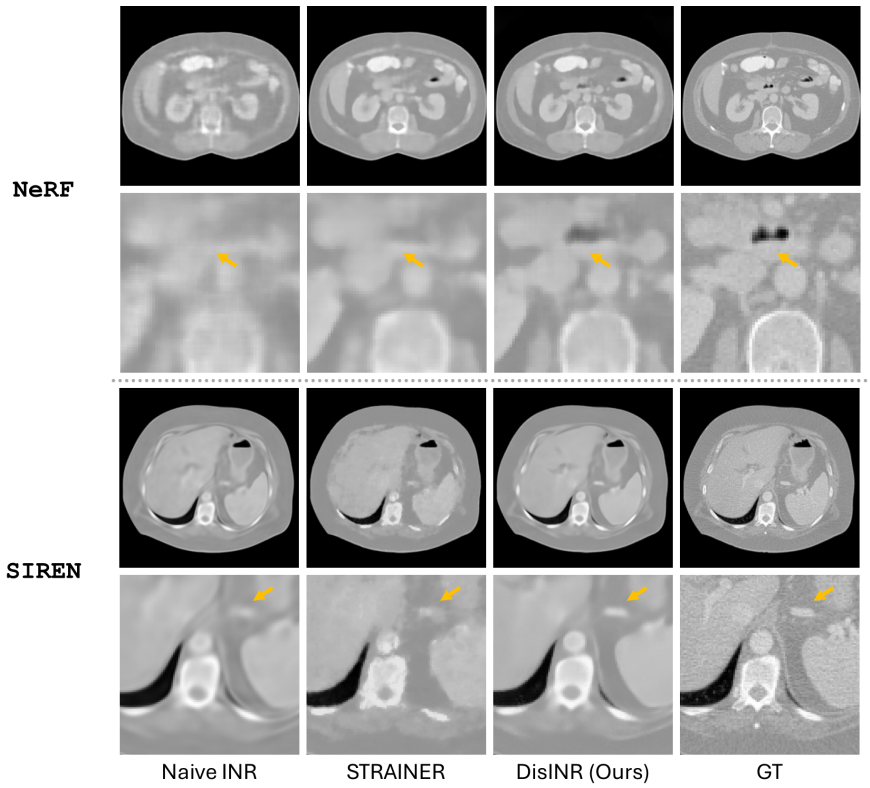

Figure 13: Qualitative comparison of naive INR, STRAINER, and DisINR with NeRF and SIREN for 3D volume fitting on AAPM (McCollough et al., 2017). Each backbone group shows a representative slice and enlarged details; the two groups depict different slices.

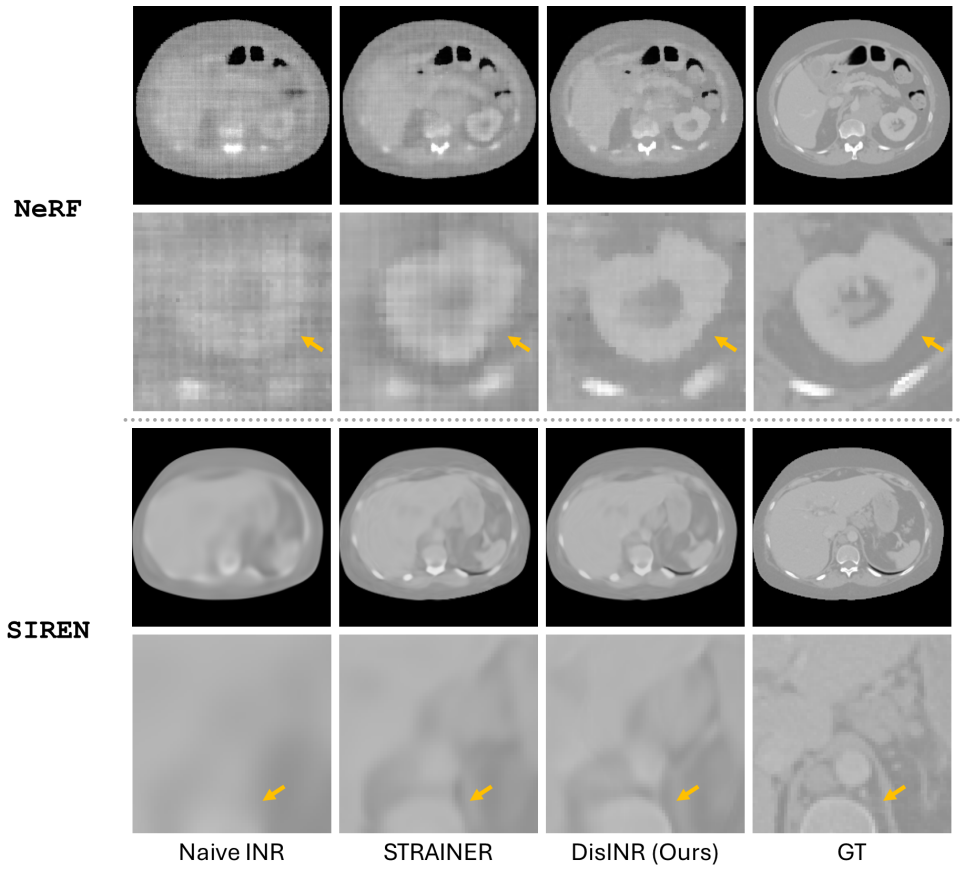

Figure 14: Qualitative comparison of naive INR, STRAINER, and DisINR with NeRF and SIREN for 60-view SVCT on DeepLesion (Yan et al., 2018). Each backbone group shows a representative slice and enlarged details; the two groups depict different slices.

A.9 Effect of Pre-training Data Size

On fastMRI-T2w (Knoll et al., 2020) with Cartesian AF = 6×, we vary the number of pre-training cases ($N = 5, 10, 50$) and compare DisINR with Meta, STRAINER, and the scan-specific IMJENSE baseline. IMJENSE does not use pre-training and therefore serves as a fixed reference.

Fig. 15 shows that DisINR achieves the highest PSNR at each tested data size, increasing from 47.92 dB with five samples to 50.41 dB with 50. STRAINER changes from 44.85 to 42.20 dB, while Meta varies little under these settings. This experiment demonstrates effective use of a small pre-training set; it does not establish large-scale scaling behavior or sensitivity to different pre-training compositions.

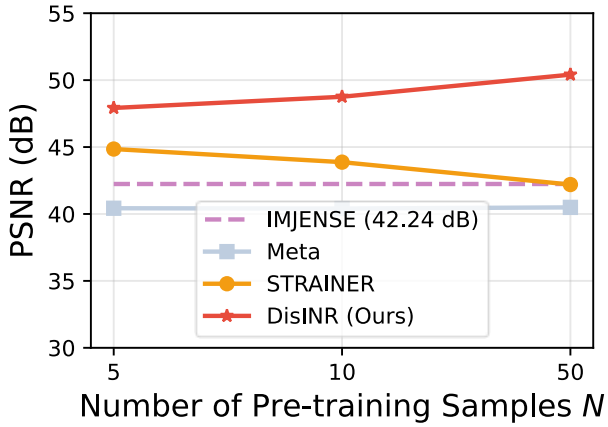

Figure 15: Performance curves of three SOTA INR baselines and DisINR under different numbers of pre-training samples N for undersampled MRI with AF = 6× on the fastMRI-T2w dataset (Knoll et al., 2020).

A.10 Additional Visual Results

Figs. 16 and 17 show additional qualitative comparisons of DisINR and the baselines for undersampled MRI and sparse-view CT. Visually, our DisINR produces the best reconstructions, which are closest to the GT samples.

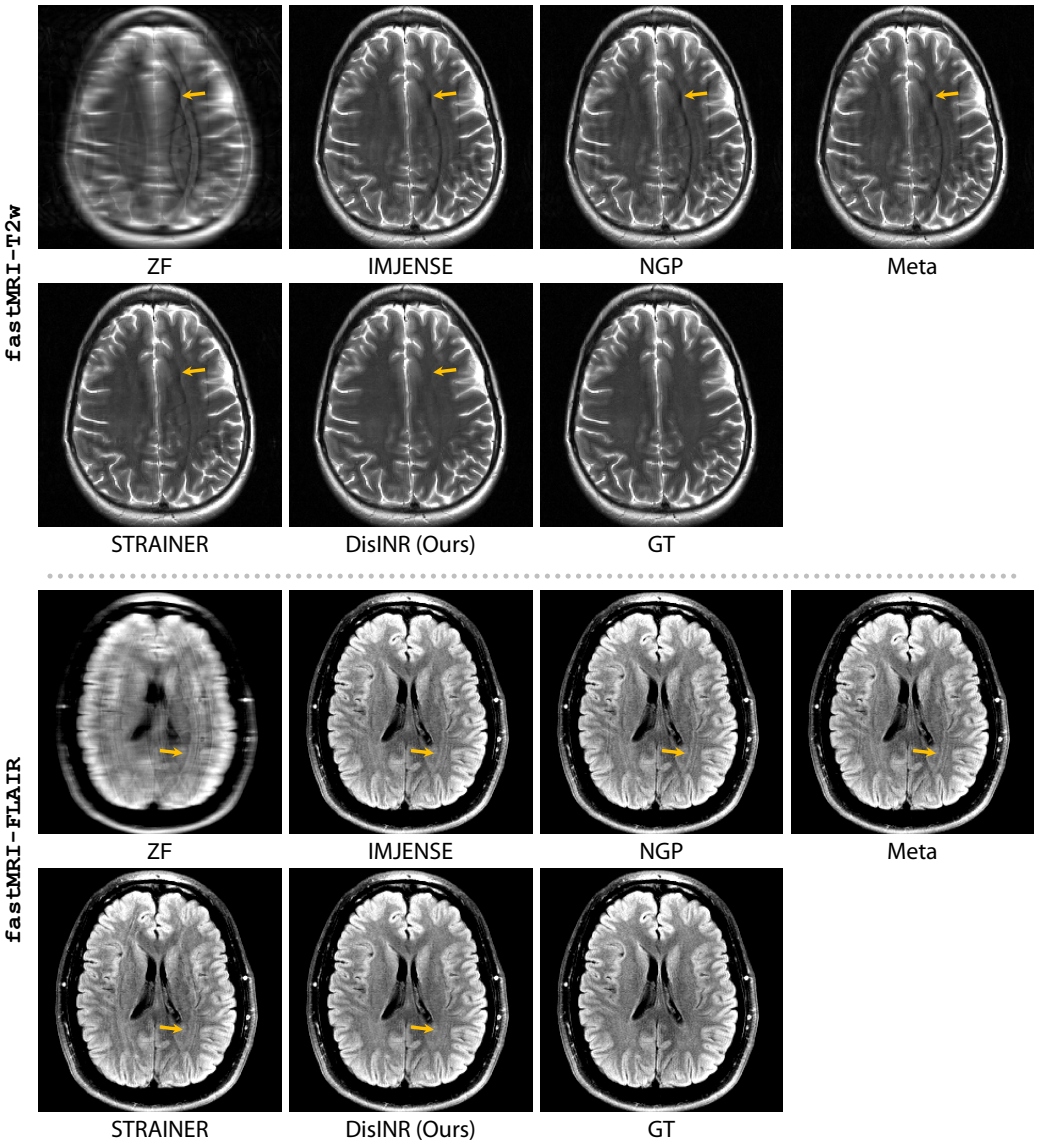

Figure 16: Qualitative comparison of five baselines and DisINR for undersampled MRI with a Cartesian pattern of AF = 6× on two representative samples of the fastMRI-T2w and fastMRI-FLAIR datasets (Knoll et al., 2020).

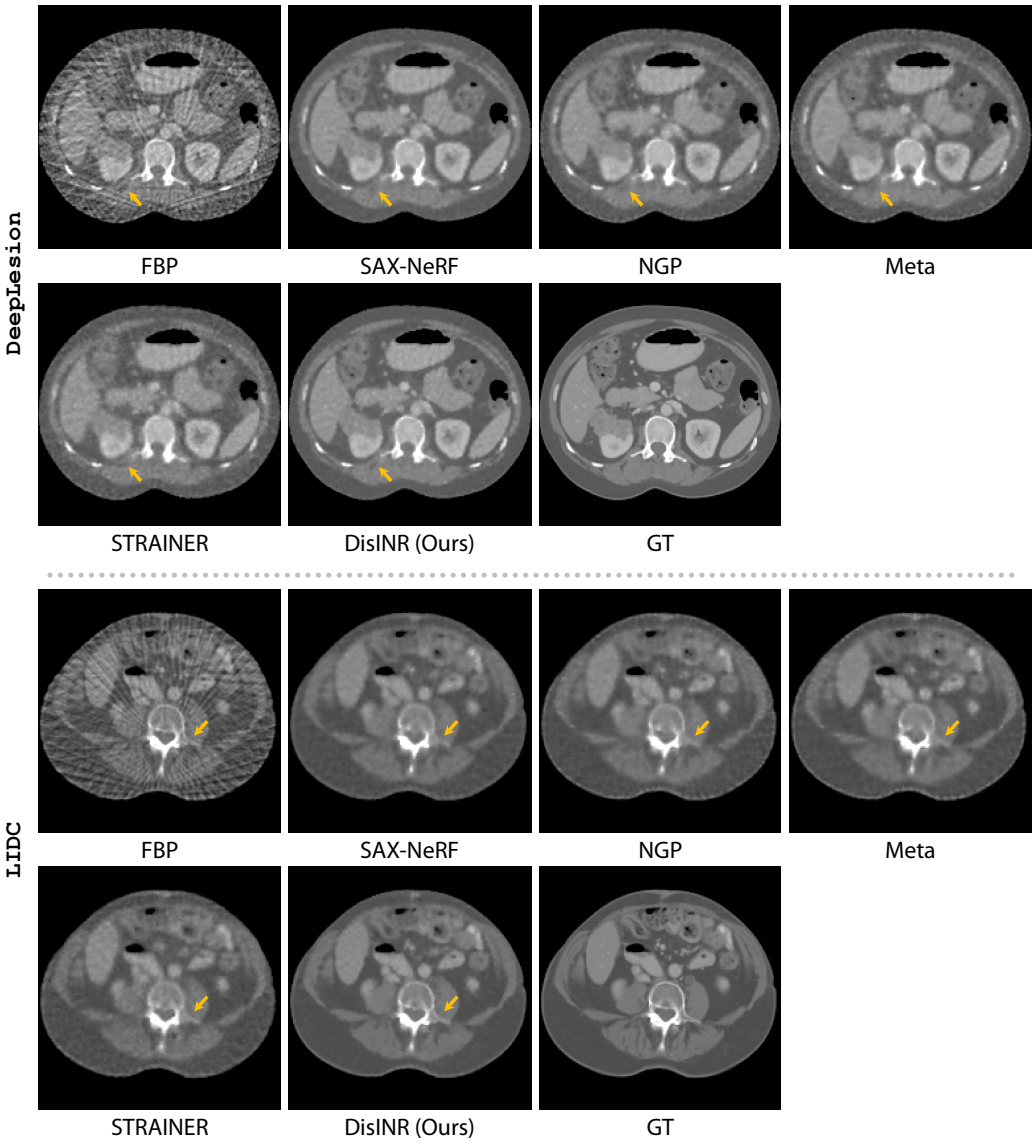

Figure 17: Qualitative results of five baselines and DisINR for sparse-view CT with 60 projection views on two representative samples of the DeepLesion (Yan et al., 2018) and LIDC (Armato III et al., 2011) datasets.