Extend STRAINER to Inverse Imaging Tasks

Alan Huang
Department of Electrical & Computer Engineering
Rice University
Houston, TX 77005, United States
ah212@rice.edu

Kushal Vyas
Department of Electrical & Computer Engineering
Rice University
Houston, TX 77005, United States
kv30@rice.edu

Guha Balakrishnan
Department of Electrical & Computer Engineering
Rice University
Houston, TX 77005, United States
guha@rice.edu

***Abstract*—Inverse problems aim to recover hidden signals from incomplete or corrupted measurements and arise in tasks such as super-resolution, denoising, and CT reconstruction. While data-driven models like CNNs, GANs, and Transformers have shown success in solving inverse problems, they suffer from limitations such as fixed resolution, poor geometric consistency, and high computational cost. Recently, Implicit Neural Representations (INRs) have emerged as flexible alternatives, modeling signals as continuous functions. However, traditional INRs require training from scratch for each signal, limiting practicality. STRAINER, a framework for learning transferable INR features, has shown promise in accelerating signal reconstruction. In this work, we extend STRAINER to inverse imaging tasks, leveraging its transferable features for fast, high-quality reconstructions. By evaluating STRAINER on super-resolution, denoising, and CT reconstruction, we aim to explore its effectiveness as a flexible prior for inverse problems and better understand its potential to address the limitations of existing data-driven methods.**

***Index Terms*—INR, Inverse Problem, Denoise, Super resolution, tomography (CT) image reconstruction**

I. Introduction

Inverse problems are fundamental challenges in science and engineering, where the goal is to recover hidden or missing information from limited or indirect measurements. In many real-world scenarios, we rarely have direct access to a complete signal; instead, we observe noisy, incomplete, or indirect observations, from which we aim to reconstruct the underlying data. Classic examples of inverse problems include image inpainting (filling in missing parts of an image), super-resolution [6] (recovering high-resolution images from low-resolution inputs), compressed sensing (reconstructing signals from sparse measurements), and medical imaging techniques such as MRI and CT [4], where the task is to reconstruct images from projection data.

These problems are typically ill-posed, meaning that an infinite number of possible solutions may exist for a given set of measurements. Therefore, incorporating strong priors is essential to guide the solution toward plausible and accurate reconstructions.

One prominent approach to address inverse problems is through data-driven priors, where models learn the statistical structure of natural signals from large datasets. These models can then map corrupted or incomplete measurements directly to high-quality reconstructions. Convolutional Neural Networks (CNNs) and Transformers have become widely used for this purpose [10]. CNN-based denoisers are effective for noise removal, super-resolution networks upscale low-resolution inputs, and Generative Adversarial Networks (GANs) are often used for generating realistic outputs in inpainting and restoration tasks. More recently, Vision Transformers (ViTs) have demonstrated powerful capabilities in capturing long-range dependencies and global context in images [9].

While data-driven models have achieved impressive results and offer fast inference once trained, they come with notable limitations. First, they are typically constrained to a fixed resolution — for instance, a model trained on 256×256 256×256 images struggles to generalize to higher resolutions. Second, they may exhibit poor geometric consistency, especially in complex tasks like 3D reconstruction or structured inverse problems. Additionally, these models are often large, containing millions of parameters, which makes them expensive to train and deploy. Finally, they suffer from dataset bias, as they often fail to generalize beyond the data distribution on which they were trained.

To address these limitations, Implicit Neural Representations (INRs) [1] have recently gained attention as a flexible alternative. INRs model signals as continuous functions parameterized by neural networks, typically Multi-Layer Perceptrons (MLPs), that can represent high-frequency details and arbitrary resolutions. Trained directly on discrete measurements, INRs can capture the underlying continuous signal without requiring a fixed grid. They have demonstrated success in a variety of tasks, including inverse problems and neural rendering. However, traditional INRs are trained from scratch for each new signal, making them computationally expensive and time-consuming to adapt to new instances. Moreover, because they are often trained on single signals, INRs are assumed to lack transferability to other signals from the same distribution.

To overcome these challenges, Kushal Vyas et al. recently proposed STRAINER [2], an INR training framework that

learns transferable features to accelerate and improve INR fitting on new signals. STRAINER leverages the sequential layer-wise structure of MLP-based INRs by sharing initial encoder layers across multiple INRs while allowing independent decoder layers. At test time, the shared encoder provides a powerful initialization that significantly reduces the training time and improves reconstruction quality for new signals from the same distribution. STRAINER achieves faster convergence and up to +10 dB gain in reconstruction quality early in training compared to untrained INRs.

In this work, we extend STRAINER to a variety of inverse imaging [5]tasks, aiming to leverage its transferable feature representations for fast and high-quality image reconstructions from corrupted or incomplete measurements. Our goal is to demonstrate that STRAINER's learned features can act as strong, flexible priors for a broad class of inverse problems, effectively addressing the limitations of traditional data-driven models, while preserving the adaptability and resolution-agnostic nature of implicit neural representations (INRs). To evaluate the generality and effectiveness of STRAINER in the inverse problem setting, we focus on three representative tasks: super-resolution, image denoising, and computed tomography (CT) image reconstruction. Through these experiments, we aim to investigate both the practical benefits and limitations of STRAINER's transferable features and to gain deeper insights into its behavior and potential as a unified solution for diverse inverse imaging problems.

II. Method

We adopt *Implicit Neural Representations (INRs)* [1]as a flexible framework to solve various inverse imaging problems [5]. An INR models a continuous signal, such as an image, as a function $f_\theta : \mathbf{x} \to \mathbf{y}$, parameterized by a neural network (normally choose multiple layer of MLP as structure), where $\mathbf{x}$ is the coordinate input and $\mathbf{y}$ is the corresponding pixel intensity or feature value. Given *incomplete or corrupted observations*, we optimize the parameters θ of the INR to reconstruct the underlying clean signal.

A. General Inverse Problem Formulation

Formally, let $\mathbf{y}_{\text{obs}}$ denote the observed corrupted measurement of the true signal $\mathbf{y}^*$, and let $\mathcal{A}$ represent the observation or corruption operator (e.g., noise addition, downsampling, projection). The goal is to find INR parameters θ such that

$$\mathcal{A}(f_\theta(\mathbf{x})) \approx \mathbf{y}_{\text{obs}}. \tag{1}$$

To achieve this, we define an appropriate loss function (e.g., Mean Squared Error) and optimize θ via gradient descent.

B. Application to Different Inverse Problems

1) Image Denoising: For denoising, we corrupt a clean image $\mathbf{y}^*$ by adding Gaussian noise to obtain $\mathbf{y}_{\text{obs}} = \mathbf{y}^* + \mathbf{n}$, where $\mathbf{n} \sim \mathcal{N}(0, \sigma^2)$. The INR is directly trained to fit this noisy observation, and we minimize the pixel-wise *Mean Squared Error (MSE)* between the INR's output and the noisy input:

$$\mathcal{L}_{\text{denoise}} = \|f_\theta(\mathbf{x}) - \mathbf{y}_{\text{obs}}\|^2. \tag{2}$$

By fitting this loss, the INR learns to approximate the underlying clean signal, effectively removing the noise.

2) Super-Resolution: [6]

For super-resolution, we start from a high-resolution image $\mathbf{y}^*$ and simulate a low-resolution observation $\mathbf{y}_{\text{obs}}$ by downsampling $\mathbf{y}^*$ using a factor (e.g., $4\times$ downsampling). During training, we supervise the INR by comparing the *downsampled version of its output* with the low-resolution observation:

$$\mathcal{L}_{\text{SR}} = \|\mathcal{D}(f_\theta(\mathbf{x})) - \mathbf{y}_{\text{obs}}\|^2 \tag{3}$$

where $\mathcal{D}$ is the downsampling operator. This encourages the INR to produce a high-resolution output whose downsampled version matches the observed low-resolution image.

3) CT Image Reconstruction: For CT reconstruction, the measurement $\mathbf{y}_{\text{obs}}$ is a set of *projections* obtained via a Radon transform or similar projection operator $\mathcal{P}$. The INR is optimized to generate an image $f_\theta(\mathbf{x})$ such that its projections match the observed projection data:

$$\mathcal{L}_{\text{CT}} = \|\mathcal{P}(f_\theta(\mathbf{x})) - \mathbf{y}_{\text{obs}}\|^2. \tag{4}$$

By minimizing this loss, the INR learns to reconstruct an image consistent with the projection measurements.

C. Training Strategy for STRAINER Examination

To further examine the effectiveness of STRAINER [2], we train four different types of INRs and compare their performance across various inverse imaging tasks. The configurations of these models are as follows:

1) **SIREN without Initialization**: This is a baseline model where we use a *SIREN* network (Sinusoidal Representations) [11] trained from scratch with no initial feature representations. This model is trained directly on the task-specific corrupted or incomplete data, allowing us to observe the performance of a standard INR without any pre-trained knowledge or initialization.
2) **Fine-Tuned SIREN**: In this case, we initialize the SIREN network with random signals, and then fine-tune it on the specific inverse task at hand. This model serves as a comparison to test how pre-initialization with random signals can affect the convergence speed and quality of the reconstruction.
3) **STRAINER with One Image**: Here, we implement the STRAINER training strategy by initializing the encoder layers of the INR with features learned from a single image. This method aims to test the efficacy of transferring knowledge from a single image, allowing for rapid adaptation to the inverse task while leveraging prior knowledge of a specific image domain.
4) **STRAINER with Ten Images**: In this configuration, the STRAINER strategy is extended by initializing the INR with encoder layers trained on a set of 10 images. This approach evaluates how a small collection of images can provide a richer initialization and whether it leads to better transferability and faster convergence compared to using a single image.

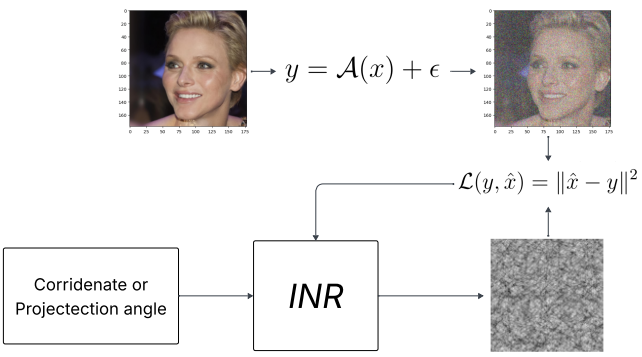

Fig. 1. Experiment process for training an Implicit Neural Representation (INR) model on inverse problem. Using denoising as an example: the input image is first corrupted with noise, then the INR model—initialized with a random signal—predicts pixel values, and the Mean Squared Error (MSE) between the output and noisy image is computed. The model is optimized iteratively until convergence. For CT reconstruction, the process is similar, except the model is conditioned on projection angles instead of spatial coordinates.

Each model is trained using the corresponding loss function for the task as Fig 1 and evaluated on various inverse imaging tasks to assess the benefits of each initialization strategy.

D. Evaluation Metrics

To quantitatively assess the quality of reconstructed images, we employ three complementary evaluation metrics:

- **Peak Signal-to-Noise Ratio (PSNR):** PSNR measures the pixel-level fidelity between the reconstructed image $\hat{y}$ and the ground-truth image y^*. It is defined as:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{MAX^2}{\text{MSE}} \right),$$

  where MAX is the maximum possible pixel value (e.g., 1.0 or 255) and MSE is the mean squared error between $\hat{y}$ and y^*.
- **Structural Similarity Index Measure (SSIM):** SSIM evaluates perceptual similarity by considering luminance, contrast, and structural information. Unlike PSNR, it correlates better with human perception. SSIM values range from 0 to 1, where 1 indicates perfect structural similarity.
- **Learned Perceptual Image Patch Similarity (LPIPS):** LPIPS [7] computes perceptual similarity using deep features extracted from pretrained neural networks (In this paper we use VGG). Lower LPIPS scores indicate better perceptual similarity between the reconstructed and ground-truth images.

These metrics together provide a holistic evaluation, balancing low-level numerical accuracy (PSNR), structural consistency (SSIM), and perceptual realism (LPIPS) across the inverse problems studied.

III. Results

A. Testing on Different Inverse Problems

We evaluate the performance of the proposed STRAINER [2] framework on the CelebA-HQ dataset, focusing on two key inverse imaging tasks: super-resolution and denoising [5]. For CT reconstruction, we utilize the MICCAI FLARE 2022 Challenge dataset.

To benchmark performance, we compare STRAINER with the baseline model *SIREN* and a fine-tuned variant of SIREN [11]. All INR-based models share similar architectural capacity to ensure a fair comparison. Specifically, SIREN uses a 6-layer Multi-Layer Perceptron (MLP), which also serves as the base architecture for STRAINER. STRAINER replaces direct training with an encoder-decoder setup: a 5-layer MLP encoder and a 1-layer MLP decoder, maintaining a comparable number of total parameters.

We evaluate the following configurations:

1) **SIREN**: Trained from scratch without pre-initialization.
2) **SIREN (Fine-Tuned)**: Initialized with random signals and fine-tuned for each task.
3) **STRAINER (One Image)**: STRAINER initialized with features learned from a single image.
4) **STRAINER (Ten Images)**: STRAINER initialized with features learned from ten images.

We report Peak Signal-to-Noise Ratio (PSNR) scores on the denoising and super-resolution tasks, along with training trends across iterations. The results demonstrate that STRAINER significantly improves reconstruction quality and convergence speed, highlighting the benefits of its feature-transfer mechanism. Detailed PSNR trends are presented in Figures 2, 3, and 4.

B. Testing on Large-Scale Data

To further validate STRAINER's generalization ability, we evaluate the same tasks and settings from the previous section on a larger testing set of 100 examples. This experiment aims to confirm whether observed improvements are statistically significant rather than isolated artifacts. The aggregated results are shown in the figures 5, 6 below.

C. Testing with Different Decoder Configurations

Finally, we explore the impact of encoder-decoder configurations on image denoising by varying the number of shared encoder layers. This experiment investigates how different model structures influence the quality of reconstruction. The results are included in figure 7.

IV. Discussion

The experimental results highlight several key observations regarding the performance of STRAINER [2] across inverse imaging tasks. First, STRAINER consistently demonstrated faster convergence, reaching comparable reconstruction quality approximately 1,000 iterations earlier than the SIREN [11] baselines. In the early stages of training—within the first 1,000 iterations—STRAINER outperformed SIREN by 2–3 dB in

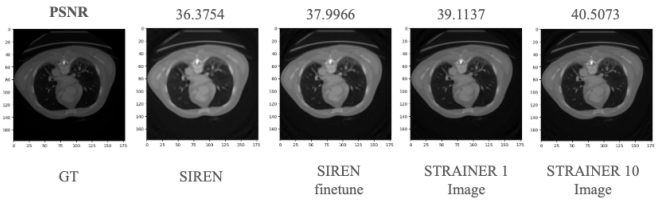

Fig. 2. CT image reconstruction results using SIREN, fine-tuned SIREN, STRAINER (1 image), and STRAINER (10 images), evaluated in terms of PSNR.

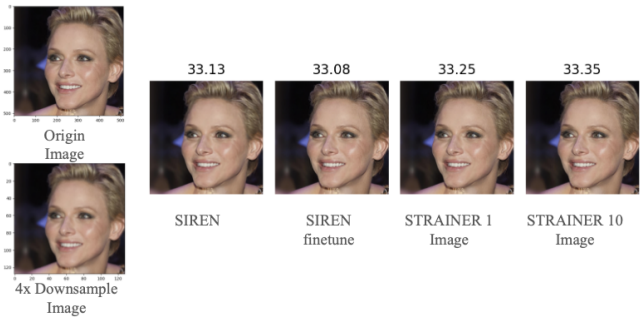

Fig. 3. 4x Super-Resolution results using SIREN, fine-tuned SIREN, STRAINER (1 image), and STRAINER (10 images), evaluated in terms of PSNR.

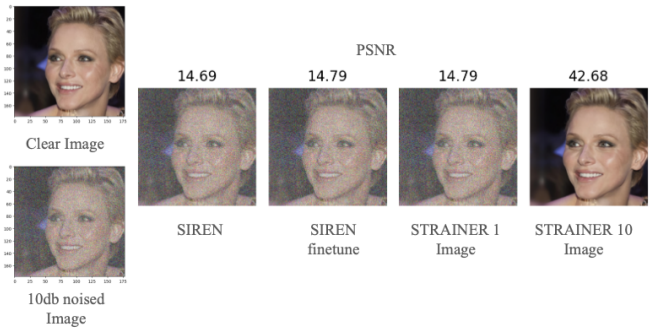

Fig. 4. Denoising results using SIREN, fine-tuned SIREN, STRAINER (1 image), and STRAINER (10 images), evaluated in terms of PSNR.

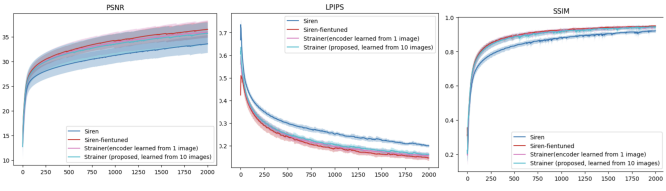

Fig. 5. CT reconstruct results on 100 test cases, evaluated using PSNR, LPIPS, and SSIM for SIREN, STRAINER (1 image), and STRAINER (10 images).

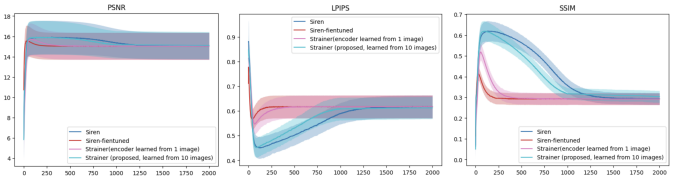

Fig. 6. Denoising results on 100 test cases, evaluated using PSNR, LPIPS, and SSIM for SIREN, STRAINER (1 image), and STRAINER (10 images).

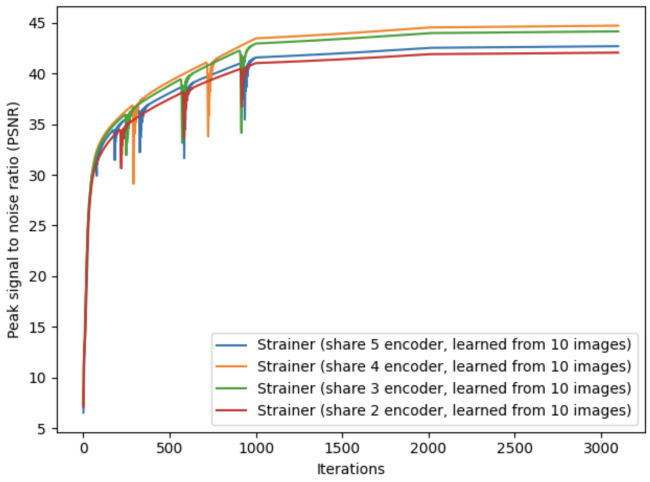

Fig. 7. STRAINER denoising results with varying numbers of encoder layers, ranging from 5 to 2.

PSNR, indicating more efficient initial learning. However, when training was extended, the advantage diminished, with final PSNR, SSIM, and LPIPS scores showing no significant differences among STRAINER and baseline models. Additionally, varying the number of encoder layers in STRAINER's architecture did not noticeably impact reconstruction quality, suggesting limited sensitivity to decoder depth in simpler tasks like denoising. On the other hand, STRAINER struggled with the more complex CT reconstruction task, failing to learn meaningful feature hierarchies. The resulting outputs were consistently blurry and lacked contrast. Visualizations support this finding, illustrating that despite its early learning advantage, STRAINER is less effective on inverse problems with higher complexity and structure dependency.

A. *Future Work*

Given that INR models are highly sensitive to activation functions, a promising direction for future research is to replace traditional MLP layers with Kolmogorov–Arnold Network (KAN) [8]layers. KANs offer several advantages, including fewer parameters, improved function approximation capabilities, stronger inductive biases, and the elimination of hyperparameter tuning for activation functions. These traits make them an attractive alternative for enhancing the expressiveness and generalization of INR-based models. However, adopting KANs also introduces challenges. Inference time is significantly longer, memory usage increases due to the

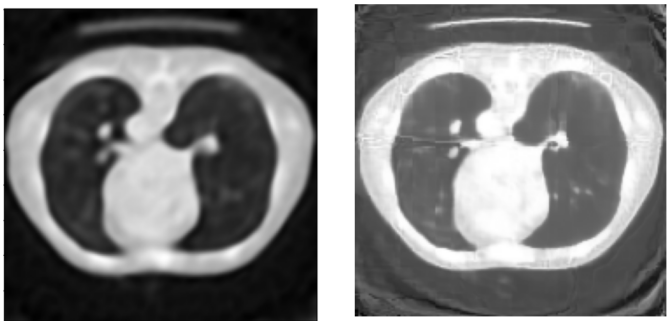

Fig. 8. CT reconstruction results generated by INR models using MLP layers (left) and KAN layers [8] (right). The KAN-based model shows improved structural approximation but with higher computational cost.

functional composition of basis functions, and the architecture is more difficult to parallelize efficiently. Addressing these limitations will be critical to making KANs practical for real-time or large-scale inverse imaging tasks. Figure 8 is a example of using 5 layer of KAN to reconstruct CT image.

ACKNOWLEDGMENT

I would like to thank our mentor, Kushal Vyas, and Professor Guha Balakrishnan for their invaluable guidance and support throughout this research. I also appreciate the help and contributions from other members of the lab, whose insights and feedback greatly enhanced the quality of this work. Their expertise and encouragement have been instrumental in the completion of this project. I acknowledge the use of ChatGPT-4o for assistance with editing, stylistic suggestions, and input on structuring the document.

REFERENCES

[1] Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. 2020. Implicit Neural Representations with Periodic Activation Functions. arXiv:2006.09661 [cs, eess] (June 2020).

[2] Vyas, Kushal Kardam, et al. "Learning Transferable Features for Implicit Neural Representations." Advances in Neural Information Processing Systems 37 (2024): 42268-42291.

[3] Saragadam, Vishwanath, et al. "Wire: Wavelet implicit neural representations." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

[4] Heinrich, Mattias P., Maik Stille, and Thorsten M. Buzug. "Residual U-net convolutional neural network architecture for low-dose CT denoising." Current Directions in Biomedical Engineering 4.1 (2018): 297-300

[5] Chen, Hanting, et al. "Pretrained Image Processing Transformer." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

[6] Shocher, Assaf, Nadav Cohen, and Michal Irani. "Zero-Shot Super-Resolution Using Deep Internal Learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[7] Zhang, Richard, et al. "The unreasonable effectiveness of deep features as a perceptual metric." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[8] Liu, Ziming, et al. "KAN: Kolmogorov-Arnold Networks." arXiv preprint arXiv:2404.19756 (2024).

[9] Dosovitskiy, Alexey, et al. "An image is worth 16x16 words: Transformers for image recognition at scale." *International Conference on Learning Representations (ICLR)*, 2021.

[10] Liu, Jiaming, et al. "When Transformers meet image restoration: A comprehensive review." *International Journal of Computer Vision*, 2023.

[11] Sitzmann, Vincent, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. "Implicit neural representations with periodic activation functions." *Advances in Neural Information Processing Systems* 33 (2020): 7462–7473.