DYNAMICAL IMPLICIT NEURAL REPRESENTATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Implicit Neural Representations (INRs) provide a powerful continuous framework for modeling complex visual and geometric signals, yet achieving high expressivity with compact architectures remains challenging. We introduce **Dynamical Implicit Neural Representations (DINR)**, a model-agnostic framework that induces richer function classes from existing INRs by changing both *what the network learns* and *how the signal is represented*. Rather than directly parameterizing the target function with a *static* neural network, DINR learns a neural vector field that determines how a latent state evolves, while the target representation is given by the resulting evolution of that state. By augmenting *static* INRs with *latent dynamics*, this formulation allows compact neural parameterizations to induce substantially richer mappings without proportionally increasing the number of trainable parameters. We provide theoretical insights into the representational and optimization behavior of DINR through Rademacher complexity and Neural Tangent Kernel analyses. To balance representation capacity and generalization, we further introduce kinetic energy regularization that regulates the complexity of the learned dynamics. Across image representation, 3D field reconstruction, and scientific data compression, DINR consistently improves reconstruction fidelity, parameter efficiency, and robustness over conventional *static* INRs.

1 INTRODUCTION

Implicit Neural Representations (INRs) have become a powerful paradigm for modeling complex signals across vision and geometry, representing data continuously as functions of input coordinates (Sitzmann et al., 2020; Park et al., 2019; Mescheder et al., 2019; Xie et al., 2022; Ashkenazi & Treister, 2024). By replacing discrete representations with neural functions, INRs enable continuous querying, arbitrary-scale reconstruction, and compact storage of diverse data modalities (Mildenhall et al., 2021; Tancik et al., 2020; Chen et al., 2021; Ben-Shabat et al., 2024; Jayasundara et al., 2025). Despite their flexibility, achieving high expressivity with compact neural architectures remains challenging. In conventional INRs, the complexity of the represented signal is tightly coupled to that of the parameterized network, as the network parameters directly encode the signal within a fixed computational structure. This coupling is particularly restrictive for intricate or high-frequency signals, which are further hindered by spectral bias (Rahaman et al., 2019; Tancik et al., 2020; Sitzmann et al., 2020). Existing approaches primarily address this limitation by modifying components within the same *static coordinate-to-signal parameterization*, input encodings (Tancik et al., 2020; Li et al., 2025), activation functions (Sitzmann et al., 2020; Liu et al., 2024; Saragadam et al., 2023), normalization (Cai et al., 2024), or regularization (Li et al., 2023). While effective, increasing expressivity still requires enriching the underlying parameterization itself.

This leads to a more fundamental question: *can we expand the class of functions represented by an INR without proportionally increasing the complexity of its parameterization?* This perspective shifts the focus from directly encoding the target signal in the network parameters to designing a compact representation that can induce a substantially richer mapping. This question is particularly relevant for implicit representations, where the goal is not merely to fit a finite set of samples, but to encode an entire continuous signal within a limited parameter budget.

To address this question, we introduce **Dynamical Implicit Neural Representations (DINR)**, a model-agnostic mechanism that augments existing INRs with *latent dynamics*. Instead of producing the latent representation through a single static mapping, DINR allows the latent state to evolve over time according to a shared neural function that determines how the state changes. The same neural

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

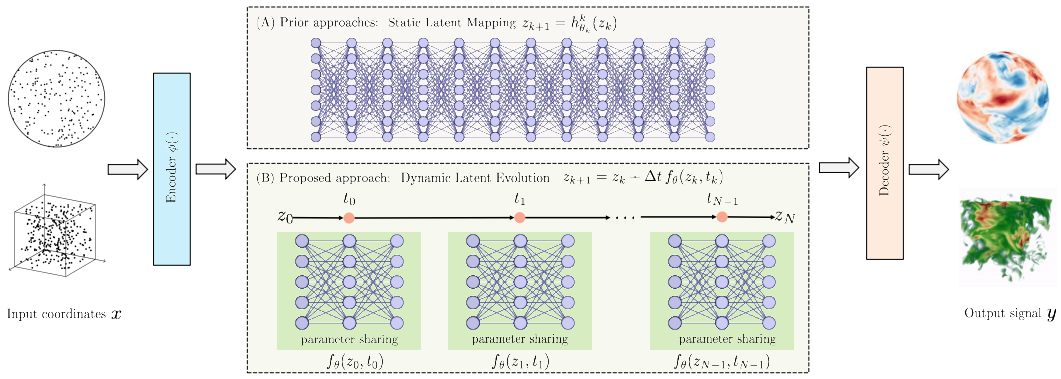

Figure 1: Comparison between static latent mapping and the proposed dynamic latent evolution.

function is repeatedly evaluated along this evolving trajectory, allowing simple local changes to accumulate into a substantially more complex mapping. Thus, DINR introduces an additional source of representational capacity through the *evolution of the latent state*, rather than through increasing the complexity of the underlying network. Importantly, this mechanism is orthogonal to the specific design of the underlying INR and can be incorporated into existing architectures by replacing their static latent mapping with shared dynamics (Figure 1).

We provide theoretical support for this dynamical formulation, showing that latent evolution improves both expressivity and trainability. By accumulating representational changes along the latent trajectory, DINR enables the same neural network to induce a richer class of mappings, resulting in higher Rademacher complexity than static INRs and allowing compact networks to represent more intricate functions (Wei & Ma, 2019; Bartlett & Mendelson, 2002; Mohri et al., 2018). The evolution also produces more diverse gradient directions, increasing the effective rank of the neural tangent kernel (NTK), which facilitates the learning of high-frequency modes and improves optimization (Jacot et al., 2018; Cao et al., 2019). Moreover, controlling the dynamical complexity allows this expressivity to be balanced with generalization. We evaluate DINR across diverse datasets spanning 2D image representation, 3D field reconstruction, and scientific data compression. Across these settings, DINR consistently improves reconstruction fidelity and parameter efficiency over conventional static INRs.

2 Preliminaries

The proposed DINR is a *model-agnostic* framework that augments existing INRs through continuous latent dynamics. To properly situate our contributions, we review the most relevant literature from these areas in Appendix C. We begin by briefly reviewing INRs, followed by a motivating example illustrating how dynamical flows can lift their induced function classes.

Implicit Neural Representations (INRs). INRs model signals as continuous mappings from coordinates $\mathbf{x} \in \mathbb{R}^{d_x}$ to values $\mathbf{y} \in \mathbb{R}^{d_y}$ via a neural network $\hat{\mathbf{y}}_{\text{INR}}(\mathbf{x}) \approx \mathbf{y}$. A typical INR consists of an input embedding $\phi(\,\cdot\,;\theta_{\text{in}}) : \mathbb{R}^{d_x} \to \mathbb{R}^{d_z}$, a latent transformation $f(\,\cdot\,;\theta_{\text{stat}}) : \mathbb{R}^{d_z} \to \mathbb{R}^{d_z}$, and an output decoder $\psi(\,\cdot\,;\theta_{\text{out}}) : \mathbb{R}^{d_z} \to \mathbb{R}^{d_y}$:

$$\mathbf{z}_0 = \phi(\mathbf{x};\theta_{\text{in}}), \quad \mathbf{z}_1 = f(\mathbf{z}_0;\theta_{\text{stat}}), \quad \hat{\mathbf{y}}_{\text{INR}} = \psi(\mathbf{z}_1;\theta_{\text{out}}). \tag{1}$$

Given a dataset of M coordinate-signal pairs $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^{M}$, the parameters $\theta = \{\theta_{\text{in}}, \theta_{\text{stat}}, \theta_{\text{out}}\}$ are obtained by solving the following training problem:

$$\min_{\theta} \left[\frac{1}{M} \sum_{i=1}^{M} \mathcal{L}_{\text{data}} \left(\hat{\mathbf{y}}_{\text{INR}}^{(i)}(\theta), \mathbf{y}^{(i)} \right) \right], \tag{2}$$

where $\mathcal{L}_{\text{data}}$ is a suitable loss function for the task at hand (e.g., the ℓ_2 loss for image regression).

Standard INRs perform a *static* feedforward transformation of the input, where the latent feature $\mathbf{z}_1$ is obtained in a single forward pass through f. Consequently, their representational capacity is strictly bounded by the expressivity of f's fixed-depth architecture, while ϕ and ψ mainly serve as

lightweight input and output mappings. This fixed-depth formulation can limit the ability of compact INRs to represent complex, particularly high-frequency, signals and contributes to the well-known *spectral bias* of neural networks (Rahaman et al., 2019; Yüce et al., 2022).

Motivating Example. A central goal of INRs is to maximize the diversity of representable functions under a fixed parameter budget. We motivate our approach by showing that a function class can be systematically *lifted* through dynamical evolution, even when the dynamics are defined by functions from the same class.

Consider the one-dimensional quadratic function class

$$\mathcal{F} = \{f(x) = ax^2 + bx + c \mid a, b, c \in \mathbb{R}.\}$$

A static representation using $f \in \mathcal{F}$ can only directly represent functions within this quadratic class. Our key observation is that *even with the same function class $f \in \mathcal{F}$, integrating it over time induces a fundamentally different class of mappings.* Specifically, consider the solution space of dynamics:

$$\Phi_{\text{dyn}}(\mathcal{F}) = \left\{ z(T; x) \;\middle|\; \frac{dz}{dt} = f(z),\ f \in \mathcal{F},\ z(0) = x \right\}.$$

Here, f does not directly represent the final mapping; instead, it defines how the latent state z evolves over time, with the final representation obtained from its evolved state $z(T; x)$. Thus, Φ_{dyn} can be viewed as a *lifting operator* that transforms a function class $\mathcal{F}$ to the richer class of mappings induced by its flows. Even for a simple choice such as $f(z) = z - z^2$, the solution is

$$z(t; x) = \frac{xe^t}{1 - x(e^t - 1)} = x + x(1 + x)t + x\left(\frac{1}{2} + \frac{3}{2}x + x^2\right)t^2 + \ldots,$$

which expands into an infinite power series. Thus, although the instantaneous vector field is quadratic, its finite-time flow induces a substantially richer mapping from x to $z(T; x)$. More generally, such ODEs can produce rational or more complex nonlinear functions (Ince, 2012); see Figure 2.

This example illustrates the central idea of *dynamical lifting*: the same function can induce a substantially richer mapping through its evolution over time. Even simple function classes can therefore generate more expressive mappings without enriching the underlying parameterization. Importantly, this mechanism addresses a key limitation of standard INRs, where expressivity is tied to the complexity of a static mapping. This suggests that *latent evolution* provides a principled way to expand the represented function space without directly increasing parameterization complexity.

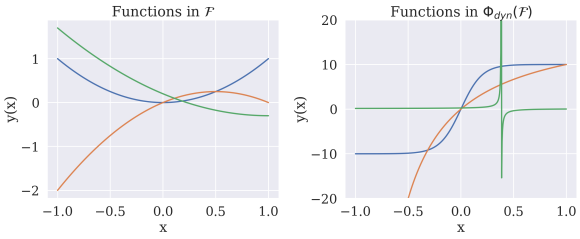

Figure 2: Three example functions from the original class $\mathcal{F}$ (left) and the expanded class $\Phi_{\text{dyn}}(\mathcal{F})$ (right), with different colors representing distinct functions.

3 Methodology

3.1 Dynamic Implicit Neural Representations

Motivated by the observation that dynamical flows can expand a static function class into a richer one, we propose *DINR*. Unlike standard INRs, which directly parameterize the target mapping in a static feedforward manner (1), DINR represents the target as the *solution of a dynamical system* in latent space. Specifically, DINR models the latent state $\mathbf{z}(t)$ as evolving according to:

$$\frac{d\mathbf{z}(t)}{dt} = f(\mathbf{z}(t), t; \theta_{\text{dyn}}), \quad \text{s.t.} \quad \mathbf{z}(0) = \phi(\mathbf{x}; \theta_{\text{in}}), \tag{3}$$

where f is a learnable vector field defining the instantaneous dynamics. The target representation is then defined implicitly as the solution trajectory of this ODE at time T (typically $T = 1$), and the final output is decoded from the terminal latent state:

$$\hat{\mathbf{y}} = \psi\left(\mathbf{z}(T)\right) = \psi\left(\mathbf{z}(0) + \int_0^T f(\mathbf{z}(t), t; \theta_{\text{dyn}})\, dt\right). \tag{4}$$

Table 1: Conceptual comparison highlighting the shift from static function parameterization (INR) to dynamical system-based representation (DINR).

	Network output	How represented	Computation style	How input used
INR	Target function	Direct mapping	Static feedforward	Single pass
DINR	Instantaneous dynamics	ODE solution (integral)	Dynamical evolution	Iterative trajectory updates

Thus, f does not directly parameterize the target mapping; instead, it parameterizes a transformation that expands the function class induced by the underlying static INR without a proportional increase in trainable parameters.

In practice, the dynamics is approximated via numerical ODE solvers, leading to the discrete iteration:

$$\mathbf{z}_0 = \phi(\mathbf{x}; \theta_{\text{in}}), \;\; \mathbf{z}_{k+1} = \mathbf{z}_k + \Delta t \cdot f(\mathbf{z}_k, t_k; \theta_{\text{dyn}}) \text{ for } 0 \leq k \leq N-1, \;\; \hat{\mathbf{y}}_{\text{DINR}} = \psi(\mathbf{z}_N; \theta_{\text{out}}), \quad (5)$$

where $T = N\Delta t$. Unlike deep residual networks, where each layer typically employ independent parameters, DINR shares the vector field f across the trajectory. The resulting repeated application of the same transformation produces a richer mapping through continuous composition, while keeping the underlying parameterization compact.

DINR has the following distinctive features, which are also summarized in Table 1.

- **A new way to enhance expressivity:** DINR expands the function class of a static INR by evolving latent features under a shared vector field, enabling richer mappings without proportionally increasing the number of trainable parameters.
- **Compact high-fidelity representation:** By expanding the effective function class through continuous latent evolution, DINR enables compact networks to represent complex signals with fewer parameters (see Section 5). Although ODE integration introduces additional computational overhead, this is offset by the reduced network size.
- **Model-agnostic augmentation:** DINR can be applied to existing INR architectures by replacing their static latent transformation with a shared dynamical system, preserving the original input and output components while introducing trajectory-based representation (see Section 5).

3.2 Kinetic Energy Regularization

Given that the final representation is derived from the trajectory's evolution, it is desired to avoid unnecessarily irregular or oscillatory trajectories as they can lead to unstable feature evolution and poor generalization. We therefore adopt a *kinetic energy (KE) regularizer* to encourage economical latent dynamics: the overall training objective, given a dataset $\{\mathbf{x}^{(i)}, \mathbf{y}^{(i)}\}_{i=1}^{M}$, becomes

$$\min_{\theta} \left[\frac{1}{M} \sum_{i=1}^{M} \mathcal{L}_{\text{data}} \left(\hat{\mathbf{y}}_{\text{DINR}}^{(i)}(\theta), \mathbf{y}^{(i)} \right) + \alpha \mathcal{L}_{\text{KE}}(\theta) \right], \quad \mathcal{L}_{\text{KE}} = \sum_{k=0}^{N-1} \| f(\mathbf{z}_k^{(i)}, t_k) \|_2^2 \, \Delta t, \quad (6)$$

where $\alpha > 0$ is a regularization parameter. By penalizing the magnitude of the velocity accumulated along the trajectory, $\mathcal{L}_{\text{KE}}$ encourages smoother and more controlled latent evolution, reducing tendency to fit training data through unnecessarily complex trajectories and thereby improving generalization (see Section 5). Further theoretical analysis of its effect on generalization is provided in Appendix A.3. Unlike weight decay and gradient clipping, which act on parameters or optimization updates, the KE regularizer directly regularizes the latent dynamics, making it particularly suited to DINR's trajectory-based representation.

4 Theoretical Analyses

Building on the DINR framework, we now provide theoretical insights into why latent dynamics enhance both expressivity and trainability. Intuitively, these benefits arise from lifting a single static transformation (1) with continuous dynamical latent updates (4), allowing for the emergence of complex functions and richer gradient patterns.

4.1 Expressivity via Rademacher Complexity

Standard INRs compute a latent feature $\mathbf{z}_1 = f(\phi(\mathbf{x}))$ in a single forward pass, constraining the range of functions they can represent. In contrast, DINR evolves the latent features $\mathbf{z}(t)$ along a trajectory under a learnable vector field f, effectively composing multiple incremental transformations. Even simple f can generate highly complex mappings when integrated over time, as illustrated in Section 2.

Formally, the Rademacher complexity $\mathcal{R}_n(\mathcal{F})$ quantifies the capacity of a function class $\mathcal{F}$, where n denotes the number of training samples. For DINR, the recursive evolution of latent features enlarges the hypothesis space relative to a static INR:

Theorem 4.1 (Expressivity, Informal)**.** *Let $\mathcal{F}_{\text{INR}}$ denote the function class of a standard INR, and $\mathcal{F}_{\text{DINR}}$ mark the function class induced by DINR with integration time T. Then,*

$$\mathcal{R}_n(\mathcal{F}_{\text{DINR}}) \gg \mathcal{R}_n(\mathcal{F}_{\text{INR}}).$$

Formal statements and proofs are provided in Appendix A. This result shows that DINR can represent a broader function class than INR and, as demonstrated in Section 5, achieve superior approximations with smaller networks than INR. Intuitively, this enhanced expressivity stems from the cumulative effect of incremental latent updates, where each integration step compounds the underlying vector field to produce more intricate mappings.

Regularizing Feature Dynamics. We also analyze the effect of the KE regularizer in (6), which discourages overly complex latent dynamics, thereby improving generalization while preserving the model's expressive capacity; see Appendix A.3 for details.

4.2 Trainability via Gradient Diversity

DINR improves optimization by increasing gradient diversity introduced by the dynamical trajectory of the latent states. INR relies on a single evaluation point, whereas DINR evaluates f along a dynamical trajectory $\{\mathbf{z}_k\}_{k=0}^{N}$, which leads to the following distinction, roughly speaking:

$$\frac{\partial \hat{\mathbf{y}}_{\text{INR}}}{\partial \theta_{\text{stat}}}(x) \in \operatorname{span}\left\{\frac{\partial f}{\partial \theta}(\mathbf{z}_0)\right\}, \quad \frac{\partial \hat{\mathbf{y}}_{\text{DINR}}}{\partial \theta_{\text{dyn}}}(x) \in \operatorname{span}\left\{\frac{\partial f}{\partial \theta}(\mathbf{z}_0), \ldots, \frac{\partial f}{\partial \theta}(\mathbf{z}_N)\right\}.$$

While INR computes the gradient at a single point $\mathbf{z}_0$, the gradients at subsequent states $\mathbf{z}_k$ in DINR explore different regions of the latent space, providing more diverse directional information. The key insight is that the gradients at each $\mathbf{z}_k$ are generally not aligned with each other, which expand the effective gradient space, directly increasing the rank of the associated Neural Tangent Kernel (NTK).

Theorem 4.2 (Gradient Diversity, Informal)**.** *Under mild regularity and non-degeneracy conditions,*

$$\operatorname{rank}(\Theta_{\text{DINR}}) > \operatorname{rank}(\Theta_{\text{INR}}),$$

where Θ_{DINR} and Θ_{INR} denote the NTK of DINR and INR, respectively.

Formal statements and proofs are provided in Appendix B. This result suggests that latent evolution increases gradient diversity, providing a richer subspace of descent directions and mitigating gradient stagnation in flat regions. Empirically, we observe faster convergence and improved recovery of high-frequency components (Section 5), though this analysis primarily highlights expanded gradient diversity rather than providing a direct guarantee of convergence rates.

5 Experiments

5.1 Experimental Setup

Data & Backbones. We evaluate continuous representations of (1) natural images (DIV2K (Agustsson & Timofte, 2017) and Kodak (Company, 1999)), (2) scientific fields (Turbulence (Holmes, 2012), Global Weather (Hersbach et al., 2020)), and (3) sparse three-dimensional volumes (Cloud-topped Boundary Layer (Boucher et al., 2013) and Cryo-EM (Gu et al., 2022)). The image benchmarks test broad INR reconstruction performance; the scientific data test representation of multiscale fields; and cryo-EM provides a domain-specific measure of recovered spatial resolution. We use FFNet (Tancik

et al., 2020), SIREN (Sitzmann et al., 2020), WIRE (Saragadam et al., 2023), and FINER (Liu et al., 2024) as static INR backbones. For each backbone, DINR replaces the static latent transformation with the shared, time-conditioned vector field in (4), while retaining the backbone's input encoding and output decoder. Further details are provided in Appendices D.

Evaluation Protocol. All models are implemented in PyTorch and optimized on a single NVIDIA RTX A6000 (48 GB) GPU. Unless otherwise specified, we train all models using the Adam optimizer (Kingma, 2014) with an initial learning rate of η=10^{-3} under identical settings. For each trained representation, we evaluate (1) *storage* (parameter count and serialized model size), (2) *reconstruction quality* (MSE, PSNR, and SSIM), and (3) *computation* (forward FLOPs, inference latency, and peak training memory). All models are serialized using the same numerical precision. Replication code is available at `https://anonymous.4open.science/r/DINR_Code-3EB2`.

5.2 Main Results

Our main findings regarding the proposed dynamical formulation of INRs are twofold:

① **Superior Fidelity–Storage Efficiency.** *DINRs improve the fidelity–storage Pareto frontier.*

② **Enhanced NTK Properties.** *Latent dynamics induce more favorable empirical NTK geometry.*

For ①, Figure 3 compares reconstruction fidelity with stored model size on DIV2K, Kodak, and Turbulence. Each marker pairs a static INR with its DINR counterpart under the same backbone; moving upward indicates higher PSNR and moving left indicates less storage. Across all three domains and all four backbones, the dynamical point lies above and to the left of its static counterpart. Thus, the result is not restricted to one activation, coordinate encoding, or signal class: the shared latent flow improves the operating point of Fourier-feature, sinusoidal, wavelet, and FINER-based INRs. The magnitude of the vertical improvement varies with the backbone and dataset, as it should for a model-agnostic augmentation. The consistent horizontal displacement is nevertheless equally consequential: higher fidelity is obtained with a smaller representation, rather than by simply increasing the number of stored weights.

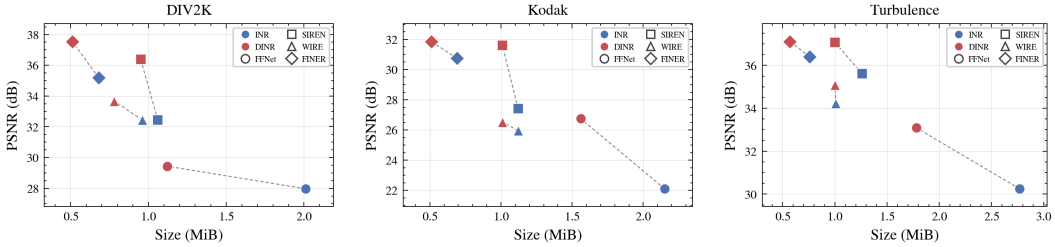

Figure 3: Fidelity–storage trade-off. Each line joins a static INR to its DINR counterpart for the same backbone; upper-left is preferable.

For ②, we analyze DINR's optimization behavior via the empirical NTK, which characterizes training behavior in the infinite-width limit and provides meaningful insights into its inductive bias. Following standard practice, we construct the empirical NTK $K_{ij} = \sum \left\langle \frac{\partial \hat{y}(x_i)}{\partial \theta}, \frac{\partial \hat{y}(x_j)}{\partial \theta} \right\rangle$ for each model using a random subset of coordinates and examine its eigenspectrum. The kernel matrix is formed by the inner product of the gradients of the network's outputs with respect to all trainable parameters θ. As shown in Figure 4, the dynamical models give better results supporting Theorem 4.2. Specifically:

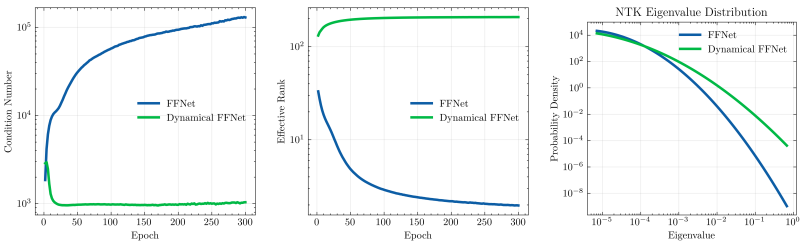

Figure 4: Empirical NTK geometry for the representative FFNet experiment.

1. A *higher effective rank* indicates that DINR has access to a richer set of gradient directions, enhancing trainability. This allows the model to learn a more diverse set of functions in practice, helping it capture complex, high-frequency details that static INRs may miss due to spectral bias.
2. The *lower condition number* implies a better-conditioned optimization landscape, leading to more stable and faster convergence. This aligns with our observation that Dynamical SIREN often succeeds where standard SIREN struggles, and DINRs generally train more robustly.
3. The *slower decay of leading eigenvalues* suggests that more gradient directions contribute meaningfully during training, allowing DINR to learn from a broader set of data modes rather than focusing primarily on the dominant low-frequency components. These findings are consistent across different models. Additional results are provided in Appendix E.1.

To verify that the preceding improvements are not artifacts of favorable initialization, selected instances, or asymmetric tuning, we independently tune each model family under the same validation protocol and evaluate the selected FFNet configurations across multiple instances and random seeds. Figure 5 shows paired comparisons, with each connection corresponding to the same image (averaged across seeds) or, for Turbulence, the same random seed. DINR consistently outperforms INR across all evaluated DIV2K and Kodak images, as well as across the 20 paired Turbulence runs. These results indicate that the favorable fidelity–storage trade-off observed above is reproducible across instances and random initialization rather than being driven by a small number of favorable runs. Complete hyperparameter search spaces, selection protocols, per-seed results, and additional multi-run evaluations are provided in the Appendix E.2.

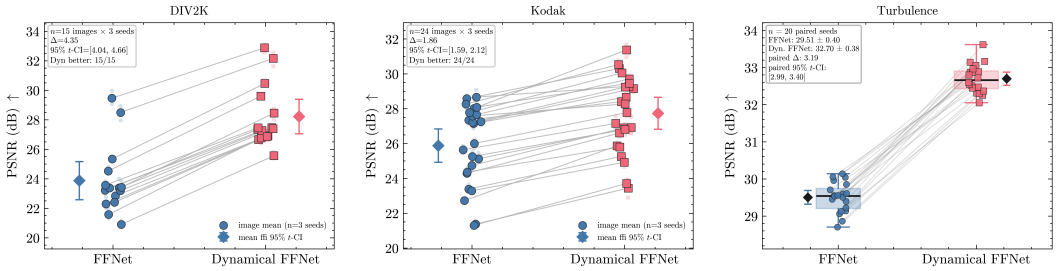

Figure 5: Paired multi-run FFNet results. Diamonds and bars denote the mean and 95% t-interval.

5.3 Model Analysis

We next isolate the source of DINR's improvements and rule out alternative explanations such as increased computation, effective depth, or insufficiently tuned static baselines. We conduct two complementary sets of controlled experiments:

① **Compute-Matched Comparisons.** *We compare DINR against static INRs under matched computational budgets, using both FLOP-matched architectures and wall-clock-time matching.*

② **Architectural Ablations.** *We systematically ablate key components of DINR, including the recurrent latent dynamics and kinetic-energy regularization, to quantify their individual contributions.*

For ①, we consider two complementary compute-matched controls across representative baselines: (i) *FLOP matching*, where static architectures are scaled to match DINR's forward-FLOP budget, and (ii) *time matching*, where static INRs are trained for additional epochs under the same wall-clock budget. Due to space constraints, we present the FLOP-matched results here and defer the time-matched study to the Appendix E.3. For (i), we construct a FLOP-matched FFNet by searching over static depth–width configurations under DINR's forward-FLOP budget and selecting the best configuration on validation data. Figure 6 compares DINR with the original FFNet and the independently tuned FLOP-matched FFNet. Scaling the static model improves reconstruction over the original FFNet, confirming that additional computation is beneficial. Nevertheless, DINR consistently achieves better reconstruction quality while retaining a substantially smaller stored representation. Its training trajectory also reaches and maintains lower reconstruction error under the common training schedule. These results indicate that DINR's gains cannot be explained solely by additional forward computation, supporting the effectiveness of its dynamical lifting formulation.

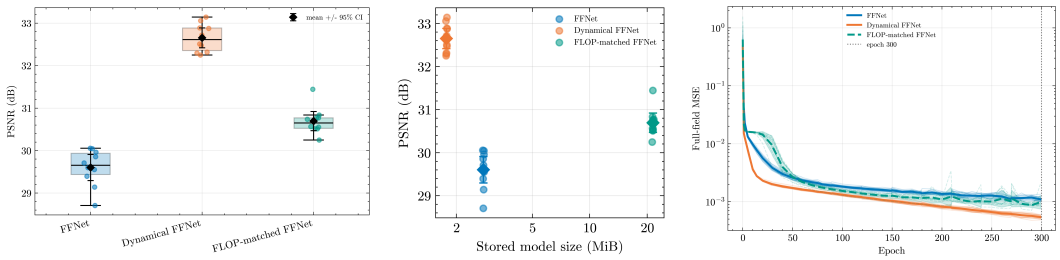

Figure 6: FLOP-matched experiments. Scaling a static FFNet to match DINR's forward FLOPs improves the static baseline but does not recover DINR's reconstruction quality, and it requires a substantially larger stored model.

For ②, we ask whether DINR's gains arise specifically from *time-conditioned latent dynamics*, rather than repeated application of a shared network. We conduct a controlled recurrent ablation on Turbulence, keeping the architecture, parameter count, number of function evaluations, optimizer, and training schedule fixed while varying only the update rule. The scaled-residual and DINR variants share the same Δt, isolating the contribution of explicit time conditioning. As shown in Table 2, repeated evaluation alone does not reproduce DINR's performance. Across both FFNet and SIREN, DINR consistently outperforms all tied-weight controls, including the scaled-residual variant.

Table 2: Recurrent Design Ablation.

Backbone	Update rule	MSE ↓	PSNR ↑	SSIM ↑
FFNet	$z_{k+1} = f(z_k)$	9.61×10^{-3}	20.17	0.50
	$z_{k+1} = z_k + f(z_k)$	1.88×10^{-2}	17.26	0.42
	$z_{k+1} = z_k + \Delta t\, f(z_k)$	6.63×10^{-3}	21.78	0.57
	$z_{k+1} = z_k + \Delta t\, f(z_k, t_k)$	$\mathbf{4.50 \times 10^{-4}}$	**33.46**	**0.88**
SIREN	$z_{k+1} = f(z_k)$	1.63×10^{-2}	17.89	0.52
	$z_{k+1} = z_k + f(z_k)$	2.91×10^{-2}	15.36	0.07
	$z_{k+1} = z_k + \Delta t\, f(z_k)$	2.80×10^{-3}	25.78	0.67
	$z_{k+1} = z_k + \Delta t\, f(z_k, t_k)$	$\mathbf{1.94 \times 10^{-4}}$	**38.00**	**0.96**

Still for ②, having established the efficacy of *latent dynamics*, we next evaluate whether *KE regularization* in Equation 6 confers synergistic robustness. Specifically, we systematically disentangle the individual contribution of continuous-time dynamics (DINR vs. INR) from that of kinetic energy regularization (DINR+KE vs. DINR); complete experimental specifications are provided in Appendix E.4. Furthermore, Figure 7 investigates whether the performance gains attributed to KE can be explained away by conventional parameter penalties. Across an extensive weight-decay sweep, DINR consistently surpasses the static FFNet baseline, with DINR+KE yielding superior reconstruction fidelity. These empirical findings demonstrate that KE imparts an inductive bias, operating as a trajectory-level regularizer, that cannot be emulated by standard weight decay alone.

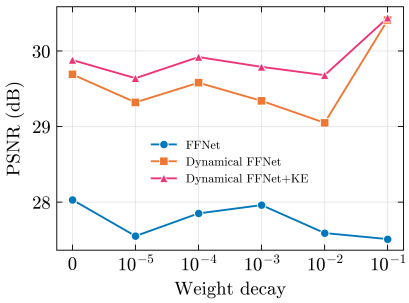

Figure 7: KE Regularization Ablation.

DINR introduces two key hyperparameters: the number of integration steps (N) and the kinetic-energy regularization coefficient λ_{KE}. Figure 8 reveals two consistent trends. First, increasing (N) improves reconstruction as the latent dynamics are more finely resolved, with diminishing returns beyond a moderate integration resolution. Second, weaker KE regularization generally yields better clean-data performance: overly large λ_{KE} restricts the latent dynamics and reduces expressivity, whereas small values provide regularization without substantially limiting reconstruction. These results suggest that DINR is relatively insensitive once the trajectory is sufficiently resolved, while KE regularization should remain modest when robustness is not the primary objective.

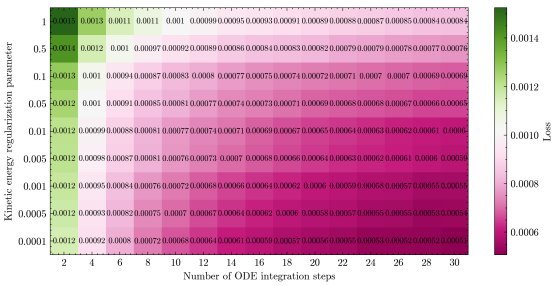

Figure 8: Sensitivity analysis.

5.4 Scientific Applications

Given its superior fidelity–storage properties, a natural application for the proposed DINR beyond standard computer vision benchmarks is scientific data compression. Consequently, we systematically evaluate DINR across a broad spectrum of scientific datasets, with complete results provided in Appendix E.5, E.6, and E.7. Here, we highlight two representative case studies:

① **High Frequency Data.** *Compression of weather and turbulence data.*

② **3D Sparse Data.** *Representation learning for Cryo-EM data.*

For ①, the 2D power spectra reveal that while all models effectively capture low-frequency components, static INRs exhibit a pronounced drop off in the high-frequency regions, a clear manifestation of the spectral bias. In stark contrast, DINR models maintain significantly more energy in the outer regions of the spectrum, indicating a superior ability to represent high-frequency details.

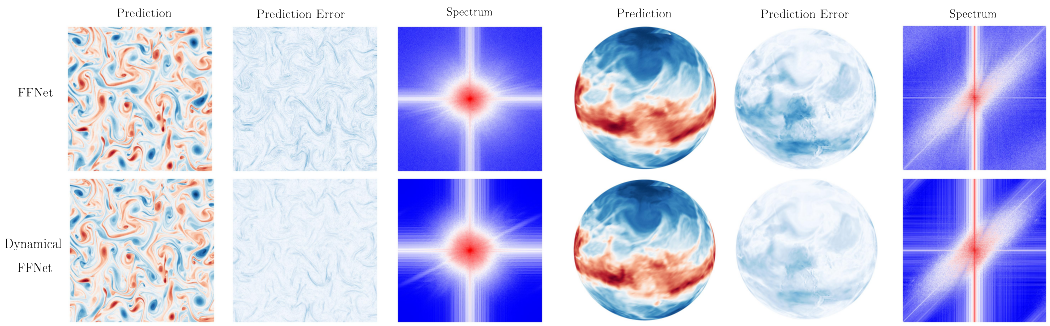

Figure 9: Qualitative evaluation of INR/DINR for turbulence and global weather data.

For ②, DINR produces a reconstruction that is visibly closer to the ground-truth density than its static counterpart, recovering finer protrusions and local structural irregularities while reducing the bulky, fragmented artifacts seen in FFNet (Figure 10). This improvement is also reflected in the Fourier domain: the FSC curve for DINR stays above that of FFNet over nearly the entire frequency range.

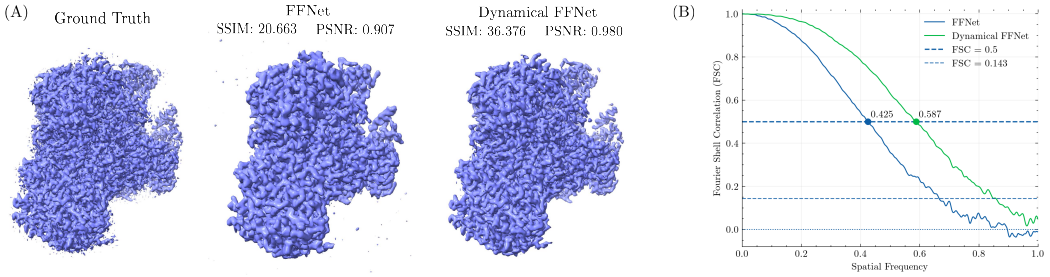

Figure 10: (A) Cryo-EM reconstruction results and (B) Fourier shell correlation curves.

6 Conclusion

We introduced DINR, *Dynamical Implicit Neural Representations*, a framework that models latent features as a continuous-time dynamical system. By evolving features via incremental updates, DINR enhances the expressivity of compact networks, improves trainability, and enables high-fidelity reconstruction of fine details. Our theoretical analysis supports this stepwise evolution boosts representational power and gradient diversity, explaining the empirical gains. This work provides a principled approach to network design that is valuable for applications where parameter efficiency and signal fidelity are critical. Looking forward, several key research directions, including improving computational efficiency through adaptive ODE integration, deepening the theoretical understanding of NTK behavior, scaling DINR to multi-modal domains, and establishing a unified framework for dynamical neural architectures, are worth exploring.

AI Disclosure. Generative AI tools were used only to identify grammatical errors and to improve the clarity and fluency of the manuscript's English. The authors independently reviewed all AI-generated suggestions and made the final revisions themselves. No generative AI tools were used for any of the required-disclosure tasks listed in the ICLR 2027 AI policy. We take responsibility for the final content of this work.

REFERENCES

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 126–135, 2017.

Akio Arakawa. Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. part i. *Journal of computational physics*, 135(2): 103–114, 1997.

Sanjeev Arora, Rong Ge, Behnam Neyshabur, and Yi Zhang. Stronger generalization bounds for deep nets via a compression approach. In *International conference on machine learning*, pp. 254–263. PMLR, 2018.

Sanjeev Arora, Simon S Du, Wei Hu, Zhiyuan Li, Russ R Salakhutdinov, and Ruosong Wang. On exact computation with an infinitely wide neural net. *Advances in neural information processing systems*, 32, 2019.

Maor Ashkenazi and Eran Treister. Towards croppable implicit neural representations. *Advances in Neural Information Processing Systems*, 37:31473–31503, 2024.

Michael Baldauf, Axel Seifert, Jochen Förstner, Detlev Majewski, Matthias Raschendorfer, and Thorsten Reinhardt. Operational convective-scale numerical weather prediction with the cosmo model: Description and sensitivities. *Monthly Weather Review*, 139(12):3887–3905, 2011.

Peter L Bartlett and Shahar Mendelson. Rademacher and gaussian complexities: Risk bounds and structural results. *Journal of machine learning research*, 3(Nov):463–482, 2002.

Peter L Bartlett, Dylan J Foster, and Matus J Telgarsky. Spectrally-normalized margin bounds for neural networks. *Advances in neural information processing systems*, 30, 2017.

Yizhak Ben-Shabat, Chamin Hewa Koneputugodage, Sameera Ramasinghe, and Stephen Gould. Neural experts: Mixture of experts for implicit neural representations. *Advances in Neural Information Processing Systems*, 37:101641–101670, 2024.

Nils Bjorck, Carla P Gomes, Bart Selman, and Kilian Q Weinberger. Understanding batch normalization. *Advances in neural information processing systems*, 31, 2018.

Blake Bordelon, Abdulkadir Canatar, and Cengiz Pehlevan. Spectrum dependent learning curves in kernel regression and wide neural networks. In *International Conference on Machine Learning*, pp. 1024–1034. PMLR, 2020.

Olivier Boucher, David Randall, Paulo Artaxo, Christopher Bretherton, Graham Feingold, Piers Forster, V-M Kerminen, Yutaka Kondo, Hong Liao, Ulrike Lohmann, et al. Clouds and aerosols. In *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*, pp. 571–657. Cambridge University Press, 2013.

Zhicheng Cai, Hao Zhu, Qiu Shen, Xinran Wang, and Xun Cao. Batch normalization alleviates the spectral bias in coordinate networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 25160–25171, 2024.

Yuan Cao, Zhiying Fang, Yue Wu, Ding-Xuan Zhou, and Quanquan Gu. Towards understanding the spectral bias of deep learning. *arXiv preprint arXiv:1912.01198*, 2019.

Guikun Chen and Wenguan Wang. A survey on 3d gaussian splatting. *ACM Computing Surveys*, 2024.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8628–8638, 2021.

Eastman Kodak Company. Kodak lossless true color image suite. In *https://r0k.us/graphics/kodak/*, 1999.

Simon S Du, Xiyu Zhai, Barnabas Poczos, and Aarti Singh. Gradient descent provably optimizes over-parameterized neural networks. *arXiv preprint arXiv:1810.02054*, 2018.

Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam Oberman. How to train your neural ODE: the world of jacobian and kinetic regularization. In *International conference on machine learning*, pp. 3154–3164. PMLR, 2020.

Tomer Galanti, Mengjia Xu, Liane Galanti, and Tomaso Poggio. Norm-based generalization bounds for sparse neural networks. *Advances in Neural Information Processing Systems*, 36:42482–42501, 2023.

Will Grathwohl, Ricky TQ Chen, Jesse Bettencourt, Ilya Sutskever, and David Duvenaud. Ffjord: Free-form continuous dynamics for scalable reversible generative models. *arXiv preprint arXiv:1810.01367*, 2018.

Jinke Gu, Tianya Liu, Runyu Guo, Laixing Zhang, and Maojun Yang. The coupling mechanism of mammalian mitochondrial complex i. *Nature Structural & Molecular Biology*, 29(2):172–182, 2022.

Tobias Günther, Alexander Kuhn, and Holger Theisel. Mcftle: Monte carlo rendering of finite-time lyapunov exponent fields. In *Computer Graphics Forum*, volume 35, pp. 381–390. Wiley Online Library, 2016.

Boris Hanin and Mihai Nica. Finite depth and width corrections to the neural tangent kernel. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=SJgndT4KwB.

Joshua Hanson and Maxim Raginsky. Rademacher complexity of neural odes via chen-fliess series. In *6th Annual Learning for Dynamics & Control Conference*, pp. 758–769. PMLR, 2024.

Bo He, Xitong Yang, Hanyu Wang, Zuxuan Wu, Hao Chen, Shuaiyi Huang, Yixuan Ren, Ser-Nam Lim, and Abhinav Shrivastava. Towards scalable neural representation for diverse videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6132–6142, 2023.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Hans Hersbach, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, Carole Peubey, Raluca Radu, Dinand Schepers, et al. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020.

Philip Holmes. *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge university press, 2012.

Jiri Hron, Yasaman Bahri, Jascha Sohl-Dickstein, and Roman Novak. Infinite attention: Nngp and ntk for deep attention networks. In *International Conference on Machine Learning*, pp. 4376–4386. PMLR, 2020.

Edward L Ince. *Ordinary differential equations*. Courier Corporation, 2012.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. *Advances in neural information processing systems*, 31, 2018.

Weronika Jakubowska, Mikołaj Zieliński, Rafał Tobiasz, Krzysztof Byrski, Maciej Zięba, Dominik Belter, and Przemysław Spurek. Gainer: Geometry-aware implicit network representation. *arXiv preprint arXiv:2511.20924*, 2025.

Dhananjaya Jayasundara, Sudarshan Rajagopalan, Yasiru Ranasinghe, Trac D Tran, and Vishal M Patel. Sinr: Sparsity driven compressed implicit neural representations. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 3061–3070, 2025.

Kelvin Kan, Xingjian Li, Benjamin J Zhang, Tuhin Sahai, Stanley Osher, and Markos A Katsoulakis. Optimal control for Transformer architectures: Enhancing generalization, robustness and efficiency. *arXiv preprint arXiv:2505.13499*, 2025.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, George Drettakis, et al. 3d gaussian splatting for real-time radiance field rendering. *ACM Trans. Graph.*, 42(4):139–1, 2023.

Taeyoung Kim and Myungjoo Kang. Bounding the rademacher complexity of fourier neural operators. *Machine Learning*, 113(5):2467–2498, 2024.

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

Andreas Klamt. The cosmo and cosmo-rs solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):699–709, 2011.

Chamin Hewa Koneputugodage, Yizhak Ben-Shabat, Sameera Ramasinghe, and Stephen Gould. Viˆ 3nr: Variance informed initialization for implicit neural representations. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 13477–13486, 2025.

Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W Mahoney. Characterizing possible failure modes in physics-informed neural networks. *Advances in Neural Information Processing Systems*, 34:26548–26560, 2021.

Jaehoon Lee, Yasaman Bahri, Roman Novak, Samuel S Schoenholz, Jeffrey Pennington, and Jascha Sohl-Dickstein. Deep neural networks as gaussian processes. In *International Conference on Learning Representations*, 2018.

Jaehoon Lee, Lechao Xiao, Samuel Schoenholz, Yasaman Bahri, Roman Novak, Jascha Sohl-Dickstein, and Jeffrey Pennington. Wide neural networks of any depth evolve as linear models under gradient descent. *Advances in neural information processing systems*, 32, 2019.

Zhemin Li, Hongxia Wang, and Deyu Meng. Regularize implicit neural representation by itself. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10280–10288, 2023.

Zhemin Li, Liyuan Ma, Hongxia Wang, Yaoyun Zeng, and Xiaolong Han. Learning input encodings for kernel-optimal implicit neural representations. In *Forty-second International Conference on Machine Learning*, 2025.

Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, and Xun Cao. Finer: Flexible spectral-bias tuning in implicit neural representation by variable-periodic activation functions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2713–2722, 2024.

Pierre Marion. Generalization bounds for neural ordinary differential equations and deep residual networks. *Advances in Neural Information Processing Systems*, 36:48918–48938, 2023.

Stefano Massaroli, Michael Poli, Jinkyoo Park, Atsushi Yamashita, and Hajime Asama. Dissecting neural ODEs. *Advances in neural information processing systems*, 33:3952–3963, 2020.

Andreas Maurer. A vector-contraction inequality for rademacher complexities. In *International Conference on Algorithmic Learning Theory*, pp. 3–17. Springer, 2016.

Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4460–4470, 2019.

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.

Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning*. MIT press, 2018.

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.

Derek Onken, Samy Wu Fung, Xingjian Li, and Lars Ruthotto. OT-Flow: Fast and accurate continuous normalizing flows via optimal transport. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 9223–9232, 2021.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 165–174, 2019.

Youngsuk Park, Danielle Maddix, François-Xavier Aubet, Kelvin Kan, Jan Gasthaus, and Yuyang Wang. Learning quantile functions without quantile crossing for distribution-free time series forecasting. In *International conference on artificial intelligence and statistics*, pp. 8127–8150. PMLR, 2022.

Maithra Raghu, Ben Poole, Jon Kleinberg, Surya Ganguli, and Jascha Sohl-Dickstein. On the expressive power of deep neural networks. In *international conference on machine learning*, pp. 2847–2854. PMLR, 2017.

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.

Sameera Ramasinghe and Simon Lucey. Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps. In *European Conference on Computer Vision*, pp. 142–158. Springer, 2022.

Pu Ren, N Benjamin Erichson, Shashank Subramanian, Omer San, Zarija Lukic, and Michael W Mahoney. Superbench: A super-resolution benchmark dataset for scientific machine learning. *arXiv preprint arXiv:2306.14070*, 2023.

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G Baraniuk. Wire: Wavelet implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18507–18516, 2023.

Hemanth Saratchandran, Sameera Ramasinghe, and Simon Lucey. From activation to initialization: Scaling insights for optimizing neural fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 413–422, 2024.

Kexuan Shi, Hai Chen, Leheng Zhang, and Shuhang Gu. Inductive gradient adjustment for spectral bias in implicit neural representations. *arXiv preprint arXiv:2410.13271*, 2024.

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, 2020.

Daniel Soudry, Elad Hoffer, Mor Shpigel Nacson, Suriya Gunasekar, and Nathan Srebro. The implicit bias of gradient descent on separable data. *Journal of Machine Learning Research*, 19(70):1–57, 2018.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.

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.

Jacob Trauger and Ambuj Tewari. Sequence length independent norm-based generalization bounds for transformers. In *International Conference on Artificial Intelligence and Statistics*, pp. 1405–1413. PMLR, 2024.

Vladimir N Vapnik and Alexey Ya Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16(2):264–280, 1971.

Alexander Vidal, Samy Wu Fung, Luis Tenorio, Stanley Osher, and Levon Nurbekyan. Taming hyperparameter tuning in continuous normalizing flows using the JKO scheme. *Scientific reports*, 13(1):4501, 2023.

Nikolaos N Vlassis and WaiChing Sun. Sobolev training of thermodynamic-informed neural networks for interpretable elasto-plasticity models with level set hardening. *Computer Methods in Applied Mechanics and Engineering*, 377:113695, 2021.

Joanna Waczyńska, Tomasz Szczepanik, Piotr Borycki, Sławomir Tadeja, Thomas Bohné, and Przemysław Spurek. Mirage: Editable 2d images using gaussian splatting. *arXiv preprint arXiv:2410.01521*, 2024.

Colin Wei and Tengyu Ma. Data-dependent sample complexity of deep neural networks via lipschitz augmentation. *Advances in neural information processing systems*, 32, 2019.

Tong Wu, Yu-Jie Yuan, Ling-Xiao Zhang, Jie Yang, Yan-Pei Cao, Ling-Qi Yan, and Lin Gao. Recent advances in 3d gaussian splatting. *Computational Visual Media*, 10(4):613–642, 2024.

Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, volume 41, pp. 641–676. Wiley Online Library, 2022.

Greg Yang. Tensor programs ii: Neural tangent kernel for any architecture. *arXiv preprint arXiv:2006.14548*, 2020.

Liu Yang and George Em Karniadakis. Potential flow generator with l_2 optimal transport regularity for generative models. *IEEE Transactions on Neural Networks and Learning Systems*, 33(2): 528–538, 2020.

Dong Yin, Ramchandran Kannan, and Peter Bartlett. Rademacher complexity for adversarially robust generalization. In *International conference on machine learning*, pp. 7085–7094. PMLR, 2019.

Wentao Yuan, Qingtian Zhu, Xiangyue Liu, Yikang Ding, Haotian Zhang, and Chi Zhang. Sobolev training for implicit neural representations with approximated image derivatives. In *European Conference on Computer Vision*, pp. 72–88. Springer, 2022.

Gizem Yüce, Guillermo Ortiz-Jiménez, Beril Besbinar, and Pascal Frossard. A structured dictionary perspective on implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19228–19238, 2022.

Xinjie Zhang, Xingtong Ge, Tongda Xu, Dailan He, Yan Wang, Hongwei Qin, Guo Lu, Jing Geng, and Jun Zhang. Gaussianimage: 1000 fps image representation and compression by 2d gaussian splatting. In *European Conference on Computer Vision*, pp. 327–345. Springer, 2024.

Zelin Zhao, Fenglei Fan, Wenlong Liao, and Junchi Yan. Grounding and enhancing grid-based models for neural fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19425–19435, 2024.

A Rademacher Complexity Analysis

A.1 Background: Empirical Rademacher Complexity

We begin by recalling the formal definition of empirical Rademacher complexity used to quantify the expressiveness of a hypothesis class.

The empirical Rademacher complexity is a standard measure of the expressiveness of a hypothesis class with respect to a fixed sample. It quantifies how well functions in the class can fit random noise and serves as a key tool for deriving generalization bounds.

Definition A.1 (Empirical Rademacher Complexity)**.** Let $\mathcal{F}$ be a class of real-valued functions defined on $\mathcal{X}$, and let $S = \{x_1, \ldots, x_n\} \subset \mathcal{X}$ be a fixed sample of size n. The empirical Rademacher complexity of $\mathcal{F}$ with respect to S is defined as:

$$\mathcal{R}_n(\mathcal{F}) := \mathbb{E}_{\boldsymbol{\sigma}} \left[\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \sigma_i f(x_i) \right],$$

where $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_n)$ are independent and identically distributed (i.i.d.) Rademacher variables uniformly sampled from $\{-1, +1\}$.

This definition captures the ability of the function class to align with arbitrary binary noise on a given sample. A higher Rademacher complexity indicates greater capacity and potentially greater risk of overfitting.

A.2 Theorems and Proofs

This appendix restates the theoretical results on the Rademacher complexity of standard INRs and dynamical implicit neural representations (DINRs) that were presented informally in Section 4.1, now providing precise assumptions and constants.

We begin by presenting the main result for DINRs in Corollary A.4, along with its supporting lemma. These results demonstrate the exponential dependence on trajectory length, arising from the recursive nature of the DINR dynamics. For comparison, we also provide an upper bound on the Rademacher complexity of standard INRs in Proposition A.5.

Remark on Assumptions. The assumptions made in this section concern only the Lipschitz continuity of the involved mappings. These assumptions are standard and widely satisfied in practice for neural networks. In particular, conventional networks equipped with typical activation functions, such as rectified linear unit (ReLU), sigmoid, or tanh, are known to be Lipschitz continuous with explicitly computable Lipschitz constants. Therefore, these conditions do not impose overly restrictive constraints and reflect realistic scenarios in modeling INRs.

Lemma A.2 (Lipschitz Continuity of the Discrete Flow Map)**.** *Consider the discrete dynamical system induced by the OC-FFN model:*

$$z_{k+1} = z_k + \Delta t \cdot f(z_k, t_k), \quad k = 0, 1, \ldots, N-1,$$

with initial condition $z_0 = \phi(x)$. Suppose that

- *$\phi : \mathbb{R}^{d_x} \to \mathbb{R}^{d_z}$ is L_ϕ-Lipschitz,*
- *For each t_k, $f(\cdot, t_k)$ is L_f-Lipschitz in the first argument (uniformly in k).*

Then, the discrete-time flow map $x \mapsto z_N$ is Lipschitz continuous, and one may bound

$$\|z_N(x_1) - z_N(x_2)\| \;\le\; L_\phi \, (1 + \Delta t \, L_f)^N \, \|x_1 - x_2\|.$$

In particular, when $N\Delta t = T$, we have the bound

$$(1 + \Delta t \, L_f)^N \le e^{L_f T},$$

resulting in

$$\|z_N(x_1) - z_N(x_2)\| \;\le\; L_\phi \, e^{L_f T} \, \|x_1 - x_2\|.$$

Proof. Let $x_1, x_2 \in \mathcal{X}$. Denote the corresponding trajectories by $\{z_k^{(1)}\}$ and $\{z_k^{(2)}\}$, initialized as

$$z_0^{(i)} = \phi(x_i), \quad i = 1, 2.$$

We prove by induction a one-step Lipschitz inequality. For the base case $k = 0$, we have

$$\|z_0^{(1)} - z_0^{(2)}\| = \|\phi(x_1) - \phi(x_2)\| \leq L_\phi \|x_1 - x_2\|.$$

Assume for some $k \geq 0$,

$$\|z_k^{(1)} - z_k^{(2)}\| \leq L_\phi (1 + \Delta t\, L_f)^k \|x_1 - x_2\|.$$

Then, at the next step,

$$\begin{aligned}
&\|z_{k+1}^{(1)} - z_{k+1}^{(2)}\| \\
&= \|z_k^{(1)} + \Delta t\, f(z_k^{(1)}, t_k) \; - \; (z_k^{(2)} + \Delta t\, f(z_k^{(2)}, t_k))\| \\
&\leq \|z_k^{(1)} - z_k^{(2)}\| + \Delta t \|f(z_k^{(1)}, t_k) - f(z_k^{(2)}, t_k)\| \\
&\leq \|z_k^{(1)} - z_k^{(2)}\| + \Delta t\, L_f \|z_k^{(1)} - z_k^{(2)}\| \\
&= (1 + \Delta t\, L_f) \|z_k^{(1)} - z_k^{(2)}\| \\
&\leq (1 + \Delta t\, L_f) \left(L_\phi (1 + \Delta t\, L_f)^k\right) \|x_1 - x_2\| \\
&= L_\phi (1 + \Delta t\, L_f)^{k+1} \|x_1 - x_2\|.
\end{aligned}$$

By induction, this inequality holds for all $k = 0, 1, \ldots, N$. In particular, for $k = N$,

$$\|z_N(x_1) - z_N(x_2)\| = \|z_N^{(1)} - z_N^{(2)}\| \leq L_\phi (1 + \Delta t\, L_f)^N \|x_1 - x_2\|.$$

This completes the proof. □

Proposition A.3 (Rademacher Complexity of DINRs)**.** *Let $\mathcal{F}_{\text{DINR}} = \{y : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y} \mid y(x) = \psi(z(T; x))\}$ be a hypothesis class defined on inputs $x \in \mathcal{X} \subset \mathbb{R}^{d_x}$ and let n denote the sample size. Suppose:*

- $\operatorname{diam}(\mathcal{X}) \leq D$,
- $\phi : \mathbb{R}^{d_x} \to \mathbb{R}^{d_z}$ *is a fixed, B_ϕ-bounded feature map,*
- $z(t) \in \mathbb{R}^{d_z}$ *solves $\dot{z}(t) = f(z(t), t)$ with $z(0) = \phi(x)$, and the vector field $f(z, t)$ is L_f-Lipschitz in z,*
- $\psi : \mathbb{R}^{d_z} \to \mathbb{R}^{d_y}$ *is L_ψ-Lipschitz.*

Then, the empirical Rademacher complexity with respect to samples of size n satisfies

$$\mathcal{R}_n(\mathcal{F}_{\text{DINR}}) \leq \frac{C \cdot L_\psi L_\phi D \cdot e^{L_f T} \cdot \sqrt{d_z d_y}}{\sqrt{n}},$$

for some universal constant C.

Proof. By Lemma A.2, the mapping $x \mapsto z(T; x)$ is Lipschitz with constant at most

$$L_{\text{flow}} := L_\phi e^{L_f T},$$

and the reachable set of the dynamics $\mathcal{Z}_T := \{z(T; x) : x \in \mathcal{X}\}$ admits an ϵ-covering with covering number:

$$\mathcal{N}(\epsilon, \mathcal{Z}_T, \|\cdot\|) \leq \left(\frac{L_{\text{flow}} D}{\epsilon}\right)_z^d.$$

Because $\psi : \mathbb{R}^{d_z} \to \mathbb{R}^{d_y}$ is L_ψ-Lipschitz, the composed function $y(x) = \psi(z(T; x))$ is $L_\psi L_{\text{flow}}$-Lipschitz in x. Thus,

$$\|y(x_1) - y(x_2)\| \leq L_\psi L_\phi e^{L_f T} \|x_1 - x_2\| =: L_{\text{comp}} \cdot \|x_1 - x_2\|.$$

As such, the diameter of the function class $\mathcal{F}_{\text{DINR}}$ satisfies:

$$\operatorname{diam}(\mathcal{F}_{\text{DINR}}) \leq L_{\text{comp}} D =: C_0.$$

To upper bound the empirical Rademacher complexity of vector-valued function classes $\mathcal{F}_{\text{DINR}} \subset \{f : \mathcal{X} \to \mathbb{R}^{d_y}\}$, we apply a standard vector-valued Dudley entropy integral bound (Bartlett & Mendelson, 2002; Maurer, 2016):

$$\mathcal{R}_n(\mathcal{F}_{\text{DINR}}) \leq \frac{12\sqrt{d_y}}{\sqrt{n}} \int_0^{\operatorname{diam}(\mathcal{F}_{\text{DINR}})} \sqrt{\log \mathcal{N}(\epsilon, \mathcal{F}_{\text{DINR}}, \|\cdot\|)}\, d\epsilon.$$

From the Lipschitz property of $y(x)$, the covering number of $\mathcal{F}_{\text{DINR}}$ satisfies:

$$\mathcal{N}(\epsilon, \mathcal{F}_{\text{DINR}}) \leq \left(\frac{L_{\text{comp}} D}{\epsilon}\right)_z^d.$$

Taking logarithms:

$$\log \mathcal{N}(\epsilon, \mathcal{F}_{\text{DINR}}) \leq d_z \cdot \log\left(\frac{L_\psi L_\phi D \cdot e^{L_f T}}{\epsilon}\right) = d_z \cdot \log\left(\frac{C_0}{\epsilon}\right).$$

Substituting into the entropy integral:

$$\mathcal{R}_n(\mathcal{F}_{\text{DINR}}) \leq \frac{12\sqrt{d_y d_z}}{\sqrt{n}} \int_0^{C_0} \sqrt{\log\left(\frac{C_0}{\epsilon}\right)}\, d\epsilon.$$

Now, use the change of variable $\epsilon = C_0 e^{-u}$, yielding:

$$\int_0^{C_0} \sqrt{\log\left(\frac{C_0}{\epsilon}\right)}\, d\epsilon = \int_0^{\infty} \sqrt{u} \cdot C_0 e^{-u}\, du = C_0 \int_0^{\infty} u^{1/2} e^{-u}\, du = C_0 \cdot \Gamma(3/2) = C_0 \cdot \frac{\sqrt{\pi}}{2}.$$

Combining all terms, we obtain:

$$\mathcal{R}_n(\mathcal{F}_{\text{DINR}}) \leq \frac{12\sqrt{d_y d_z}}{\sqrt{n}} \cdot C_0 \cdot \frac{\sqrt{\pi}}{2} = \frac{C \cdot L_\psi L_\phi D \cdot e^{L_f T} \cdot \sqrt{d_z d_y}}{\sqrt{n}}$$

for a universal constant $C = 6\sqrt{\pi}$. □

Corollary A.4. *Under the same assumptions as Proposition A.3, suppose additionally that the vector field $f : \mathbb{R}^{d_z} \to \mathbb{R}^{d_z}$ is implemented as a depth-ℓ feedforward neural network of the form*

$$f = f_\ell \circ f_{\ell-1} \circ \cdots \circ f_1,$$

where each layer f_i is Lipschitz continuous with constant at most L_0, i.e.,

$$\|f_i(u) - f_i(v)\| \leq L_0 \|u - v\| \quad \text{for all } u, v \in \mathbb{R}^{d_z}.$$

Then, the empirical Rademacher complexity of the corresponding hypothesis class satisfies

$$\mathcal{R}_n(\mathcal{F}_{\text{DINR}}) \leq \frac{C \cdot L_\psi L_\phi D \cdot e^{L_0^\ell T} \cdot \sqrt{d_z d_y}}{\sqrt{n}}$$

for some universal constant C.

Proof. By composition of Lipschitz functions, the network $f = f_\ell \circ \cdots \circ f_1$ is $L_f = L_0^\ell$-Lipschitz. Substituting this bound into Proposition A.3 yields the stated result. □

Proposition A.5 (Rademacher Complexity of INRs)**.** *Let $\mathcal{F}_{\text{INR}} = \{y : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y} \mid y(x) = \psi(f(\phi(x)))\}$ be a hypothesis class defined on inputs $x \in \mathcal{X} \subset \mathbb{R}^{d_x}$ and let n denote the sample size. Suppose:*

- $\operatorname{diam}(\mathcal{X}) \leq D$,

- $\phi : \mathbb{R}^{d_x} \to \mathbb{R}^{d_z}$ *is a fixed,* B_ϕ*-bounded feature map,*
- $f : \mathbb{R}^{d_z} \to \mathbb{R}^{d_z}$ *is a depth-*ℓ *feedforward neural network of the form*
$$f = f_\ell \circ f_{\ell-1} \circ \cdots \circ f_1,$$
where each layer f_i *is Lipschitz continuous with constant at most* L_0*, i.e.,*
$$\|f_i(u) - f_i(v)\| \le L_0 \|u - v\| \quad \textit{for all } u, v \in \mathbb{R}^{d_z},$$
- $\psi : \mathbb{R}^{d_z} \to \mathbb{R}^{d_y}$ *is* L_ψ*-Lipschitz.*

Then, the empirical Rademacher complexity satisfies:

$$\mathcal{R}_n(\mathcal{F}_{\text{INR}}) \le \frac{\tilde{C} \cdot L_\psi L_\phi L_0^\ell D \cdot \sqrt{d_z d_y}}{\sqrt{n}}$$

for some universal constant $\tilde{C}$*.*

Proof. By composition of Lipschitz functions, the overall Lipschitz constant of the mapping

$$x \mapsto \psi(f(\phi(x)))$$

is at most $L_\psi L_0^\ell L_\phi$. As the input space $\mathcal{X}$ has diameter at most D, the image of $\mathcal{X}$ under this composition lies in a subset of $\mathbb{R}$ with diameter at most $L_\psi L_0^\ell L_\phi D$.

Applying Dudley's entropy integral bound (as in the proof of Proposition A.3), we obtain:

$$\mathcal{R}_n(\mathcal{F}_{\text{INR}}) \le \frac{\tilde{C} \cdot L_\psi L_\phi L_0^\ell D \cdot \sqrt{d_z d_y}}{\sqrt{n}}$$

for some universal constant $\tilde{C}$ as claimed. □

We would like to note that since these results provide upper bounds without corresponding lower bounds, they do not by themselves imply a strict inequality between the capacities of DINR and INR in general. Rather, our analysis characterizes the capacity of the induced function classes and suggests that latent evolution provides an additional mechanism for expanding the expressive space beyond static parameterization.

A.3 Impact of Kinetic Energy Regularization on Generalization

In this section, we theoretically observe that regularizing the kinetic energy (6) in DINRs improves their generalization and mitigates overfitting. We first recall a standard generalization bound based on the Rademacher complexity (Bartlett & Mendelson, 2002; Mohri et al., 2018).

Theorem A.6 (Generalization Bound via Rademacher Complexity (Bartlett & Mendelson, 2002; Mohri et al., 2018))**.** *Let* $\mathcal{F}$ *be a class of functions mapping from* $\mathcal{X}$ *to* $[-B, B]$ *and let* $\{(x_i, y_i)\}_{i=1}^n$ *be i.i.d. samples drawn from a distribution* $\mathcal{D}$*. Then, for any* $\delta > 0$*, with probability at least* $1 - \delta$*, the following holds uniformly for all* $f \in \mathcal{F}$*:*

$$\mathbb{E}_{(x,y)\sim\mathcal{D}}\left[(f(x) - y)^2\right] \le \frac{1}{n}\sum_{i=1}^n (f(x_i) - y_i)^2 + 4B \cdot \mathcal{R}_n(\mathcal{F}) + 3B^2\sqrt{\frac{\log(2/\delta)}{2n}},$$

where $\mathcal{R}_n(\mathcal{F})$ *denotes the empirical Rademacher complexity of* $\mathcal{F}$*.*

This bound implies that a DINR with larger Rademacher complexity has greater expressive power but also higher risk of overfitting. In DINRs, the complexity of the latent dynamics is influenced by the length of the trajectory. The following proposition shows that kinetic energy regularization constrains this trajectory length and, thus, reduces the Rademacher complexity.

Proposition A.7 (Rademacher Complexity Bound for Regularized Class)**.** *Let* $\mathcal{F}_C$ *be the hypothesis class induced by the OC-regularized architecture with discrete kinetic energy bounded by* E*:*

$$\mathcal{F}_{\text{DINR}}^E := \left\{ x \mapsto \psi(z(T)) \;\middle|\; \sum_{k=0}^{N-1} \|f(z_k, t_k; \theta_{\text{dyn}})\|^2 \, \Delta t \le E \right\}.$$

Assume:

- $\mathrm{diam}(\mathcal{X}) \le D$,
- $\phi : \mathbb{R}^{d_x} \to \mathbb{R}^{d_z}$ *is a fixed,* B_ϕ*-bounded feature map,*
- $\psi : \mathbb{R}^{d_z} \to \mathbb{R}^{d_y}$ *is* L_ψ*-Lipschitz.*

Then, the empirical Rademacher complexity of $\mathcal{F}^E_{\mathrm{DINR}}$ *satisfies*

$$C' \cdot L_\psi \cdot \left(B_\phi + \sqrt{TE}\right) \cdot \sqrt{d_z d_y},$$

for some universal constant C'.

Proof. Starting from the dynamics,

$$z_{k+1} = z_k + \Delta t \cdot f(z_k, t_k), \quad k = 0, 1, \ldots, N-1,$$

with initial points bounded as $\|z(0)\| = \|\phi(x)\| \le B_\phi$.

By discrete Cauchy–Schwarz,

$$\Big\| \sum_{k=0}^{N-1} f(z_k^{(i)}, t_k)\, \Delta t \Big\| \le \sum_{k=0}^{N-1} \Delta t\, \|f(z_k^{(i)}, t_k)\| \le \sqrt{\Big(\sum_{k=0}^{N-1} \Delta t\Big)\Big(\sum_{k=0}^{N-1} \|f(z_k^{(i)}, t_k)\|^2\, \Delta t\Big)} \le \sqrt{TE}.$$

Hence, the reachable set at time T is contained in a ball of radius

$$R := B_\phi + \sqrt{TE}.$$

Given that ϕ is L_ϕ-Lipschitz and the input space $\mathcal{X}$ has diameter D, the initial points $z(0) = \phi(x)$ lie within a set of diameter at most $L_\phi D$.

The reachable set radius R bounds the size of the image of $\mathcal{X}$ under the flow. Since $\mathcal{Z}_T \subset \mathbb{R}^{d_z}$ is contained in a ball of radius R, we have

$$\mathcal{N}(\epsilon, \mathcal{Z}_T) \le \left(\frac{2R}{\epsilon}\right)^{d_z}.$$

By the L_ψ-Lipschitz property of ψ, it follows that

$$\mathcal{N}(\epsilon, \psi(\mathcal{Z}_T)) \le \left(\frac{2L_\psi R}{\epsilon}\right)^{d_z}.$$

Therefore, we have

$$\mathcal{R}_n(\mathcal{F}^E_{\mathrm{DINR}}) \le C' \cdot L_\psi \cdot R \cdot \sqrt{d_z d_y}.$$

□

This result provides a theoretical justification for the effect of kinetic energy regularization. As the kinetic energy bound E decreases, the Rademacher complexity is reduced, leading to better generalization as indicated by Theorem A.6. In practice, training a DINR without kinetic energy regularization can lead to a large effective E, while adding this regularization controls E and balances expressive power and generalization.

B Neural Tangent Kernel (NTK) Analysis

B.1 Background: NTK and Gradient Dynamics

We briefly review the NTK and explain how its spectrum and rank influence neural network training dynamics.

Definition of NTK. Let $\hat{y}(x;\theta)$ be a neural network with parameters θ. The NTK is defined as

$$\Theta(x,x') = \nabla_\theta \hat{y}(x)^\top \nabla_\theta \hat{y}(x'), \quad \Theta \in \mathbb{R}^{n\times n},$$

where $\Theta_{ij} = \nabla_\theta \hat{y}(x_i)^\top \nabla_\theta \hat{y}(x_j)$ for a dataset $\{x_i\}_{i=1}^n$. Equivalently, let

$$J = \begin{bmatrix} \nabla_\theta \hat{y}(x_1)^\top \\ \vdots \\ \nabla_\theta \hat{y}(x_n)^\top \end{bmatrix} \in \mathbb{R}^{n\times P},$$

then $\Theta = JJ^\top$ and, therefore,

$$\operatorname{rank}(\Theta) = \operatorname{rank}(J).$$

The NTK captures how changes in network parameters affect outputs across the dataset. Importantly, the *rank* of Θ quantifies the number of independent directions in function space that the network can influence through parameter updates, directly reflecting the network's capacity to learn diverse components of the target function.

Gradient Flow Dynamics. Consider training $\hat{y}(x;\theta)$ via gradient flow on a dataset $\{(x_i,y_i)\}_{i=1}^n$ using the mean squared error (MSE) loss

$$\mathcal{L}(\theta) = \frac{1}{2}\sum_{i=1}^n (\hat{y}(x_i;\theta) - y_i)^2.$$

The gradient flow dynamics are then

$$\frac{d}{dt}\theta_t = -\nabla_\theta \mathcal{L}(\theta_t), \quad \hat{y}_t = [\hat{y}(x_1;\theta_t),\dots,\hat{y}(x_n;\theta_t)]^\top,$$

where θ_t denotes the network parameters evolving continuously over training time $t \geq 0$.

In the *NTK regime*, we assume $\nabla_\theta \hat{y}(x)$ remains approximately constant during training, resulting in the linearized dynamics

$$\frac{d}{dt}\hat{y}_t = -\Theta(\hat{y}_t - y), \quad y = [y_1,\dots,y_n]^\top.$$

Spectral Decomposition and Convergence. Because Θ is symmetric positive semi-definite, it admits the eigendecomposition $\Theta = U\Lambda U^\top$, where $\Lambda = \operatorname{diag}(\lambda_1,\dots,\lambda_n)$ with $\lambda_i \geq 0$ and U orthonormal. The solution to the linearized dynamics is

$$\hat{y}_t - y = \sum_{i=1}^n e^{-\lambda_i t}\langle u_i, \hat{y}_0 - y\rangle\, u_i.$$

Taking norms gives

$$\|\hat{y}_t - y\|^2 = \sum_{i=1}^r e^{-2\lambda_i t}\langle u_i, \hat{y}_0 - y\rangle^2, \quad r = \operatorname{rank}(\Theta).$$

This decomposition shows that each eigendirection u_i is learned at a rate determined by its eigenvalue λ_i, while directions corresponding to zero eigenvalues are never learned. Thus, the NTK rank directly measures the number of directions in the target function space that the network can effectively learn. Higher NTK rank implies a broader span of learnable directions, allowing the network to fit a more diverse set of signal components simultaneously, which leads to faster and more robust convergence.

In summary, the NTK rank quantifies the effective capacity of the network under gradient descent. In the following, we show the DINR architecture can strictly increase the NTK rank compared to a standard INR, enabling richer and more expressive learning dynamics.

B.2 Gradient Derivation of INR and DINR

Before analyzing the NTK rank, we derive the gradients of the standard INR and the proposed DINR with respect to their parameters. This subsection introduces a convenient notation for cumulative Jacobians along the latent trajectory of the DINR, which will be used in subsequent theoretical analysis.

Standard INR. Consider first the standard INR defined as:

$$\hat{y}_{\text{INR}}(x) = \psi\big(f(\phi(x))\big),$$

where ϕ, k, f, and ψ are as previously defined.

The gradient of the standard INR output with respect to parameters $\theta = (\theta_\psi, \theta_f, \theta_\phi)$ is given by:

$$\nabla_\theta \hat{y}_{\text{INR}}(x) = \begin{bmatrix} \frac{\partial \psi}{\partial \theta_\psi}(z_1) \\ \frac{\partial \psi}{\partial z_1} \cdot \frac{\partial f}{\partial \theta_f}(z_0) \\ \frac{\partial \psi}{\partial z_1} \cdot \frac{\partial f}{\partial z_0} \cdot \frac{\partial \phi}{\partial \theta_\phi}(x) \end{bmatrix},$$

where $z_0 := \phi(x)$ and $z_1 := f(z_0)$.

DINR. Now, consider the proposed DINR, where the latent state evolves recursively as

$$z_{k+1} = z_k + \Delta t \cdot f(z_k, t_k; \theta_f), \quad k = 0 \ldots, N-1,$$

with initial state $z_0 = \phi(x; \theta_\phi)$ and output

$$\hat{y}_{\text{DINR}}(x) = h(z_N; \theta_\psi).$$

For notational convenience, define the Jacobian matrices

$$J_k := \frac{\partial f(z_k, t_k; \theta_f)}{\partial z_k} \in \mathbb{R}^{d \times d}, \quad k = 0, \ldots, N-1,$$

and define the partial cumulative Jacobians for any $a \leq b$ as

$$P_{a:b} := \prod_{j=a}^{b} (I + \Delta t \cdot J_j)$$

with the convention that $P_{a:b} = I$ if $a > b$. The cumulative Jacobian along the latent trajectory is then:

$$P := P_{0:N-1}.$$

Using this notation, the gradient of $\hat{y}_{\text{DINR}}(x)$ with respect to parameters $\theta = (\theta_\psi, \theta_f, \theta_\phi)$ is expressed as

$$\nabla_\theta \hat{y}_{\text{DINR}}(x) = \begin{bmatrix} \frac{\partial \psi}{\partial \theta_\psi}(z_N) \\ \frac{\partial \psi}{\partial z_N} \sum_{k=0}^{N-1} P_{k+1:N-1} \cdot \Delta t \cdot \frac{\partial f}{\partial \theta_f}(z_k, t_k) \\ \frac{\partial \psi}{\partial z_N} P \frac{\partial \phi}{\partial \theta_\phi}(x) \end{bmatrix}.$$

B.3 Theorems and Proofs

We restate here the theoretical results of the NTK presented in Section 4.2. These results establish that under mild non-degeneracy conditions, the NTK associated with the DINR architecture has higher rank than that of a standard INR. We begin by providing several supporting lemmas required for the proof, followed by the main statement in Theorem B.3.

Since the input embedding ϕ and the output projection ψ are shared between INR and DINR, the difference in their Jacobians arises solely from the derivatives with respect to the parameters θ_f of the core network f. Therefore, it suffices to focus on the gradients with respect to θ_f.

For notational convenience, for each input x, we denote the gradient (restricted to θ_f) by

$$g_{\text{INR}}(x) = \frac{\partial \psi}{\partial z_1}(x) \frac{\partial f}{\partial \theta}(z_0(x), t_0),$$

and

$$g_{\text{DINR}}(x) = \frac{\partial \psi}{\partial z_N}(x) \sum_{k=0}^{N-1} B_k(x) \Delta t, \quad B_k(x) := P_{k+1:N-1} \frac{\partial f}{\partial \theta}(z_k(x), t_k).$$

Hence,

$$\text{Row}(g_{\text{INR}}) \subseteq \text{span}\left\{ \frac{\partial f}{\partial \theta}(z_0(x), t_0) \,\middle|\, i = 1, \ldots, n \right\},$$

and

$$\text{Row}(g_{\text{DINR}}(x)) \subset \text{span}\{B_0(x), \ldots, B_{N-1}(x)\}.$$

Lemma B.1 (Expansion of reachable gradient directions)**.** *Assume*

1. *There exists* $k \in \{1, \dots, N-1\}$ *and an input* x *such that*
$$\frac{\partial f}{\partial \theta}(z_k(x), t_k) \notin \operatorname{span}\left\{\frac{\partial f}{\partial \theta}(z_0(x), t_0)\right\}.$$
2. *There exists* $M > 0$ *such that*
$$\|J_k\| \leq M \quad \textit{for all } k = 0, \dots, N-1.$$

.

Define
$$B_k(x) := P_{k+1:N-1} \frac{\partial f}{\partial \theta}(z_k(x), t_k).$$

Then, for sufficiently small Δt*,*
$$\operatorname{span}\{B_0(x)\} \subsetneq \operatorname{span}\{B_0(x), \dots, B_{N-1}(x)\}.$$

Proof. For any step size $0 < \Delta t < \frac{1}{M}$, we have
$$\|\Delta t J_k\| \leq \Delta t M < 1,$$
for each k. Hence, the Neumann series
$$(I + \Delta t J_k)^{-1} = \sum_{m=0}^{\infty} (-\Delta t J_k)^m$$
converges, implying that $I + \Delta t J_k$ is invertible. Since a product of invertible matrices is invertible, each $P_{k+1:N-1}$ is invertible for $0 < \Delta t < \frac{1}{M}$.

Therefore,
$$\operatorname{span}(B_k(x)) = \operatorname{span}\left(\frac{\partial f}{\partial \theta}(z_k)\right).$$
Hence,
$$\operatorname{span}\{B_0(x), \dots, B_{N-1}(x)\} = \operatorname{span}\left\{\frac{\partial f}{\partial \theta}(z_0), \dots, \frac{\partial f}{\partial \theta}(z_{N-1})\right\}.$$
By the non-degeneracy assumption, there exists $k \geq 1$ such that
$$\frac{\partial f}{\partial \theta}(z_k) \notin \operatorname{span}\left\{\frac{\partial f}{\partial \theta}(z_0)\right\}.$$
Therefore,
$$\operatorname{span}\{B_0(x)\} \subsetneq \operatorname{span}\{B_0(x), \dots, B_{N-1}(x)\}.$$
This implies that, for at least one input x, the DINR gradient spans a strictly larger subspace than the INR gradient.

□

Lemma B.1 highlights a key structural difference between INR and DINR. While the standard INR evaluates the network only at a single latent point z_0, the DINR evolves the latent state along a dynamical trajectory $\{z_k\}_{k=0}^{N}$ and evaluates the vector field f at multiple states. As a consequence, the DINR can access gradient information from multiple points in latent space, which increases the likelihood of generating directions that are not present at the initial state z_0.

The boundedness assumption on the Jacobians is mild and typically satisfied in practice, as f is typically Parametrized by a multilayer perceptron with Lipschitz activation functions. Under such architectures, the network is locally Lipschitz, and its Jacobian remains bounded on compact subsets of the latent space.

Together, these conditions ensure that the transported gradients $B_k(x)$ retain the directional diversity induced along the trajectory, leading to a strict expansion of the reachable gradient span.

Lemma B.2 (Genericity of row space escape)**.** *Let $\{x_i\}_{i=1}^n$ be a fixed dataset. Define the INR Jacobian*

$$J_{\mathrm{INR}}(\theta) = \begin{bmatrix} g_{\mathrm{INR}}(x_1;\theta) \\ \vdots \\ g_{\mathrm{INR}}(x_n;\theta) \end{bmatrix} \in \mathbb{R}^{n\times P}.$$

Fix an index $i^\star \in \{1,\dots,n\}$ and define

$$v(\theta) := g_{\mathrm{DINR}}(x_{i^\star};\theta) \in \mathbb{R}^P.$$

Assume

1. *All mappings $\theta \mapsto g_{\mathrm{INR}}(x_i;\theta)$ are real analytic.*
2. *There exists θ_0 such that*
$$v(\theta_0) \notin \mathrm{Row}(J_{\mathrm{INR}}(\theta_0)).$$

Then, the set

$$\mathcal{Z} := \{\theta : v(\theta) \in \mathrm{Row}(J_{\mathrm{INR}}(\theta))\}$$

has Lebesgue measure zero.

Proof. Let $A(\theta) := J_{\mathrm{INR}}(\theta)$ and define the augmented matrix

$$\tilde{J}(\theta) := \begin{bmatrix} A(\theta) \\ v(\theta) \end{bmatrix} \in \mathbb{R}^{(n+1)\times P}.$$

Then

$$v(\theta) \in \mathrm{Row}(A(\theta)) \quad \Longleftrightarrow \quad \mathrm{rank}(\tilde{J}(\theta)) = \mathrm{rank}(A(\theta)).$$

Let

$$r_{\max} := \max_{\theta\in\mathbb{R}^P} \mathrm{rank}(A(\theta)) \le n.$$

If $v(\theta) \in \mathrm{Row}(A(\theta))$, then

$$\mathrm{rank}(\tilde{J}(\theta)) = \mathrm{rank}(A(\theta)) \le r_{\max},$$

which implies that all $(r_{\max}+1)\times(r_{\max}+1)$ minors of $\tilde{J}(\theta)$ vanish.

Let $\{f_j(\theta)\}_{j=1}^M$ denote the collection of all such minors. Since there are finitely many $(r_{\max}+1)\times(r_{\max}+1)$ submatrices, $M<\infty$, and each f_j is a determinant of a submatrix of $\tilde{J}(\theta)$. Because all entries of $A(\theta)$ and $v(\theta)$ are real analytic, each $f_j(\theta)$ is a real analytic function on $\mathbb{R}^P$. Therefore,

$$\mathcal{Z} \subseteq \bigcap_{j=1}^M \{\theta : f_j(\theta) = 0\}.$$

By assumption, there exists θ_0 such that

$$v(\theta_0) \notin \mathrm{Row}(A(\theta_0)),$$

hence

$$\mathrm{rank}(\tilde{J}(\theta_0)) > \mathrm{rank}(A(\theta_0)),$$

which implies that at least one $(r_{\max}+1)\times(r_{\max}+1)$ minor of $\tilde{J}(\theta_0)$ is nonzero. Therefore, there exists $j^\star$ such that

$$f_{j^\star}(\theta_0) \neq 0, \quad \Rightarrow \quad f_{j^\star} \not\equiv 0.$$

A standard result in real analytic geometry states that the zero set of a non-identically-zero real analytic function on $\mathbb{R}^P$ has Lebesgue measure zero. Hence,

$$\{\theta : f_{j^\star}(\theta) = 0\}$$

has Lebesgue measure zero.

Since

$$\mathcal{Z} \subseteq \{f_{j^\star} = 0\},$$

we conclude that $\mathcal{Z}$ has Lebesgue measure zero.

□

Theorem B.3 (NTK rank increase for DINR)**.** *Let $\{x_i\}_{i=1}^n$ be a fixed dataset. Define*

$$J_{\mathrm{INR}}(\theta) = \begin{bmatrix} g_{\mathrm{INR}}(x_1;\theta) \\ \vdots \\ g_{\mathrm{INR}}(x_n;\theta) \end{bmatrix}, \quad J_{\mathrm{DINR}}(\theta) = \begin{bmatrix} g_{\mathrm{DINR}}(x_1;\theta) \\ \vdots \\ g_{\mathrm{DINR}}(x_n;\theta) \end{bmatrix}.$$

Assume:

1. *All mappings $\theta \mapsto g_{\mathrm{INR}}(x_i;\theta)$ and $\theta \mapsto g_{\mathrm{DINR}}(x_i;\theta)$ are real analytic.*
2. *For each $i \in \{1,\dots,n\}$, there exists θ such that*
$$g_{\mathrm{DINR}}(x_i;\theta) \notin \mathrm{Row}(J_{\mathrm{INR}}(\theta)).$$

Then, for almost all $\theta \in \mathbb{R}^P$,

$$\mathrm{rank}(J_{\mathrm{DINR}}(\theta)) > \mathrm{rank}(J_{\mathrm{INR}}(\theta)).$$

Proof. For each i, define

$$\mathcal{Z}_i := \{\theta : g_{\mathrm{DINR}}(x_i;\theta) \in \mathrm{Row}(J_{\mathrm{INR}}(\theta))\}.$$

By Lemma B.2, each $\mathcal{Z}_i$ has Lebesgue measure zero.

Let

$$\mathcal{Z} := \bigcap_{i=1}^{n} \mathcal{Z}_i.$$

Then $\mathcal{Z}$ also has Lebesgue measure zero.

Thus, for almost all θ, there exists at least one i such that

$$g_{\mathrm{DINR}}(x_i;\theta) \notin \mathrm{Row}(J_{\mathrm{INR}}(\theta)).$$

This implies that

$$\mathrm{Row}(J_{\mathrm{DINR}}(\theta)) \not\subseteq \mathrm{Row}(J_{\mathrm{INR}}(\theta)),$$

and therefore

$$\mathrm{rank}(J_{\mathrm{DINR}}(\theta)) > \mathrm{rank}(J_{\mathrm{INR}}(\theta)).$$

□

Theorem B.3 establishes that, under mild regularity conditions, the Jacobian of DINR generically attains a strictly higher rank than that of INR. Since the NTK is given by the Gram matrix of the Jacobian, this result implies that DINR induces a strictly richer kernel in the linearized NTK regime.

The key structural difference underlying this result is captured by Lemma B.1, which shows that the dynamical trajectory of DINR gives rise to additional gradient directions that are not available at the initial latent state. While the DINR gradient is expressed as a sum over these transported directions, the event that this sum lies entirely within the span of the initial gradient requires precise cancellations across multiple terms. Such cancellations impose nontrivial algebraic constraints on the parameters and are therefore non-generic. Consequently, it is reasonable to expect that the non-degeneracy condition in Theorem B.3 holds in typical settings.

We note, however, that the present analysis does not provide a fully constructive characterization of when this condition holds, nor does it completely rule out pathological parameter configurations. Nevertheless, the combination of trajectory-induced directional expansion and the non-generic nature of exact cancellations provides strong theoretical support for the empirical observation that DINR exhibits a higher NTK rank than INR.

C Related Work

The proposed DINR is a model-agnostic framework that synthesizes principles from INRs and continuous-time modeling. To properly situate our contributions, which include a rigorous theoretical analysis, we now review the most relevant literature from these domains.

Implicit Neural Representations. INRs constitute a novel class of models that encode data as continuous functions via neural networks (Xie et al., 2022). INRs' modality-agnostic nature has facilitated a broad range of applications across diverse fields, such as signal representation, e.g., 2D images (Chen et al., 2021), 3D scenes (Park et al., 2019), videos (He et al., 2023), and neural rendering (Mildenhall et al., 2021; Müller et al., 2022). However, widespread adoption of INRs is impeded by the following challenges.

- ***(C1) Accuracy Challenge.*** One major challenge in deploying INRs using MLP architectures is their inherent *spectral bias* (Rahaman et al., 2019; Yüce et al., 2022). This bias leads neural networks to preferentially learn low-frequency components of functions, resulting in smoother approximations that may lack high-frequency details crucial for accurate representations. To combat this issue, various strategies have been introduced, such as embedding inputs with multiple orthogonal Fourier or Wavelet bases as part of positional encoding (Tancik et al., 2020; Saragadam et al., 2023), employing special periodic or non-periodic activation functions (Sitzmann et al., 2020; Ramasinghe & Lucey, 2022), and applying learning-based regularization techniques (Li et al., 2023; Krishnapriyan et al., 2021).
- ***(C2) Training Challenge.*** Prior work has addressed training-dynamics issues (e.g., stability) via meta-learned initialization to curb vanishing/exploding signals (Koneputugodage et al., 2025; Tancik et al., 2021; Saratchandran et al., 2024), normalization to stabilize and speed training through spectral conditioning (Cai et al., 2024; Bjorck et al., 2018), gradient transformations to alleviate spectral bias via reweighting/preconditioning (Shi et al., 2024; Bjorck et al., 2018), and Sobolev training to match derivatives when available to damp oscillations (Vlassis & Sun, 2021; Yuan et al., 2022).

Among the current literature, the proposed DINR is a plug-and-play component orthogonal to current methods.

Gaussian Splatting. In parallel with MLP-based INRs, Gaussian splatting has emerged as a fast and editable representation paradigm. Originating from 3DGS for real-time radiance-field rendering (Kerbl et al., 2023), recent methods adapt Gaussian primitives to 2D signals (Chen & Wang, 2024; Wu et al., 2024). For example, GaussianImage (Zhang et al., 2024) uses compact 2D Gaussians with efficient rasterization, MiraGe (Waczyńska et al., 2024) enables 3D-aware image editing with flat Gaussian components, and GaINeR (Jakubowska et al., 2025) combines trainable Gaussians with an INR decoder. Unlike these explicit or hybrid Gaussian representations, DINR targets the spectral-bias and optimization challenges of coordinate-based INRs, making it orthogonal to existing encodings, activations, training strategies, and Gaussian/INR hybrids.

Dynamical Neural Networks. Dynamical neural networks (DNNs), also known as *continuous-time neural networks*, describe latent dynamics that evolve continuously over time (Chen et al., 2018). In these models, the hidden dynamics are governed by an ordinary differential equation (ODE) parameterized by neural networks, which can be viewed as a continuous-time generalization of residual connections (He et al., 2016). A major advantage of DNNs is their *parameter efficiency* as they share weights across continuous trajectories, allowing compact models to approximate complex mappings (Chen et al., 2018; Grathwohl et al., 2018; Massaroli et al., 2020). They have been widely adopted in generative modeling through continuous normalizing flows (CNFs) (Chen et al., 2018; Grathwohl et al., 2018; Onken et al., 2021), where kinetic energy-based regularizers are often introduced to promote smooth trajectories and reduce integration cost (Finlay et al., 2020; Yang & Karniadakis, 2020; Vidal et al., 2023). Recently, dynamical formulations also have been explored in other architectures, such as Transformers (Kan et al., 2025). However, these studies mainly focus on regularization and control-theoretic perspectives. In contrast, our work investigates how a dynamical formulation can expand the expressive power of INRs and mitigate spectral bias.

Theoretical Analysis of Neural Networks. The theoretical understanding of deep learning relies on a diverse set of tools to explain the success of modern neural networks (Soudry et al., 2018; Arora

et al., 2018; Bartlett et al., 2017; Raghu et al., 2017; Lee et al., 2018). Among them, two tools are particularly relevant to our study: 1) capacity measures, such as Rademacher complexity, which provide distribution-agnostic generalization bounds, and 2) infinite-width kernel theories, e.g., NTK, that characterize training dynamics.

- ***(T1) Rademacher Complexity.*** Rademacher complexity (Bartlett & Mendelson, 2002; Mohri et al., 2018) is a data-dependent measure of a model's expressivity by quantifying how well a hypothesis class, i.e., the set of functions representable by the network, can fit random label noise. Unlike distribution-free metrics, such as Vapnik–Chervonenkis dimension (Vapnik & Chervonenkis, 1971), it adapts to empirical data and yields tighter generalization guarantees. This framework has been applied to various models, including INRs (Zhao et al., 2024), computer vision architectures (Yin et al., 2019; Galanti et al., 2023; Trauger & Tewari, 2024), and other models (Park et al., 2022; Marion, 2023; Kim & Kang, 2024), to characterize the relationship between model capacity and generalization. Recent work (Hanson & Raginsky, 2024) has derived Rademacher complexity bounds for dynamical networks via Chen–Fliess series expansions. However, in contrast to our work, that analysis focuses on mappings from an initial condition to a terminal scalar output in control-affine dynamical systems, which studies latent feature dynamics across the full coordinate domain of INRs.
- ***(T2) Neural Tangent Kernel.*** The NTK (Jacot et al., 2018) provides a principled tool to study optimization dynamics. It reveals that in the infinite-width limit, gradient descent on the squared loss corresponds to kernel regression with a fixed kernel, thereby linking neural network training to classical kernel methods. This insight enables formal analysis of convergence and generalization and has been extended to diverse architectures, such as MLPs, CNNs, RNNs, and Transformers (Lee et al., 2019; Arora et al., 2019; Yang, 2020; Hron et al., 2020). In finite-width regimes, the NTK becomes data-dependent and stochastic (Hanin & Nica, 2020), while its spectrum determines convergence rates and biases learning toward low-frequency components (Du et al., 2018; Bordelon et al., 2020). To the best of our knowledge, the NTK of DNNs has not been systematically characterized, highlighting the novelty of our analysis for DINR.

D Dataset Details

D.1 Turbulent Flow

This dataset features a two-dimensional simulation of Kraichnan turbulence, a canonical problem in fluid dynamics characterized by a wide cascade of energy across a broad range of spatial frequencies. The resulting flow fields are chaotic and contain intricate, high-frequency structures, making this an ideal benchmark for evaluating the ability of implicit neural representations (INRs) to capture complex, multiscale signals. The data are generated via a direct numerical simulation (DNS) of the incompressible Navier-Stokes equations, which govern the motion of viscous fluid substances (Holmes, 2012). The equations enforce the conservation of mass (incompressibility) and momentum:

$$\nabla \cdot \mathbf{u} = 0, \quad \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla \mathbf{p} + \nu \nabla^2 \mathbf{u},$$

where $\mathbf{u}$ is the velocity field, $\mathbf{p}$ is the pressure, ρ is the fluid density, and ν is the kinematic viscosity. The simulation is performed in a doubly periodic square domain defined by $[0, 2\pi]^2$ and is discretized on a high-resolution 2048×2048 grid. The numerical integration employs a second-order energy-conserving Arakawa scheme for the nonlinear Jacobian term (Arakawa, 1997) and a second-order finite-difference scheme for the Laplacian of the vorticity (Ren et al., 2023). For our experiments, we use a single snapshot of the resulting vorticity field, which is normalized to the range $[-1, 1]$ to serve as the ground truth signal.

D.2 Global Weather Pattern Dataset

The global weather pattern data used in our study are derived from the ERA5 atmospheric reanalysis dataset, a state-of-the-art global climate product (Hersbach et al., 2020). ERA5 provides a physically consistent and comprehensive record of the Earth's atmosphere by assimilating vast quantities

of historical observations from sources including satellites and ground stations into an advanced numerical model. The resulting data fields are characterized by complex, multiscale spatial patterns that span from large-scale circulation features to fine-scale local variations, making it a challenging benchmark for generative and representational models. The dataset offers global coverage with a native spatial resolution of 0.25 degrees, which translates to a Cartesian grid of 721×1440 pixels. While ERA5 provides a long-term, hourly record of numerous atmospheric variables across 37 vertical pressure levels, our experiments focus on a single, representative snapshot of the total column water vapor (TCWV). This specific variable is chosen for its intricate and detailed structures. For our experimental setup, the scalar TCWV field is extracted and normalized to the range $[0, 1]$ to serve as the ground truth signal.

D.3 Cloud-topped Boundary Layer

This dataset comprises high-resolution, three-dimensional simulation data of a cloud-topped boundary layer, a critical component of the Earth's climate system that plays a significant role in cloud formation and radiative transfer. The data was generated using the University of California, Los Angeles Large-Eddy Simulation (UCLA-LES) model, a widely recognized tool for simulating atmospheric turbulence and cloud dynamics with high fidelity (Boucher et al., 2013; Günther et al., 2016). The simulation captures the complex, multiscale interactions between turbulent air motion, water vapor, and liquid water content, making it an excellent benchmark for evaluating the ability of neural representations to model intricate, high-frequency spatial phenomena. The simulation was configured to model a stratocumulus-topped boundary layer. It operates under double-periodic boundary conditions, meaning the domain is continuous in the horizontal directions, which is a standard setup for idealized atmospheric studies. The model is driven by homogeneous surface forcing and incorporates large-scale meteorological information from the COSMO-DE numerical weather prediction model to ensure physical realism (Klamt, 2011; Baldauf et al., 2011). For our experiments, we employ a snapshot from the simulation, focusing on a scalar field that represents the liquid water content. This field is defined on a three-dimensional Cartesian grid of $384 \times 384 \times 130$ voxels. The data exhibits fine-grained structures, such as wispy cloud tendrils and sharp gradients at cloud boundaries, which present a significant challenge for implicit neural representation models. The scalar values are normalized to the range $[0, 1]$ to serve as the ground truth signal for our evaluation.

D.4 Cryo-EM

The cryogenic electron microscopy (cryo-EM) data used in our experiments is the EMD-32218 entry from the Electron Microscopy Data Bank (EMDB) (Gu et al., 2022). This dataset contains the volumetric three-dimensional density map of the matrix arm of mammalian mitochondrial complex I, captured in its deactive state. Complex I is a crucial enzyme in cellular respiration, and understanding its structure is of significant biological importance. The original study provides a structural basis for the regulation of this enzyme's deactive state. The volumetric map was determined using single-particle analysis, a technique that averages thousands of two-dimensional projection images of individual protein particles to reconstruct a high-resolution three-dimensional volume. The sample organism is *Bos taurus* (cattle). The reconstructed map has a reported resolution of 2.5 Å (Angstroms), providing a detailed view of the protein's atomic structure. The data are provided as a three-dimensional volumetric grid with dimensions of $480 \times 480 \times 480$ voxels. Each voxel is isotropic with a spacing of 1.05 Å along each axis. For our experiments, the volumetric data, originally in CCP4 map format, is normalized to the range $[0, 1]$ to serve as the ground truth signal for evaluating the reconstruction fidelity of the INR models.

E Additional Results

E.1 Additional NTK Results

To gain a theoretical understanding of DINR's advantages, we analyze its function space and optimization landscape via the empirical NTK, which characterizes training behavior in the infinite-width limit and reveals inductive biases (Yüce et al., 2022). Following standard practice, we construct the empirical NTK $K_{ij} = \sum \left\langle \frac{\partial \hat{y}(x_i)}{\partial \theta}, \frac{\partial \hat{y}(x_j)}{\partial \theta} \right\rangle$ for each model using a random subset of coordinates

and examine its eigenspectrum through: 1) the effective rank that measures the spectral complexity and the model's capacity to learn diverse functions; 2) the condition number that indicates the optimization problem's stability; and 3) the decay rate of the leading eigenvalues, which reflects how the model prioritizes different functional modes. Our analysis reveals that DINRs consistently induce a more favorable kernel structure on the tested datasets. We compare standard and dynamical parameterizations using three neural tangent kernel diagnostics: condition number, effective rank, and eigenvalue spectrum. Across architectures (FFNet, SIREN) and initializations (lognormal, Weibull), the dynamical variants consistently avert spectral collapse, maintaining a better-conditioned and higher-rank kernel throughout training.

FFNet. Standard FFNet suffers from extreme spectral degeneration during training. Its condition number monotonically degrades from $\approx 10^3$ to $> 10^5$, while its effective rank collapses from ≈ 30 to near 1, indicating that optimization becomes bottlenecked by a single principal direction. In stark contrast, Dynamical FFNet rapidly stabilizes the condition number near 10^3 and actively *increases* effective rank to ≈ 200. Eigenvalue distributions confirm this mechanism: whereas the standard model's spectrum sharply concentrates, the dynamical variant preserves substantial mass across intermediate and large eigenvalues. Compared with the baseline, Dynamical FFNet retains substantially more mass in the intermediate and large-eigenvalue regime as training progresses, resulting in a broader and less concentrated spectrum.

SIREN. While baseline SIREN is inherently more stable, the dynamical approach still yields measurable optimization benefits. It successfully suppresses the baseline's characteristic early-training condition number spike and maintains a consistently higher effective rank (≈ 600 compared to the baseline's 400–500). Furthermore, Dynamical SIREN sustains a broader eigenspectrum with slower tail decay (most prominently under Weibull initialization), avoiding excessive spectral concentration.

Discussion & Implications. These diagnostics empirically demonstrate that dynamical parameterization fundamentally improves optimization conditioning. By mitigating NTK spectral collapse, the proposed dynamical reformulation prevents the learning trajectory from degrading into a low-dimensional subspace, ensuring optimization proceeds along a broad set of active directions. This yields transformative improvements for poorly conditioned networks (FFNet) while providing robust, consistent refinements for already-stable architectures (SIREN).

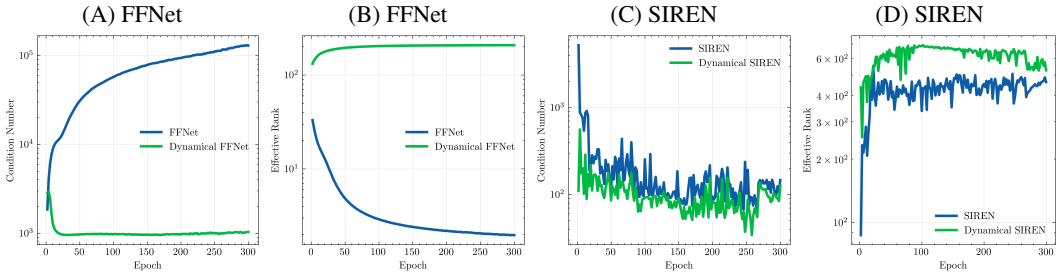

Figure 11: Empirical NTK properties for FFNet and SIREN.

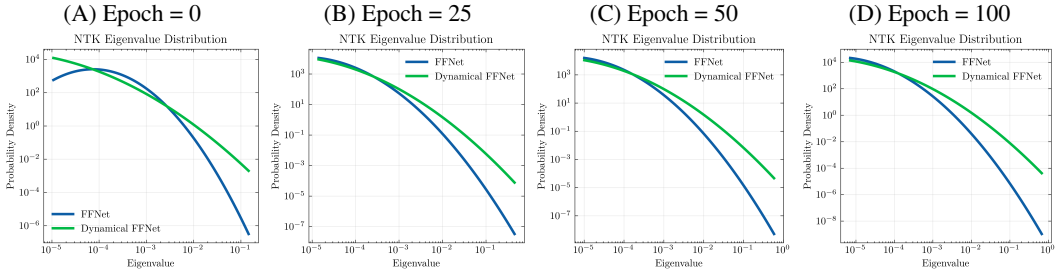

Figure 12: NTK eigenvalue distributions for FFNet under the lognormal setting at different epochs.

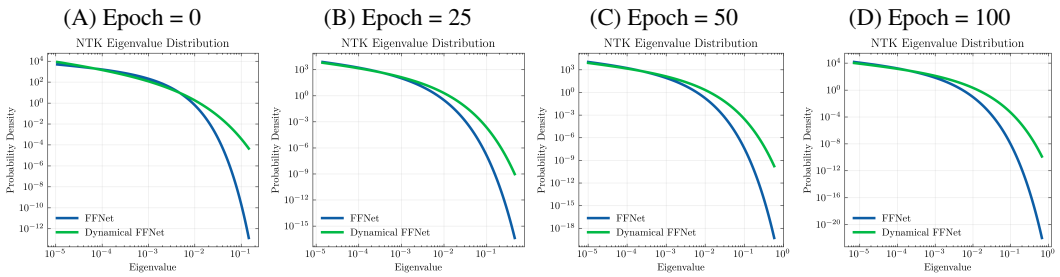

Figure 13: NTK eigenvalue distributions for FFNet under the weibull setting at different epochs.

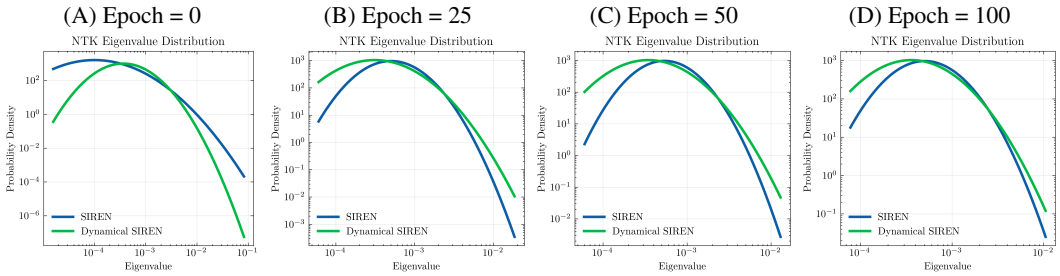

Figure 14: NTK eigenvalue distributions for SIREN under the lognormal setting at different epochs.

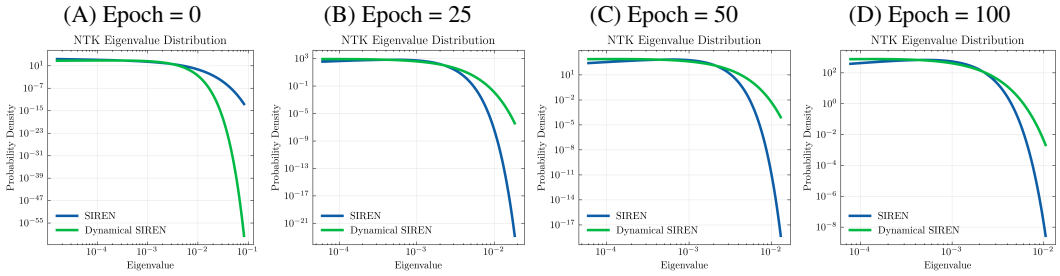

Figure 15: NTK eigenvalue distributions for SIREN under the weibull setting at different epochs.

E.2 Additional Multi Run Results

We assess the statistical robustness of the dynamical formulation across both scientific turbulence fields and natural-image benchmarks. On Turbulence, we train FFNet and Dynamical FFNet independently over 20 paired random seeds and evaluate seed-wise performance differences. On DIV2K and Kodak, we evaluate all available images using three seeds per image; paired differences are first averaged across seeds within each image, and the resulting image-level means form the inferential sample. These results demonstrate that the observed gains persist across random initializations, image instances, and data modalities.

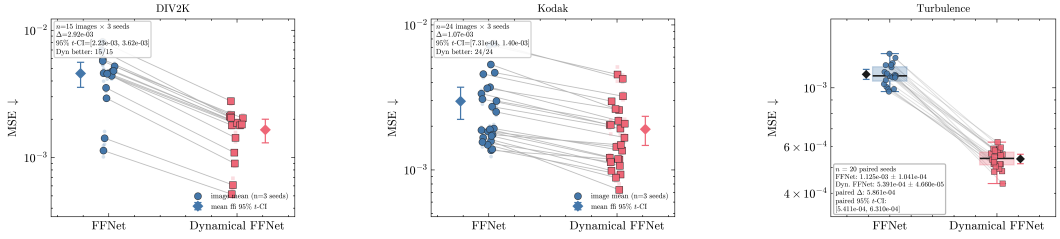

Figure 16: Paired multi-run FFNet MSE results.

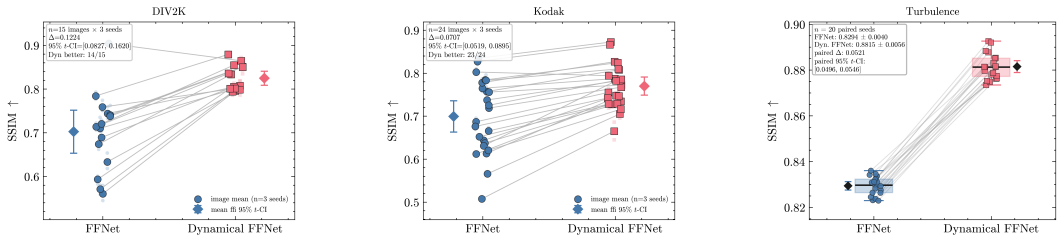

Figure 17: Paired multi-run FFNet SSIM results.

E.3 Additional Compute-Matched Results

Increasing the capacity of static INRs under the same computational budget improves reconstruction quality, but does not recover DINR's parameter efficiency. Despite an extensive search over static architectures, the best FLOP-matched FFNet requires 5.57M parameters. This result highlights the central motivation of DINR: rather than increasing model capacity, latent dynamics improve expressivity by repeatedly evolving a compact shared representation. Consequently, DINR achieves a more favorable fidelity–storage trade-off, providing substantially greater representational efficiency without scaling the underlying network size.

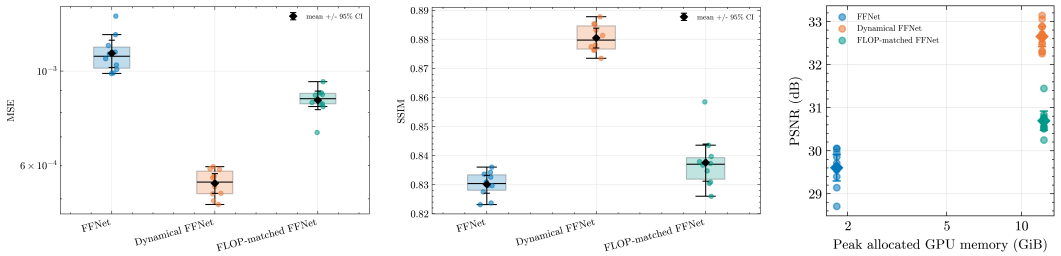

Figure 18: Additional FLOP-matched results.

In parallel, we conducted time-matched experiments. Across all three reconstruction metrics, Dynamical FFNet consistently outperforms both the original FFNet and the time-matched FFNet. Allowing the static FFNet to train for the same wall-clock budget improves performance over the original baseline, indicating that additional optimization time is beneficial; however, it does not close the gap to the dynamical formulation. Dynamical FFNet achieves substantially lower MSE and higher PSNR and SSIM, with relatively tight variability across runs, whereas the time-matched baseline exhibits both weaker average performance and greater variation. These results suggest that DINR's gains cannot be explained solely by additional training time, providing further evidence that the dynamical latent evolution itself contributes to the improved reconstruction quality.

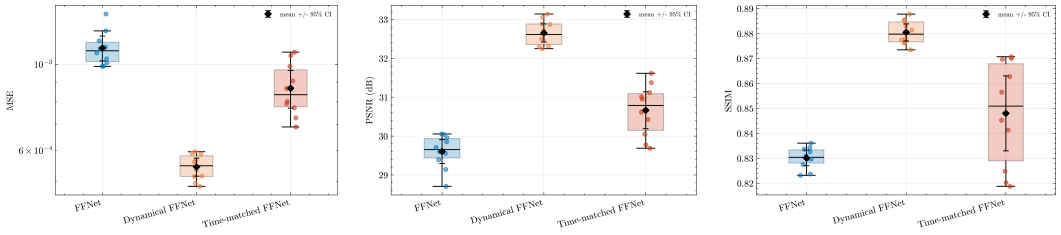

Figure 19: Additional time-matched results.

E.4 Additional KE Regularizer Ablation Results

We conduct a targeted ablation study to validate the effectiveness of our KE regularizer. By penalizing the kinetic energy of learned trajectories, the regularizer directly encourages smoother latent dynamics

and stabilizes the underlying ODE, enhancing robustness and data efficiency. Our hypothesis is that *by encouraging smoother learned dynamics, the regularizer will enhance the model's robustness and data efficiency*. We evaluate this in two challenging scenarios: reconstruction from noisy data and generalization from scarce data.

Regularization Complementary to Weight Decay. Results are provided in Section 5.3.

Robustness to Noise. First, we test the model's ability to reconstruct the turbulence signal from observations corrupted by additive independent and identically distributed (i.i.d.) noise. As shown in Figure 20, the KE regularizer consistently and significantly reduces reconstruction error across all noise levels.

1. For the FFNet backbone, the regularizer reduces MSE when compared to the static baseline. More importantly, its benefit over the unregularized dynamical model grows as noise intensifies, indicating its role in actively combating noise.
2. The effect is even more dramatic for the expressive SIREN backbone. The KE regularizer slashes MSE by a remarkable improvement at noisy learning settings compared to the static baseline. Crucially, while the error of the unregularized dynamical SIREN explodes under heavy noise, the KE-regularized version maintains a much flatter error curve. This demonstrates that the regularizer successfully damps the model's tendency to fit spurious high-frequency components introduced by noise, promoting a smoother, more plausible underlying function.

Generalization from Scarce Data. Next, we evaluate the regularizer's impact on generalization when trained on a reduced number of samples. This tests the model's ability to learn a good inductive bias for interpolation and extrapolation.

1. The KE regularizer again provides substantial gains.
2. This result shows the KE term acts as a powerful inductive bias, guiding the model toward smoother solutions that generalize better to unseen regions of the domain. By discouraging overly complex feature trajectories, the regularizer prevents the model from overfitting, a critical property for real-world scenarios where data are often limited.

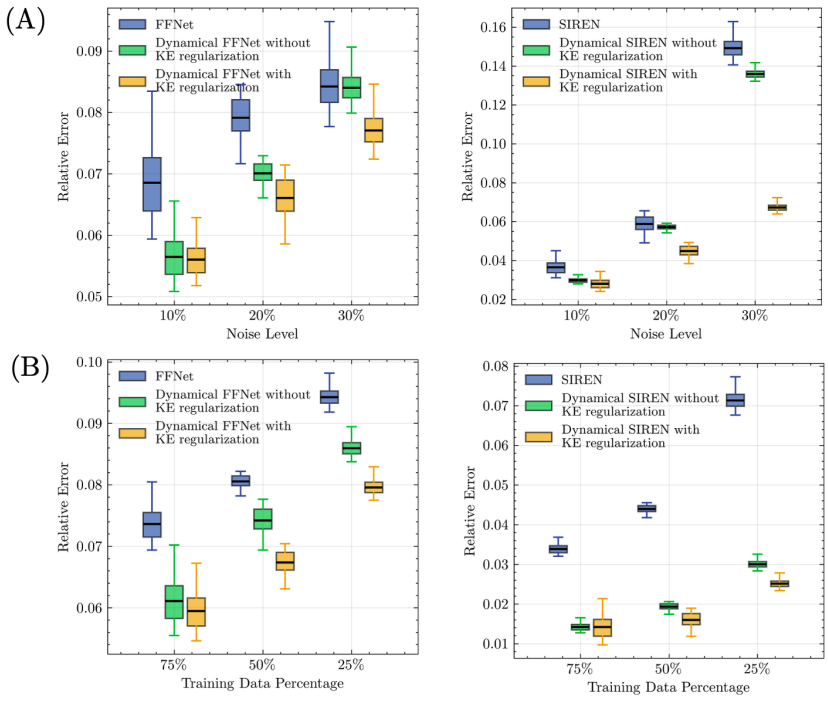

Figure 20: Performance comparison on (A) noisy signals and (B) under reduced training data.

E.5 Additional Natural Imaging Learning Results

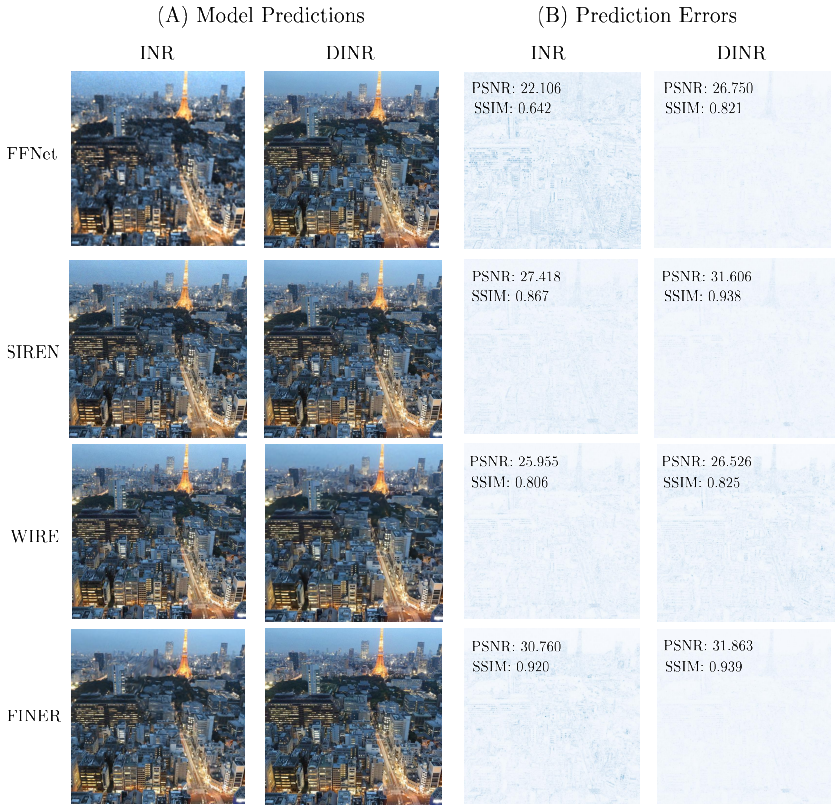

Figure 21: Results of INRs and DINRs on DIV2K datasets.

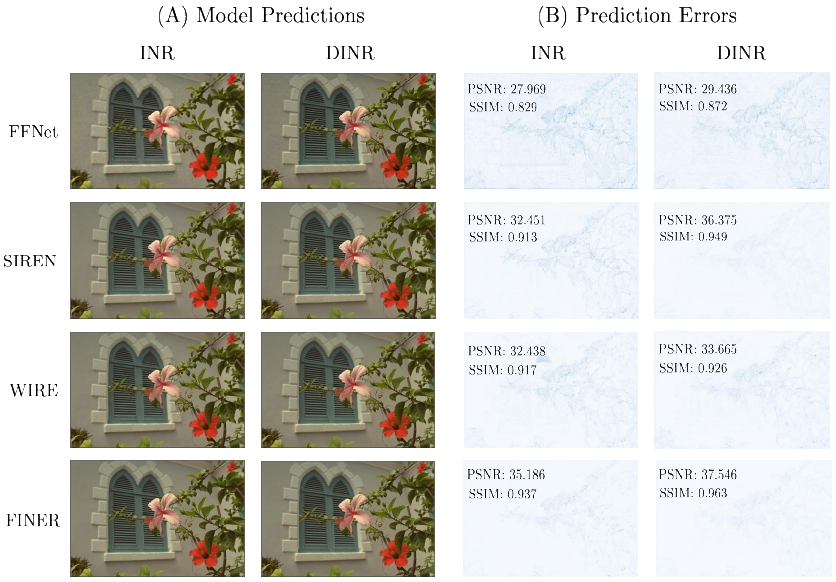

Figure 22: Results of INRs and DINRs on Kodak datasets.

E.6 Additional Data Compression Results

A primary application of the proposed DINR is the compression of large-scale scientific data, where storage and transmission costs are a significant bottleneck. Fields, e.g., climate modeling, generate petabytes of data from high-resolution simulations, making efficient representation critical.

Qualitative Results. Visual and spectral evidence further explains why DINR successfully maintains fidelity at higher compression rates. In both FFNet and SIREN comparisons (Figure 23), dynamical models yield cleaner reconstructions with less structured residual error. For turbulence, this manifests as better recovery of coherent vortical filaments rather than simple smoothing, a suppression of error that is visibly pronounced for SIREN. For weather, the dynamical variants accurately track large-scale circulation patterns while yielding lighter, more diffuse residual maps without obvious over-sharpening artifacts.

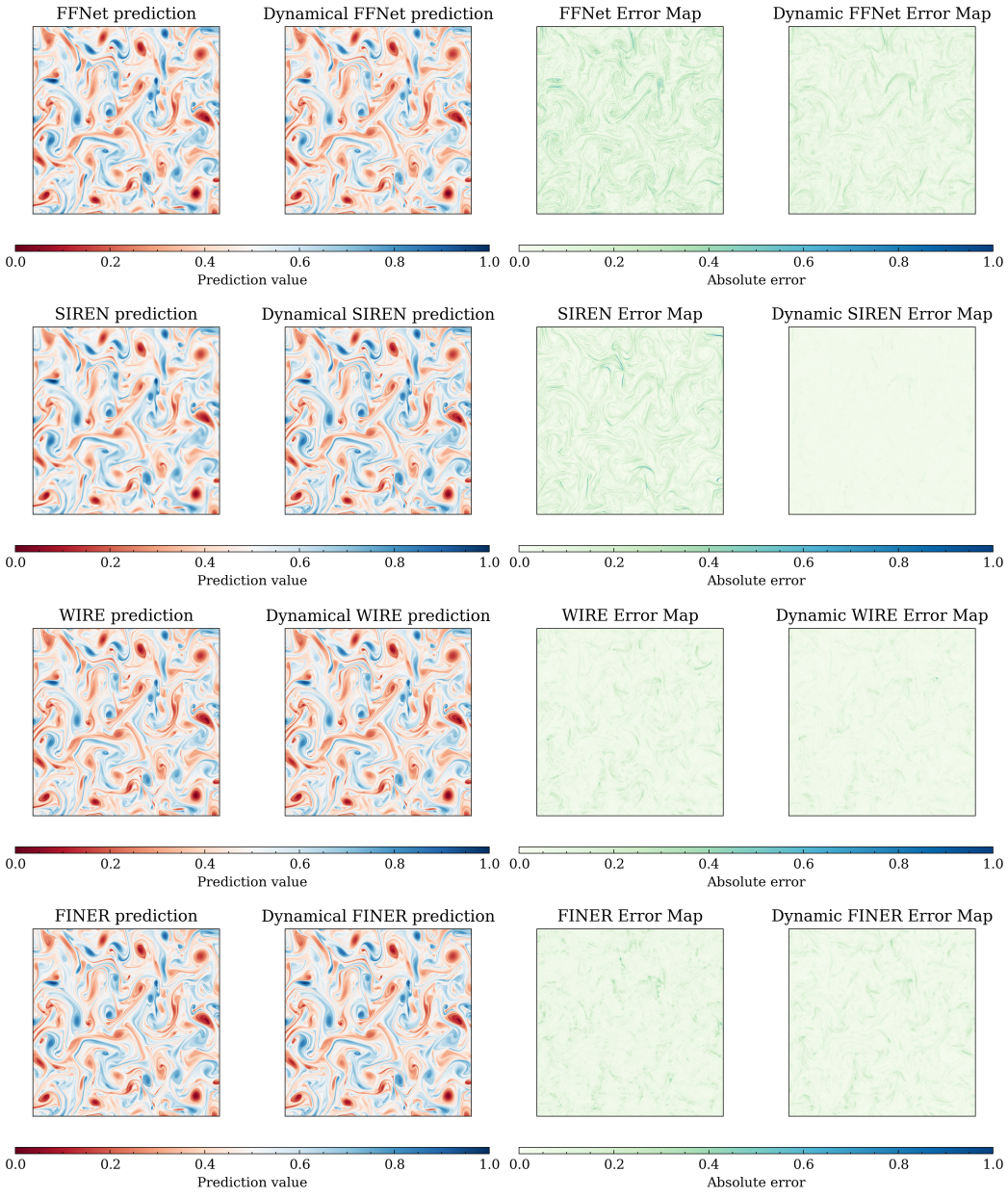

Figure 23: Results of INRs and Dynamical INRs.

Physical Fidelity and Spectral Preservation. Pointwise reconstruction metrics often mask underlying physical distortions. To rigorously evaluate whether the compressed representations retain the multiscale energy distribution of the underlying fluid dynamics, we compare the reconstructed power spectral density (PSD) against the ground truth. As shown in Figure 24, dynamical parameterizations consistently yield higher spectral fidelity than their static counterparts. In the turbulence regime, both Dynamical FFNet and Dynamical SIREN track the ground-truth PSD across nearly all wavenumbers. They accurately capture the macroscopic low-wavenumber content and intermediate decay, while significantly improving alignment in the high-wavenumber dissipative tail where the static FFNet prematurely diverges. This advantage is even starker in the more challenging CTBL case, where static baselines suffer from severe spectral distortions. Static FFNet fundamentally mischaracterizes the scale distribution, substantially overestimating low-to-mid wavenumber energy and exhibiting a non-physical flattening in the high-wavenumber tail. Dynamical FFNet effectively suppresses these artifacts, restoring both the accurate macroscopic energy distribution and a realistic high-frequency decay. Similarly, Dynamical SIREN refines the baseline's already reasonable fit, achieving superior and tight agreement across the full spectrum, particularly during the transition to large wavenumbers.

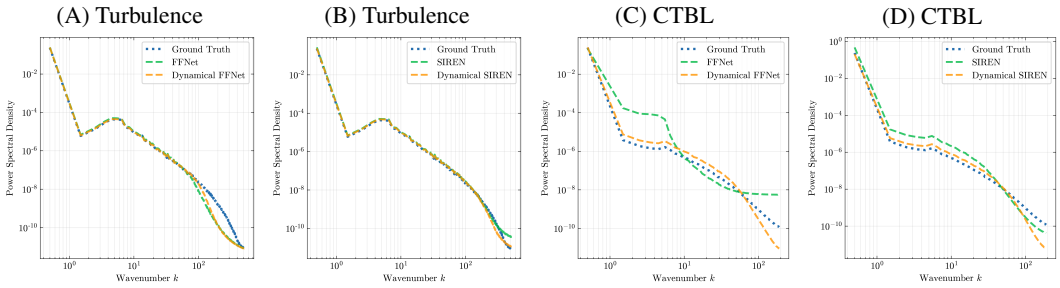

Figure 24: Power spectral density (PSD) of the reconstructed signals across wavenumbers.

Compression Results. Beyond signal fidelity, we also compare model parameters, memory footprint, and the resulting compression ratios on the 2D turbulence dataset. Table 3 shows our dynamical models are significantly more parameter-efficient, achieving a higher compression ratio with a smaller model size. This efficiency stems from the inherent weight sharing across the continuous depth of the DINR formulation, leading to a more compact yet powerful representation.

Table 3: Model Sizes, Compression Ratios, and Gains for Turbulence and Weather Datasets

	Turbulence Data			*Weather Data*		
Model	Parameters	Comp. Ratio	Size (MiB)	Parameters	Comp. Ratio	Size (MiB)
FFNet	727,297	1.44x	2.77	892,417	1.16x	3.40
+Dynamical	465,939	2.25x	1.78	599,065	1.73x	2.29
Gain	*-261,358 (-35.9%)*	*1.56x rel.*	*-1.00*	*-293,352 (-32.9%)*	*1.49x rel.*	*-1.11*
SIREN	329,985	3.18x	1.26	329,985	3.14x	1.26
+Dynamical	262,669	3.99x	1.00	262,669	3.95x	1.00
Gain	*-67,316 (-20.4%)*	*1.26x rel.*	*-0.26*	*-67,316 (-20.4%)*	*1.26x rel.*	*-0.26*
WIRE	264,453	3.97x	1.01	443,637	2.34x	1.69
+Dynamical	261,265	4.01x	1.00	338,994	3.06x	1.29
Gain	*-3,188 (-1.2%)*	*1.01x rel.*	*-0.01*	*-104,643 (-23.6%)*	*1.31x rel.*	*-0.40*
FINER	198,401	5.29x	0.76	329,985	3.14x	1.26
+Dynamical	150,343	6.97x	0.57	262,669	3.95x	1.00
Gain	*-48,058 (-24.2%)*	*1.32x rel.*	*-0.19*	*-67,316 (-20.4%)*	*1.26x rel.*	*-0.26*

E.7 Additional 3D Sparse Data Learning Results

Cryo-EM Reconstructions. The cryo-EM visualizations make the quantitative trends particularly clear (Figure 25), demonstrating that DINR improves not only global shape recovery but also effective structural resolution:

- **FFNet:** The surface reconstructed by Dynamical FFNet is significantly closer to the ground truth than its static counterpart. While the static reconstruction is noticeably coarse and exhibits disconnected artifacts, the dynamical variant produces a denser, more faithful structure that successfully recovers fine geometric detail and irregular local topology. The associated Fourier Shell Correlation (FSC) plot supports this: the entire FSC curve shifts upward and to the right, and the FSC $= 0.5$ crossing improves from 0.425 to 0.587, indicating that the dynamical model retains correlation at markedly higher spatial frequencies (Figure 26).

- **SIREN:** Static SIREN reconstructs an overly simplified global geometry that departs substantially from the ground truth. Dynamical SIREN mitigates this, restoring the overall shape more convincingly and recovering more local structure. Correspondingly, the FSC $=$ 0.5 cutoff increases from 0.226 to 0.308. However, both qualitative and quantitative results for SIREN remain clearly behind the FFNet-based models.

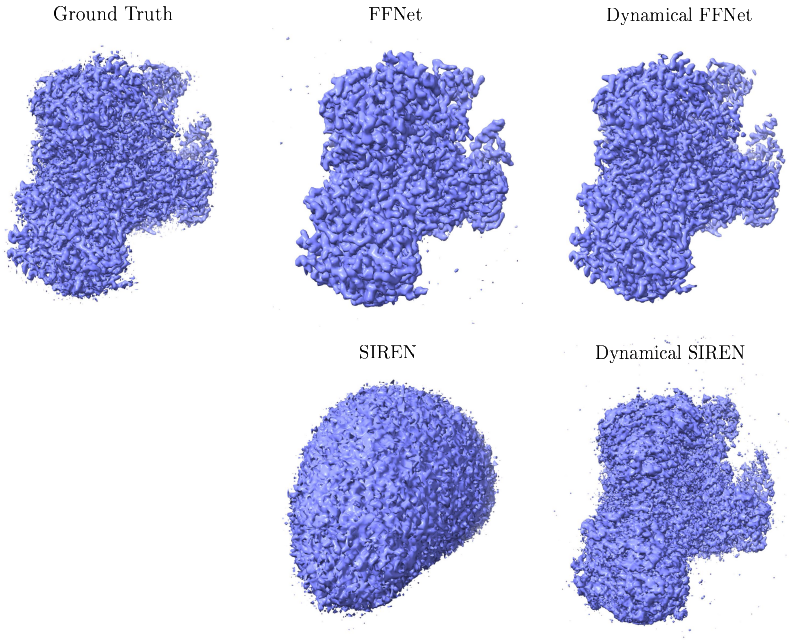

Figure 25: Qualitative comparison of cryo-EM reconstructions across static and dynamical INRs. Visual comparison of 3D reconstructions on sparse cryo-EM data for FFNet, Dynamical FFNet, SIREN, and Dynamical SIREN. The dynamical variants produce reconstructions that are closer to the ground truth, with clearer structural detail and fewer shape distortions than their static counterparts.

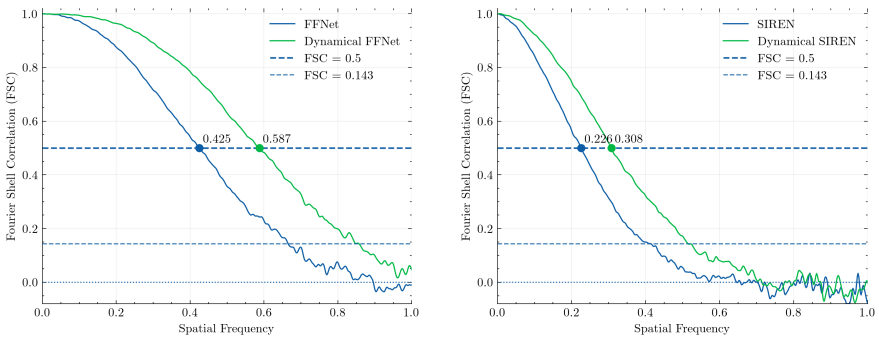

Figure 26: FSC analysis for cryo-EM reconstruction using INRs and DINRs.

Cloud Simulations. The cloud simulation experiments (Figure 27) further demonstrate DINR's robustness and highlight its capacity to model highly localized, multi-scale variations that typically cause static INRs to oversmooth or collapse to incorrect global statistics:

- **FFNet:** The static model produces an oversmoothed volume dominated by broad, low-frequency structures. In contrast, Dynamical FFNet faithfully recovers the complex, filamentary, and anisotropic patterns visible in the ground truth.
- **SIREN:** The contrast here is even sharper. The static SIREN reconstruction is severely distorted and fails to preserve the target field's qualitative appearance. Dynamical SIREN successfully restores the overall texture and structure to a much more plausible form. These visual improvements align perfectly with the near-saturation of quantitative metrics.

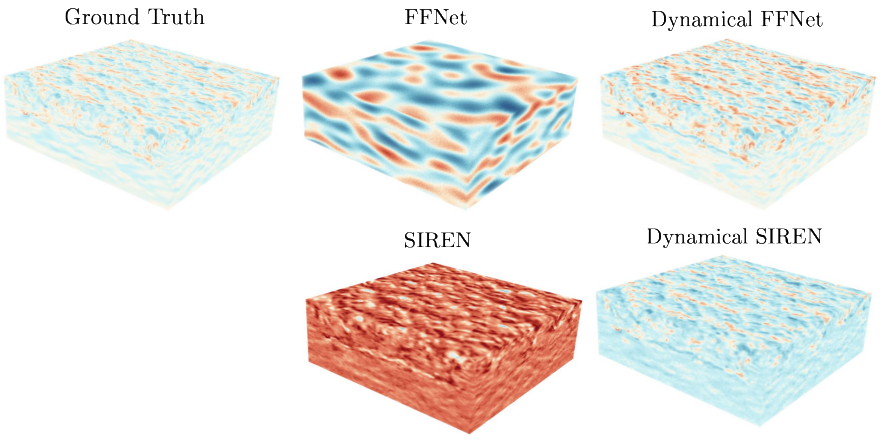

Figure 27: Qualitative comparison on sparse 3D cloud simulation data. Visual comparison of reconstructions for FFNet, Dynamical FFNet, SIREN, and Dynamical SIREN on cloud simulation data. Static baselines tend to oversmooth or distort the volumetric structure, whereas the dynamical variants better preserve the filamentary and spatially localized patterns present in the ground truth.