Iterative Refinement Neural Operators are Learned Fixed-Point Solvers: A Principled Approach to Spectral Bias Mitigation[1]

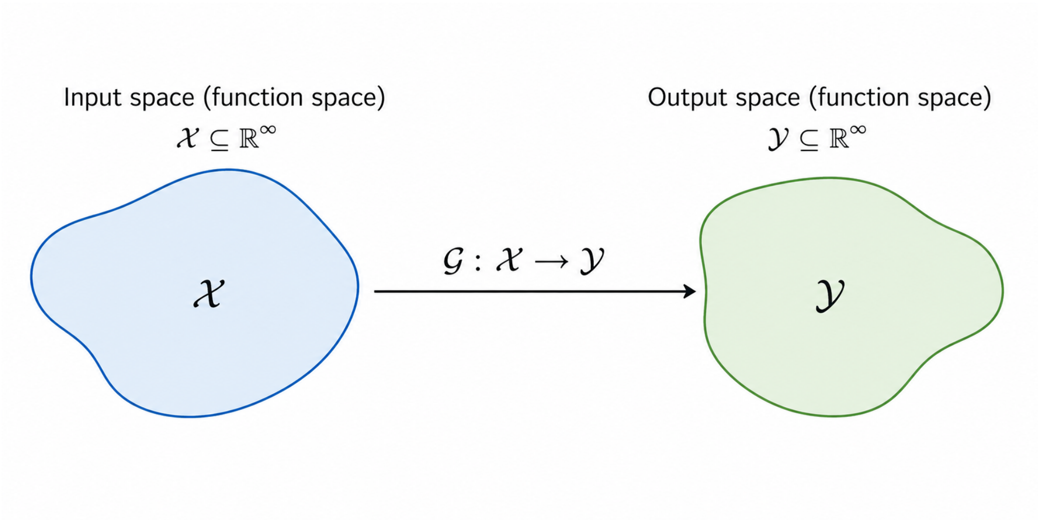
May 11, 2026

Presenter: Shuyuan Shang

[1] X. Liu et al. *Accepted by ICML 2026.*

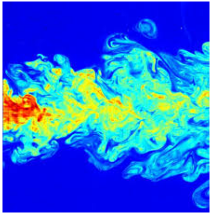
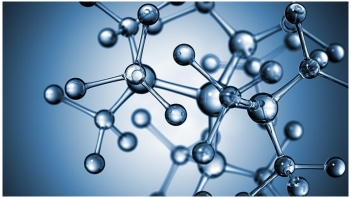
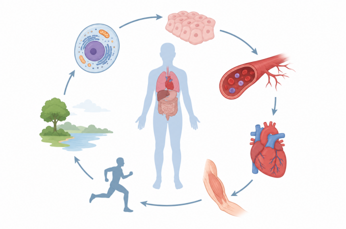
Operator Learning

- From learning functions to learning operators (mappings between function spaces).



Operator Learning

- **Application:** Fluid Dynamics, Material Science, Biophysics...

Neural Operator and its Bottlenecks

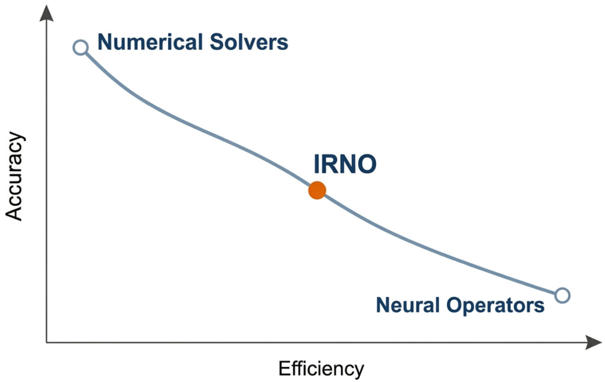
- **Neural Operator:** Utilizes spectral filters (FNO) or basis expansions (DeepONet), etc.
- **Bottlenecks**:
  1. **Single-pass Limit:** Accuracy is tied to model capacity and data scaling.
  2. **Spectral Bias:** Capture global structure but struggle with fine-scale details.

Speed comes at the cost of high-fidelity physical resolution.

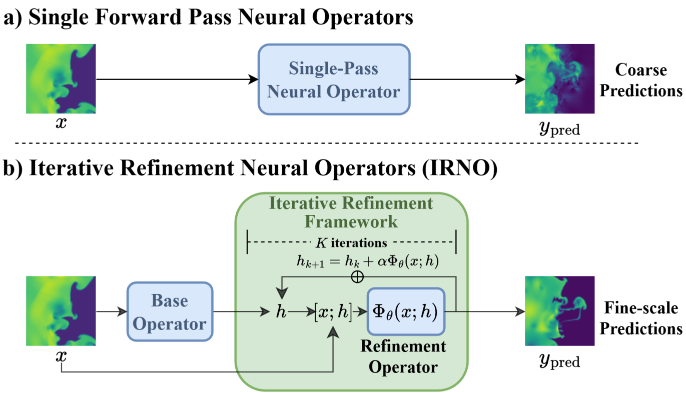
The Accuracy-Efficiency Gap

- **Numerical Solvers:** Iterative error correction; Too slow for real-time applications.
- **Neural Operators:** Faster while missing fine-scale physics

IRNO: Mimics a numerical solver in iterations while maintaining the efficiency of neural operators.



IRNO



- **IRNO:**
  1. A Base Operator for coarse initialization.
  2. A *shared-weight* refinement operator iteratively corrects the output at inference time.
  3. The network concatenates the original input x with the current estimate h_k to predict the residual update.

Learning Objective

Learning Objective: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{spatial}} + \beta_{\text{spectral}} \mathcal{L}_{\text{spectral}} + \beta_{\text{fp}} \mathcal{L}_{\text{fp}}$

- **Multi-step Supervision:** $\mathcal{L}_{\text{spatial}} = \frac{1}{K} \sum_{k=1}^{K} \|h_k - y\|^2$.
- **Fixed-point Regularization:** $\mathcal{L}_{\text{fp}} = \|\Phi_\theta(x, y)\|^2$.
- **Progressive Spectral Loss:** Let $\hat{h}_k = \mathcal{F}(h_k)$ be the Fourier transform of the estimate.

$$\mathcal{L}_{\text{spectral}}^{(k)} = \frac{1}{HW} \frac{1}{\bar{\rho}_k} \sum_{\omega} \rho(\omega, \lambda_k) \cdot \left| |\hat{h}_k(\omega)| - |\hat{y}(\omega)| \right|^2$$

where $\rho(\omega, \lambda_k) = 1 + (|\omega| / |\omega|_{\text{nyq}})^{\lambda_k}$. λ_k increases linearly from λ_{start} to λ_{end} across iterations.

$$\mathcal{L}_{\text{spectral}} = \frac{1}{K} \sum_{k} \mathcal{L}_{\text{spectral}}^{(k)}$$

IRNO

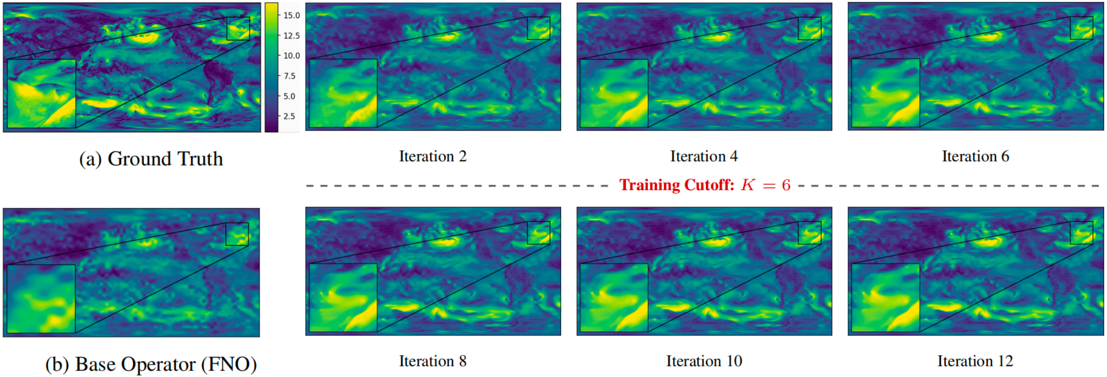


Figure: ERA5 16× super-resolution

Theoretical Outline

- **Theoretical Foundations:**
  - **Convergence Mechanism:** Fixed-point iteration
  - **Underlying Assumptions:** Three key conditions
- **Main Results:**
  - **Convergence Analysis:** Error decay
  - **Generalization Performance:** Error floor

Assumptions 1

We model IRNO as learning a fixed-point iteration:

$$h_{k+1} = h_k + \alpha \cdot \Phi_\theta\left(x, h_k\right), \quad k = 0, \ldots, K-1,$$

- **Local affine approximation:**

$$\Phi(x, h) = b(x) + A(x, h)e + R(x, h),$$

where $e = y - h$, $A(x, h)$ is a bounded linear operator, $\|R(x, h)\| \leq \frac{L}{2}\|e\|^2$. Ideally, a perfectly learned refinement operator would satisfy $||b|| = 0$.

Assumptions 2

- **Lipschitz continuity and strong monotonicity:** Assume there exists $\mu > 0$ such that for all $h \in B_\delta(y)$,

  $$\|A(x,h) - A(x,y)\|_{\text{op}} \leq \mu \|e\|$$

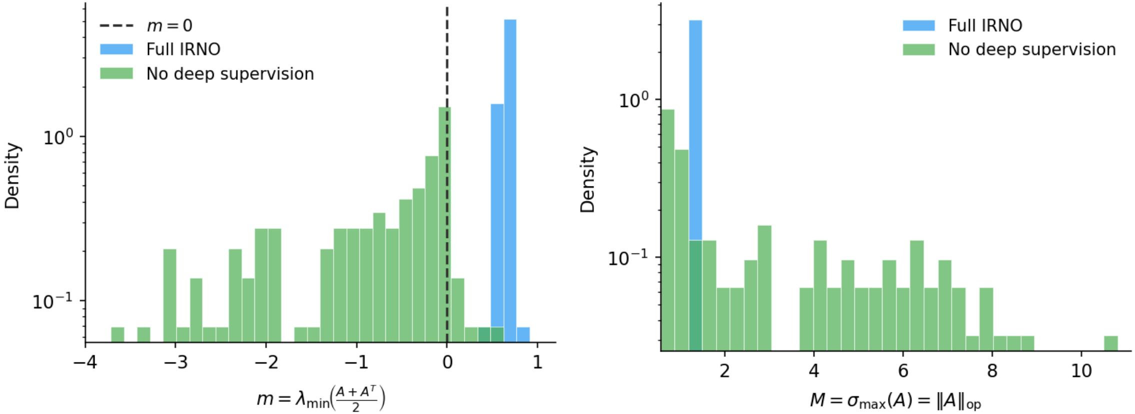
  Furthermore, $A(x,y)$ is locally bounded and strongly monotone, i.e., there exist constants $0 < m \leq M < \infty$ such that

  $$\langle A(x,y)e, e\rangle \geq m\|e\|^2, \quad \|A(x,y)\|_{\text{op}} \leq M.$$

  Under this condition, choosing $0 < \alpha < \frac{2m}{M^2}$ guarantees $\|I - \alpha A(x,y)\|_{\text{op}} = q < 1$.

Assumptions 2

1D synthetic experiment



Multi-step (Deep) Supervision: $\mathcal{L}_{\text{spatial}} = \frac{1}{K}\sum_{k=1}^{K}\|h_k - y\|^2.$

Assumption 2

SciML Benchmarks: TR-2D & Active Matter

Dataset	m (mean $\pm$ std)	$m > 0$
TR-2D U-Net (power iter.)	13.71 ± 2.41	100%
Active Matter U-Net (power iter.)	3.01 ± 3.12	99.5%

Assumptions 3

- **Initialization quality and invariant-ball:** The base operator provides a sufficiently accurate initialization $\|e_0\| < \min\left\{\delta, \frac{1-q}{2c}\right\}, \quad c = \alpha\left(\frac{L}{2} + \mu\right)$
  Further assume there exists r such that $\|e_0\| \le r < \min\left\{\delta, \frac{1-q}{2c}\right\}$ and

$$\alpha\|b\| \le r(1 - q - cr),$$

  which is a small-bias condition ensuring the iterates remain in $B_r(y) \subset B_\delta(y)$.

Fixed-point Regularization: $\mathcal{L}_{\text{fp}} = \|\Phi_\theta(x, y)\|^2$.

Main Results

Theorem (Quadratic–Linear Convergence)

In the linear-dominant regime ($b = 0$), the error decays geometrically:

$$\|e_k\| \lesssim q^k \|e_0\|$$

To achieve $\|e_k\| \leq \varepsilon$, the required iteration complexity is:

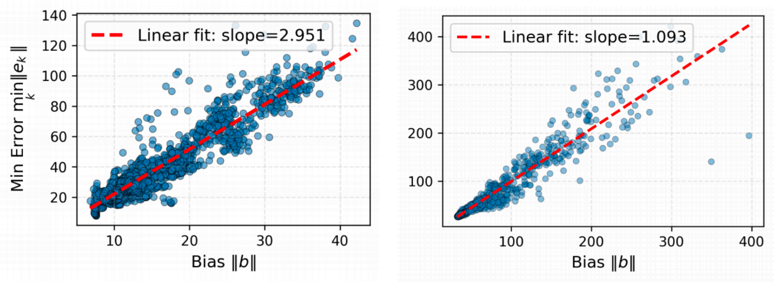
$$k = O\left(\frac{\log\left(\|e_0\| / \varepsilon\right)}{\log(1/q)}\right)$$

Main Results

Theorem (Convergence with Bias)

If $b = \Phi(x, y) \neq 0$ and $\|b\|$ is sufficiently small, there exists a unique fixed point h^* satisfying $h^* = h^* + \alpha\Phi\left(x, h^*\right)$. The iteration converges linearly to h^* with limiting error:

$$\|e^*\| \leq \frac{\alpha\|b\|}{1-q} + O\left(\|b\|^2\right)$$



(a) Active Matter (FNO) (b) TR-2D (TFNO)

Experimental Setup

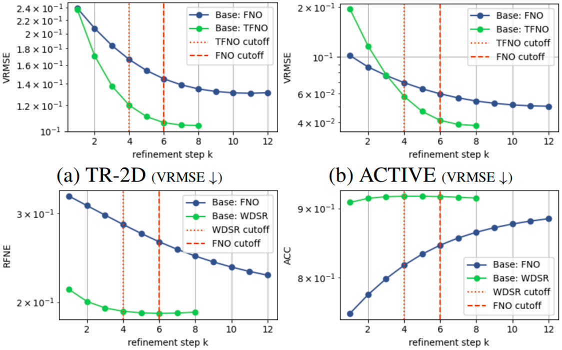
- **SciML benchmarks:**
  - Turbulent Radiative Layer (TR-2D)
  - Active Matter from the Well [2, 3, 4, 5]
  - ERA5 global weather $16\times$ super-resolution from SuperBench [6, 7, 8]
  - CE-Gauss from the RIGNO dataset [9]
- **Metrics:**
  - **TR-2D and Active Matter:** Variance-scaled RMSE (VRMSE)
  - **ERA5:** Anomaly correlation coefficient (ACC) and relative Frobenius norm error (RFNE)
  - **CE-Gauss:** L^2 error

Experimental Setup

- **Base Operators:**
  - **TR-2D and Active Matter:** FNO [10] and the Tucker-Factorized FNO (TFNO) [11].
  - **ERA5:** FNO and the Wide-Activated Deep Super-Resolution network (WDSR) [12].
  - **CE-Gauss:** RIGNO [9]
- **Baselines:**
  - **Standard Base Operator**
  - **Capacity-Matched Residual Model:** single-shot residual correction networks with parameter count and FLOPs exceed those of IRNO executed for K refinement steps.
  - **SOTA Models for Spectral Bias Mitigation:**
    1. High Frequency Scaling (HFS) [13]
    2. Hierarchical Neural Operator Transformer (HiNOTE) [14]

Global Convergence Behavior

Dataset	Metric	Base Model	Initial	IRNO (Ours)	Improvement(%)
TR-2D	VRMSE ↓	FNO	0.2394	**0.1309**	45.32%
		TFNO	0.2371	**0.1042**	56.05%
Active Matter	VRMSE ↓	FNO	0.1017	**0.0501**	50.73%
		TFNO	0.1981	**0.0387**	80.46%
ERA5	ACC ↑	FNO	0.7523	**0.8871**	17.92%
		WDSR	0.9091	**0.9104**	0.143%
	RFNE ↓	FNO	0.3247	**0.2345**	27.78%
		WDSR	0.2119	**0.1953**	7.83%



(a) TR-2D (VRMSE ↓) (b) ACTIVE (VRMSE ↓) (c) ERA5 (RFNE ↓) (d) ERA5 (ACC ↑)

- IRNO achieves 45–56% reductions on TR-2D, 51–80% on Active Matter in VRMSE and 27% in RFNE while increasing 18% ACC on ERA5.
- Stable extrapolation

State-of-the-Art Models Comparison

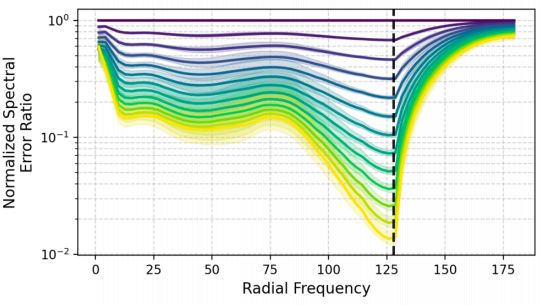
Table: Baseline comparisons on ERA5

Method	ACC ↑	RFNE ↓
ResUNet-HFS	0.892	0.225
HiNOTE	0.906	0.222
IRNO (WDSR, Ours)	**0.910**	**0.195**

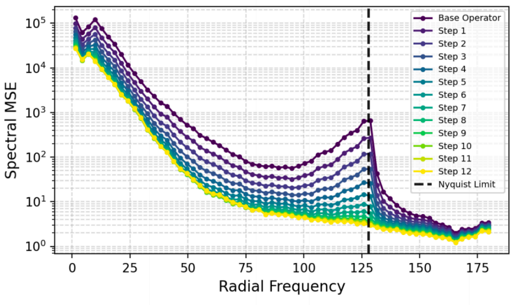
Table: HFS + IRNO on Active Matter

	Base (HFS)	$k=1$	$k=2$	$k=3$	$k=4$	$k=6$	$k=8$
VRMSE	0.0631	0.0518	0.0504	0.0495	0.0490	0.0486	0.0487

Spectral Dynamics



(a) Dataset-level median normalized spectral error ratios $\tilde{E}^{(k)}(\omega)$.



(b) Instance-level spectral MSE trajectories.

- Refinement mainly attenuates error in mid-to-high frequency bands.
- The largest error reductions are observed near the Nyquist limit.

Transferability Across Base Operators

Dataset	Metric	Base Model	Initial	IRNO (Ours)	Improvement(%)
TR-2D	VRMSE ↓	FNO	0.2394	**0.1309**	45.32%
		TFNO	0.2371	**0.1042**	56.05%
Active Matter	VRMSE ↓	FNO	0.1017	**0.0501**	50.73%
		TFNO	0.1981	**0.0387**	80.46%
ERA5	ACC ↑	FNO	0.7523	**0.8871**	17.92%
		WDSR	0.9091	**0.9104**	0.143%
	RFNE ↓	FNO	0.3247	**0.2345**	27.78%
		WDSR	0.2119	**0.1953**	7.83%

Table: Original performances

Dataset	Metric	Base + $\text{IRNO}_{\text{train}}$	Initial	IRNO	Improvement(%)
TR-2D	VRMSE ↓	FNO + $\text{IRNO}_{\text{TFNO}}$	0.2396	**0.0994**	58.53%
		TFNO + IRNO_{FNO}	0.2366	**0.1345**	43.15%
Active Matter	VRMSE ↓	FNO + $\text{IRNO}_{\text{TFNO}}$	0.1004	**0.0445**	55.66%
		TFNO + IRNO_{FNO}	0.1955	**0.1127**	42.36%
ERA5	ACC ↑	FNO + $\text{IRNO}_{\text{WDSR}}$	0.7523	**0.8022**	6.22%
		WDSR + IRNO_{FNO}	0.9091	**0.9219**	1.39%
	RFNE ↓	FNO + $\text{IRNO}_{\text{WDSR}}$	0.3247	**0.2823**	13.06%
		WDSR + IRNO_{FNO}	0.2119	**0.1935**	8.68%

Table: Transferbility performances

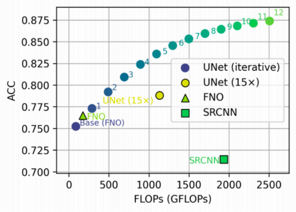
Generalization to Irregular Meshes

Table: IRNO ($K = 4$) with RIGNO on CE-Gauss: autoregressive rollout over 7 timesteps.

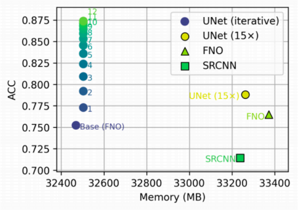
Model	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$
Base (RIGNO)	3.351	5.017	7.808	10.923	12.732	14.433	16.617
IRNO	2.931	4.359	6.605	8.990	10.371	11.450	13.080
Improvement	12.5%	13.1%	15.4%	17.7%	18.5%	20.7%	21.3%

- Early refinement holds up error accumulation in autoregressive rollout.

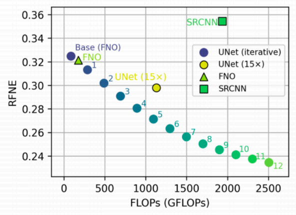
Cost–Performance Pareto Frontier



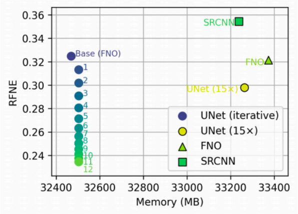
(a) ACC vs. FLOPs



(b) ACC vs. Memory



(c) RFNE vs. FLOPS



(d) RFNE vs. Memory

Performance gains are driven by the refinement mechanism rather than by increased model capacity.

Conclusion

- **IRNO:** A learned iterative refinement operator progressively reducing errors
- **Theoretical Foundations:** Contraction-based analysis
- **Experimental results:**
  1. Consistent spectral error reduction
  2. Stable extrapolation
  3. Improved cost–performance trade-offs relative to capacity-matched baselines

References I

[1] X. Liu et al. *Accepted by ICML 2026.*

[2] Ruben Ohana et al. "The well: a large-scale collection of diverse physics simulations for machine learning". In: *Advances in Neural Information Processing Systems* 37 (2024), pp. 44989–45037.

[3] Rudy Morel, Jiequn Han, and Edouard Oyallon. "DISCO: learning to DISCover an evolution Operator for multi-physics-agnostic prediction". In: *Forty-second International Conference on Machine Learning*. 2025. URL: https://openreview.net/forum?id=6EZ3MDDf6p.

[4] Zhikai Wu et al. "COAST: Intelligent Time-Adaptive Neural Operators". In: *2nd AI for Math Workshop @ ICML 2025*. 2025. URL: https://openreview.net/forum?id=cz99ALd3yv.

[5] Benjamin Holzschuh et al. "PDE-Transformer: Efficient and Versatile Transformers for Physics Simulations". In: *Forty-second International Conference on Machine Learning*. 2025. URL: https://openreview.net/forum?id=3BaJMRaPSx.

References II

[6] Pu Ren et al. "SuperBench: A Super-Resolution Benchmark Dataset for Scientific Machine Learning". In: *Journal of Data-centric Machine Learning Research* (2025). Dataset Certification, Reproducibility Certification. URL: https://openreview.net/forum?id=OJ6zUcWldW.

[7] Sheikh Md Shakeel Hassan et al. "BubbleML: A Multiphase Multiphysics Dataset and Benchmarks for Machine Learning". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh et al. Vol. 36. Curran Associates, Inc., 2023, pp. 418–449.

[8] Wuyang Chen et al. "Data-Efficient Operator Learning via Unsupervised Pretraining and In-Context Learning". In: *Advances in Neural Information Processing Systems*. Ed. by A. Globerson et al. Vol. 37. Curran Associates, Inc., 2024, pp. 6213–6245. DOI: 10.52202/079017-0201.

[9] Sepehr Mousavi et al. "RIGNO: A Graph-based framework for robust and accurate operator learning for PDEs on arbitrary domains". In: *Advances in Neural Information Processing Systems*. Vol. 38. 2025.

References III

[10] Zongyi Li et al. "Fourier Neural Operator for Parametric Partial Differential Equations". In: *International Conference on Learning Representations (ICLR)*. See also arXiv:2010.08895 (2020). 2021. URL: https://openreview.net/forum?id=c8P9NQVtmnO.

[11] Jean Kossaifi et al. "A Library for Learning Neural Operators". In: *arXiv preprint arXiv:2412.10354* (2025).

[12] Jiahui Yu et al. "Wide activation for efficient and accurate image super-resolution". In: *arXiv preprint arXiv:1808.08718* (2018).

[13] Siavash Khodakarami et al. "Mitigating spectral bias in neural operators via high-frequency scaling for physical systems". In: *Neural Networks* 193 (2026), p. 108027. ISSN: 0893-6080. DOI: https://doi.org/10.1016/j.neunet.2025.108027. URL: https://www.sciencedirect.com/science/article/pii/S0893608025009074.

References IV

[14] Xihaier Luo, Xiaoning Qian, and Byung-Jun Yoon. "Hierarchical Neural Operator Transformer with Learnable Frequency-aware Loss Prior for Arbitrary-scale Super-resolution". In: *arXiv preprint arXiv:2405.12202* (2024).