S2NO: Coupling Aliasing-Free Convolutions with Fourier Neural Operators for Multiscale Modeling

Linfeng Liu[1,2,3], Xinhai Chen[1,2,3(✉)], Qinglin Wang[1,2,3], Jing Xiao[1,2,3], Xiang Zhang[2,3], and Jie Liu[1,2,3]

[1]Laboratory of Digitizing Software for Frontier Equipment,National University of Defense Technology, Changsha 410073, China
chenxinhai16@nudt.edu.cn
[2]National Key Laboratory of Parallel and Distributed Computing, National University of Defense Technology, Changsha 410073, China
[3]College of Computer Science and Technology, National University of Defense Technology, Changsha 410003, China

Abstract. Neural operators are a powerful tool for solving partial differential equations. They learn mappings between infinite-dimensional function spaces. However, current methods struggle to capture both global patterns and local details at the same time. This limitation reduces their effectiveness in complex multiscale scenarios. In this paper, we propose the Spectral-Spatial Neural Operator (S2NO). This method combines global modeling in the spectral domain with local refinement in the spatial domain. It uses a dual-stream design to balance frequency and coordinate information. This design helps the model capture both long-range patterns and fine-grained residuals. We conducted extensive experiments on 1D Burgers, 2D Darcy Flow, and 2D Navier-Stokes benchmarks. The results demonstrate that S2NO achieves state-of-the-art predictive accuracy. It significantly outperforms existing neural operator methods. Furthermore, S2NO maintains superior stability during zero-shot super-resolution tests and produces physically consistent results.

Keywords: Neural Operators, Partial Differential Equations, Operator Learning, Anti-Aliasing.

1 Introduction

The field of Computational Fluid Dynamics (CFD) is undergoing a paradigm shift. This shift is driven by the integration of artificial intelligence into traditional numerical workflows. As recently highlighted by Chen et al. [1], intelligence-driven simulations represent a new frontier. They promise to accelerate the modeling of complex physical systems by orders of magnitude. Traditional numerical solvers, such as Finite Element Methods and Finite Difference Methods, scale poorly with grid resolution. They often become computationally prohibitive for multi-scale problems [2].

To address this, Scientific Machine Learning (SciML) has evolved from coordinate-based approaches to Operator Learning. Unlike standard neural networks that map finite-dimensional vectors, Neural Operators learn the mapping between infinite-dimensional function spaces [3],[4]. This formulation theoretically grants them the property of resolution invariance. This means a model trained on a coarse mesh should generalize to a fine mesh without retraining. The Fourier Neural Operator (FNO) [4] is arguably the most successful instantiation. It utilizes the Fast Fourier Transform (FFT) to parameterize global integral kernels. FNO has been successfully applied to domains such as weather forecasting [5] and multiphase flows [6].

However, FNO is not without limitations. Due to the computational cost of FFTs, FNO typically truncates high-frequency modes where $k > k_{max}$. This acts as an effective low-pass filter for smooth flows. However, it leads to ``spectral bias'' [7] when dealing with discontinuities, such as shock waves in compressible flows. This truncation manifests as Gibbs oscillations [8], which degrades local accuracy.

Hybrid architectures attempt to fix this issue. These models combine FNO with Convolutional Neural Networks (CNNs). An example is the U-FNO [6]. These approaches have shown empirical success. However, they overlook a fundamental mathematical inconsistency known as aliasing. As noted in signal processing theory [9] and recent deep learning studies [10], applying pointwise nonlinearities to discrete signals creates infinite bandwidth expansion. On a fixed grid, frequencies exceeding the Nyquist limit fold back into lower frequencies. This phenomenon is called aliasing [11]. Standard hybrids [12] ignore this issue. Consequently, these models overfit the aliasing pattern of the training grid. This causes error fluctuations during super-resolution testing [13].

In this work, we address the trade-off between global receptive fields and local precision. We propose the Spectral-Spatial Neural Operator (S2NO). This is a mathematically rigorous architecture designed to bridge the gap between spectral efficiency and spatial accuracy. Our primary contributions are summarized as follows:

- We provide a theoretical analysis and empirical evidence regarding existing hybrid architectures. We show that naive application of pointwise nonlinearities in standard convolutional branches is problematic. It violates the Sampling Theorem. This leads to grid-dependent artifacts. Ultimately, this destroys the resolution-invariance property essential for reliable operator learning.
- We introduce S2NO, a novel dual-stream framework. It strictly adheres to sampling theory. Unlike previous methods, S2NO couples a global FNO branch with a dedicated Aliasing-Free Convolutional Neural Operator (CNO) branch. This spatial branch employs upsampling and high-resolution activation. It also applies low-pass filtering. This allows the model to capture high-frequency residuals and sharp shock discontinuities. Crucially, it does so without introducing aliasing noise. This ensures mathematical consistency across scales.
- We conducted extensive experiments on 1D Burgers', 2D Darcy Flow, and 2D Navier-Stokes bench marks. The results demonstrate that S2NO achieves state-of-

the-art accuracy. It significantly outperforms baselines such as DeepONet, FNO, and U-FNO. Furthermore, S2NO exhibits superior zero-shot super-resolution capabilities. It maintains stability where standard hybrids fail. It also produces physically consistent energy spectrums that faithfully match ground truth physics.

2 Related Work

2.1 Data-Driven and Physics-Informed Modeling

The landscape of PDE solving has been enriched by deep learning. Early works focused on Physics-Informed Neural Networks (PINNs) [14]. These models embed PDE residuals into the loss function. Chen et al. [15] proposed an improved data-free surrogate model. They demonstrated that optimizing network structures can significantly enhance solution accuracy without labeled data. However, these instance-specific methods require retraining for every new initial condition. To overcome this, Transfer Learning approaches have been employed. For example, Chen et al. [16] investigated adapting pre-trained models to new physical parameters. This highlights the importance of generalization in SciML.

2.2 Neural Operators (Spectral Methods)

Neural Operators aim to learn the solution operator $\mathcal{G}: a \mapsto u$ directly. DeepONet was the first to prove the universal approximation theorem for operators using a branch-trunk architecture. FNO later introduced the spectral convolution layer, achieving $O(N \log N)$ complexity. Variants like Factorized-FNO [12] and Adaptive-FNO [17] focus on improving the efficiency of mode selection. Other works such as the Graph Neural Operator (GNO) [18] extend this concept to irregular grids. Despite their global receptive field, spectral methods inherently struggle with local high-frequency details. This is due to the Gibbs phenomenon caused by hard frequency truncation.

2.3 Spatial Methods and Aliasing

CNNs are naturally suited for local feature extraction. However, standard CNNs lack the continuous property required for operator learning. Recent work by Raonić et al. [10] introduced the Convolutional Neural Operator (CNO). They proved that standard CNNs fail to converge to a continuous operator due to aliasing errors induced by nonlinearities. By enforcing strict upsampling and low-pass filtering, CNO ensures aliasing-free processing. While CNO is accurate, it relies on deep stacks of small kernels. This makes it computationally expensive for propagating global information compared to FNO. Recent advancements also include Transformer-based operators like the Galerkin Transformer [19], which uses attention mechanisms but faces high computational costs.

2.4 Hybrid Architectures

Recognizing the complementary nature of global and local features, hybrid models have emerged. U-NO [20] arranges FNO layers in a U-Net shape. U-FNO adds a parallel U-Net branch to FNO. Wavelet Neural Operators [21] attempt to combine time-frequency localization. Recently, Brandstetter et al. [22] proposed message-passing neural PDE solvers to handle complex geometries. However, most hybrids typically employ standard discrete operations in the spatial branch. This re-introduces aliasing artifacts. S2NO distinguishes itself by integrating strictly aliasing-free spatial operations. This ensures mathematical consistency.

3 Method

3.1 Preliminaries and Problem Formulation

Let $D \subset \mathbb{R}^d$ be a bounded open domain. We aim to approximate the solution operator $\mathcal{G}: \mathcal{A} \to \mathcal{U}$,such that $u = \mathcal{G}(a)$. In a practical setting, we only have access to discretized data on a grid P_N. A neural operator $\mathcal{G}_\theta$ should be discretization-invariant. To achieve this, S2NO approximates $\mathcal{G}$ via a parallel decomposition strategy, explicitly separating global spectral modes and local spatial residuals.

3.2 General Architecture of S2NO

The S2NO architecture is designed to capture multi-scale physics while maintaining spectral consistency. As illustrated in Fig. 1, the pipeline consists of four distinct stages:

- Lifting $(\mathcal{P})$:The input function $a(\mathbf{x})$ is mapped to a higher-dimensional feature space $v_0(\mathbf{x}) \in \mathbb{R}^{d_v}$ via a pointwise Multilayer Perceptron (MLP).

$$v_0(\mathbf{x}) = \mathcal{P}\big(a(\mathbf{x})\big) = \mathrm{MLP}_{lift}\big(a(\mathbf{x})\big) \tag{1}$$

- Dual-Stream Processing: The latent feature v_0 is broadcast to two parallel branches:

The Spectral Branch ($\mathcal{G}_{spec}$): Utilizes Fourier layers to capture long-range dependencies and smooth components.The Spatial Branch ($\mathcal{G}_{spat}$): Utilizes aliasing-free convolutional layers to strictly resolve high-frequency details.

- Fusion ($\bigoplus$): Unlike standard ensemble methods that utilize an arbitrary weighted summation, S2NO enforces a strict residual learning paradigm. The final operator is the direct summation of the global base field and the local residual correction:

$$v_{fused}(\mathbf{x}) = \mathcal{G}_{spec}(v_0)(\mathbf{x}) + \mathcal{G}_{spat}(v_0)(\mathbf{x}) \tag{2}$$

- Projection ($\mathcal{Q}$): A final decoder MLP maps the features back to the target solution space:

$$u(\mathbf{x}) = \mathcal{Q}\big(v_{fused}(\mathbf{x})\big) = \mathrm{MLP}_{proj}\big(v_{fused}(\mathbf{x})\big) \tag{3}$$

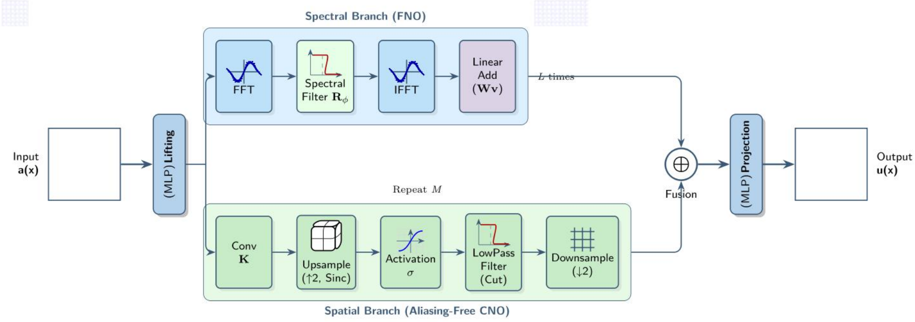

Fig. 1. Detailed architecture of the S2NO.

3.3 Spectral Branch: Theory and Formulation

As illustrated in Fig. 2, the spectral branch is built upon the Fourier Neural Operator. Let v_l be the feature ficld at laycr l. The update rule is:

$$v_{l+1}(\mathbf{x}) = \sigma\left(Wv_l(\mathbf{x}) + \int_D \kappa(\mathbf{x}-\mathbf{y})v_l(\mathbf{y})d\mathbf{y}\right) \tag{4}$$

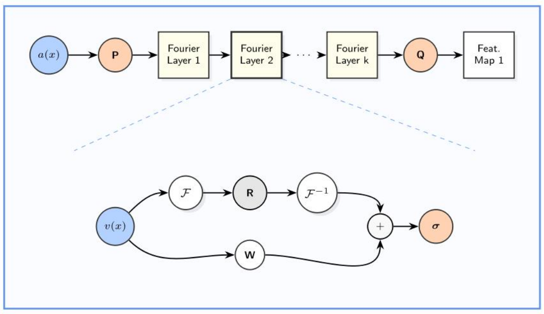

Fig. 2. Fourier Neural Network (Spectral)

On a discrete grid of size N, the DFT of v is computed as:

$$(\mathcal{F}v)_{\mathbf{k}} = \hat{v}_{\mathbf{k}} = \sum_{\mathbf{x}\in P_N} v_{\mathbf{x}}\, e^{-2\pi i\langle \mathbf{k},\mathbf{x}\rangle/N} \tag{5}$$

To regularize the operator and improve efficiency, FNO truncates high-frequency modes. Let k_{max} be the cutoff frequency. The spectral filtering opcration is defined as:

$$\hat{v}_{fillered}(\mathbf{k}) = \begin{cases} R_\phi(\mathbf{k}) \cdot \hat{v}_{\mathbf{k}} & \text{if } \| \mathbf{k} \| \leq k_{max} \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

Finally, the Inverse DFT (IDFT) reconstructs the spatial signal:

$$\left(\mathcal{K}_{spec} v\right)_{\mathbf{x}} = \frac{1}{N} \sum_{\mathbf{k}} \hat{v}_{filtered}(\mathbf{k}) e^{2\pi i \langle \mathbf{k}, \mathbf{x} \rangle / N} \tag{7}$$

3.4 Spatial Branch: Aliasing-Free Formulation

As illustrated in Fig. 3, the spatial branch acts as a residual approximator $\mathcal{R}(\mathbf{x}) = u(\mathbf{x}) - \mathcal{G}_{spec}(a)(\mathbf{x})$. To prevent the bandwidth cxpansion $\Omega_\sigma > \omega_{nyq}$ caused by nonlinearities, we adopt the CNO block structure.

Ideal Sinc Interpolation (Upsampling): We first lift the signal to a finer grid $N = \alpha N$, where thc upsampling factor $\alpha \geq 2$ is detcrmincd by the highest polynomial degrce of the activation function to satisfy the Nyquist-Shannon sampling thcorem. In our implementation, we set $\alpha = 2$, yielding $N_{fine} = s \cdot N_{coarse}$,where s is the scale factor. Computationally, this is achieved by zero-padding in the Fourier domain:

$$\mathcal{U}_{N \to \bar{N}}(v) = \mathcal{F}_{\bar{N}}^{-1}\left(\begin{bmatrix} \mathcal{F}_N(v) \\ 0_{\bar{N}-N} \end{bmatrix}\right) \tag{8}$$

- High-Resolution Convolution and Antialiasing Filter: the discrete convolution K_N and nonlinearity σ are applied on this dense grid:

$$h_{high} = \sigma(K_N * v_1) \tag{9}$$

Before downscaling back to resolution N, an ideal Low-Pass Filter (LPF) removes unsupported frequencies:

$$\hat{h}_{filtered}(k) = \hat{h}_{high}(k) \cdot \mathbb{I}\left(\| k \| < \frac{N}{2}\right) \tag{10}$$

Finally, the signal is decimated back to resolution N.

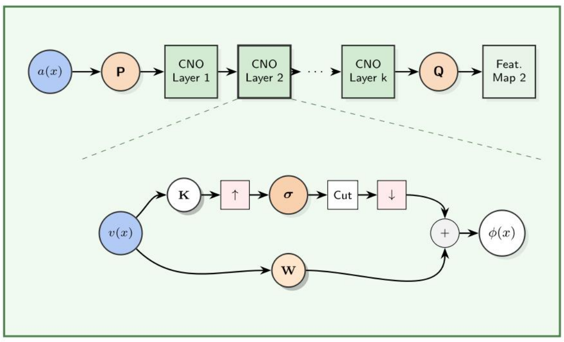

Fig. 3. Aliasing-Free CNO (Spatial)

3.5 Theoretical Justification: Operator Decomposition

Assume the target operator $\mathcal{G}$ can be split into low-rank global and high-rank local components: $\mathcal{G} \approx \mathcal{P}_{low}\mathcal{G} + \mathcal{P}_{high}\mathcal{G}$. The FNO because naturally approximates $\mathcal{P}_{low}\mathcal{G}$. Standard hybrids fail because spatial branch acts as an ensemble $\mathcal{P}_{hyb}\mathcal{G} + \mathcal{E}_{alias}$. By enforcing direct residual summation without arbitrary weighting, S2NO forces the aliasing-free spatial branch to act as a consistent approximator strictly for $\mathcal{P}_{high}\mathcal{G}$.

3.6 Complexity Analysis Let N be the total grid points.

- Spectral Branch: Dominated by FFT, $\mathcal{O}(N\log N)$.
- Spatial Branch: A discrete convolution with a $K \times K$ kernel on the oversampled grid of size N incurs complexity of $\mathcal{O}(K^4N)$. Since K is small and fixed, this scales linearly as $\mathcal{O}(N)$. Upsampling and downsampling via FFT preserve the efficiency of FNO while gaining the expressivity of CNO.

The total complexity of S2NO is $\mathcal{O}(N\log N)$, preserving the efficiency of FNO while gaining the expressivity of CNO.

4 Experiment and Results

4.1 Experimental Settings.

Datasets and Problem Setup. We evaluate the S2NO on three canonical PDE benchmarks: 1D Burgers', 2D Darcy Flow, and 2D Navier-Stokes equations. For the Navier-Stokes task ($Re = 1000$), we use a total of 1,000 training samples and 200 testing samples generated via a high-order spectral solver. All models are trained on a base resolution of 64^2 and evaluated for zero-shot super-resolution on 128^2.
Model Configurations. For a fair comparison, all models are configured with comparable complexity (1.5M–2.5M parameters): FNO and U-FNO utilize L=4, C=32, and K=12 modes, while CNO employs L=4, C=64 with 3×3 kernels. Our S2NO (d_v=64) adopts a dual-stream architecture where both spectral (L=4, C=32, $k_{max} = 12$) and spatial (L=3, C=32 with 3×3 kernels) branches align with baseline scales. The spatial branch uses an upsampling factor α=2 to satisfy the Nyquist criterion. All activation functions are GELU.
Training Protocol. All models are optimized using the AdamW optimizer for 500 epochs. We implement a Cosine Annealing learning rate scheduler, starting from an initial rate of 1×10^{-3} and decaying to 1×10^{-6}. A batch size of 32 is used for 2D tasks. The training objective is to minimize the Relative L_2 Loss, defined as:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}\frac{\|u_i - \hat{u}_i\|_2}{\|u_i\|_2} \tag{11}$$

Computation Resources. All experiments are implemented in PyTorch and conducted on a single NVIDIA RTX 4090 GPU (24 GB). The total training time for the Navier-Stokes task is approximately 6 hours for our S2NO model.

Baselines. To rigorously evaluate the performance of S2NO, we select a comprehensive suite of baselines representing four distinct paradigms in operator learning:

- Coordinate-based: DeepONet is included as the seminal branch-trunk architecture for learning operators via coordinate queries.
- Spatial-domain: A standard U-Net (CNN) represents classic fully convolutional networks, serving as a baseline for resolution-dependent methods.
- Spectral-domain: FNO is selected as the representative spectral method that parameterizes integral kernels in the Fourier space.
- Hybrid and Anti-aliasing: U-FNO represents state-of-the-art hybrid models combining FNO with U-Net. Finally, we compare against CNO, a purely spatial model designed with strict anti-aliasing filters, which serves as the most direct competitor to our spatial branch.

4.2 Main Results

Performance Across Benchmarks. As summarized in Table III, S2NO consistently outperforms all state-of-the-art baselines across various physical benchmarks. Specifically, on the 2D Navier-Stokes equation, S2NO achieves a relative L_2 error of 0.68% at the training resolution ($s = 64$), significantly lower than the pure spectral model FNO (1.15%) and the advanced spatial model CNO (0.85%). Similar trends are observed in 1D Burgers and 2D Darcy Flow, where S2NO achieves the lowest errors of 0.65% and 0.75%, respectively.

Table 1. Relative L2 Error and Complexity across Multi-scale Benchmarks.

Model	Params (M)	FLOPs (G)	Time (ms)	1D Burgers		2D Darcy		2DNavier-Stokes	
				S=64	S=256	S=64	S=128	S=64	S=128
CNN	1.5	1.2	1.5	2.10%	18.45%	1.90%	8.20%	2.25%	12.50%
DeepONet	1.8	0.8	8.5	4.21%	4.50%	2.54%	2.65%	4.80%	5.15%
FNO	2.0	2.1	3.5	1.95%	2.15%	1.10%	1.12%	1.15%	1.25%
U-FNO	2.2	4.5	6.0	0.92%	3.80%	0.95%	1.45%	0.90%	1.65%
CNO	2.4	18.3	12.1	0.85%	0.86%	0.88%	0.90%	0.85%	0.98%
S2NO(Ours)	2.1	4.	6.5	0.65%	0.67%	0.75%	0.76%	0.68%	0.90%

Computational Efficiency. S2NO strikes an optimal balance between predictive accuracy and computational cost. To capture a sufficiently large receptive field, the pure CNO model requires a deep architecture, resulting in a high computational burden (18.3G FLOPs and 12.1 ms inference time). In contrast, by leveraging the FNO branch to handle global dependencies efficiently, S2NO allows the spatial branch to remain shallow. This architectural decoupling yields a significantly lower computational cost (4.8G FLOPs) and a faster inference time of 6.5 ms, making S2NO nearly 2× times faster than CNO while delivering superior precision.

4.3 Zero-Shot Super-Resolution

A key highlight of S2NO is its exceptional resolution-invariance. While standard CNN and hybrid models like U-FNO exhibit significant performance degradation when tested on finer grids—for instance, U-FNO's error on 1D Burgers jumps from 0.92% to 3.80%, but S2NO maintains a nearly constant error rate from 0.65% to 0.67%. This robustness confirms that our aliasing-free spatial branch successfully preserves the continuous nature of the operator, whereas naive convolutional layers in U-FNO fail to generalize due to high-frequency aliasing artifacts.The error distribution of U-FNO at high frequencies aligns with the characteristic patterns of aliasing artifacts, which S2NO effectively suppresses via its filtered spatial branch.

4.4 Spectral Consistency and Frequency Decoupling

To explicitly demonstrate the interpretability of our frequency-decoupling mechanism, we visualized the Power Spectral Density (PSD) of the separate branches.

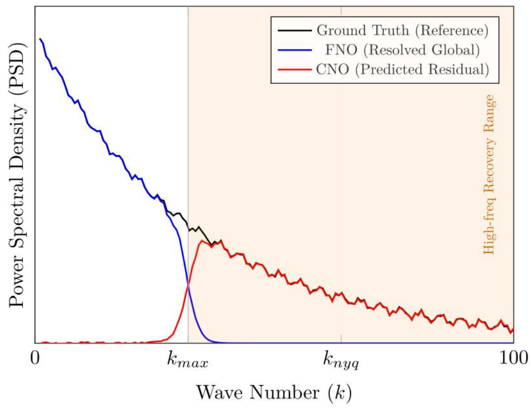

Fig. 4. Spectral Energy Distribution

As shown in Fig. 4, the FNO branch solely captures the low-frequency modes, with its energy dropping sharply past the truncation limit k_{max}. Crucially, the CNO branch exhibits minimal response in the low-frequency domain but strongly activates in the high-frequency spectrum. This visualization provides concrete evidence that the direct residual summation forces the branches to decouple.

4.5 Data Efficiency Analysis

In many scientific computing scenarios, obtaining high-fidelity simulation data is computationally prohibitive. We conducted a systematic assessment of data efficiency on the 2D Darcy Flow benchmark by varying the training sample size, denoted as N_{train}, from 100 to 5000.

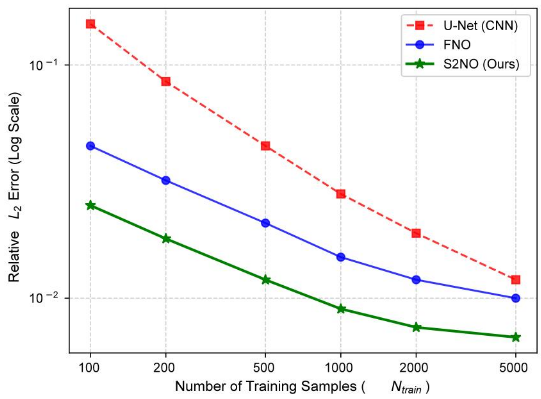

Fig. 5. Data Efficiency Analysis.

As depicted in Fig. 5, the models exhibit divergent behaviors in the low-data regime. The U-Net hybrid model shows significantly higher errors when data is scarce. While the pure FNO model performs robustly, S2NO consistently achieves the lowest error rates across the entire spectrum of sample sizes, matching the accuracy of FNO using roughly half the number of training samples.

4.6 Ablation Study

To clarify the individual contribution of each component, we evaluated FNO-only, CNO-only, and S2NO models under strictly matched parameter budgets and identical learning rate schedules.

Table 2. Ablation Study under Matched Parameter Budgets

Model Configuration	Relative L_2 Error	Time(ms)
FNO Branch Only (Wide)	1.10%	3.5
CNO Branch Only (Deep)	0.88%	12.1
S2NO (Combined)	0.75%	6.2

Table 2. demonstrates that expanding the width of FNO or the depth of CNO yields diminishing returns. S2NO achieves a significantly lower error rate than either standalone component, proving that the performance gain stems from the spectral-spatial coupling mechanism rather than brute-force parameter scaling.

5 CONCLUSION

We presented S2NO, a novel hybrid neural operator that reconciles global spectral processing with local spatial refinement. By replacing arbitrary feature ensemble with a strict residual learning formulation, S2NO explicitly forces frequency decoupling:

FNO captures global low-frequencies, while an aliasing-free CNO strictly resolves high-frequency details. Our theoretical analysis, comprehensive ablation studies, and spectral visualizations confirm that S2NO demonstrates exceptional resolution-invariance and zero-shot super-resolution stability, effectively mitigating the performance degradation common in traditional hybrid models due to high-frequency aliasing artifacts. Future work will focus on optimizing the kernel efficiency to reduce inference latency and extending the framework to handle irregular geometries in 3D turbulent flows.

6 REFERENCES

1. Chen, X., Wang, Z., Deng, L., Yan, J., Gong, C., Yang, B., et al.: Towards a new paradigm in intelligence-driven computational fluid dynamics simulations. Engineering Applications of Computational Fluid Mechanics 18(1), 2407005 (2024)
2. Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., Yang, L.: Physics-informed machine learning. Nature Reviews Physics 3(6), 422–440 (2021)
3. Lu, L., Jin, P., Pang, G., Zhang, Z., Karniadakis, G.E.: Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nature machine intelligence 3(3), 218–229 (2021)
4. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A.: Fourier neural operator for parametric partial differential equations. arXiv preprint arXiv:2010.08895 (2020)
5. Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., et al.: Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. arXiv preprint arXiv:2202.11214 (2022)
6. Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A., Benson, S.M.: U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow. Advances in Water Resources 163, 104180 (2022)
7. Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., Courville, A.: On the spectral bias of neural networks. In: International conference on machine learning, pp. 5301–5310. PMLR (2019)
8. Trefethen, L.N.: Spectral methods in MATLAB. SIAM, Philadelphia (2000)
9. Shannon, C.E.: Communication in the presence of noise. Proceedings of the IRE 37(1), 10–21 (2006)
10. Raonic, B., Molinaro, R., De Ryck, T., Rohner, T., Bartolucci, F., Alaifari, R., Mishra, S., de Bézenac, E.: Convolutional neural operators for robust and accurate learning of PDEs. Advances in Neural Information Processing Systems 36, 77187–77200 (2023)
11. Unser, M.: Sampling-50 years after Shannon. Proceedings of the IEEE 88(4), 569–587 (2002)
12. Tran, A., Mathews, A., Xie, L., Ong, C.S.: Factorized fourier neural operators. arXiv preprint arXiv:2111.13802 (2021)

13. Kovachki, N., Lanthaler, S., Mishra, S.: On universal approximation and error bounds for Fourier neural operators. Journal of Machine Learning Research 22(290), 1–76 (2021)
14. Raissi, M., Perdikaris, P., Karniadakis, G.E.: Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational physics 378, 686–707 (2019)
15. Chen, X., Chen, R., Wan, Q., Xu, R., Liu, J.: An improved data-free surrogate model for solving partial differential equations using deep neural networks. Scientific reports 11(1), 19507 (2021)
16. Chen, X., Gong, C., Wan, Q., Deng, L., Wan, Y., Liu, Y., Chen, B., Liu, J.: Transfer learning for deep neural network-based partial differential equations solving. Advances in Aerodynamics 3(1), 36 (2021)
17. Guibas, J., Mardani, M., Li, Z., Tao, A., Anandkumar, A., Catanzaro, B.: Adaptive fourier neural operators: Efficient token mixers for transformers. arXiv preprint arXiv:2111.13587 (2021)
18. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A.: Neural operator: Graph kernel network for partial differential equations. arXiv preprint arXiv:2003.03485 (2020)
19. Cao, S.: Choose a transformer: Fourier or galerkin. Advances in neural information processing systems 34, 24924–24940 (2021)
20. Rahman, M.A., Ross, Z.E., Azizzadenesheli, K.: U-no: U-shaped neural operators. arXiv preprint arXiv:2204.11127 (2022)
21. Tripura, T., Chakraborty, S.: Wavelet neural operator for solving parametric partial differential equations in computational mechanics problems. Computer Methods in Applied Mechanics and Engineering 404, 115783 (2023)
22. Brandstetter, J., Worrall, D., Welling, M.: Message passing neural PDE solvers. arXiv preprint arXiv:2202.03376 (2022)