PHYSICS-INFORMED SHEARLET NEURAL OPERATOR (PI-SHEARLETNO) FOR PARAMETRIC PARTIAL DIFFERENTIAL EQUATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper introduces the Physics-Informed Shearlet Neural Operator (PI-ShearletNO), a framework for learning solution operators of parametric partial differential equations. The model combines neural operator learning with the geometric sensitivity of shearlet transforms, which provide a multiscale and directional representation. By embedding the governing physical laws directly into the learning objective as constraints, PI-ShearletNO produces predictions that fit observed data while remaining consistent with the underlying PDE and boundary conditions. This physics-informed formulation improves generalization and accuracy compared with purely data-driven operator learning. We demonstrate the efficacy of PI-ShearletNO through numerical experiments on benchmark problems, highlighting both accuracy and computational efficiency when learning mappings between function spaces.

1 INTRODUCTION

Machine learning for scientific computing has evolved from approximating individual PDE solutions to learning discretization-independent operators acting on function spaces. Physics-Informed Neural Networks (PINNs) (Raissi et al., 2019) incorporate governing equations into the training loss via automatic differentiation, but they typically require retraining for new parameters or boundary conditions, limiting their use in parametric settings.

This motivated neural operator learning. Deep Operator Networks (DeepONets) (Lu et al., 2021) introduced a branch–trunk architecture with universal approximation guarantees, while the Fourier Neural Operator (FNO) (Li et al., 2021) learned integral operators in Fourier space using FFTs and achieved strong performance on benchmark PDEs such as Navier–Stokes, Darcy flow, and Burgers' equations. A unified theoretical framework was later established by Kovachki et al. (2023), formalizing neural operators as mappings between Banach spaces and introducing variants such as graph-based and low-rank operators.

Physics-aware extensions further improved generalization. Physics-Informed Neural Operators (PINOs) (Li et al., 2023) regularize operator learning with PDE residuals, while hybrid multigrid neural operators (Hu et al., 2025) combine classical solvers with neural corrections for stability.

A key limitation of neural networks is spectral bias, which favors low-frequency components and leads to oversmoothing and poor resolution of sharp gradients and shocks (John Xu et al., 2020; Rahaman et al., 2019). This bias carries over to neural operators. Multi-scale FNOs (You et al., 2024) and DeepONets (Liu & Cai, 2021) mitigate this effect through hierarchical representations, while Wavelet Neural Operators improve spatial localization but remain largely isotropic.

In a different context, shearlets (Guo et al., 2005) provide a multiscale, directional representation with (near-)optimal sparsity for anisotropic features such as edges and curvilinear singularities. Unlike prior uses of shearlets as fixed feature extractors (Wiatowski et al., 2018), we introduce the Physics-Informed Shearlet Neural Operator (PI-ShearletNO), which embeds a differentiable shearlet transform within the operator-learning framework. This design combines global operator learning with directional adaptivity, making it well suited for PDEs whose solutions develop shocks, layers, and sharp gradients.

2 BACKGROUND

2.1 NEURAL OPERATORS

Many parametric PDE problems can be written as learning a *solution operator* $\mathcal{G}$ mapping an input function (e.g., coefficients, forcing terms, initial/boundary data, or parameters encoded on a grid) to an output solution field:

$$\mathcal{G} : \mathcal{A} \to \mathcal{U}, \qquad a(\cdot) \mapsto u(\cdot).$$

Neural operators aim to approximate $\mathcal{G}$ directly, rather than learning a map between finite-dimensional vectors tied to one mesh. This yields *resolution invariance*: after training on one discretization, the same model can be evaluated on different meshes by sampling the input/output functions accordingly (Kovachki et al., 2023).

A common architecture is an iterative feature evolution

$$\nu_{t+1}(x) = \sigma\big((\mathcal{K}_\theta \nu_t)(x) + (W_\theta \nu_t)(x)\big),$$

where ν_t denotes a latent feature field defined on Ω, W_θ is a pointwise (local) linear map, and $\mathcal{K}_\theta$ is a nonlocal operator capturing long-range interactions.

In many neural operators, $\mathcal{K}_\theta$ is parameterized as an integral operator

$$(\mathcal{K}_\theta u)(x) = \int_\Omega \kappa_\theta(x, y)\, u(y)\, dy,$$

with the kernel κ_θ learned from data. In practice, the integral is implemented by a discrete summation on a grid, and different parameterizations (Fourier, wavelet, low-rank, graph-based) correspond to different structured approximations of κ_θ (Kovachki et al., 2023).

To interface with data, neural operators typically include *lifting* and *projection* maps: the input is lifted to a higher-dimensional latent channel space, several operator layers are applied, and the result is projected back to the physical output dimension.

2.1.1 FOURIER NEURAL OPERATOR (FNO)

The Fourier Neural Operator (FNO) parameterizes the kernel in Fourier space (Li et al., 2021). Given an input function u, it:

1. Transforms u to Fourier coefficients $\widehat{u}(k)$.
2. Multiplies low-frequency modes by a learned filter $R(k)$:
$$\widehat{v}(k) = R(k)\widehat{u}(k), \quad |k| \le m.$$
3. Inverse transforms to physical space, producing a nonlocal update $v(x)$.

An FNO layer combines this spectral convolution with a local linear transformation W and a nonlinearity σ:

$$u_{j+1}(x) = \sigma\Big(\mathcal{F}^{-1}[R_j \widehat{u}_j](x) + W_j u_j(x)\Big).$$

Stacked layers form a deep operator network.

FNO's spectral approach provides a global receptive field, ensures mesh invariance, and achieves strong performance in PDE-based applications.

2.2 PHYSICS-INFORMED LEARNING FOR PARAMETRIC PDES

Physics-informed learning incorporates the governing PDE as an inductive bias, typically by penalizing the residual of the differential operator evaluated at collocation points. In Physics-Informed Neural Networks (PINNs) (Raissi et al., 2019), this residual is computed via automatic differentiation, enabling training with limited labeled data but often leading to optimization challenges and limited transfer across parameter regimes.

For operator learning, the same principle can be applied by regularizing the learned mapping with physics constraints. Physics-Informed Neural Operators (PINOs) (Li et al., 2023) augment a data-driven operator loss with a PDE residual term, encouraging the network to generalize across discretizations and parameter values while remaining consistent with the underlying dynamics. Related physics-aware hybrids combine neural components with classical numerical structure (e.g., multigrid-inspired designs) to improve stability and sample efficiency (Hu et al., 2025).

2.3 Spectral bias and the need for multiscale representations

Neural networks often exhibit *spectral bias*, learning low-frequency components more readily than high-frequency ones (John Xu et al., 2020; Rahaman et al., 2019). In PDE settings, this can manifest as oversmoothing and loss of sharp gradients, thin layers, and shocks. For neural operators, the same effect can appear as insufficient resolution of fine-scale structures unless the architecture explicitly supports multiscale processing.

Several approaches address this limitation by enriching the representation across scales. Multi-scale operator architectures introduce hierarchical pathways or explicit scale separation to improve recovery of high-frequency content (You et al., 2024; Liu & Cai, 2021). Wavelet-based neural operators further improve spatial localization compared to pure Fourier representations, but their standard constructions remain largely isotropic and may be suboptimal for strongly directional features.

2.4 Shearlets and anisotropic directional structure

Shearlets provide a multiscale system with *directional selectivity* and *parabolic scaling*, yielding near-optimal sparse approximations of functions with curvilinear singularities (Guo et al., 2005). Compared to isotropic multiresolution transforms, shearlets adapt angular resolution with scale, producing wedge-shaped frequency tilings that are well matched to transport-dominated PDE phenomena (e.g., wavefronts and shock curves).

Beyond their role as fixed feature extractors (Wiatowski et al., 2018), shearlet representations can be embedded as differentiable modules, enabling end-to-end learning that explicitly mixes information across orientations and scales. This motivates shearlet-based neural operator designs that target anisotropic PDE solution manifolds.

3 Numerical methodology

Physics-informed Fourier Neural Operators (PI-FNOs) (Li et al., 2023) rely on isotropic global convolutions, which can be inefficient for anisotropic PDE features such as shocks, edges, and multi-oriented structures. We introduce the **Physics-Informed Shearlet Neural Operator (PI-ShearletNO)**, which replaces Fourier filters with a multiscale, directional shearlet system, combining spectral learning with geometric adaptivity (Guo et al., 2005). In this section, we describe the proposed methodology and compare it with its counterpart, PI-FNO. The next subsections provide the key definitions.

3.1 Shearlet Frequency Windows

A discrete shearlet (Guo et al., 2005) system is constructed on the FFT grid using smooth radial and angular windows:

$$\bar{W}_j(k_x, k_y) = \phi\left(a_j \frac{\sqrt{k_x^2 + k_y^2}}{R_{\max}}\right), \tag{1}$$

$$A_{j,k}(k_x, k_y) = \psi(\theta(k_x, k_y) - \theta_{j,k}), \tag{2}$$

where ϕ is a radial bump, ψ is an angular window, $a_j = 2^{-(j+1)}$ sets the scale, and $\theta_{j,k}$ defines shear orientation. The full shearlet window is:

$$\Xi_{j,k}(k_x, k_y) = \bar{W}_j(k_x, k_y)\, A_{j,k}(k_x, k_y), \tag{3}$$

normalized to satisfy an approximate tight-frame condition:

$$\sum_{j,k} |\Xi_{j,k}(k_x, k_y)|^2 \approx 1. \tag{4}$$

This yields a directional tiling of the frequency plane: isotropic at low frequencies and decomposed into sheared sectors at higher frequencies.

3.2 Shearlet Spectral Convolution

Let $\{\Xi_m\}$ be the set of shearlet windows. A ShearletNO layer performs windowed spectral convolution:

$$\widehat{u}_m(k_x, k_y) = \widehat{u}(k_x, k_y)\, \Xi_m(k_x, k_y), \tag{5}$$

followed by learned complex mixing across channels:

$$\widehat{v}_o(k_x, k_y) = \sum_{m,c} \bar{W}_{m,c,o}\, \widehat{u}_m^{(c)}(k_x, k_y). \tag{6}$$

A learnable scale gate $g_j = \sigma(\gamma_j)$ modulates each scale's importance. The final spatial output is:

$$v(x, y) = \mathcal{F}^{-1}\left[\sum_m g_{\text{scale}(m)}\, \widehat{v}_{o,m}(k_x, k_y)\right]. \tag{7}$$

Stacked layers form the full shearletNO:

$$u_{j+1}(x) = \sigma(v(x) + W_j u_j(x))\,. \tag{8}$$

In this context, the advantages of Shearlet Neural Operators follow directly from the *parabolic scaling* and *directional tiling* induced by the shearlet windows $\{\Xi_{j,k}\}$:

1. **Directional sensitivity (shear parameter k)**: Each window $\Xi_{j,k}$ concentrates energy in a narrow wedge of the frequency plane, so coefficients respond strongly to features aligned with a specific orientation. This yields sparse, orientation-aware representations of edges, layers, and shock fronts.
2. **Multiscale anisotropy (parabolic scaling in j)**: The scale index j produces a hierarchy of frequency bands, while the associated anisotropic (parabolic) scaling refines angular resolution at higher frequencies. As a result, coarse scales capture global smooth components, whereas fine scales isolate localized high-frequency singularities.
3. **Stable global processing (tight-frame behavior)**: The approximate partition of unity $\sum_{j,k} |\Xi_{j,k}|^2 \approx 1$ implies that the shearlet decomposition is well conditioned, so the representation preserves energy up to controlled constants. This supports stable learning and reconstruction when mixing coefficients across bands.
4. **Adaptive anisotropy control (learnable scale gates)**: The learnable gates g_j reweight the contribution of each scale in a data-dependent manner, allowing the operator to emphasize transport-dominated fine scales when needed while suppressing irrelevant high-frequency content in smoother regimes.

Unlike FNO's isotropic low-pass filtering, ShearletNO performs learned filtering within directional shearlet bands, enabling high-fidelity modeling of transport-dominated and anisotropic phenomena.

3.3 Physics-informed formulation

The Physics-Informed Shearlet Neural Operator (PI-ShearletNO) addresses the problem of learning *solution operators* for parametric PDEs, where the mapping from physical parameters to the corresponding solution is nonlinear and may exhibit multiscale, anisotropic features.

To make the setting precise, let $\Omega \subset \mathbb{R}^d$ be a bounded domain and let $\lambda \in \Lambda$ denote a (possibly vector-valued) parameter. For each λ, consider the boundary value problem

$$\mathcal{N}_\lambda[u](\boldsymbol{x}) = f_\lambda(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega, \tag{9}$$

with boundary and/or initial conditions $\mathcal{B}_\lambda[u](\boldsymbol{x}) = g_\lambda(\boldsymbol{x})$ for $\boldsymbol{x} \in \partial\Omega$. Assuming well-posedness, this defines a solution operator

$$\mathcal{G} : \Lambda \to \mathcal{U}, \qquad \lambda \mapsto u(\cdot;\lambda), \tag{10}$$

where $\mathcal{U}$ is an appropriate function space (e.g., $L^2(\Omega)$ or $H^1(\Omega)$, depending on the PDE).

Given a parametric dataset $\{(\lambda_i, u_{\mathrm{ref}}(\cdot;\lambda_i))\}_{i=1}^N$, we learn an approximation $\mathcal{G}_\theta$ by minimizing a weighted sum of a data misfit and a physics residual. Using squared L^2 norms, we define

$$\begin{aligned}\mathcal{L}_{\text{PI-ShearletNO}}(\theta) = \mathbb{E}_{\lambda\sim p(\lambda)}\Big[\|\mathcal{G}_\theta(\lambda) - u_{\mathrm{ref}}(\cdot;\lambda)\|^2_{L^2(\Omega)}\Big] \\ + \beta\,\mathbb{E}_{\lambda\sim p(\lambda)}\Big[\|\mathcal{N}_\lambda[\mathcal{G}_\theta(\lambda)] - f_\lambda\|^2_{L^2(\Omega)} + \|\mathcal{B}_\lambda[\mathcal{G}_\theta(\lambda)] - g_\lambda\|^2_{L^2(\partial\Omega)}\Big],\end{aligned} \tag{11}$$

where $\beta > 0$ balances the contributions (and the expectations are approximated in practice by empirical averages over the training set).

4 Numerical Benchmarks

We evaluate PI-shearletNO against PI-FNO on four PDE benchmarks designed to stress-test directional modeling. All problems are solved on a 256×256 grid with periodic boundaries.

4.1 Test Cases

4.1.1 Multi-Angle Shock Waves

We consider a 2D viscous Burgers' equation with two shock fronts propagating at angles $\theta_1 = \arctan(0.8)$ and $\theta_2 = \arctan(-1.2)$, motivated by supersonic and combustion-like flow patterns.

$$\frac{\partial u}{\partial t} + u\nabla u = \nu\nabla^2 u, \quad \nu = 8 \times 10^{-4}. \tag{12}$$

$$u_0(x,y) = \tanh\left[10(x + 0.8y - 3)\right] + \tanh\left[12(x - 1.2y - 1)\right]. \tag{13}$$

This benchmark stresses nonlinear interaction of directional discontinuities and the ability to preserve sharp fronts without spurious oscillations.

4.1.2 Multi-Orientation Texture

This benchmark is a superposition of multiple wave components with different orientations, wavenumbers, and phases undergoing anisotropic diffusion, motivated by texture-like patterns in biological tissues.

$$\frac{\partial u}{\partial t} = D_x\frac{\partial^2 u}{\partial x^2} + D_y\frac{\partial^2 u}{\partial y^2}, \quad D_x = 0.4, \quad D_y = 12. \tag{14}$$

$$u_0(x,y) = \sum_{k=1}^{4}\left[\sin\left(10kx + 6ky\right) + \sin\left(12kx - 8ky\right)\right]. \tag{15}$$

The goal is to test separation of multiple frequency-direction components and robustness to interference among orientations.

4.1.3 Spiral Shock Pattern

This is a geometrically complex test with an Archimedean spiral shock pattern, where the local orientation varies continuously with the polar angle.

$$\frac{\partial u}{\partial t} + u\nabla u = \nu\nabla^2 u, \quad \nu = 5 \times 10^{-4}. \tag{16}$$

$$u_0(r,\theta) = \tanh\left[10\left(r - 0.7 - 0.15\theta\right)\right]. \tag{17}$$

It tests whether models can track smoothly varying directional information while maintaining sharp transition regions.

4.1.4 POLYGONAL SHOCK PATTERN

This toy model features a hexagonal shock pattern with sixfold symmetry, challenging models with angular periodicity and discrete directional symmetries.

$$\frac{\partial u}{\partial t} + c_x \frac{\partial u}{\partial x} + c_y \frac{\partial u}{\partial y} = 0, \quad c_x = 2, \quad c_y = 1.3. \tag{18}$$

$$u_0(r, \theta) = \tanh\left[12\left(\cos(6\theta) - 0.3\right)\right]. \tag{19}$$

4.2 TRAINING PROTOCOL

Both models were trained under identical settings: AdamW optimizer (learning rate 10^{-4}, weight decay 10^{-4}), batch size 32, and early stopping with a patience equal to 50% of the total number of epochs. Architecturally, the Shearlet-based neural operator (ShearletNO) employs 4 scales with 16 directions per scale, while the Fourier neural operator (FNO) uses 4×4 Fourier modes. The FNO has approximately 3x the number of parameters of the SNO. All results use a fine grid of 256x256. The networks were initialized with the xavier initialization algorithm.

The ground truths were simulated with runge-kutta of order 3 and simulated for 250*5 time steps with a step of 0.001s, where the 5 represents the temporal stride, i.e., in order to avoid selecting several near identical frames we simulate the equations with more time steps and take 1 out of every five to train. This yields a dataset of 250 samples of 256^2. The training, validation and testing sets were separated temporally to avoid data leakage into tests, so the first 70% of frames were used to train, the next 15% for validation and the remaining for testing. Note that there are a considerable amount of steps between train and test. To avoid any possibility of leakage a gap was introduced between train-val and val-test datasets. All results shown below were acquired on the test sets.

Each model begins with a projection layer followed by a lifting layer. The core neural operator consists of 4 layers, which we found sufficient to capture the essential dynamics of the underlying equations.

5 RESULTS

In the following subsections, we present qualitative comparisons for the selected benchmark problems. Prior to that, Table 1 summarizes the quantitative results across all evaluated cases displaying MSE, MAE and SSIM values.

Dataset	$\|e\|_2^S$	$\|e\|_2^F$	S/F	MSE_S	SSIM_S	SSIM_F
multi-angle	8.1×10^{-6}	8.4×10^{-5}	0.10	1.1×10^{-4}	0.999	0.958
multi-orient	7.2×10^{-6}	4.2×10^{-5}	0.17	3.1×10^{-6}	0.998	0.954
spiral	8.3×10^{-6}	1.7×10^{-5}	0.48	1.1×10^{-4}	0.998	0.991
polygonal	8.4×10^{-6}	2.2×10^{-5}	0.37	6.0×10^{-5}	0.998	0.986

Table 1: Comparison between SNO (S) and FNO (F). Values < 1 in the S/F column indicate lower relative L^2 error for SNO. SSIM is reported with three decimals to highlight structural differences.

MULTI-ANGLE SHOCK WAVES

The multi-angle shock field contains sharp transitions and interacting anisotropic fronts at different orientations. While both models reproduce the global structure, the error maps reveal a substantial difference in reconstruction quality. The SNO exhibits significantly lower absolute error across most of the domain, with errors primarily localized along high-curvature shock regions. In contrast, the FNO displays larger, spatially diffuse residuals and stronger artifacts near discontinuities and curved fronts. The error difference map (SNO - FNO) further highlights regions where SNO achieves markedly improved accuracy, particularly near multi-directional shock interactions. This result was expected, as FNO-based models typically struggle with discontinuities.

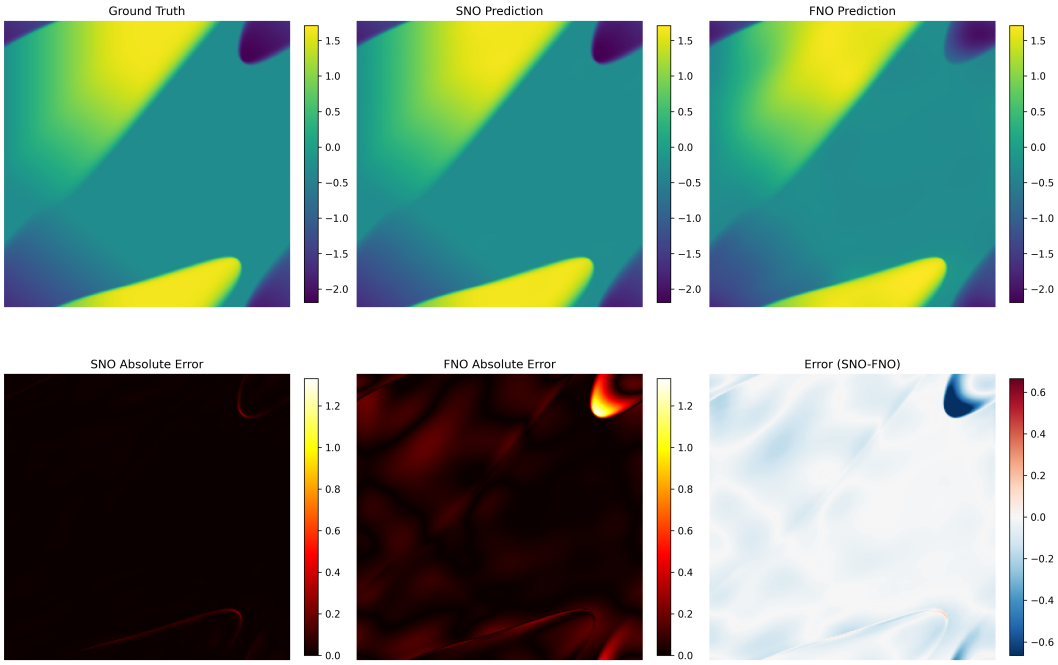

Figure 1: Comparison between SNO and FNO on the multi-angle schock benchmark. Top row: Ground truth, SNO prediction, and FNO prediction. Bottom row: Absolute error for SNO, absolute error for FNO, and pointwise error difference (—SNO - Target— - —FNO - Target—).

Multi-Orientation Texture

The multi-orientation texture field contains superimposed periodic components at multiple directions and comparable spatial frequencies. Both SNO and FNO accurately reconstruct the global texture pattern, producing visually indistinguishable predictions from the ground truth. The absolute error maps indicate low-magnitude residuals primarily structured along the dominant periodic directions. The error difference map is nearly uniform showing a dominance for the SNO over the entire field. Despite PI-FNO having three times the number of parameters as PI-SNO, the anisotropic effect remains the predominant source of error for PI-FNO due to the global nature of the Fourier transform.

Spiral Shock Pattern

The spiral shock field contains curved, high-gradient fronts with strong anisotropy and non-uniform curvature. Both operators reproduce the overall shape of the spiral structure. However, the absolute error maps reveal distinct differences: SNO errors are more localized and closely aligned with the shock interface, whereas FNO exhibits broader residual bands and stronger smoothing artifacts around the curved front.The error difference map (blue = SNO lower error, red = FNO lower error) indicates that SNO achieves lower error along substantial portions of the spiral boundary and in high-curvature regions, while FNO performs comparably or slightly better in limited interior areas.

Polygonal Shock Pattern

The polygonal shock field consists of piecewise-linear fronts meeting at sharp vertices, generating strong anisotropy and non-smooth transitions. Both operators reproduce the global geometry of the shock pattern. However, the absolute error maps show clear differences: SNO errors remain more localized along shock edges and vertices, whereas FNO exhibits broader residual regions and more diffuse oscillatory artifacts in the surrounding domain. The error difference map highlights coherent

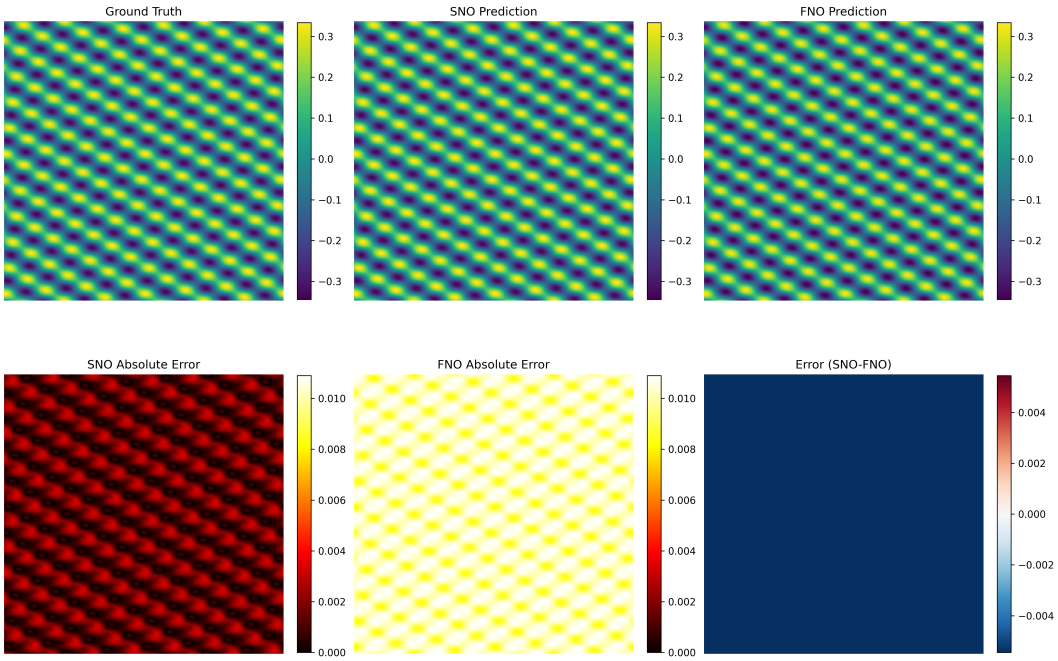

Figure 2: Comparison between SNO and FNO on the multi-orientation texture benchmark. Top row: Ground truth, SNO prediction, and FNO prediction. Bottom row: Absolute error for SNO, absolute error for FNO, and pointwise error difference (—SNO - Target— - —FNO - Target—).

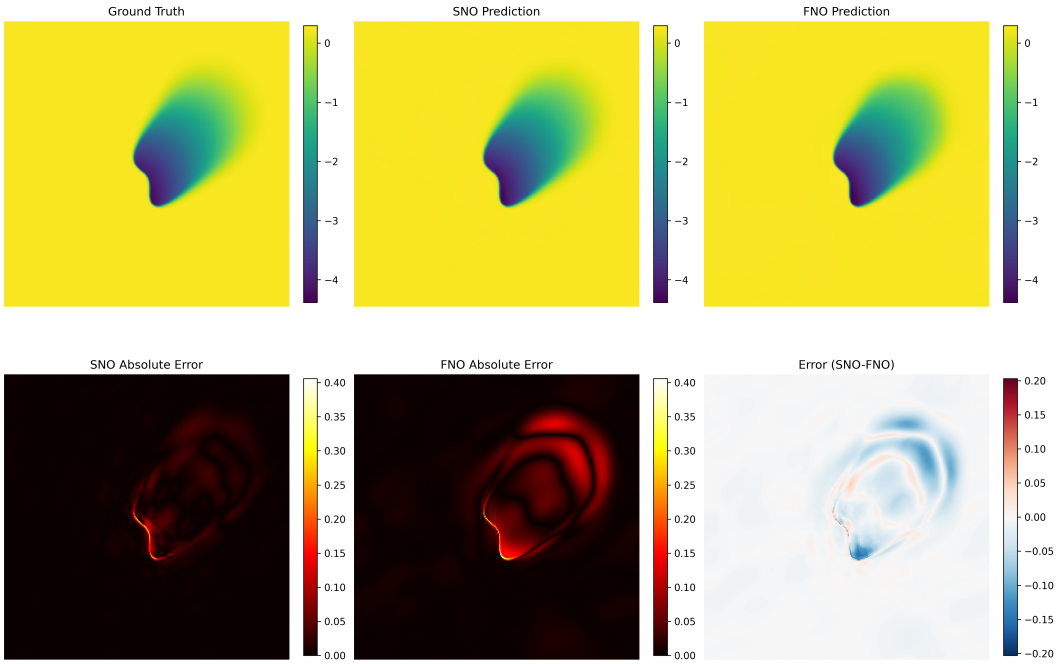

Figure 3: Comparison between SNO and FNO on the spiral shock benchmark. Top row: Ground truth, SNO prediction, and FNO prediction. Bottom row: Absolute error for SNO, absolute error for FNO, and pointwise error difference (—SNO - Target— - —FNO - Target—).

regions (predominantly blue) where SNO achieves lower error, particularly near sharp corners and along intersecting shock interfaces.

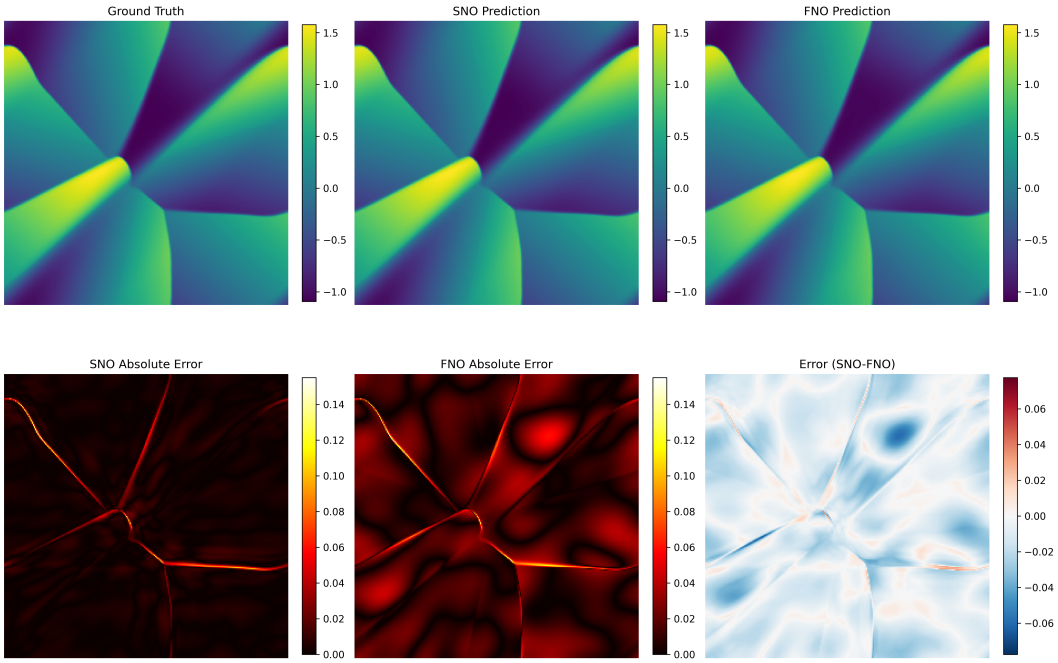

Figure 4: Comparison between SNO and FNO on the polygonal shock benchmark. Top row: Ground truth, SNO prediction, and FNO prediction. Bottom row: Absolute error for SNO, absolute error for FNO, and pointwise error difference (—SNO - Target— - —FNO - Target—).

6 Conclusion

In conclusion, the Physics-Informed Shearlet Neural Operator (PI-ShearletNO) advances scientific machine learning by combining neural operator approximation with the multiscale, directional structure of shearlet transforms. By explicitly enforcing the governing PDE and boundary/initial conditions through a residual-based regularization term, the learned operator produces outputs that are both data-consistent and physically admissible. In our benchmarks, this integration improves generalization and accuracy relative to PI-FNO while retaining competitive computational cost. These results suggest PI-ShearletNO as a promising direction for learning solution operators of parametric PDEs with anisotropic or shock-like features.

Acknowledgments

This work was supported by the Serrapilheira Institute (grant number Serra – R-2111-39718)

7 References

References

Kanghui Guo, Gitta Kutyniok, and Demetrio Labate. Sparse multidimensional representations using anisotropic dilation and shear operators. 05 2005.

Yifan Hu, Weimin Zhang, Fukang Yin, and Jianping Wu. Hmgno: Hybrid multigrid neural operator with low-order numerical solver for partial differential equations. *Neural Networks*, 190:107649,

2025. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2025.107649. URL https://www.sciencedirect.com/science/article/pii/S0893608025005295.

Zhi-Qin John Xu, Yaoyu Zhang, Tao Luo, Yanyang Xiao, and Zheng Ma. Frequency principle: Fourier analysis sheds light on deep neural networks. *Communications in Computational Physics*, 28(5):1746–1767, June 2020. ISSN 1991-7120. doi: 10.4208/cicp.oa-2020-0085. URL http://dx.doi.org/10.4208/cicp.OA-2020-0085.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: learning maps between function spaces with applications to pdes. *J. Mach. Learn. Res.*, 24(1), January 2023. ISSN 1532-4435.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations, 2021. URL https://arxiv.org/abs/2010.08895.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations, 2023. URL https://arxiv.org/abs/2111.03794.

Lizuo Liu and Wei Cai. Multiscale deeponet for nonlinear operators in oscillatory function spaces for building seismic wave responses, 2021. URL https://arxiv.org/abs/2111.04860.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, March 2021. ISSN 2522-5839. doi: 10.1038/s42256-021-00302-5. URL http://dx.doi.org/10.1038/s42256-021-00302-5.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks, 2019. URL https://arxiv.org/abs/1806.08734.

M. Raissi, P. Perdikaris, and G.E. Karniadakis. 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. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.10.045. URL https://www.sciencedirect.com/science/article/pii/S0021999118307125.

Thomas Wiatowski, Philipp Grohs, and Helmut Bölcskei. Energy propagation in deep convolutional neural networks, 2018. URL https://arxiv.org/abs/1704.03636.

Zhilin You, Zhenli Xu, and Wei Cai. Mscalefno: Multi-scale fourier neural operator learning for oscillatory function spaces, 2024. URL https://arxiv.org/abs/2412.20183.