

PeSAnet: Physics-encoded Spectral Attention Network for Simulating PDE-Governed Complex Systems

Han Wan^{1,†}, Rui Zhang^{1,†}, Qi Wang¹, Yang Liu², Hao Sun^{1,*}

¹Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China

²School of Engineering Science, University of Chinese Academy of Sciences, Beijing, China

{wanhan2001, rayzhang, qi_wang, haosun}@ruc.edu.cn, liuyang22@ucas.ac.cn

Abstract

Accurately modeling and forecasting complex systems governed by partial differential equations (PDEs) is crucial in various scientific and engineering domains. However, traditional numerical methods struggle in real-world scenarios due to incomplete or unknown physical laws. Meanwhile, machine learning approaches often fail to generalize effectively when faced with scarce observational data and the challenge of capturing local and global features. To this end, we propose the Physics-encoded Spectral Attention Network (PeSAnet), which integrates local and global information to forecast complex systems with limited data and incomplete physical priors. The model consists of two key components: a physics-encoded block that uses hard constraints to approximate local differential operators from limited data, and a spectral-enhanced block that captures long-range global dependencies in the frequency domain. Specifically, we introduce a novel spectral attention mechanism to model inter-spectrum relationships and learn long-range spatial features. Experimental results demonstrate that PeSAnet outperforms existing methods across all metrics, particularly in long-term forecasting accuracy, providing a promising solution for simulating complex systems with limited data and incomplete physics.

1 Introduction

Accurately modeling and simulating complex systems is crucial in various scientific and engineering fields, such as fluid dynamics, meteorology, and biology [Anderson and Wendt, 1995; Holton and Hakim, 2013; Ghergu and Radulescu, 2011]. These systems are typically described by partial differential equations (PDEs) that govern their spatiotemporal evolution. However, forecasting these systems efficiently and accurately in real-world scenarios presents significant challenges. Firstly, the physical laws that govern these systems can be incomplete or unknown, making it difficult to formulate the correct PDEs. Secondly, observational data

are often scarce, limiting the ability to learn the underlying physics. Additionally, the characteristics of complex systems are multi-faceted, with local variants and global dependencies that need to be captured for reliable long-term predictions.

Over the years, various methods have been developed to address these challenges. Traditional numerical methods [Moukalled *et al.*, 2016; Karniadakis and Sherwin, 2005; Zienkiewicz *et al.*, 2005] enjoy good convergence and stability when PDEs are well-defined, but become difficult to apply when the system’s governing laws are incomplete. To overcome this, data-driven methods such as DeepONet [Lu *et al.*, 2021] and Fourier Neural Operators (FNOs) [Li *et al.*, 2021] use neural networks as surrogate models, aiming to learn the underlying physical evolution from data. These methods do not require explicit knowledge of the physical laws and achieve significant speedup compared to traditional numerical solvers. However, they often fail to offer meaningful predictions when the training data is limited, which is a common issue in real-world scenarios.

To address the limitations of purely data-driven approaches, researchers have developed a series of physics-aware methods that incorporate prior physical knowledge into neural networks. These methods can be classified into physics-informed and physics-encoded, depending on how they integrate the physical priors. Physics-informed methods typically use soft constraints, such as PDE residual-based loss functions, to enforce the model that aligns with the governing PDEs. Representative methods include Physics-Informed Neural Networks (PINNs) [Raissi *et al.*, 2019] and Physics-Informed Neural Operators (PINOs) [Li *et al.*, 2024b]. However, due to their soft-constrained nature, these methods encounter challenges in hyperparameter tuning and training stability, and require sufficient prior knowledge of the governing PDEs [Krishnapriyan *et al.*, 2021]. To this end, physics-encoded methods use a hard-constrained strategy to incorporate physical prior into the model architecture. For instance, the Physics-encoded Recurrent Convolutional Neural Networks (PeRCNN) [Rao *et al.*, 2023; Rao *et al.*, 2022] use specially designed convolutional kernels to approximate finite difference stencils. These approaches can automatically learn local differential operators that match the observed data, even when the governing PDE is unknown. However, due to their limited receptive field, they can only capture local variations and struggle to learn long-range spa-

*Corresponding author (Email: haosun@ruc.edu.cn).

tial dependencies in complex systems.

To overcome these limitations, we propose the Physics-encoded Spectral Attention Network (PeSANet). The model consists of the physics-encoded and spectral-enhanced blocks, which learn the local and global evolution of the complex system, respectively. The physics-encoded block employs hard constraints to approximate local differential operators. This hard-constrained approach does not require sufficient physical prior knowledge and can achieve generalization from the limited observed data. On the other hand, the spectral-enhanced block handles spectral information in the frequency domain to capture global features. In particular, we proposed the spectral attention mechanism to exploit the inter-spectrum relationship, further enhancing the ability to learn long-range spatial dependencies. In summary, we make the following contributions:

- We propose the Physics-encoded Spectral Attention Network (PeSANet) to integrate both local and global learning mechanisms, which can achieve long-term forecasting for complex systems with limited data and incomplete physical priors.
- We propose a spectral attention mechanism in the frequency domain to capture long-range spatial features. By using all frequency components and capturing interactions across spectral layers, this approach can enhance the model’s ability to describe global system features.
- PeSANet achieves superior long-term prediction performance across multiple complex systems. Furthermore, it outperforms other baseline models in all metrics, showing robust generalization with limited training data and transfer learning ability across varying conditions.

2 Related Work

The numerical solution of PDEs has been a central focus of scientific computing for decades. Traditional numerical methods, such as finite difference and finite element methods, have achieved remarkable success in many applications. However, these methods can be computationally expensive and require significant domain knowledge. In recent years, deep learning has emerged as a powerful tool for solving PDEs, offering the potential for more efficient and flexible solutions. Existing AI-driven methods can be categorized into data-driven, physics-aware, and hybrid models. The key distinction lies in the incorporation of prior physical knowledge.

2.1 Data-driven learning methods

When sufficient data is available, many works have investigated neural networks for learning the dynamics governed by time-dependent PDEs. These methods generally fall into two categories. The first category leverages well-established architectures from the computer vision domain, such as classical convolutional neural networks (CNNs) [Stachenfeld *et al.*, 2021; Bar-Sinai *et al.*, 2019], U-Net [Gupta and Brandstetter, 2023], ResNet [Lu *et al.*, 2018], and ConvLSTM [Shi *et al.*, 2015], to capture intricate spatiotemporal patterns. These architectures excel at extracting local features and are naturally suited for data represented on regular grids or meshes.

The second category involves neural operators, which focus on learning mappings between function spaces. This class includes models like DeepONet [Lu *et al.*, 2021] and its variations [Seidman *et al.*, 2022; Venturi and Casey, 2023; Lee *et al.*, 2023], as well as the FNO [Li *et al.*, 2021] and its derivatives [Tran *et al.*, 2023; Rahman *et al.*, 2023; Wen *et al.*, 2022; Zhang *et al.*, 2024]. Neural operators offer the advantage of mesh independence, enabling generalization across different discretizations of the same class of PDEs. Additionally, advances in other data-driven neural PDE solvers, such as graph neural networks (GNNs) [Sanchez-Gonzalez *et al.*, 2020; Pfaff *et al.*, 2021] and transformer-based architectures [Wu *et al.*, 2024; Hang *et al.*, 2024; Janny *et al.*, 2023; Li *et al.*, 2024a], provide powerful tools to extend neural operators to various complex geometries.

Despite the progress achieved by these data-driven methods, they still face several limitations. One significant drawback is their heavy reliance on large and high-quality datasets, which are often time-consuming to generate or infeasible to many scientific and engineering domains. Moreover, these models are prone to overfitting and fail to generalize to unseen data, which is a key challenge in complex physical systems due to their inherent complexity and diversity.

2.2 Physics-aware learning methods

When data is limited, integrating physical knowledge into neural networks becomes important. This integration typically follows two strategies: physics-informed (soft constraints) and physics-encoded (hard constraints). Physics-informed methods incorporate physical laws, such as the PDE formulations, into the loss function. Representative works include PINN [Raissi *et al.*, 2019], PINO [Li *et al.*, 2024b], PhyCRNet [Ren *et al.*, 2022], MCNP [Zhang *et al.*, 2025], and PI-DeepONet [Wang *et al.*, 2021]. These methods can be data-efficient but face challenges such as hyperparameter tuning, training instability, and requiring substantial prior knowledge of the governing equations [Krishnapriyan *et al.*, 2021].

On the other hand, physics-encoded approaches directly embed physical constraints into the network architecture, such as PDE operators, conservation laws, and boundary conditions [Long *et al.*, 2019; Rao *et al.*, 2023; Mi and Sun, 2024; Zeng *et al.*, 2025]. These methods can be highly effective when working with limited data and incomplete physics priors. However, hard-constrained methods like PerCNN, which use specially designed convolutional kernels to approximate local differential operators, often struggle to capture global information, limiting their performance on problems that require long-range spatial dependencies.

2.3 Hybrid learning methods

Hybrid models provide an effective acceleration strategy for solving PDEs when the governing equations are fully known. These models couple neural networks with traditional numerical methods (e.g., FDM, FVM) to combine their advantages. Neural networks approximate computationally expensive components of the PDE solution, such as complex nonlinearities or fine-scale features, while established numerical methods handle the more tractable parts. This coupling enhances computational efficiency without sacrificing accuracy.

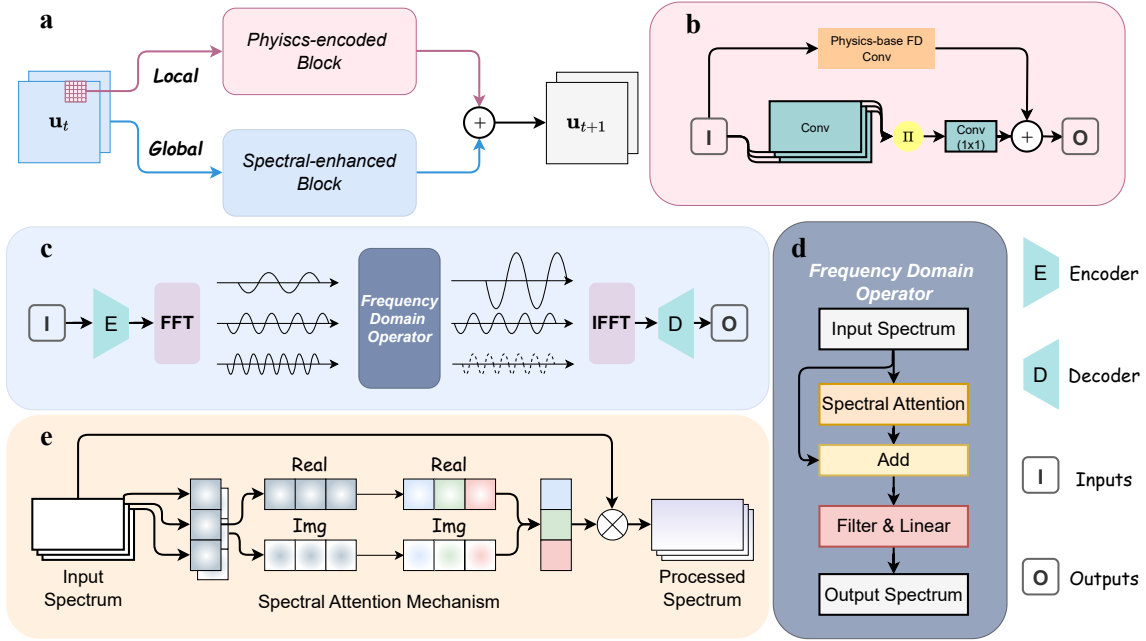


Figure 1: **The architecture of PeSANet.** (a) Core blocks: physics-encoded and spectral-enhanced blocks to learn local and global features, respectively. (b) Physics-based convolution for known PDE terms (such as $\nabla^2 \mathbf{u}$) and II-block for unknown terms. (c) Spectral-enhanced block: including the FFT, encoder, frequency domain operator, IFFT, and decoder. (d) Frequency domain operator. (e) Spectral attention.

Examples include multi-scale hybrid models [Vlachas *et al.*, 2022] and other couplings with FVM (learned interpolation model [Kochkov *et al.*, 2021] and temporal stencil modeling [Sun *et al.*, 2023]) and FDM (P²C²Net [Wang *et al.*, 2024]). However, these models require complete knowledge of the PDEs, restricting their application to systems with unknown or partially unknown physics.

3 Methodology

3.1 Problem Formulation

Let us consider spatiotemporal complex systems governed by the following time-dependent PDEs:

$$\frac{\partial \mathbf{u}}{\partial t} = \mathcal{F}(t, \mathbf{x}, \mathbf{u}, \nabla \mathbf{u}, \mathbf{u} \cdot \nabla \mathbf{u}, \nabla^2 \mathbf{u}, \dots; \boldsymbol{\mu}), \quad (1)$$

where $\mathcal{F}(\cdot)$ represents an unknown operator that involves spatial derivatives of various orders, such as the gradient $\nabla \mathbf{u}$ and the Laplacian $\nabla^2 \mathbf{u}$. The parameter $\boldsymbol{\mu}$ governs the physical properties of the system. Additionally, the initial conditions (ICs) and boundary conditions (BCs) are denoted as $\mathcal{I}[\mathbf{u}](\mathbf{x}, t = 0) = 0$ and $\mathcal{B}[\mathbf{u}](\mathbf{x}, t) = 0$, respectively.

In this paper, we aim to use the neural network as a surrogate model to simulate the complex systems based on incomplete physical prior of \mathcal{F} and limited training data (e.g., 2-5 trajectories). The trained model is expected to learn the underlying dynamics of the physical system, which can accurately predict long-term system dynamics and generalize across different ICs. Traditional numerical and hybrid methods are not applicable to this scenario because they require sufficient prior knowledge of the physical system. On

the other hand, data-driven methods demand large numbers of training data to achieve reasonable generalization performance. To this end, we propose the PeSANet.

3.2 PeSANet Architecture

In this section, we introduce the overall architecture of the proposed PeSANet (Figure 1(a)). The model has two key components: the physics-encoded block and the spectral-enhanced block, which are designed to learn the local and global features of the complex system, respectively. The physics-encoded block uses multiplicative convolution to approximate the finite difference (FD) stencil, which extracts the local spatial information in a hard-constrained manner. The spectral-enhanced block conducts spectral attention operations in the frequency domain, which can further enhance the performance via capturing global features, such as long-range spatial dependencies. This synergistic combination allows the model to learn local details and global structures of the complex system, resulting in a thorough understanding of the underlying physical phenomena.

3.3 Physics-encoded Block

The physical-encoded block (Figure 1(b)) aims to learn the underlying local differential operators that govern the complex system via a hard-constrained approach. Similar to PeRCNN [Rao *et al.*, 2023], we adopt the II-block and physics-based convolutional (PyConv) layer to handle known and unknown operators, respectively.

Considering the forward Euler scheme, the updated rule of the system in Equation 1 is given by:

$$\mathbf{u}_{(k+1)\delta t} = \mathcal{F}(\mathbf{u}_{k\delta t}) \cdot \delta t + \mathbf{u}_{k\delta t}, \quad (2)$$

where δt denotes the step size. The operator \mathcal{F} is then approximated using the following Π -block $\hat{\mathcal{F}}$:

$$\hat{\mathcal{F}}(\mathbf{u}_{k\delta t}) = \sum_{c=1}^{N_c} W_c \cdot \left[\prod_{l=1}^{N_l} (K_{c,l} \star \hat{\mathbf{u}}_{k\delta t} + b_l) \right], \quad (3)$$

where N_c represents the number of channels, and N_l is the total number of parallel convolutional layers. The symbol \star denotes the convolution operation. For each layer l and channel c , $K_{c,l}$ denotes the filter weight, while b_l represents the bias for layer l . W_c is a 1×1 convolutional layer assigned to the c^{th} channel.

The main idea of the Π -block is to use multiplicative convolutional layers to approximate the FD stencil. Because of its multiplicative structure, the Π -block can efficiently approximate unknown nonlinear terms, such as $\mathbf{u} \cdot \nabla \mathbf{u}$. Theorem 1 in [Rao *et al.*, 2023] demonstrates that the Π -block can serve as a universal polynomial approximator for the unknown nonlinear operator.

Furthermore, to effectively incorporate the partially known knowledge of the PDE system, we use the physics-based convolutional (PyConv) layer as a complementary strategy. When a term in the governing PDE is known, such as the Laplacian term $\nabla^2 \mathbf{u}$, we directly set the kernel of the PyConv to align with the corresponding FD stencil. As a result, the PyConv layer is constructed to incorporate known physical information, while the Π -block is designed to learn the unknown operators. The combination of the Π -block and PyConv forms the physics-encoded block, which fully utilizes the physics structure of the differential operator and can efficiently model nonlinear systems from limited training data.

3.4 Spectral-enhanced Block

Although the physics-encoded block can efficiently approximate local operators within the complex system, it still fails to capture long-range spatial dependencies due to its limited receptive field. To overcome this limitation, we propose the spectral-enhanced block to process global information in the frequency domain, enabling a more comprehensive and holistic understanding of the complex system.

The proposed spectral-enhanced block (Figure 1(c)) contains the following components. An encoder (E) is first employed to map the spatial physical field into the frequency domain via the Fast Fourier Transform (FFT). After that, a frequency domain operator processes the transformed information in the frequency domain. Subsequently, the data is returned to the spatial domain using an inverse FFT (IFFT), followed by a decoder (D) to reconstruct the spatial field.

Frequency Domain Operator

The frequency domain operator (Figure 1(d)) is specifically designed to process information in the frequency domain. The block takes in the input spectrum and applies the following series of operations.

Firstly, a spectral attention mechanism is introduced to capture inter-spectrum relationships, as detailed in the next section. This mechanism allows the model to identify and emphasize key spectral features critical for the task, enhancing its ability to capture dependencies across different spectral

layers. Secondly, the processed spectrum is aggregated with the original spectrum via element-wise addition. This skip connection strategy helps ensure robustness during training and preserves the original frequency information. Finally, the aggregated spectrum undergoes filtering and a linear transformation, aligning with the traditional neural Fourier layer [Li *et al.*, 2021].

These operations work together to handle the frequency signals, enhance the model’s representational capacity, and ultimately produce the output spectrum.

Spectral Attention Mechanism

Motivated by the channel attention module [Wang *et al.*, 2020], we introduce the spectral attention mechanism (Figure 1(e)) within the frequency domain operator, which effectively exploits the inter-spectrum relationships among spectral features. Given a complex feature denoted as $\mathbf{X} + i\mathbf{Y} \in \mathbb{C}^{c \times k_1 \times k_2}$, where c denotes the number of channels and $k_1 \times k_2$ represents the number of frequency components, we first aggregate the frequency information of each feature map via the following average-pooling (Avg) and max-pooling (Max) operations:

$$\begin{aligned} (\mathbf{X}_{\text{avg}}, \mathbf{Y}_{\text{avg}}) &= (\text{Avg}(\mathbf{X}), \text{Avg}(\mathbf{Y})) \in \mathbb{R}^c \times \mathbb{R}^c, \\ (\mathbf{X}_{\text{max}}, \mathbf{Y}_{\text{max}}) &= (\text{Max}(\mathbf{X}), \text{Max}(\mathbf{Y})) \in \mathbb{R}^c \times \mathbb{R}^c. \end{aligned} \quad (4)$$

We handle the real and imaginary parts separately because the complex field is not well-ordered, i.e., there is no natural ordering between complex numbers. As a result, directly applying the max-pooling operation to the complex-valued features would lead to meaningless results. After the pooling operations, we apply four multi-layer perceptrons (MLPs) with one hidden layer to capture the inter-spectrum relationships as follows:

$$\begin{aligned} \text{att}(\mathbf{X}) &= \sigma(\text{MLP}_1(\mathbf{X}_{\text{avg}}) + \text{MLP}_2(\mathbf{X}_{\text{max}})) \in \mathbb{R}^c, \\ \text{att}(\mathbf{Y}) &= \sigma(\text{MLP}_3(\mathbf{Y}_{\text{avg}}) + \text{MLP}_4(\mathbf{Y}_{\text{max}})) \in \mathbb{R}^c, \end{aligned} \quad (5)$$

where σ represents an activation function. Applying the attention coefficients to re-weight the input spectrum, the processed spectrum $\mathbf{X}_p + i\mathbf{Y}_p \in \mathbb{C}^{c \times k_1 \times k_2}$ can be obtained as:

$$\mathbf{X}_p + i\mathbf{Y}_p = (\mathbf{X} + i\mathbf{Y}) \otimes (\text{att}(\mathbf{X}) + i\text{att}(\mathbf{Y})), \quad (6)$$

where \otimes denotes the element-wise multiplication.

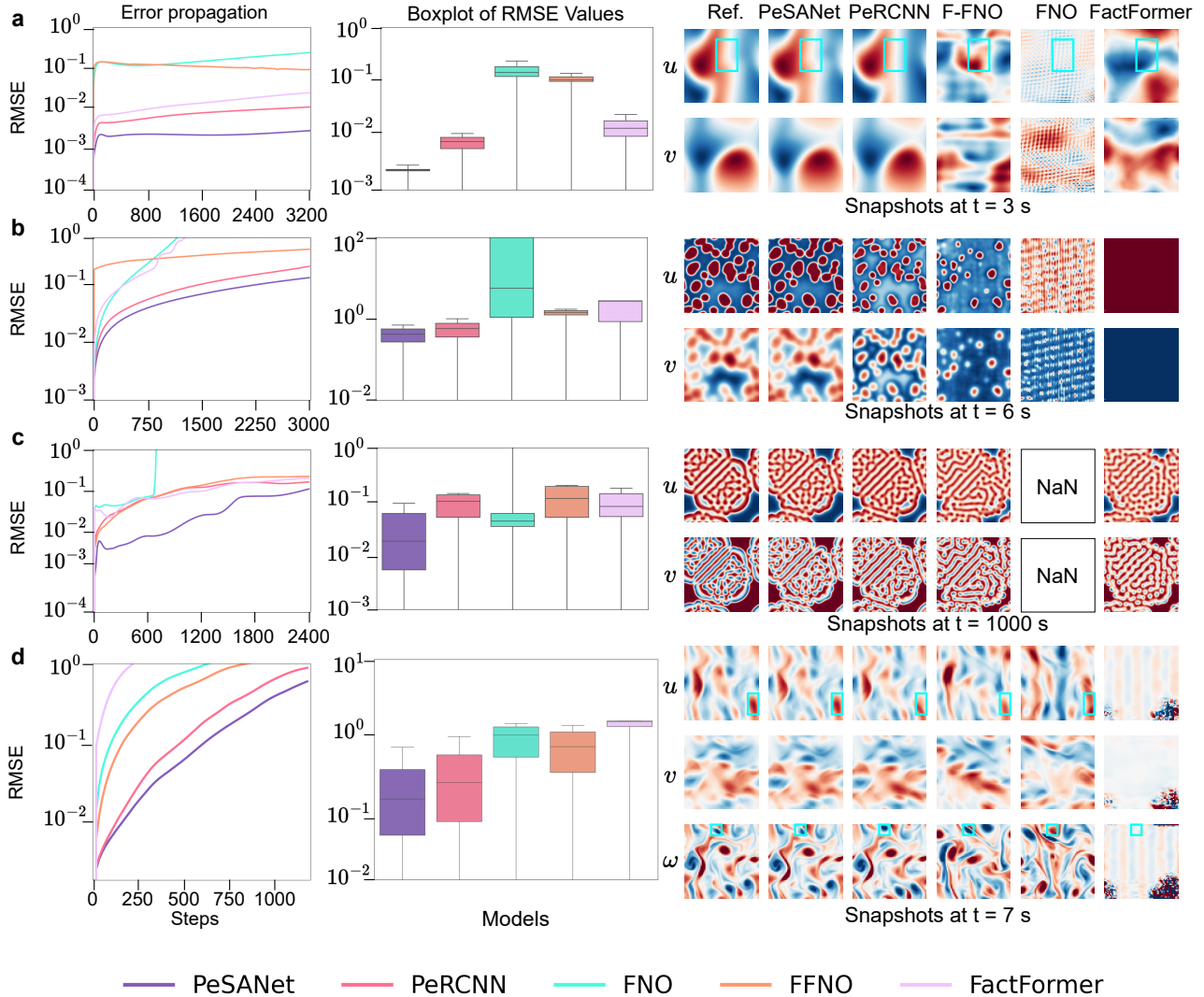
Compared to the traditional neural Fourier layer [Li *et al.*, 2021], the spectral attention mechanism further enhances the representation capability for the following reasons. On the one hand, the spectral attention mechanism uses all frequency components to calculate the attention coefficients, instead of directly truncating high-frequency signals. This comprehensive consideration ensures all critical frequency information in the spectral domain is leveraged effectively. On the other hand, the attention mechanism in Equation 5 enhances the interactions between spectral layers by dynamically capturing their internal relationship. This interaction allows information to be processed jointly across spectrum layers, resulting in a more expressive representation of spectral features.

4 Experiments

To rigorously evaluate the efficacy and generalizability of the proposed PeSANet, we conducted a series of comprehensive experiments on a wide range of complex systems, such

Case	Numerical Methods	Spatial Domain	Temporal Domain	Training Trajectories	Test Trajectories
Burgers	FDM	$(0, 1)^2$	$[0, 3.216]$	5	5
FN	FDM	$(0, 128)^2$	$[0, 10]$	5	5
GS	FDM	$(0, 1)^2$	$[0, 1900]$	2	5
NSE	FVM	$(0, 2\pi)^2$	$[0, 10]$	5	14

Table 1: Summary of experimental settings for different cases.


 Figure 2: **An overview comparison between PeSAnet and other baselines:** error propagation curves (left), error boxplots (middle), and final prediction plots (right). Figures (a-d) respectively show the qualitative results for the 2D Burgers, 2D FN, 2D GS, and 2D NSE cases.

as fluid dynamics and reaction-diffusion systems. Specifically, the model is evaluated on the two-dimensional Burgers' equation, the two-dimensional FitzHugh-Nagumo (FN) system, the two-dimensional Gray-Scott (GS) system, and the two-dimensional Navier-Stokes equations (NSE). The experimental results demonstrate that the PeSAnet consistently achieves superior accuracy compared to existing methodologies across all challenging benchmarks. These experiments not only showcase the model's robust adaptability to a diverse

set of systems but also confirm its ability to accurately predict long-term dynamics.

4.1 Experimental Setup

This section details the experimental setup, including the datasets, baselines, metrics, and the training procedure.

Datasets. We considered several two-dimensional PDE-governed nonlinear complex systems, including the Burgers' equation, the FitzHugh-Nagumo (FN) system, the Gray-

Case	Model	RMSE ↓	MAE ↓	HCT (s) ↑
Burgers	FactFormer	0.0142	0.0109	3.216
	FNO	0.1565	0.1301	0.104
	F-FNO	0.1097	0.0877	0.088
	PeRCNN	<u>0.0075</u>	<u>0.0058</u>	<u>3.216</u>
	PeSANet (Ours)	0.0069	0.0051	3.216
	Promotion (↑)	8.00%	12.07%	0.00%*
FN	FactFormer	0.8664	0.3472	1.62
	FNO	3196737	1892978	1.77
	F-FNO	0.5328	0.3828	0.06
	PeRCNN	<u>0.2281</u>	<u>0.1526</u>	<u>6.27</u>
	PeSANet (Ours)	0.1577	0.0784	8.13
	Promotion (↑)	30.86%	48.62%	29.66%
GS	FactFormer	0.1095	0.0704	982.5
	FNO	NaN	NaN	330.0
	F-FNO	0.1265	0.0829	832.5
	PeRCNN	<u>0.0993</u>	<u>0.0595</u>	<u>1110.0</u>
	PeSANet (Ours)	0.0481	0.0307	1327.5
	Promotion (↑)	51.56%	48.4%	19.59%
NSE	FactFormer	0.8971	0.6032	0.896
	FNO	0.8625	0.6617	2.464
	F-FNO	0.6992	0.5171	3.808
	PeRCNN	<u>0.3533</u>	<u>0.1993</u>	<u>7.197</u>
	PeSANet (Ours)	0.2308	0.1573	7.728
	Promotion (↑)	34.67%	21.07%	7.37%

Table 2: **Quantitative results of PeSANet and baselines.** The best results are in bold, and the second-best results are underlined. The * indicates the inference process has reached the end of the test data.

Scott (GS) system, and the Navier-Stokes equations (NSE). These systems represent diverse physical phenomena and dynamic behaviors, providing a comprehensive assessment of the model’s performance and generalization capabilities. In this paper, we target the data-scare scenario, and each experiment includes 2-5 trajectories in the training set. We summarize the datasets in Table 1, with a more detailed introduction in Appendix Dataset Informations.

Baseline Models. To demonstrate the effectiveness of PeSANet, we compared it with the following baseline models:

- **FNO** [Li *et al.*, 2021]: A widely recognized high-performing data-driven neural operator, which utilizes the Fourier transform to handle PDE information in the frequency domain.
- **PeRCNN** [Rao *et al.*, 2023]: A model that embeds physical knowledge in a hard-constrained manner and demonstrates excellent long-term prediction ability.
- **F-FNO** [Tran *et al.*, 2023]: The Factorized Fourier Neural Operator, which employs a separable Fourier layer and enhanced residual connections to effectively reduce computational complexity and improve scalability.
- **FactFormer** [Li *et al.*, 2024a]: A model that addresses the computational expense and numerical instability of applying Transformers to PDEs with large grids by employing an axial factorized kernel integral.

Evaluation Metrics. We use three types of metrics to

quantitatively assess the model performance: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and High Correction Time (HCT). Specifically, RMSE quantifies the average magnitude of prediction errors in the original data units, offering a general measure of model performance. MAE, conversely, measures the average absolute difference between predicted and ground truth values, exhibiting robustness to outliers compared to RMSE. Finally, HCT quantifies the temporal duration required for the model’s predictions to converge to a high level of accuracy, a critical consideration for real-time applications. The complete mathematical formulations of these metrics can be found in Appendix Evaluation Metrics.

Model Training. We train our model autoregressively, where the model predicts an output at each iteration, and the output is used as input for the next prediction. The mean squared error loss function is utilized for training, optimized using the Adam optimizer.

4.2 Main Results

The performance characteristics of the proposed PeSANet are comprehensively evaluated and presented in Figure 2 and Table 2. Demonstrating robust performance despite limited training data, Figure 2 provides visualizations of error distribution, error propagation characteristics, and predicted trajectories. These qualitative results are further substantiated by the quantitative metrics presented in Table 2.

2D Burgers Equation. The efficacy of the proposed PeSANet framework is demonstrated in the context of the 2D Burgers equation (Figure 2(a)). While PeRCNN exhibited some success in the beginning, it struggled with long-term predictions, as evidenced by the errors observed in Figure 2(a). In contrast, the PeSANet framework demonstrates superior performance in capturing the underlying physical dynamics, particularly in long-term forecasting. This qualitative advantage is further corroborated by the quantitative metrics presented in Table 2, which show significant improvements in both RMSE and MAE. The divergence of PeRCNN is also evident in the final time step snapshot.

2D FitzHugh-Nagumo Equation. Figure 2(b) demonstrates the difficulties encountered by data-driven methods (F-FNO, FNO, and FactFormer) when trained on limited data, with FactFormer showing near-complete failure. Even F-FNO, the strongest among these, exhibits limitations in the long-term prediction of the FN dynamics. Although PeRCNN, incorporating physics-informed structures, generates predictions with similar contours to the ground truth, its long-term accuracy remains limited. In contrast, our model achieves at least a 30% improvement over the best baseline, demonstrating the benefits of our approach for small datasets and the crucial role of physics-encoded block in enhancing long-term prediction accuracy.

2D Gray-Scott Equation. According to Table 2, PeSANet achieves substantial gains of approximately 50% in both RMSE and MAE. The complex, labyrinthine patterns of the GS-RD system are visualized in the snapshots presented in Figure 2 (c). A clear visual comparison reveals that PeSANet maintains a reasonable resemblance to the ground truth even after long-term prediction, with only minor devi-

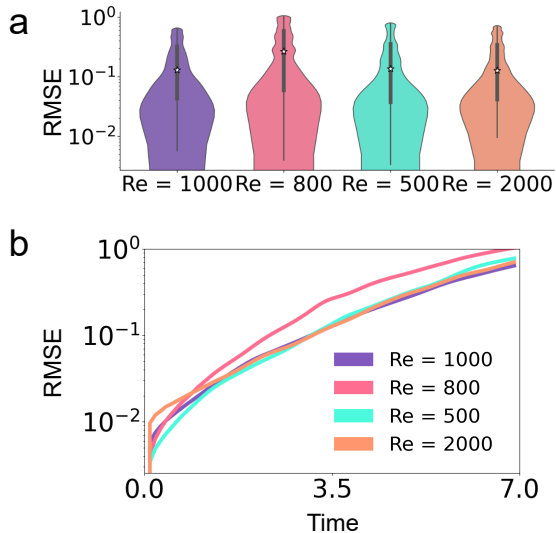


Figure 3: **Generalization test.** The error distribution (a) and propagation (b) for generalization over different Reynolds numbers.

ations. In contrast, other baseline models fail to capture the intricate patterns or even the basic contours of the structures.

2D Navier-Stokes Equation. For complex fluid dynamics, we evaluate the performance of the proposed PeSAnet with other baseline models on a Navier-Stokes Equations (NSE) dataset with a Reynolds number of 1000. This task requires models that can accurately capture both local variations and global features to ensure reliable long-term predictions. Compared to other baseline, our model achieves significant improvements of 20% and 30% in RMSE and MAE (Table 2). The error propagation curves in Figure 2(d) further validate the superiority of PeSAnet, with its error consistently lower than that of other models. The snapshot at $t = 7s$ shows near-perfect agreement between PeSAnet’s predictions and the ground truth, while other models exhibit significant deviations. This collective evidence establishes PeSAnet as the superior and more robust model for long-term prediction tasks in complex fluid dynamics systems.

4.3 Generalization Test

Our model leverages incorporated physical knowledge to enable effective fine-tuning and robust generalization with scarce training data. Using the NSE as a representative example, PeSAnet demonstrates excellent performance at the training Reynolds number ($Re = 1000$) and further exhibits strong generalization capabilities by fine-tuning with limited data to new Reynolds numbers ($Re = 500, 800,$ and 2000). The error distributions across these Reynolds numbers (Figure 3) confirm the generalization ability of PeSAnet. The error curves maintain similar characteristics to the $Re = 1000$ case, with error propagation well controlled, demonstrating the model’s reliability across different flow regimes. It is worth mentioning that fine-tuning with only two data samples enables PeSAnet to generate predictions closely approximating the ground truth, further validating its robustness and generalization capacity. In summary, PeSAnet effectively cap-

Model	RMSE ↓	MAE ↓	HCT (s) ↑
PeSAnet	0.1577	0.0784	8.13
PeSAnet w/o SA	0.1980	0.1073	7.65
PeSAnet w/o Pe	NaN	NaN	0.18
Pe+FNO	NaN	NaN	0.002
Pe+F-FNO	0.5118	0.3432	2.25

Table 3: Results for ablation study on the FN equation.

tures the underlying physical principles of the system, achieving accurate predictions for NSE at varying Reynolds numbers with minimal fine-tuning data.

4.4 Ablation Study

To evaluate the contributions of different components in our PeSAnet, we designed a series of ablation studies based on the FN equation. Specifically, we conducted experiments by (1) PeSAnet w/o SA, removing the spectral attention component, (2) PeSAnet w/o Pe, removing the physical-encoded block in the physical module, (3) Pe+FNO, replacing the spectral-enhanced block with FNO, and (4) Pe+F-FNO, replacing the spectral-enhanced block with F-FNO.

As shown in Table 3, experimental results demonstrate that removing the spectral attention module leads to a significant decline in network performance, highlighting the crucial role of the spectral attention mechanism in capturing spectral information and enhancing the model’s representational capacity. The removal of the physics-encoded block results in poor performance on small datasets, indicating the importance of prior physical knowledge in improving the model’s generalization ability, especially when data is limited. Furthermore, completely replacing the spectral-enhanced block with FNO or F-FNO also results in a significant performance drop.

5 Conclusion

This paper introduces PeSAnet, a predictive model designed to enhance the prediction accuracy of complex systems governed by PDEs by integrating global and local information, leveraging available physical data, and emphasizing frequency-domain feature processing. The PeSAnet is designed to not only capture the overall system dynamics but also exploit local information through hard-coding physical knowledge. Furthermore, it incorporates a carefully designed spectral attention module to capture key spectral features, effectively handling the global information in the frequency domain. These strategies significantly improve PeSAnet’s data efficiency, enabling superior long-term prediction performance and generalization capabilities even with limited training data. Experimental results demonstrate that PeSAnet achieves state-of-the-art predictive performance on four distinct nonlinear dynamical systems with small datasets, exhibiting strong adaptability and flexibility even with limited data and varying initial conditions. The generalization test also demonstrates PeSAnet’s transfer learning ability. In conclusion, PeSAnet offers a novel approach to long-term forecasting by integrating physics and spectral attention, demonstrating superior performance in data-scarce scenarios.

Acknowledgments

The work is supported by the National Natural Science Foundation of China (No. 62276269, No. 92270118), the Beijing Natural Science Foundation (No. 1232009), and the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDB0620103). Y. Liu would like to thank the support from the Fundamental Research Funds for the Central Universities (No. E2EG2202X2). The Appendix of this manuscript can be found in the preprint version: <https://www.arxiv.org/abs/2505.01736>. The source code and data are found at <https://github.com/intell-sci-comput/PeSAnet>.

Contribution Statement

Han Wan and Rui Zhang contributed equally to this work.

References

- [Anderson and Wendt, 1995] John David Anderson and John Wendt. *Computational Fluid Dynamics*, volume 206. Springer, 1995.
- [Bar-Sinai *et al.*, 2019] Yohai Bar-Sinai, Stephan Hoyer, Jason Hickey, and Michael P Brenner. Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences*, 116(31):15344–15349, 2019.
- [Ghergu and Radulescu, 2011] Marius Ghergu and Vicentiu Radulescu. *Nonlinear PDEs: Mathematical models in biology, chemistry and population genetics*. Springer Science & Business Media, 2011.
- [Gupta and Brandstetter, 2023] Jayesh K Gupta and Johannes Brandstetter. Towards multi-spatiotemporal-scale generalized PDE modeling. *Transactions on Machine Learning Research*, 2023.
- [Hang *et al.*, 2024] Zhou Hang, Yuezhou Ma, Haixu Wu, Haowen Wang, and Mingsheng Long. Unisolver: PDE-Conditional transformers are universal PDE solvers. *arXiv preprint arXiv:2405.17527*, 2024.
- [Holton and Hakim, 2013] James R Holton and Gregory J Hakim. *An introduction to dynamic meteorology*, volume 88. Academic Press, 2013.
- [Janny *et al.*, 2023] Steeven Janny, Aurélien Benetteau, Nicolas Thome, Madiha Nadri, Julie Digne, and Christian Wolf. Eagle: Large-scale learning of turbulent fluid dynamics with mesh transformers. In *International Conference on Learning Representations*, pages 1–9, 2023.
- [Karniadakis and Sherwin, 2005] George Karniadakis and Spencer J. Sherwin. *Spectral/HP Element Methods for Computational Fluid Dynamics*. Oxford University Press, USA, 2005.
- [Kochkov *et al.*, 2021] Dmitrii Kochkov, Jamie A Smith, Ayya Alieva, Qing Wang, Michael P Brenner, and Stephan Hoyer. Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, 118(21):e2101784118, 2021.
- [Krishnapriyan *et al.*, 2021] Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W Mahoney. Characterizing possible failure modes in physics-informed neural networks. In *Advances in Neural Information Processing Systems*, volume 34, pages 26548–26560, 2021.
- [Lee *et al.*, 2023] Jae Yong Lee, SungWoong CHO, and Hyung Ju Hwang. HyperdeepONet: learning operator with complex target function space using the limited resources via hypernetwork. In *International Conference on Learning Representations*, 2023.
- [Li *et al.*, 2021] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *International Conference on Learning Representations*, 2021.
- [Li *et al.*, 2024a] Zijie Li, Dule Shu, and Amir Barati Farihani. Scalable transformer for PDE surrogate modeling. *Advances in Neural Information Processing Systems*, 36, 2024.
- [Li *et al.*, 2024b] 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. *ACM/JMS Journal of Data Science*, 1(3):1–27, 2024.
- [Long *et al.*, 2019] Zichao Long, Yiping Lu, and Bin Dong. PDE-net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network. *Journal of Computational Physics*, 399:108925, 2019.
- [Lu *et al.*, 2018] Yiping Lu, Aoxiao Zhong, Quanzheng Li, and Bin Dong. Beyond finite layer neural networks: Bridging deep architectures and numerical differential equations. In *International Conference on Machine Learning*, pages 3276–3285. PMLR, 2018.
- [Lu *et al.*, 2021] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deepnet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.
- [Mi and Sun, 2024] Yuan Mi and Hao Sun. Spatiotemporal learning on cell-embedded graphs. *arXiv preprint arXiv:2409.18013*, 2024.
- [Moukalled *et al.*, 2016] Fadl Moukalled, Luca Mangani, and Marwan Darwish. *The finite volume method in computational fluid dynamics*. Springer, 2016.
- [Pfaff *et al.*, 2021] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter W Battaglia. Learning Mesh-Based Simulation with Graph Networks. In *International Conference on Learning Representations*, pages 1–9, 2021.
- [Rahman *et al.*, 2023] Md Ashiqur Rahman, Zachary E Ross, and Kamyar Azizzadenesheli. U-no: U-shaped neural operators. *Transactions on Machine Learning Research*, 2023.

- [Raissi *et al.*, 2019] Maziar Raissi, Paris Perdikaris, and George 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.
- [Rao *et al.*, 2022] Chengping Rao, Pu Ren, Yang Liu, and Hao Sun. Discovering nonlinear PDEs from scarce data with physics-encoded learning. In *International Conference on Learning Representations*, pages 1–9, 2022.
- [Rao *et al.*, 2023] Chengping Rao, Pu Ren, Qi Wang, Oral Buyukozturk, Hao Sun, and Yang Liu. Encoding physics to learn reaction–diffusion processes. *Nature Machine Intelligence*, 5(7):765–779, 2023.
- [Ren *et al.*, 2022] Pu Ren, Chengping Rao, Yang Liu, Jian-Xun Wang, and Hao Sun. Phycnet: Physics-informed convolutional-recurrent network for solving spatiotemporal PDEs. *Computer Methods in Applied Mechanics and Engineering*, 389:114399, 2022.
- [Sanchez-Gonzalez *et al.*, 2020] Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International Conference on Machine Learning*, pages 8459–8468. PMLR, 2020.
- [Seidman *et al.*, 2022] Jacob H Seidman, Georgios Kissas, Paris Perdikaris, and George J. Pappas. NOMAD: Nonlinear manifold decoders for operator learning. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022.
- [Shi *et al.*, 2015] Xingjian Shi, Zhourong Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 2015.
- [Stachenfeld *et al.*, 2021] Kimberly Stachenfeld, Drummond B Fielding, Dmitrii Kochkov, Miles Cranmer, Tobias Pfaff, Jonathan Godwin, Can Cui, Shirley Ho, Peter Battaglia, and Alvaro Sanchez-Gonzalez. Learned coarse models for efficient turbulence simulation. *arXiv preprint arXiv:2112.15275*, 2021.
- [Sun *et al.*, 2023] Zhiqing Sun, Yiming Yang, and Shinjae Yoo. A neural pde solver with temporal stencil modeling. In *International Conference on Machine Learning*, pages 33135–33155. PMLR, 2023.
- [Tran *et al.*, 2023] Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. In *The Eleventh International Conference on Learning Representations*, 2023.
- [Venturi and Casey, 2023] Simone Venturi and Tiernan Casey. Svd perspectives for augmenting deeponet flexibility and interpretability. *Computer Methods in Applied Mechanics and Engineering*, 403:115718, 2023.
- [Vlachas *et al.*, 2022] Pantelis R Vlachas, Georgios Arampatzis, Caroline Uhler, and Petros Koumoutsakos. Multi-scale simulations of complex systems by learning their effective dynamics. *Nature Machine Intelligence*, 4(4):359–366, 2022.
- [Wang *et al.*, 2020] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11534–11542, 2020.
- [Wang *et al.*, 2021] Sifan Wang, Hanwen Wang, and Paris Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed deep-ponets. *Science Advances*, 7(40):eabi8605, 2021.
- [Wang *et al.*, 2024] Qi Wang, Pu Ren, Hao Zhou, Xin-Yang Liu, Zhiwen Deng, Yi Zhang, Ruizhi Chengze, Hongsheng Liu, Zidong Wang, Jian-Xun Wang, Ji-Rong Wen, Hao Sun, and Yang Liu. P²C²Net: PDE-Preserved Coarse Correction Network for efficient prediction of spatiotemporal dynamics. In *Advances in Neural Information Processing Systems*, volume 38, 2024.
- [Wen *et al.*, 2022] Gege Wen, Zongyi Li, Kamyar Azizzadenesheli, Anima Anandkumar, and Sally M Benson. U-fno—an enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022.
- [Wu *et al.*, 2024] Haixu Wu, Huakun Luo, Haowen Wang, Jianmin Wang, and Mingsheng Long. Transolver: A fast transformer solver for pdes on general geometries. In *International Conference on Machine Learning*, volume 2200, pages 53681 – 53705, 2024.
- [Zeng *et al.*, 2025] Bocheng Zeng, Qi Wang, Mengtao Yan, Yang Liu, Ruizhi Chengze, Yi Zhang, Hongsheng Liu, Zidong Wang, and Hao Sun. PhyMPGN: Physics-encoded message passing graph network for spatiotemporal PDE systems. In *International Conference on Learning Representations*, pages 1–9, 2025.
- [Zhang *et al.*, 2024] Rui Zhang, Qi Meng, and Zhi-Ming Ma. Deciphering and integrating invariants for neural operator learning with various physical mechanisms. *National Science Review*, 11(4):nwad336, 2024.
- [Zhang *et al.*, 2025] Rui Zhang, Qi Meng, Rongchan Zhu, Yue Wang, Wenlei Shi, Shihua Zhang, Zhi-Ming Ma, and Tie-Yan Liu. Monte carlo neural pde solver for learning pdes via probabilistic representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.
- [Zienkiewicz *et al.*, 2005] Olek C Zienkiewicz, Robert L Taylor, and Jian Z Zhu. *The Finite Element Method: Its Basis and Fundamentals*. Elsevier, 2005.