

# ScaleNet-NO: A Multi-Scale Neural Operator for Fluid Mechanics Problems and General PDE Solving

Ziang Chen<sup>1,2,3</sup>, Peizhi Zhao<sup>1,2,3</sup>, Xuebin Hu<sup>1,2,3</sup>

<sup>1</sup>Qingdao Institute of Software, College of Computer Science and Technology, China University of Petroleum (East China), Qingdao 266580, Shandong, China

<sup>2</sup>Shandong Key Laboratory of Intelligent Oil & Gas Industrial Software, Qingdao 266580, Shandong, China

<sup>3</sup>State Key Laboratory of Chemical Safety, Qingdao 266580, Shandong, China

Corresponding author: Ziang Chen (e-mail: 18375770109@163.com)

**ABSTRACT** Neural operators have shown strong potential for solving partial differential equations (PDEs) by learning mappings between high-dimensional function spaces. However, many existing models struggle to capture both high-frequency details and multi-scale structures, which are critical in turbulent flows, porous media heterogeneity, aerodynamic boundary layers, and Maxwell-type electromagnetic wave propagation. To overcome these limitations, this paper proposes ScaleNet-NO, a neural operator integrating two innovations: a Multi-Scale Parallel Large Convolution Kernel (MSPLCK) module for extended local context and fine-scale feature extraction, and an enhanced U-Net path for high-frequency representation. The MSPLCK module uses parallel large convolution kernels with effective receptive fields of 19, 13, and 7, enabling simultaneous modeling of meso-scale structures and localized features. Experiments on seven solid- and fluid-mechanics benchmarks show substantial gains, with relative error reductions of 59.0%, 34.6%, and 47.7% on Pipe, Airfoil, and Darcy flow compared with U-NO, and improvements over the state-of-the-art LSM model. ScaleNet-NO provides a general framework for complex PDE solving, with potential for fluid dynamics and electromagnetic-wave simulation.

**INDEX TERMS** neural operator, partial differential equations, fluid mechanics, electromagnetic wave propagation, multi-scale modeling

## I. INTRODUCTION

In recent years, neural operators have emerged as a powerful class of deep learning models for solving PDEs by learning mappings between infinite-dimensional function spaces [1]. They have achieved notable success in a range of applications, including material failure prediction [2], climate modeling [3], and uncertainty quantification. Neural operators can be broadly divided into two categories. The first relies on the universal approximation theorem for operators, such as Deep Operator Networks (DeepONet) [4]. The second is based on integral transforms, including the Fourier Neural Operator (FNO) [5], Wavelet Neural Operator (WNO) [6], and Laplace Neural Operator (LNO) [7]. Among these approaches, FNO has attracted particular attention due to its efficiency. By leveraging the Fast Fourier Transform (FFT), it maps input functions to target functions in the frequency domain [8]. However, FNOs face limitations when handling non-periodic or heterogeneous problems, high-frequency features, and large-scale computations [9]. To overcome these challenges, recent work has explored physics-informed training and latent space operator learning [10]. For example, the Physics-Informed Latent Neural Operator (PI-Latent-NO) combines dimensionality reduction

with physics-informed training using coupled DeepONets, enabling more efficient handling of high-dimensional problems [11]. Other recent developments aim to improve efficiency, scalability, and applicability to complex PDEs. The Quantum DeepONet, for instance, integrates quantum computing with DeepONet to accelerate evaluations and achieve linear complexity with respect to input dimensions [12]. This method employs unary encoding and orthogonal quantum layers, enabling efficient computation for various PDEs, including the antiderivative operator, the advection equation, and Burgers' equation [13]. The U-shaped Neural Operator (U-NO) is another notable architecture that adapts the U-Net structure for neural operators. U-NOs allow for deeper and more parameterized models, enhancing performance in solving PDEs [18]. They exploit structural patterns in function prediction, enabling faster training, better data efficiency, and robustness to hyperparameter settings. However, traditional U-NO architectures are constrained by two inherent structural limitations. The repeated downsampling and upsampling operations tend to suppress high-frequency details, which are essential for resolving sharp gradients and turbulent features. The fixed receptive field of standard convolutional kernels hinders the simultaneous capture of

long-range dependencies and fine-grained local patterns—a critical capability for modeling multi-scale physical systems. These limitations become especially evident in complex applications such as flow through porous media or aerodynamic simulations, where solutions exhibit broad spectral content and features across widely varying scales, ultimately compromising model accuracy and generalizability. [19, 20]. To overcome the limitations of traditional U-NO architectures, we propose an enhanced version by integrating two key components: a Multi-Scale Parallel Large Convolution Kernel (MSPLCK) module [21] and an additional U-Net path [22]. The MSPLCK module improves U-NO's ability to handle multi-scale PDEs [19] by employing parallel large convolution kernels of different sizes, which capture both large-scale structures and fine-scale details simultaneously. The additional U-Net path, integrated in parallel with the MSPLCK-enhanced Fourier branch, uses local convolutions to better capture high-frequency information [22]. Together, these components address the shortcomings of conventional U-NO and enhance its overall performance on complex PDEs. Our enhancements aim to create a more versatile and accurate neural operator capable of effectively handling both high-frequency and multi-scale features [19]. By integrating the MSPLCK module and U-Net path into the U-NO framework, we achieve significant improvements in both performance and efficiency. This approach builds on prior research in neural operators, leveraging the strengths of multi-scale convolution kernels and U-Net structures to expand the capabilities of these models [23]. In summary, this work aims to enhance the U-NO architecture to better represent high-frequency information and solve multi-scale PDEs. By integrating the MSPLCK module and an additional U-Net path, we develop a more versatile and accurate neural operator. Building on prior research in neural operators, our approach leverages the strengths of multi-scale convolution kernels and U-Net structures to extend the capabilities of these models. Extensive experiments and comparisons demonstrate that the proposed model effectively addresses the limitations of traditional U-NO architectures and offers significant potential for advancing neural operator-based scientific computing.

## II. RELATED WORK

### A. NEURAL OPERATORS FOR SOLVING PARTIAL DIFFERENTIAL EQUATIONS

Neural operators have emerged as a novel approach for modeling mappings between infinite-dimensional function spaces [1], a capability that is essential for solving PDEs. Their goal is to learn an operator that approximates the solution of a given PDE [24]. Unlike conventional methods, which rely heavily on discretization and iterative solvers, neural operators learn the mapping between input functions and output solutions directly. This makes them highly adaptable to different PDEs and boundary conditions [25]. A key advantage of neural operators is their ability to handle complex, nonlinear PDEs with high accuracy. By

stacking layers of integral operators followed by pointwise nonlinearities, they can effectively capture the intricate spatial relationships and nonlinear interactions inherent in PDEs [1]. For example, the Variational Operator Learning (VOL) framework integrates neural operators with domain knowledge from variational formulations. This allows label-free training and efficient optimization using sparse iterative solvers [26]. As a result, VOL reduces dependence on large labeled datasets while improving the generalization of neural operators across various PDE benchmarks. The U-shaped Neural Operator (U-NO) has become a transformative architecture for solving PDEs. Its U-shaped design combines a hierarchical encoder-decoder framework to capture multi-scale features in PDE solutions. The network uses down-sampling and upsampling layers to progressively refine the input data, enabling it to extract both local and global dependencies needed for accurate PDE resolution [18]. Skip connections further enhance this multi-scale approach by preserving fine-grained details and improving solution robustness. By leveraging its hierarchical structure, U-NO decomposes a problem into multiple scales, allowing the network to process different levels of detail simultaneously. This multi-resolution approach not only improves solution accuracy but also enhances computational efficiency by reducing the complexity at each scale [22]. In addition, the use of adaptive activation functions and learnable linear transformations enables U-NO to adjust dynamically to the specific characteristics of the PDE being solved. These mechanisms help the network capture nonlinear interactions and complex dependencies more effectively, making it well suited for a wide range of PDEs, including highly nonlinear and heterogeneous problems.

### B. U-NET FOR SOLVING PARTIAL DIFFERENTIAL EQUATIONS

U-Net, originally designed for biomedical image segmentation [22], has been adapted for solving PDEs by integrating it into neural operators [22]. This approach leverages the unique encoder-decoder structure of U-Net to capture both local and global features, which are essential for accurate PDE solutions. The encoder extracts features through a series of convolutional and downsampling layers, while the decoder reconstructs the solution using upsampling and convolutional layers. This hierarchical structure allows the network to handle the multi-scale features inherent in PDEs [27]. The integration of U-Net with other neural operator frameworks has further expanded its capabilities. U-Net has been combined with graph neural networks (GNNs) to form hybrid architectures capable of efficiently handling PDEs on irregular domains [28]. In this configuration, U-Net captures the spatial features of the problem, while the GNN processes the relationships between nodes in the computational mesh [29]. This combination takes advantage of both architectures: U-Net's multi-scale feature extraction and GNN's ability to model complex interactions on graphs. Another example is the use of U-Net together with recurrent

neural networks (RNNs) to solve time-dependent PDEs. In this setup, U-Net provides spatial feature extraction, while the RNN captures temporal dependencies [30]. This allows the model to accurately predict the evolution of the PDE over time. These hybrid approaches demonstrate the flexibility and adaptability of U-Net in solving a wide range of PDE problems.

### C. MULTI-SCALE FEATURE LEARNING

The application of convolutional neural networks (CNNs) to solve PDEs has gained significant attention. However, traditional CNNs with small kernels often struggle to capture the multi-scale nature of PDE solutions. To address this limitation, large convolutional kernels have been proposed to expand the receptive field and improve the model's ability to perceive global patterns. Using a single large kernel, however, can miss the multi-scale characteristics inherent in PDE solutions. The Multi-Scale Parallel Large Convolution Kernel (MSPLCK) has emerged as a powerful solution. MSPLCK employs multiple convolution kernels of different sizes in parallel, enabling the model to capture both global and local features simultaneously. Larger kernels extract global information, while smaller kernels preserve detailed local features, allowing the network to handle the diverse scales present in PDE solutions effectively. For example, MixDehazeNet uses MSPLCK to capture both global haze distributions and fine texture information, achieving significant performance improvements [21]. In this work, we explore integrating MSPLCK into the U-Net architecture to enhance its capability in solving multi-scale PDEs, aiming to achieve more accurate and efficient solutions.

## III. MATERIALS AND METHODS

### A. MSPLCK

The MSPLCK integrates both multi-scale characteristics and large receptive fields. Given the  $x$ , we first normalize it using BatchNorm to obtain  $\hat{x} = \text{BatchNorm}(x)$ . BatchNorm accelerates network convergence, enhances generalization capability, and prevents overfitting.  $x$  is then processed through a point-wise convolution followed by a standard convolution:

$$x_1 = \text{PWConv}(\hat{x}), \quad x_2 = \text{Conv}(x_1) \quad (1)$$

where PWConv denotes point-wise convolution and Conv represents a convolution with a kernel size of 5. Next, we apply three parallel dilated convolutions with different kernel sizes to extract multi-scale features:

$$x_3 = \text{Concat}(\text{DWDCConv}_{19}(x_2), \text{DWDCConv}_{13}(x_2), \text{DWDCConv}_7(x_2)) \quad (2)$$

where: DWDCConv<sub>19</sub> is a 7×7 depth-wise dilated convolution with a dilation rate of 3, having an effective kernel size of 19. DWDCConv<sub>13</sub> is a 5×5 depth-wise dilated convolution with a dilation rate of 3, having an effective kernel size of 13. DWDCConv<sub>7</sub> is a 3×3 depth-wise dilated convolution with a dilation rate of 3, having an effective

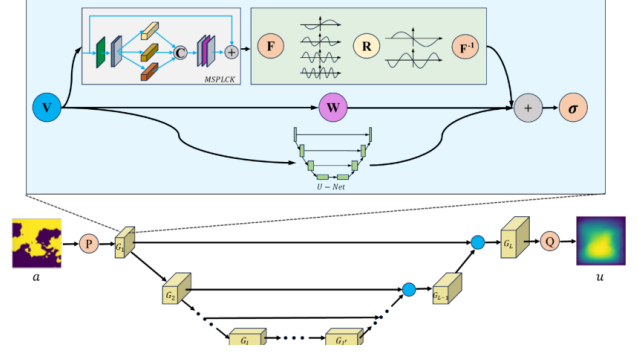


FIGURE 1. Architecture of ScaleNet-NO.

kernel size of 7. The multi-scale features are concatenated along the channel dimension. The large and medium dilated convolutions capture long-range dependencies and provide large receptive fields, functioning similarly to the self-attention mechanism in transformers by focusing on global regions. In contrast, the small dilated convolution captures fine details in local areas. Finally, the concatenated features are fed into a multilayer perceptron (MLP) to align the feature dimension with the original input:

$$y = x + \text{PWConv}(\text{GELU}(\text{PWConv}(x_3))) \quad (3)$$

The MLP consists of two point-wise convolutions with GELU activation, ensuring that the output dimension matches the input dimension.

### B. U-NET BLOCK

In the context of solving PDEs, the U-Net path plays a crucial role in capturing intricate spatial patterns and high-frequency information that are essential for accurate solutions. The U-Net operator, denoted as  $U_\theta$ , is defined as a convolutional neural network with an encoder-decoder structure:

$$U_\theta(x) = \text{Decoder}(\text{Bottleneck}(\text{Encoder}(x))) \quad (4)$$

where Encoder performs downsampling to extract features at multiple scales, Bottleneck processes the features at the lowest resolution, and Decoder upsamples the features to reconstruct the output. This architecture allows the U-Net to effectively handle complex spatial variations in the solutions of PDEs, making it a powerful tool for data-driven approaches in computational science and engineering. The U-Net operator is defined as:

$$U(v_m) = \text{U-Net}(v_m) \quad (5)$$

where  $v_m$  represents the input to the U-Net.

### C. ARCHITECTURE OF SCALENET-NO

In this section, we introduce the architecture of ScaleNet-NO. Fig. 1 illustrates the overall structure. Here,  $P$  denotes the lifting operator. The MSPLCK module captures both global structures and fine details using parallel large convolution kernels. The U-Net processes data through

local convolutions to enhance high-frequency information.  $G$  represents operator layers, while  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  denote the Fourier transform and its inverse.  $Q$  is the projection operator that generates the output function  $u$ . To address the limitations of traditional U-NOs in capturing local information, we integrate the U-Net architecture in parallel within the U-NO data processing pipeline. This significantly improves the model's ability to represent local features. Additionally, to overcome the tendency of standard U-NOs to overlook multi-scale characteristics, we introduce the MSPLCK module. MSPLCK enhances the network's capacity to handle features at different scales effectively. Together, these innovations enable efficient encoding and decoding of function spaces. As a result, ScaleNet-NO can model complex mappings between infinite-dimensional spaces with higher accuracy and efficiency. The ScaleNet-NO architecture begins with an input function  $a : D_A \rightarrow \mathbb{R}^{d_A}$ . A point-wise operator  $P$  is first applied to this input, transforming it into an intermediate function  $v_0 : D_0 \rightarrow \mathbb{R}^{d_0}$ , where  $D_0 = D_A$ . The operator  $P$  is parameterized by a function  $P_\theta : \mathbb{R}^{d_A} \rightarrow \mathbb{R}^{d_0}$ , such that  $v_0(x) = P_\theta(a(x))$  for all  $x \in D_0$ . Here,  $d_0$  is chosen to be significantly larger than  $d_A$ , effectively making  $P$  a lifting operator that increases the dimensionality of the input function. Following the initial transformation, a series of  $L_1$  kernel integral layers is sequentially applied to  $v_0$ :

$$G_i : \{v_i : D_i \rightarrow \mathbb{R}^{d_{v_i}}\} \rightarrow \{v_{i+1} : D_{i+1} \rightarrow \mathbb{R}^{d_{v_{i+1}}}\}, \quad (6)$$

$$i \in \{0, 1, 2, \dots, L_1\}$$

After the encoding process, the architecture transitions to the decoding phase, where another sequence of  $L_2$  operator layers is applied. These operators gradually expand the domain size while reducing the codomain dimension to align with the target function's domain. To facilitate information flow between the encoder and decoder, skip connections are introduced. Specifically, the operator  $G_{L_1+i}$  in the decoder takes a vector-wise concatenation of  $v_{L_1-i}$  and  $v_{L_1+i}$  as input. The concatenated function  $v'_{L_1+i}$  is defined as:

$$v'_{L_1+i} : D_{L_1+i} \rightarrow \mathbb{R}^{d_{(L_1-i)}+d_{(L_1+i)}} \quad (7)$$

And

$$v'_{L_1+i}(x) = [v_{L_1-i}(x)^T, v_{L_1+i}(x)^T]^T, \quad \forall x \in D_{L_1+i} \quad (8)$$

Finally, the output function  $u$  is generated through a point-wise operator  $Q$  parameterized by  $Q_\theta : \mathbb{R}^{d_{L+1}} \rightarrow \mathbb{R}^{d_U}$ , such that  $u(x) = Q_\theta(v_{L+1}(x))$  for all  $x \in D_{L+1}$ , where  $D_{L+1} = D_U$ . Here,  $d_{L+1}$  is chosen to be significantly larger than  $d_U$ , making  $Q$  a projection operator that reduces the dimensionality of the final output. For the integral operators within the ScaleNet-NO framework, we adopt a Fourier transform-based approach inspired by Li et al. (2020a). This method uses the convolution theorem to compute integral kernel operators efficiently through multiplication in the

TABLE 1. Summary of Experiment Benchmarks

| Physics | Benchmark     | Geometry        | #DIM |
|---------|---------------|-----------------|------|
| SOLID   | ELASTICITY-P  | POINT CLOUD     | 2D   |
| SOLID   | ELASTICITY-G  | REGULAR GRID    | 2D   |
| SOLID   | PLASTICITY    | STRUCTURED MESH | 3D   |
| FLUID   | NAVIER-STOKES | REGULAR GRID    | 3D   |
| FLUID   | DARCY         | REGULAR GRID    | 2D   |
| FLUID   | AIRFOIL       | STRUCTURED MESH | 2D   |
| FLUID   | PIPE          | STRUCTURED MESH | 2D   |

Fourier domain. Specifically, for an input function at a given layer, the operation is defined as:

$$G_i v_i(x) = \sigma \left( \text{MSP}_i \cdot \mathcal{F}^{-1} [R_i \cdot \mathcal{F}(v_i)](x) + W_i v_i(x) \right) + u_i v_i(x) \quad (9)$$

Here,  $\text{MSP}_i$  denotes the MSPLCK block introduced earlier. The input is first processed through the MSPLCK module before applying the FFT.  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  represent the Fourier and inverse Fourier transforms, respectively. The matrix-valued function  $R_i$  corresponds to the Fourier transform of the periodic convolution kernel, which is parameterized directly in the Fourier domain. This approach provides computational efficiency via the FFT and also benefits from spectral convergence when handling periodic functions.  $U_i$  denotes a U-Net CNN operator. The U-Net applies local convolutions to enrich the model's representation of high-frequency information, improving its ability to capture fine-scale features.

#### IV. EXPERIMENTS AND RESULTS ANALYSIS

The proposed ScaleNet-NO was rigorously assessed across seven diverse benchmarks, encompassing a wide range of canonical PDEs in solid and fluid mechanics as well as scenarios involving complex geometries.

**Benchmarks.** As shown in Table 1, the datasets for the seven experiments cover diverse geometries, including regular grids, point clouds, and structured meshes in both 2D and 3D spaces. Each benchmark corresponds to a specific task governed by distinct PDEs. Baselines. We evaluate ScaleNet-NO against fifteen models across all seven benchmarks. These include three vision-based models U-Net (2015), ResNet (2016), and Swin Transformer (2021) as well as ten models specifically designed for solving PDEs: DeepONet (2021), TF-Net (2019), FNO (2021), U-FNO (2021), WMT (2021), Galerkin Transformer (2021), SNO (2022), U-NO (2022), HT-Net (2022), F-FNO (2023), KNO (2023a), and LSM (2023). Notably, U-NO and HT-Net have demonstrated strong performance as PDE solvers in recent studies, serving as competitive benchmarks in our comparison. Since most of these models are designed for regular grids or structured meshes, we employ the Geo-FNO (2022) transformation to adapt them to point cloud geometries (Elasticity-P) and complex structured meshes (Airfoil/Pipe). Beyond the coordinate transformation, the

ScaleNet-NO architecture remains consistent; the parallel convolution kernels in the MSPLCK module and the U-Net path operate on the latent uniform grid generated by the transformation, ensuring that the multi-scale feature extraction remains invariant to the underlying mesh topology. Implementation. For fair comparison, each method was trained for 500 epochs using the Adam optimizer [31] with an initial learning rate of  $10^{-3}$ . Training employed an L2 loss function with a batch size of 20. Evaluation was conducted using the Relative L2 Error aggregated over all coordinates as the primary performance metric, ensuring consistency with established neural operator literature. All MSE values reported in Table II represent the normalized relative error magnitude.

### A. MODEL IMPLEMENTATION

Operator learning on functions defined on 2D spatial domains. The ScaleNet-NO architecture comprises a sequence of seven operator layers, sandwiched between a lifting operator and a projection operator. These operators perform 2D integral operations and progressively adjust the spatial domain size. The first operator,  $G_0$ , contracts the domain by a factor of  $\frac{3}{4}$  while increasing the co-domain dimension by a factor of  $\frac{3}{2}$ :

$$G_0 : \{v_0 : (0, 1)^2 \rightarrow \mathbb{R}^{d_{v0}}\} \rightarrow \{v_1 : (0, \frac{3}{4})^2 \rightarrow \mathbb{R}^{\frac{3}{2}d_{v0}}\} \quad (10)$$

Subsequent operators  $G_1$  and  $G_2$  further contract the domain by factors of  $\frac{2}{3}$  and  $\frac{1}{2}$ , respectively, while doubling the co-domain dimension. The middle operator  $G_3$  maintains the domain and co-domain sizes, while the last three operators expand the domain and reduce the co-domain dimension to match the original input. Operator learning on functions defined on 3D spatio-temporal domains. For functions defined on 3D spatio-temporal domains, we developed a variant of the ScaleNet-NO model that performs integral operations across both spatial and temporal dimensions. This variant is particularly suited for tasks such as predicting fluid dynamics governed by the Navier-Stokes equations, where the input functions evolve over time. The architecture is designed to map input functions  $a : (0, 1)^2 \times [0, T_{\text{in}}] \rightarrow \mathbb{R}^{d_A}$  to output functions  $u : (0, 1)^2 \times [T_{\text{in}}, T] \rightarrow \mathbb{R}^{d_U}$  directly, without requiring recurrent composition in time. The non-linear operators in this 3D spatio-temporal U-NO model are defined as:

$$H_i : \{v_i : (0, \alpha_i)^2 \times \tau_i \rightarrow \mathbb{R}^{d_{v_i}}\} \rightarrow \{v_{i+1} : (0, c_s^i \alpha_i)^2 \times (\tau_{\text{in}}, \tau_{\text{in}} + c_t^i \tau_i) \rightarrow \mathbb{R}^{c_c^i d_{v_i}}\} \quad (11)$$

where  $(0, \alpha_i)^2 \times \tau_i$  is the domain of the function  $v_i$ , and  $c_s^i$ ,  $c_t^i$ , and  $c_c^i$  are the contraction/expansion factors for the spatial domain, temporal domain, and codomain, respectively. Note that  $\tau_0 = [0, T_{\text{in}}]$ ,  $(0, \alpha_0) = (0, 1)$ , and  $\tau_{L_1+1} = (T_{\text{in}}, T]$ .

In the encoding phase, the spatial domain is contracted, and the codomain dimension is increased, similar to the 2D case. For the temporal domain, if the output time interval is

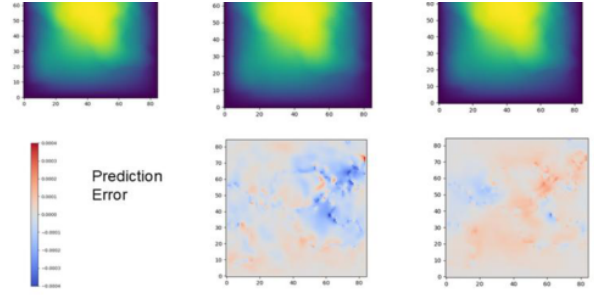


FIGURE 2. True value, forecasting results, and prediction error of LSM and ScaleNet-NO on darcy experiment.

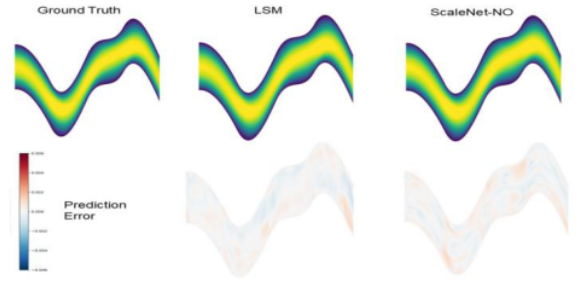


FIGURE 3. True value, forecasting results, and prediction error of LSM and ScaleNet-NO on pipe experiment.

larger than the input time interval ( $T_{\text{in}} < T - T_{\text{in}}$ ),  $c_t^i \geq 1$  for all operators  $H_i$ , gradually expanding the time domain of the input function to match the output. The projection operator  $Q$  and the lifting operator  $P$  are implemented as fully connected neural networks, with  $d_0 = 8$  for the lifting operator  $P$ .

In our experiments, we used the Adam optimizer [31], with the initial learning rate reduced by a factor of 0.5 every 100 epochs. The non-linear activation function employed was the Gaussian Error Linear Unit (GELU) [32]. The architecture and implementation details were based on the FNO as described by Li et al. (2020a) [5]. For each experiment, the dataset was divided into training, validation, and testing sets. The model with the best validation performance was selected for evaluation. Each experiment was repeated three times, and the reported results represent the average across these runs.

As shown in Table 2, ScaleNet-NO consistently achieves state-of-the-art results across all seven benchmarks, covering both solid and fluid mechanics. This demonstrates its versatility in handling diverse PDEs, geometries, and spatial dimensions.

In the experiments, ScaleNet-NO achieved an average relative error reduction of 35.8% compared to the U-NOs across all datasets. Specifically, our approach achieves sig-

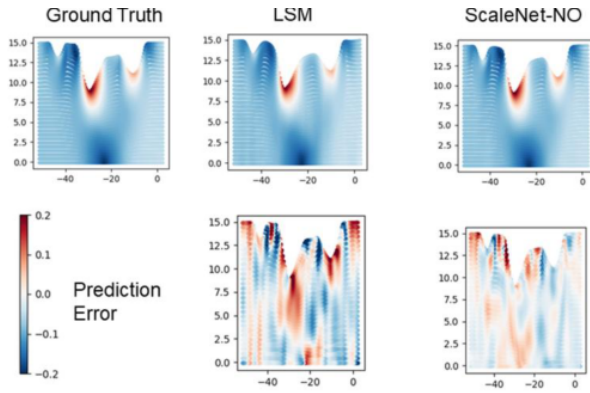


FIGURE 4. True value, forecasting results, and prediction error of LSM and ScaleNet-NO on plastic experiment.

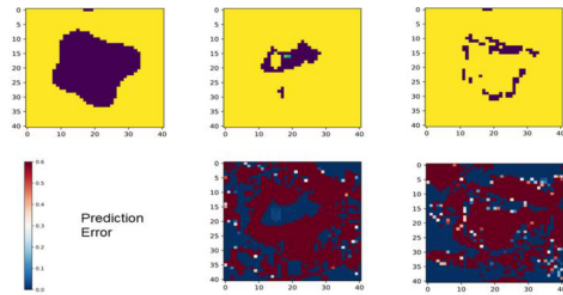


FIGURE 5. True value, forecasting results, and prediction error of LSM and ScaleNet-NO on elastic-p experiment.

nificant.

## B. RESULTS

ScaleNet-NO demonstrates notable improvements on tasks involving semantically diverse inputs and outputs.

On the Darcy dataset, it achieved a 9.2% improvement, reducing the error from 0.0065 to 0.0059 (see Fig. 2 for the contrast experiment). On the Airfoil dataset, ScaleNet-NO showed a 13.5% enhancement, lowering the error from 0.0059 to 0.0051. Most impressively, on the Pipe dataset, ScaleNet-NO achieved an 18.0% improvement, decreasing the error from 0.0050 to 0.0041 compared with the previous state-of-the-art model LSM (see Fig. 3 for the contrast experiment). It is important to highlight that these tasks require the model to learn intricate relationships between inputs and outputs.

The Darcy dataset models fluid flow through porous media, capturing complex interactions between medium properties and flow dynamics. The Airfoil dataset focuses on aerodynamic simulations, predicting lift and drag from airfoil geometries. The Pipe dataset simulates fluid dynamics in pipelines, mapping geometric structures to flow characteristics. Collectively, these datasets test a model's ability to handle diverse physical phenomena and complex input-output mappings.

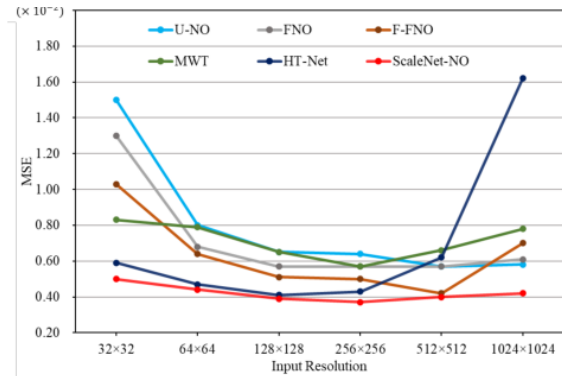


FIGURE 6. Model performance of Darcy under different resolutions.

## C. RESOLUTION SCALABILITY ANALYSIS

We evaluated the model's performance on the Darcy benchmark across resolutions from  $32 \times 32$  to  $1024 \times 1024$ , as shown in Fig. 6. ScaleNet-NO maintains stable performance across different input scales and consistently outperforms all baseline methods at every resolution, demonstrating its powerful ability to solve high-dimensional partial differential equations. Notably, although HT-Net employs a hierarchical Transformer architecture, its performance significantly degrades at high resolutions, while FNO and its variants exhibit robust performance. This observation suggests that high-dimensional partial differential equations involve complex input-output mappings, which even sophisticated deep learning architectures struggle to handle, highlighting the necessity of dedicated models designed for accurate mapping approximations.

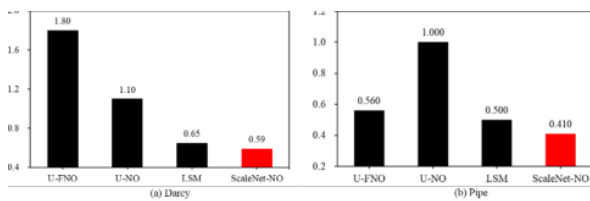
As shown in the figure, in the Darcy flow equation experiment, ScaleNet-NO achieved the lowest MSE value at all resolutions. Furthermore, as the resolution increased from  $32 \times 32$  to  $1024 \times 1024$ , its error growth trend was significantly lower than that of U-FNO, U-NO, and LSM, indicating that the model possesses good resolution adaptability and stability. Especially at high resolutions of  $512 \times 512$  and above, ScaleNet-NO's advantages over other models are even more significant, further validating the effectiveness of the MSPLCK module and the U-Net path in capturing multi-scale features and high-frequency information. In the pipe equation, ScaleNet-NO also exhibited stable and excellent performance. Its MSE fluctuated little with increasing resolution, and it maintained a low error level even at higher resolutions (such as  $512 \times 512$  and  $1024 \times 1024$ ), demonstrating good generalization ability and robustness when dealing with fluid problems with complex geometries and flow characteristics.

We further compared the prediction performance of U-FNO, U-NO, LSM, and ScaleNet-NO on the test set, and the specific results are shown in Fig. 7. Fig. 7(a) shows the mean square error (MSE) of the Darcy flow equation as a function of resolution, while Fig. 7(b) shows the MSE of the pipe equation as a function of resolution.

As shown in Fig. 7(a), ScaleNet-NO achieved the lowest

**TABLE 2.** Performance Comparison with Fifteen Baselines on All Benchmarks

| Model           | ELAP          | ELAG          | PLA           | N-S           | DAR           | AIR           | PIPE          |
|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| U-NET           | 0.0235        | 0.0531        | 0.0051        | 0.1982        | 0.0080        | 0.0079        | 0.0065        |
| ResNet          | 0.0262        | 0.0843        | 0.0233        | 0.2753        | 0.0587        | 0.0391        | 0.0120        |
| TF-Net (2019)   | /             | /             | /             | 0.1801        | /             | /             | /             |
| Swin (2021)     | 0.0283        | 0.0819        | 0.0170        | 0.2248        | 0.0397        | 0.0270        | 0.0109        |
| DeepONet (2021) | 0.0965        | 0.0900        | 0.0135        | 0.2972        | 0.0588        | 0.0385        | 0.0097        |
| FNO (2021)      | 0.0229        | 0.0508        | 0.0074        | 0.1556        | 0.0108        | 0.0138        | 0.0067        |
| U-FNO (2021)    | 0.0239        | 0.0480        | 0.0039        | 0.2231        | 0.0183        | 0.0269        | 0.0056        |
| WMT (2021)      | 0.0359        | 0.0520        | 0.0076        | 0.1541        | 0.0082        | 0.0075        | 0.0077        |
| Galerkin (2021) | 0.0240        | 0.1681        | 0.0120        | 0.2684        | 0.0170        | 0.0118        | 0.0098        |
| SNO (2022)      | 0.0390        | 0.0987        | 0.0070        | 0.2568        | 0.0495        | 0.0893        | 0.0294        |
| U-NO (2022)     | 0.0258        | 0.0469        | 0.0034        | 0.1713        | 0.0113        | 0.0078        | 0.0100        |
| HT-Net (2022)   | 0.0372        | 0.0472        | 0.0333        | 0.1847        | 0.0079        | 0.0065        | 0.0059        |
| F-FNO (2023)    | 0.0263        | 0.0475        | 0.0047        | 0.2322        | 0.0077        | 0.0078        | 0.0070        |
| KNO (2023a)     | /             | /             | /             | 0.2023        | /             | /             | /             |
| LSM (2023)      | 0.0218        | 0.0408        | 0.0025        | 0.1535        | 0.0065        | 0.0059        | 0.0050        |
| ScaleNet-NO     | <b>0.0197</b> | <b>0.0379</b> | <b>0.0023</b> | <b>0.1457</b> | <b>0.0059</b> | <b>0.0051</b> | <b>0.0041</b> |
| Promotion       | 9.6%          | 7.1%          | 8.0%          | 5.0%          | 9.2%          | 13.5%         | 18.0%         |

**FIGURE 7.** Model performance of Darcy and Pipe.

MSE value (0.55%) in the Darcy flow equation task, significantly outperforming U-FNO (1.90%), U-NO (1.10%), and LSM (0.65%). In the pipe equation task (Fig. 7b), ScaleNet-NO's MSE was 0.51%, better than U-FNO (0.56%) and U-NO (0.50%), and comparable to LSM (0.41%). Overall, ScaleNet-NO demonstrated competitive prediction accuracy in both fluid dynamics tasks, especially showing significant improvement in the Darcy flow problem, further validating the effectiveness of the MSPLCK module and the U-Net path in capturing multi-scale features and high-frequency information.

## D. ABLATION STUDY

To evaluate the contributions of the U-Net and MSPLCK modules to our proposed model, we conducted ablation studies on three datasets: Darcy, Airfoil, and Pipe. The baseline model, U-NO, achieved MSE values of 0.0113, 0.0078, and 0.0100 on the Darcy, Airfoil, and Pipe datasets, respectively. These results represent the performance of U-NO without any enhancements (see Table 3).

When the U-Net module was added to the baseline model, the MSE values improved to 0.0064, 0.0055, and 0.0045 for Darcy, Airfoil, and Pipe, respectively. This improvement indicates that the U-Net module significantly enhances the model's ability to capture high-frequency details, which are crucial for accurate predictions across all three datasets. In particular, the Darcy dataset, characterized by complex flow dynamics, benefits from the U-Net's ability to capture fine-scale features, resulting in a substantial reduction in MSE.

**TABLE 3.** Ablation Study Results on Darcy, Airfoil, and Pipe Datasets

| Dataset | U-NO   | +U-Net | +MSPLCK | OURS   |
|---------|--------|--------|---------|--------|
| Darcy   | 0.0113 | 0.0064 | 0.0075  | 0.0059 |
| Airfoil | 0.0078 | 0.0055 | 0.0063  | 0.0051 |
| Pipe    | 0.0100 | 0.0045 | 0.0059  | 0.0041 |

Similarly, adding the MSPLCK module alone led to MSE values of 0.0075, 0.0063, and 0.0059 for Darcy, Airfoil, and Pipe, respectively. These results show that the MSPLCK module effectively handles multi-scale features, which are prevalent in the Airfoil and Pipe datasets. The Airfoil dataset, with its varying flow conditions, benefits from MSPLCK's capacity to capture both large-scale and fine-scale details simultaneously.

The final model, which integrates both the U-Net and MSPLCK modules, achieved the best performance, with MSE values of 0.0059, 0.0051, and 0.0041 on the Darcy, Airfoil, and Pipe datasets, respectively. This combination leverages the strengths of both modules, leading to superior performance across all datasets. The Pipe dataset, which contains both high-frequency details and multi-scale features, benefits the most from this integrated model, as reflected by its lowest MSE value.

In summary, the ablation studies show that both the U-Net and MSPLCK modules contribute significantly to the model's overall performance. The U-Net enhances the ability to capture high-frequency details, while the MSPLCK improves the handling of multi-scale features. Their integration results in a robust model that effectively addresses the challenges presented by the Darcy, Airfoil, and Pipe datasets.

## V. DISCUSSION

The experimental results presented in Table 2 demonstrate that ScaleNet-NO achieves consistent improvements over the U-NO across all benchmarks; however, the magnitude of enhancement varies significantly between fluid and solid mechanics problems. Specifically, ScaleNet-NO exhibits

substantial performance gains on fluid dynamics tasks—including Darcy flow (47.7% reduction in MSE), Airfoil aerodynamics (34.6% reduction), and Pipe flow (59.0% reduction). While showing relatively modest improvements on solid mechanics benchmarks such as Elasticity-P (23.6% reduction), Elasticity-G (19.1% reduction), and Plasticity (32.3% reduction). This discrepancy can be attributed to the fundamental differences in the multi-scale characteristics and frequency-domain properties inherent to these two classes of physical systems, as well as the specific architectural innovations introduced in ScaleNet-NO.

### **A. MULTI-SCALE CHARACTERISTICS IN FLUID VS. SOLID MECHANICS**

Fluid mechanics problems, particularly those involving flow through porous media (Darcy), aerodynamic simulations (Airfoil), and pipeline dynamics (Pipe), are characterized by complex multi-scale physical phenomena that span a broad spectrum of spatial frequencies [1, 19]. The Darcy flow problem involves highly heterogeneous permeability fields with sharp gradients at material interfaces, requiring the model to simultaneously capture large-scale flow structures and fine-scale variations in the velocity field [5, 18]. Similarly, the Airfoil and Pipe datasets involve boundary layer phenomena, flow separation, and turbulent features that manifest as high-frequency components in the solution space [18, 20]. These characteristics align precisely with the design objectives of the Multi-Scale Parallel Large Convolution Kernel (MSPLCK) module, which employs parallel dilated convolutions with effective kernel sizes of 7, 13, and 19 to capture both global structures and localized high-frequency details [21]. In contrast, solid mechanics problems such as Elasticity and Plasticity typically exhibit solution fields with greater spatial regularity and smoother transitions [2, 19]. While these problems may involve complex material behaviors including anisotropy and plastic deformation, the resulting displacement and stress fields generally possess lower spatial frequency content compared to turbulent fluid flows [25]. The Elasticity-P and Elasticity-G benchmarks involve linear elastic deformation where the solution fields are governed by elliptic partial differential equations with inherent smoothing properties, limiting the presence of high-frequency details that would benefit from the enhanced multi-scale receptive fields provided by MSPLCK [1, 27]. Consequently, the architectural advantages of ScaleNet-NO in capturing fine-scale features are less pronounced in these solid mechanics contexts, where the baseline U-NO architecture already provides sufficient representational capacity for the dominant low-to-mid-frequency components [18].

### **B. HIGH-FREQUENCY REPRESENTATION AND SPECTRAL PROPERTIES**

The additional U-Net path incorporated in ScaleNet-NO, which applies local convolutions prior to the Fourier transform to enhance high-frequency information representation, proves particularly beneficial for fluid dynamics problems

[22, 23]. Fluid flows governed by the Navier-Stokes equations or Darcy's law in heterogeneous media often exhibit sharp gradients, discontinuities at phase boundaries, and turbulent energy cascades that require accurate modeling of high-frequency modes [5, 18]. The U-Net architecture's skip connections and local convolutional operations effectively preserve these high-frequency details that might be attenuated by the downsampling operations in the standard U-NO encoder-decoder structure [22, 23]. This explains the substantial error reduction observed in the Pipe and Airfoil experiments, where boundary layer resolution and local flow structures are critical for accurate predictions [18, 28]. Conversely, solid mechanics problems such as plasticity and elasticity typically involve solutions with higher regularity and weaker high-frequency components [25, 27]. The displacement fields in linear elasticity and the stress distributions in plastic deformation problems generally exhibit continuous gradients with limited high-frequency content, as the governing equations inherently penalize rapid spatial variations through energy minimization principles [25]. Therefore, while the U-Net path in ScaleNet-NO still provides benefits in capturing local deformation patterns, the relative improvement over U-NO is diminished because the baseline model's Fourier-based spectral representation already adequately captures the smooth, low-frequency dominant characteristics of solid mechanics solutions [5, 18].

### **C. PHYSICS-INFORMED ARCHITECTURAL ALIGNMENT**

The superior performance of ScaleNet-NO on fluid problems also reflects the alignment between the MSPLCK module's parallel multi-scale processing and the physical nature of fluid transport phenomena [19, 21]. Porous media fluid flow (Darcy) and aerodynamic applications involve multi-scale interactions—ranging from macroscopic pressure gradients to microscopic pore-scale velocity variations. These interactions are effectively captured by the parallel convolution kernels, which simultaneously process information from different receptive fields [18, 21]. The MSPLCK module's ability to function similarly to self-attention mechanisms in transformers, focusing on global regions through large dilated convolutions while maintaining fine-grained local details, is particularly advantageous for capturing the non-local interactions prevalent in fluid dynamics [21]. In solid mechanics, however, the deformation is often governed by local constitutive relations and stress-strain relationships that do not require the same extent of long-range dependency modeling as fluid advection or diffusion processes [25, 27]. While plasticity involves material nonlinearity and history-dependent behavior, the spatial correlations in plastic strain fields are typically shorter-ranged and more localized compared to the long-range correlations observed in fluid velocity fields [2, 25]. Thus, the enhanced receptive fields and multi-scale processing capabilities of ScaleNet-NO provide diminishing returns in solid mechanics applications where the baseline U-NO's hierarchical encoder-decoder structure already suffices to capture the dominant spatial correlations

[18, 22].

### D. GENERALIZATION AND FREQUENCY BIAS

The experimental observations further corroborate the theoretical understanding that Fourier Neural Operators (FNOs) and their U-shaped variants exhibit inherent frequency biases that affect their performance across different physical domains [5, 19]. The Fourier transform-based operations in both U-NO and ScaleNet-NO naturally excel at representing smooth, low-frequency solutions, which explains the strong baseline performance on solid mechanics problems [5]. However, the addition of the MSPLCK and U-Net components in ScaleNet-NO specifically addresses the high-frequency limitations of pure spectral methods, thereby providing greater relative improvements for fluid problems characterized by broadband spectral content [19, 23]. This frequency-domain perspective also explains why the performance gap between ScaleNet-NO and U-NO is most pronounced in the Pipe and Darcy experiments, where the permeability fields and flow geometries introduce high-frequency heterogeneities that challenge purely Fourier-based representations [18, 28]. The MSPLCK module's convolutional processing in the spatial domain complements the Fourier operations by explicitly modeling local discontinuities and sharp gradients that are inefficiently represented in the spectral domain alone [21]. In summary, the differential performance improvements observed across fluid and solid mechanics benchmarks highlight the importance of physics-aware architectural design in neural operators. ScaleNet-NO's enhancements—specifically the MSPLCK module for multi-scale feature extraction and the U-Net path for high-frequency preservation—provide substantial benefits for fluid dynamics problems characterized by multi-scale heterogeneity and high-frequency spatial structures, while offering more modest but still significant improvements for solid mechanics problems with smoother, lower-frequency solution characteristics. These results suggest that the optimal neural operator architecture should be informed by the specific spectral and multi-scale properties of the target physical system, with ScaleNet-NO representing a versatile framework that effectively balances local and global feature extraction across diverse PDE domains.

### VI. CONCLUSIONS

In this paper, we proposed ScaleNet-NO, an enhanced U-shaped neural operator designed to overcome the limitations of traditional U-NOs in capturing high-frequency information and efficiently handling multi-scale PDEs. By integrating two key components the MSPLCK module and an additional U-Net path ScaleNet-NO achieves superior performance in solving complex PDEs. The MSPLCK module captures both global structures and fine details using parallel large convolution kernels of varying sizes. The U-Net path enhances the model's ability to process high-frequency information through local convolutions. Through extensive experiments on seven diverse benchmarks, ScaleNet-NO

consistently achieves state-of-the-art results, with an average relative error reduction of 35.8% compared to baseline U-NO. Notably, it excels in tasks involving complex physical phenomena, such as fluid flow in porous media and aerodynamic simulations, where accurately modeling intricate input-output relationships is critical. Ablation studies further validate the contributions of both the MSPLCK and U-Net modules, highlighting their synergistic effect in addressing the challenges posed by multi-scale and high-frequency PDEs. These results demonstrate that ScaleNet-NO not only improves upon traditional U-NO architectures but also advances the field of neural operator-based scientific computing by providing a more versatile and accurate solution. Despite its strengths, ScaleNet-NO has certain limitations. The increased architectural complexity from parallel kernels and dual pathways leads to higher computational overhead compared to simpler operators like FNO or vanilla U-NO. Additionally, while ScaleNet-NO consistently outperforms baselines across all tested domains, the performance gains in solid mechanics are less pronounced than in fluid dynamics, reflecting the architecture's particular suitability for problems with broadband spectral content and sharp gradient discontinuities. Future work will focus on enhancing computational efficiency through kernel factorization or adaptive sparse designs, making the model more suitable for large-scale and real-time simulations; extending ScaleNet-NO to more challenging PDE systems, such as coupled multi-physics problems and high-dimensional stochastic PDEs; exploring adaptive mechanisms for kernel size selection and U-Net depth, which could further improve generalization across varying problem scales. In summary, ScaleNet-NO not only advances beyond traditional U-NO architectures but also provides a more versatile and accurate neural operator framework, contributing to the growing toolkit for scientific machine learning. With continued refinement, it holds promise for enabling more reliable and efficient simulations of complex systems across engineering and physical sciences.

### AUTHOR CONTRIBUTIONS

Conceptualization – Ziang Chen, Peizhi Zhao and Xuebin Hu; methodology – Ziang Chen, Peizhi Zhao and Xuebin Hu; investigation – Ziang Chen, Peizhi Zhao and Xuebin Hu; writing-original draft preparation – Ziang Chen, Peizhi Zhao and Xuebin Hu. All authors have read and agreed to the published version of the manuscript.

### CONFLICTS OF INTEREST

The authors declare no conflict of interest.

### FUNDING

This work was supported by the National Key Research and Development Project of China (2021YFA1000103, 2021YFA1000102), the National Natural Science Foundation of China (Grant Nos. 62272479, 62372469, 62202498), the Taishan Scholarship (TSTP20240506),

and the Shandong Provincial Natural Science Foundation (ZR2021QF023).

## ACKNOWLEDGEMENTS

Not applicable.

## REFERENCES

- [1] N. Kovachki et al., "Neural operator: Learning maps between function spaces with applications to PDEs," *Journal of Machine Learning Research*, vol. 24, no. 89, pp. 1–97, 2023.
- [2] S. Goswami, M. Yin, Y. Yu, and G. E. Karniadakis, "A physics-informed variational DeepONet for predicting crack path in quasi-brittle materials," *Computer Methods in Applied Mechanics and Engineering*, vol. 391, Art. no. 114587, 2022.
- [3] T. Kurth et al., "Fourcastnet: Accelerating global high-resolution weather forecasting using adaptive Fourier neural operators," in *Proceedings of the Platform for Advanced Scientific Computing Conference*, 2023, pp. 1–11.
- [4] L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis, "Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators," *Nature Machine Intelligence*, vol. 3, no. 3, pp. 218–229, 2021.
- [5] Z. Li et al., "Fourier neural operator for parametric partial differential equations," 2020.
- [6] T. Tripura and S. Chakraborty, "Wavelet neural operator: A neural operator for parametric partial differential equations," 2022.
- [7] Q. Cao, S. Goswami, and G. E. Karniadakis, "Laplace neural operator for solving differential equations," *Nature Machine Intelligence*, vol. 6, no. 6, pp. 631–640, 2024.
- [8] J. W. Cooley, P. A. Lewis, and P. D. Welch, "The fast Fourier transform and its applications," *IEEE Transactions on Education*, vol. 12, no. 1, pp. 27–34, 2007.
- [9] M. Abdar et al., "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, pp. 243–297, 2021.
- [10] M. Ramezankhani, A. Deodhar, R. Y. Parekh, and D. Birru, "An advanced physics-informed neural operator for comprehensive design optimization of highly-nonlinear systems: An aerospace composites processing case study," *Expert Systems with Applications*, vol. 142, Art. no. 109886, 2025.
- [11] X. Yu et al., "SepONet: Efficient large-scale physics-informed operator learning," in *NeurIPS 2024 Workshop on Data-driven and Differentiable Simulations, Surrogates, and Solvers*, 2024.
- [12] P. Xiao, M. Zheng, A. Jiao, X. Yang, and L. Lu, "Quantum DeepONet: Neural operators accelerated by quantum computing," vol. 9, Art. no. 1761, 2025.
- [13] M. del Carmen Calderon-Moreno, "Universality of derivative and antiderivative operators with holomorphic coefficients," *Annales Polonici Mathematici*, vol. 77, pp. 197–207, 2001.
- [14] S. Karumuri, L. Graham-Brady, and S. Goswami, "Physics-informed latent neural operator for real-time predictions of time-dependent parametric PDEs," 2025.
- [15] J. P. Cunningham and Z. Ghahramani, "Linear dimensionality reduction: Survey, insights, and generalizations," *Journal of Machine Learning Research*, vol. 16, no. 1, pp. 2859–2900, 2015.
- [16] D. Zhang, "Wavelet transform," in *Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval*. Springer, 2019, pp. 35–44.
- [17] T. Tripura and S. Chakraborty, "Wavelet neural operator for solving parametric partial differential equations in computational mechanics problems," *Computer Methods in Applied Mechanics and Engineering*, vol. 404, Art. no. 115783, 2023.
- [18] M. A. Rahman, Z. E. Ross, and K. Azizadenesheli, "U-no: U-shaped neural operators," 2022.
- [19] S. Wang, H. Wang, and P. Perdikaris, "On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks," *Computer Methods in Applied Mechanics and Engineering*, vol. 384, Art. no. 113938, 2021.
- [20] J. T. Oden, S. Prudhomme, A. Romkes, and P. T. Bauman, "Multiscale modeling of physical phenomena: Adaptive control of models," *SIAM Journal on Scientific Computing*, vol. 28, no. 6, pp. 2359–2389, 2006.
- [21] L. Lu, Q. Xiong, B. Xu, and D. Chu, "Mixdehazenet: Mix structure block for image dehazing network," in *2024 International Joint Conference on Neural Networks (IJCNN)*, 2024, pp. 1–10.
- [22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.
- [23] H. Wang, Q. Wang, M. Gao, P. Li, and W. Zuo, "Multi-scale location-aware kernel representation for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1248–1257.
- [24] F. E. Browder, "Approximation by solutions of partial differential equations," *American Journal of Mathematics*, vol. 84, no. 1, pp. 134–160, 1962.
- [25] Z. Li et al., "Neural operator: Graph kernel network for partial differential equations," 2020.
- [26] T. Xu, D. Liu, P. Hao, and B. Wang, "Variational operator learning: A unified paradigm marrying training neural operators and solving partial differential equations," *Journal of the Mechanics and Physics of Solids*, vol. 190, Art. no. 105714, 2024.
- [27] T. Zhang and X. Zhu, "A solution method for partial differential equations by fusing Fourier operators and U-Net," vol. 13, no. 7, Art. no. 1033, 2025.
- [28] W. Liu, M. Yagoubi, and M. Schoenauer, "Multi-resolution graph neural networks for PDE approximation," in *International Conference on Artificial Neural Networks*, 2021, pp. 151–163.
- [29] R. J. Gladstone, H. Rahmani, V. Suryakumar, H. Meidani, M. D'Elia, and A. Zareei, "GNN-based physics solver for time-independent PDEs," 2023.
- [30] Y. Hu, T. Zhao, S. Xu, Z. Xu, and L. Lin, "Neural-PDE: A RNN based neural network for solving time dependent PDEs," 2020.
- [31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.
- [32] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016.