VSMNO: Solving PDE by Utilizing Spectral Patterns of Different Neural Operators

Anonymous Author(s)
Affiliation
Address
email

Abstract

Seeking effective numerical approximations for partial differential equations (PDEs) is a major challenge in modern science and technology. Recently, AI-inspired data-driven solvers, such as neural operators, have achieved great success in quickly PDE solving. However, in the design of neural operators, the processing of frequency domain information is crucial, and a single processing method is difficult to comprehensively handle frequency domain information of different components. We present the V-shaped spectral mixture neural operator (VSMNO) architecture which combines spectral learning modes of different neural operators to process frequency domain information in PDE solving at various levels. For general PDE solving, we propose a residual learning structure that can transfer residuals in V-Cycle by combining frequency domain learning patterns of different neural operators to reduce high-frequency and low-frequency error. For differences in PDEs, we propose a neural operator correction strategy based on the correspondence between the PDE spectrum distribution and the neural operator spectral pattern, to correct the results by utilizing the prior knowledge of the PDE system. Experimentally, VSMNO achieves state-of-the-art and yields a relative error reduction of 22% averaged on four classical benchmarks.

1 Introduction

Many of the most fundamental laws of nature can be formulated as partial differential equations (PDEs). Nevertheless, the analytical solutions needed to comprehend these pivotal PDEs in modern science are frequently unknown, rendering the pursuit of effective approximate solutions through numerical methods one of the major challenges confronting humanity(1). However, traditional methods for solving PDE still suffer from high computational costs and expenses. Recently, with the emergence of neural operator methods(3; 4; 5), a quick solver to PDEs has been proposed by learning the complex mapping relationship between input and output in function space with AI operators. Following the research line, lots of basic works have been reported, such as DeepONet(3) with trunk network and branch network to learn the PDE initial conditions. Fourier neural operator (FNO) approximates integral operators in the frequency domain through the convolution theorem(4). Transformer models(5; 6; 7), based on attention mechanisms, are reported to capture global information of PDEs.

Building on these foundational architectures, considerable effort has been devoted to enhancing the performance of neural operators in PDE solving, particularly by improving the handling of spectral information. For example, extending FNO based methods(8; 9; 10), considering basis functions(11; 13) or aliasing phenomena(12) in the frequency domain, or combining multi-scale and spectral tools(13; 14; 15). Although these works have realized the use of frequency domain

Submitted to Foundation Models for Science Workshop, 38th Conference on Neural Information Processing Systems (NeurIPS 2024).

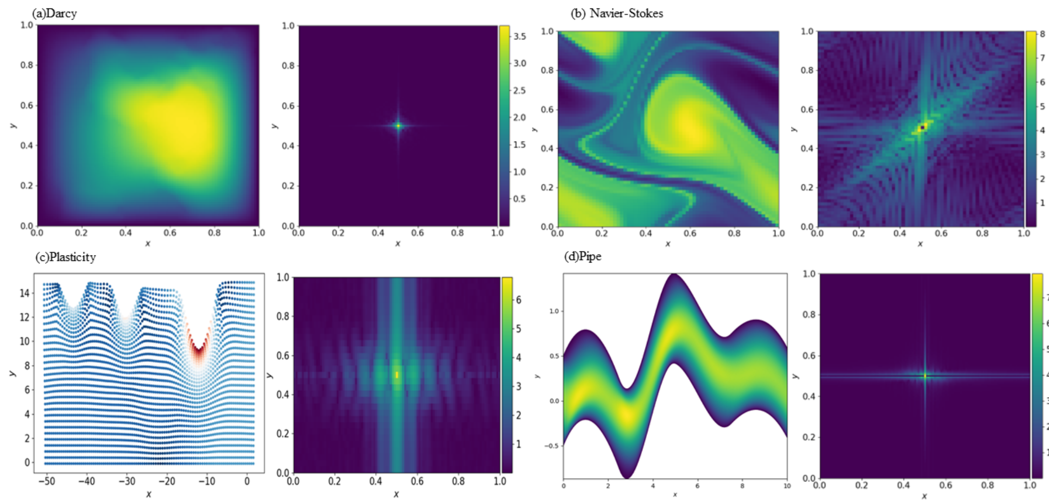

Figure 1: The figure shows the solutions and spectral distributions of different PDEs, with the low-frequency region in the middle. We can observe that the spectral distribution patterns of PDEs are both similar and different. The spectra of different PDEs are often dominated by low-frequency energy, but there are often significant differences in spectral distribution patterns and high-frequency components.

information for PDE solving, they often only focus on one aspect of spectral information, ignoring the commonalities and differences in the PDE solving process, and failing to consider the different spectral learning modes of different neural operators which leads to a lack of flexibility in the design of neural operators.

To address this challenge, we propose the VSMNO architecture. Based on the commonality of PDE spectral distribution, a residual transfer operator was designed by combining the spectral learning modes of FNOs and CNN to correctly transfer residuals in a multi-scale structure. This ensures the correct transfer of errors at different scales while improving high-frequency and low-frequency learning capabilities.

After that, we summarized the learning patterns of FNO type methods. In this process, we observed the certain similarity between the spectral distribution of solution to PDE and the spectral pattern of neural operators. And we propose the neural operator correction strategy with spectrum analysis to design the correction module to correct the output. Experimentally, VSMNO achieves state-of-the-art and yields a relative error reduction of 22% averaged on four classical benchmarks. Our contributions are summarized as follows.

- We designed VSMNO by combining spectral learning patterns of different neural operators and proposed a residual learning structure that can transfer residuals in V-cycles to comprehensively process the frequency domain information of PDE.
- Considering the correspondence between spectral learning pattern of neural operator and spectral distribution of PDE, we propose a novel neural operator correction strategy based on the spectrum analysis.
- As results, we achieved the state-of-the-art (SOTA) and an 22% average improvement in computational accuracy on four PDE benchmarks, And the effectiveness of our method was demonstrated in the ablation experiment.

2 Preliminaries

2.1 Neural operators

Neural operator is a type of data-driven models used to approximate the mapping from parameter functions to PDEs solutions (27). By learning operators in the function space, it is found that mapping relationship between input and output in finite dimensions, and thus have the ability to generalize a class of PDE. Various neural operator methods have been proposed for solving PDE. Among them, FNO(4) and DeepONet(3) are the most famous works. DeepONet is proposed by using infinite dimensional operator mapping learning based on the general approximation theorem. FNO approximates the integration operator by directly defining a kernel function in the Fourier space.

Building on these basic models, extensive research has focused on employing various frequency domain processing methods to enhance the accuracy of neural operators in solving PDEs. In 2021, MWT(13) introduces multi-wavelet basis operators to deal with the coupled mappings between the functions. SNO(11) solves PDEs by learning basis functions and the mapping between coefficients. In 2022, FFNO(9) proposes to decompose high-dimensional Fourier coefficients into multiple independent one-dimensional Fourier coefficients and it overcomes the problem of difficult convergence of deep FNO. U-FNO(28) and U-NO(14) combine the U-net architecture with FNO to process physical information in multi-scale space. MG-TFNO(10) improves model performance by combining global tensor decomposition and multigrid domain decomposition strategies. CNO(12) handles aliasing error by combining sampling theorem. Latent Spectral Models (LSM)(15) overcomes the limitations of redundant coordinate spaces in multi-scale hidden space inputs and achieves (state-of-the-art) SOTA performance on multiple datasets by combining neural spectral block.

Although these methods have achieved impressive results, due to the commonalities and differences in PDE spectral distributions (see 1), using a single frequency domain processing method is actually difficult to handle the differences in frequency domain components of different PDEs. While designing a universal PDE solver, in order to handle the spectral distribution differences of different PDEs more flexibly. Unlike the above methods, we present the VSMNO architecture by summarizing the spectral learning patterns of different neural operator methods. This architecture can comprehensively process different frequency domain information and adapt to different PDE spectral distribution patterns, thereby further improving the accuracy of neural operator solutions for PDE.

3 Method

3.1 Problem settings

We consider PDEs in a bounded open set $\mathcal{D} \subset \mathbb{R}^d$, both inputs and outputs can be rewritten as functions w.r.t. coordinates, which are in the Banach spaces $\mathcal{X}(\mathcal{D}, \mathbb{R}^{d_x})$ and $\mathcal{Y} = \mathcal{Y}(\mathcal{D}, R^{d_y})$ respectively(3; 4). $\mathbb{R}^{d_x}$ and $\mathbb{R}^{d_y}$ are the range of input and output functions. The solving process is to approximate the optimal operator $\mathcal{G} : X \to Y$ with deep model $\mathcal{G}_\theta$ to minimize the relative mean squared error loss between the prediction and data.

3.2 Overview of model architecture

To design a universal solver for neural operators based on frequency domain methods, we first need to be able to comprehensively process frequency domain information of different components. In frequency domain analysis, the frequency components of a signal are decomposed into different frequency components, each with its own specific amplitude and phase. It can be divided into high frequency and low frequency according to the size of the frequency components. In PDE solving, the low-frequency component reflects the large-scale structure of the physical system, while the high-frequency component represents more details in the physical system. In order to further utilize frequency domain information to enhance the accuracy of neural operators, we propose the VSMNO architecture, which combines the spectral patterns of different neural operators to improve low-frequency and high-frequency learning capabilities.

Here, we provide an overview of our VSMNO model, as shown in Figure 2. we could divide it into three parts, including preprocess module, cycle module, and correction module.

Preprocess module. The preprocess part includes the post-processing part of the entire preprocess operator and correction operator. The module attempt to convert input on irregular regions into input on regular regions, which can be handled by neural operators. We follow the same settings as geo-FNO(8). Among them, preprocess module could convert irregular geometric inputs into regular geometric shapes and perform mapping from physical space to computational space. And then conversion between latent space and computational space.

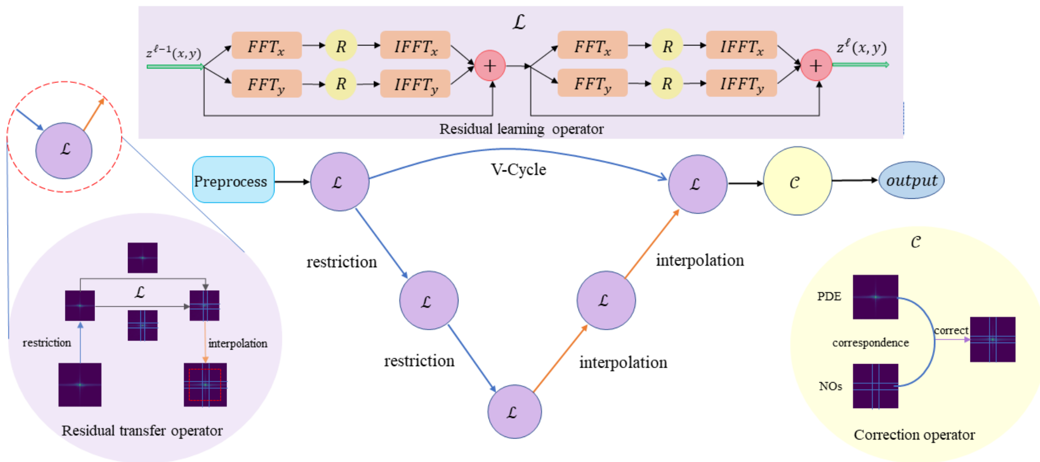

Figure 2: Overview of the model architecture. After data preprocess, we will fold and expand the low-frequency region, and use a multi residual transfer operator structure to comprehensively process the frequency domain information in V-cycle. The residual learning operator handles low-frequency error (blue line), while interpolation reconstructs high-frequency information (red line). Finally, we use correct module which designed based on a correction strategy to correct the results. The images of the correct module represent the spectral distribution pattern of the PDE data and its errors in the spectral solution process. NOs represent the spectral learning pattern of neural operators.

Cycle module. As shown in Figure 2, our cycle module is a V-shaped multi-scale structure composed of multiple residual transfer operators. In existing literature(29) on multigrid operators, we found that with the same depth, the performance differences between various cycle structures are minimal. Therefore, we chose the V-cycle architecture.

In the cycle module, the residual transfer operator expands the low-frequency learning range of the neural network through frequency domain folding. Residual transfer operators of different scales can eliminate the incoming residuals at different scales and transfer low-frequency error, while gradually reconstructing the high-frequency details of the PDE solution in the frequency domain space, completing the solution of the corresponding PDE. The residual transfer operator consists of restriction operator, residual learning operator, and interpolation operator.

Restriction Operator. Restriction operator could transfer the input information from the high scale to the low scale space, retaining key features between residual transfer operators of different scales. Meanwhile, the restriction operator can also be seen as a low-pass filter(25), which retaining the low-frequency information and enhancing the low-frequency features in the solution area. The restriction operator not only reduces computational overhead by reducing the size of the solution region, but also enhances the low-frequency learning ability of neural operators in the low-frequency region. We use average pooling with multiple convolution blocks to implement the restriction Operator.

Residual learning operator. In order to process low-frequency information in PDE and reduce low-frequency error in PDE solving. We have designed a residual learning operator in the residual transfer operator. Residual learning operator can process physical information of different scales to solve and eliminate low-frequency error in the corresponding solution region. In order to enable the residual learning operator to correctly reduce low-frequency error, we first explored the mathematical form of error e could be written in the following form:

$$e = \mathcal{Y} - \mathcal{G}_\theta\left(\mathcal{X}\right) \tag{1}$$

The correct solution $\mathcal{Y}$ of a given PDE is usually unknown. $\mathcal{G}_\theta$ is the solution operator obtained during the training process through neural operators, where theta is the parameter. $\mathcal{G}_\theta\left(\mathcal{X}\right)$ represents the PDE solution obtained by the neural operator when the input is $\mathcal{X}$. As a remedy, we consider the residual. If we record the output of the preprocess as $\mathcal{X}$ and assume that the result $H(\mathcal{X})$ is the correct mapping we want to learn, we can obtain the following form of residual r as,

$$r = H\left(\mathcal{X}\right) - \mathcal{X} \tag{2}$$

When r approaches 0, e is also the same. So operators can learn the correct mapping by correctly learning residuals. For this, we design a residual learning operators in a form with residual connections(19). Besides, we design a structure without skip connections within the same level layer to ensure the above structure could correctly learn residuals and to avoid the problem of mismatched features between different layers(24).

Furthermore, Although neural networks tend to learn low-frequency information during training(21; 22). But in order to better handle important low-frequency error, we use the FNOs model specifically for processing low-frequency error. The Fourier block of the FNO type model can generally be written in the following form,

$$(R_\theta \bullet f\,(a))\,(\omega) = \begin{cases} (R_\theta \bullet f(a))(\omega) & ,\Omega(\omega_x,\omega_y) \le T_\omega \\ 0 & ,other \end{cases} \tag{3}$$

Among them, Ω is the range control function, Among θ is a set of the network parameters and $\mathcal{F}$ is a fast Fourier transformation. If we set T_ω as a cutoff frequency. In FNOs method, it truncates the low-frequency components ($\omega{<}T_\omega$). And we will discuss it in 3.3. From the above equation, it can be seen that the FNOs model provides us with a powerful tool for addressing low-frequency errors. We have chosen the FNOs model to implement our residual learning operator, thereby better reducing low-frequency error in PDE solving.

We chose FFNO as the foundation for the residual learning operator module because, as shown in formula 6, FFNO has the largest frequency domain learning range. To better capture frequency domain features, we used FFNO as the basic framework for the residual module, incorporating shared parameters to reduce the parameter count and maintain consistent feature extraction between layers. Additionally, to ensure solution accuracy, we employed a multi-layer structure to enhance the model's expressive power.

Interpolation operator. The function of interpolation operator and restriction operator operator should be similar, that is, to transfer the residuals to residual learning operators from low scale to high scale. However, in the design of the residual learning operator, we mentioned that in order to avoid inconsistent features in inter layer learning, we cancelled the skip connection which has led to the problem of excessive smoothing of high-frequency information by the restriction operator. In addition, our residual learning operator is also limited on learning in low frequency during this process and cannot handle high-frequency information. And these problems require us to transmit high-frequency information during the error propagation process in the cycle module. So we further designed interpolation operators based on the frequency domain learning pattern of CNN.

The frequency domain patterns of CNN have some similarities with FNO(20), but CNN can capture high-frequency regions in the frequency domain, and there have been considerable works utilizing this feature to achieve super-resolution tasks(26). The pattern of CNN enables us to transmit and reconstruct high-frequency information and features in the frequency domain during the error propagation process. Based on this, we designed an interpolation operator consisting of multiple convolution blocks and deconvolution, which reduces computational overhead, preserves the low-frequency information calculated by the residual neural operator, and gradually reconstructs the high-frequency details of the PDE solution. Improving the low-frequency and high-frequency learning ability of neural operators enhances the accuracy and performance of VSMNO in solving PDE.

Correction module. The correction module includes a correction operator which composed of several residual learning operators with different model which selected based on our operator correction strategy based on spectrum analysis, and a post-processing part. The correction operator outputs the reconstructed high and low frequency information for final correction processing. And through post-processing, the output of the latent space is mapped back to the computational space and the original physical space to obtain the final result we expect. The post-processing part is also the same as the settings in geo-FNO(8). And we will now discuss our correction strategy in 3.3.

3.3 Operator correction strategy based on spectrum analysis

In order to flexibly handle PDEs with different spectral distribution patterns, we further explore the specific forms of spectral learning patterns of different neural operators based on the previous section, and construct our correction module's correction strategy according to the correspondence between PDE spectral distribution patterns and neural operator spectral learning patterns.

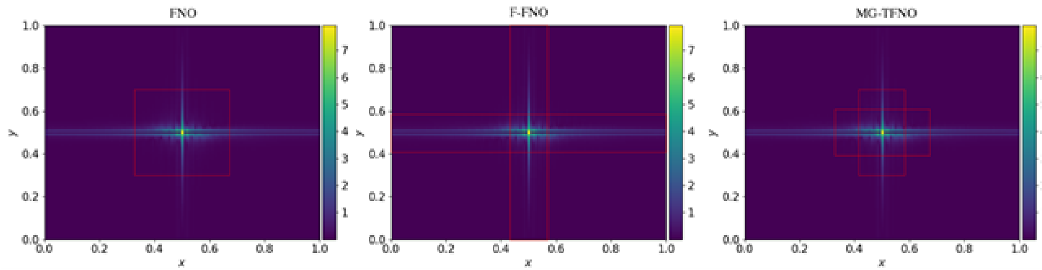

Figure 3: The red rectangles in figure show low-frequency learning pattern of two-dimensional FNO, FFNO, and MG-TFNO. The figures are plotted in the logarithmic spectrogram from the dataset of the Pipe task. The middle of the spectrum is the low-frequency region.

Here we begin to discuss the low-frequency learning mode of neural operators, using FNO as an example. Its integration kernel can be defined as,

$$\left(K\left(\theta\right)v_t\right)(x) = \mathcal{F}^{-1}\left(R_\theta \mathcal{F}\left(v_t\right)\right)(x) \tag{4}$$

From formula 3, when the model is FNO, $\Omega(x)$ is taken as $max(x)$. the integral kernel of two-dimensional FNO can also be expressed as,

$$\left(R_\theta \bullet f\left(a\right)\right)(\omega) = \begin{cases} (R_\theta \bullet f(a))(\omega) & , max(\omega_x, \omega_y) \leq T_\omega \\ 0 & , other \end{cases} \tag{5}$$

We can note that without considering the mlp component of FNO, the frequency domain learning pattern of FNO is equivalent to passing the original region through an ideal low-pass filter and then reconstructing the target PDE solution in the corresponding region. Its frequency domain learnable range f is equal to cutoff frequency T_ω.

From formula 5, we can see that one basic limitation of the FNO method in solving PDE is its frequency domain learnable range, and the different forms of this frequency domain learning range also determine the frequency domain learning modes of different operators. Next, we will summarize the frequency domain learning modes of other FNO type methods.

FFNO factorizes high-dimensional Fourier transform into multiple one-dimensional Fourier transform. Its $\Omega(x)$ is taken as $min(x)$ and its integral kernel can also be written as,

$$\left(R_\theta \bullet f\left(a\right)\right)(\omega) = \begin{cases} (R_\theta \bullet f(a))(\omega) & , min(\omega_x, \omega_y) \leq T_\omega \\ 0 & , other \end{cases} \tag{6}$$

MG-TFNO uses low-rank decomposition to process the parameter matrix of FNO. the $\Omega(x)$ of MG-TFNO increases control over the rank of the learnable matrix. We formalize it according to the above process and directly provide the approximate expression of the MG-TFNO integral kernel as,

$$(R_\theta \bullet f(a))(\omega) = \begin{cases} \left(R_\theta \bullet f\left(a\right)\right)(\omega) & , \ max(\omega_x, \omega_y) \leq T_\omega, rank(R_\theta) \leq T_\omega \\ 0 & , other \end{cases} \tag{7}$$

We plot the weight representation regions of these models on the spectrum to represent the differences among their spectral learning patterns (see Figure 3). Compared to the FNO method, CNN captures more structural features corresponding to high-frequency regions as depth increases(26). Similar patterns could exist in some PDEs, taking the heat conduction equation of an infinitely long rod as an example,

$$f(x) = \begin{cases} \frac{\partial u}{\partial x} = a^2 \frac{\partial^2 u}{\partial x^2} & , (-\infty < x < +\infty, t > 0) \\ u|_{t=0} = \Phi(x) & , (-\infty < x < +\infty) \end{cases} \tag{8}$$

the $f(x)$ is heat conduction equation. The coefficient 'a' is the thermal diffusion coefficient, and using the integral transformation method, the solution of the equation can be written as,

$$u\left(x,t\right) = \Phi\left(x\right) * \left(\frac{1}{2a\sqrt{\pi t}} \exp\left(-\frac{x^2}{4a^2 t}\right)\right) \tag{9}$$

Its Green's function is a Gaussian function which the energy almost concentrating in the central the frequency domain, which have similar shape with spectral domain learnable regions of FNO, as time varies. The above examples indicate that spectral distribution patterns of solution to some PDE could correspond to certain neural operators learning patterns. In addition, we observed in the experiment that although the equation solutions corresponding to different initial conditions differ significantly, their spectral distribution patterns are quiet similar, which could be seem as knowledge of the corresponding PDE system, implying that we could leverage this correspondence to enhance the neural operator.

For example, the spectral energy of the Pipe dataset is mainly distributed in the central cross region of spectrum, similar to the spectral learning pattern of FFNO. Therefore, we use single-layer FFNO as the correction operator to solve on the Pipe dataset. This flexible correction strategy allows us to process PDEs with different spectral distribution patterns.

Our methods guides the extraction of frequency domain features of PDE and spectral learning patterns of neural operators through prior knowledge of the physical system, enabling neural operators to more effectively focus on the truly critical frequency domain information in the problem. This method can be seen as a feature engineering based on prior knowledge of PDE frequency domain, using known physical information to assist neural operators in quickly extracting intrinsic laws and structural features in complex PDE tasks. The introduction of prior knowledge in physics constrains the mapping learning process of neural operators on solutions, and avoids overfitting problems based on correct and reasonable physical constraints. Thus, it accelerates model convergence and improves accuracy without increasing model complexity, providing a concise and reliable solution for building flexible universal PDE solvers.

4 Experiment

4.1 Benchmark

The benchmarks we use four classic tasks include Darcy, NS-equation, Pipe and Plasticity datasets. These benchmark are generated by different PDEs for different tasks. Specifically, the datasets we used comes from FNO datasets and geo-FNO datasets(8). Among them, Darcy and NS are fluid tasks for 2D regular grid input. Pipe is a fluid task with 2D irregular grid input, and Plasticity is a solid task with 3D irregular input. We show its details in table 1.

4.2 Baseline

We compared the VSMNO with 5 neural network models on 4 task benchmarks, including the classic FNO methods and its variant such as F-FNO and MWT. LSM, U-NO architecture based on multi-scale methods. Among them, LSM is the state-of-the-art (SOTA) model in this field before.

4.3 Hyperparameter

All of our calculations are performed on an Nvidia RTX4090 GPU. For fairness, we use relative l2 error as the training loss and evaluation metric. We use 500 epochs and an ADAM(23) optimizer with an initial learning rate of 10-3. Set the batch size to 20. We set the mode to 12, the width to 64, and the convolution kernel size in both the interpolation and constraint operators to 3. And We present the selection of correction strategies and the spectral distribution patterns of different datasets in Table 1.

Table 1: Details for benchmarks and the selection of correction strategies.

ATTRIBUTE/DATASET	DARCY	NAIVER-STOKES	PIPE	PLASTICITY
TYPE	FLUID	FLUID	FLUID	SOLID
MODES	12	12	12	12
MIN MODES	8	8	8	4
DISTRIBUTION	CENTER	COMPLEX	CROSS	CENTER
STRATEGY	FNO	/	FFNO	FNO
KERNAL SIZE	(4,4)	/	(12,12)	(12,12,6)
INPUT	POROUS MEDIUM	PAST VELOCIT	STRUCTURE	BOUNDARY CONDITION
OUTPUT	FLUID PRESSUR	FUTURE VELOCIT	FLUID VELOCITY	MESH DISPLACEMENT

Table 2: Our main results of operator learning on several datasets from multiple PDEs. Related-MSE is recored.

Model\Dataset	Darcy	NS	pipe	Plasticity
FNO	0.0108	0.1556	0.0067	0.0074
MWT	0.0078	0.1541	0.0070	0.0076
UNO	0.0148	0.1713	0.0100	0.0034
F-FNO	0.0077	0.1213	0.0070	0.0047
LSM	0.0065	0.1535	0.0050	0.0025
ours	**0.0051**	**0.0941**	**0.0040**	**0.0019**

Table 3: Ablation experiments on each dataset, We conduct multi removing components (w/o). Related-MSE is recored.

Model\Dataset	Darcy	NS	pipe	Plasticity
W/O correct	6.08E-3	/	4.52E-3	2.34E-3
W/O FNOs	1.52E-2	2.28E-1	6.51E-3	5.12E-3
W/O Conv	8.55E-3	1.48E-1	6.91E-3	3.17E-3
W/O Cycle	6.81E-3	1.21E-1	5.66E-3	3.36E-3
Ours	**5.06E-3**	**0.94E-1**	**4.01E-3**	**1.90E-3**

4.4 Result analysis

The main results are shown in Table 2. Based on these results, we have the following observations. Our proposed VSMNO model shows significant improvement compared to various multi-scale models and FNO variant models on benchmark datasets, with an average improvement of 22% compared to the previous best method. Overall, VSMNO has a parameter count of 0.65M for 2D tasks and 1.2M for 3D tasks, which is only 1/30 and 1/10 of the LSM parameter count of the previous SOTA model under the same hyperparameter settings. While reducing the number of parameters, VSMNO performs on average 26% better than LSM on each benchmark. In the Darcy task with significant low-frequency dominance, VSMNO showed a 21.5% improvement in performance compared to the optimal method, demonstrating the significant advantage of our method in low-frequency learning; In the 2D NS equation task with complex frequency domain distribution, it improved by 38.6% compared to LSM, demonstrating the reliable frequency domain learning ability of VSMNO. In 2D fluid pipe tasks and 3D solid-state plasticity equation tasks, which with irregular grid inputs, VSMNO improved the performance by 20% and 24% compared to LSM, demonstrating that our model can achieve good results in different geometric inputs. These experimental results validate the effectiveness of our proposed approach in improving spectral learning ability and solving PDE accuracy.

4.5 Ablation experiment

To verify the effectiveness of our proposed method, we conducted comprehensive ablation experiments by removing and replacing components (see Table 3). We can obtain the following results from Table3. In the removal experiment, we can find that all components are essential for the final performance. If without correct module, our model's performance will decrease by an average of 17% and use operators with mismatched spectral patterns for correction can result in a loss of accuracy of over 10%. Removing the convolution blocks in interpolation and constraint operators will significantly hinder the transfer of features in multi-scale frameworks, resulting in a decrease of over 30% in accuracy. Cancelling the F-FNO block of the residual learning module will degrade the architecture to U-net, resulting in an accuracy loss of over 50%. The accuracy loss is more pronounced in low-frequency tasks such as Darcy and Plasticity, which also reflects the importance of low-frequency learning for neural operators to solve PDE. By eliminating the Cycle structure and retaining the original F-FNO block, the performance of the Cycle module was reduced by an average of 29.1% and 43.5% on irregular input on plasticity and pipe tasks, respectively, which demonstrating its effectiveness. The above ablation experiments demonstrate that our proposed approach can effectively solve PDE tasks with different scales and frequency domain properties, and effectively improve the interpretability and accuracy of the model.

5 Conclusion

In this paper, in order to improve the frequency domain processing capability of PDE, We presented the VSMNO. We combined the low-frequency learning pattern of FNO type neural operators with the spectral learning ability of CNN to restruct high-frequency features, and designed a residual learning transfer operator structure using the spectral learning patterns of different operators. Comprehensively improved the frequency domain information processing capability of neural operators. Meanwhile, in response to handle the different spectral distribution patterns of PDEs, we have designed a correction strategy based on the spectral learning patterns of neural operators, making it possible to flexibly process different PDEs. Thanks to the fusion of different spectral learning paradigms, VSMNO has achieved consistent state-of-the-art performance in various benchmark tests, while significantly reducing the number of parameters. In the future, we will further explore the spectral patterns of different neural networks and combine them with the different spectral distributions of PDEs to explore more accurate and interpretable neural operator design methods and establish a universal PDE solver using neural operators.

References

[1] Wazwaz, A. M. (2002). Partial differential equations. CRC Press.

[2] Wesseling, P. (1995). Introduction to multigrid methods (No. NASA-CR-195045).

[3] Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nature machine intelligence, 3(3), 218-229.

[4] Li, Z., Kovachki, N. B., Azizzadenesheli, K., liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. Fourier neural operator for parametric partial differential equations. In ICLR, 2021.

[5] Li, Z., Meidani, K., and Farimani, A. B. Transformer for partial differential equations' operator learning. arXiv preprint arXiv:2205.13671, 2022.

[6] Cao, S. Choose a transformer: Fourier or galerkin. Advances in Neural Information Processing Systems, 34:24924–24940, 2021.

[7] Hao, Z., Wang, Z., Su, H., Ying, C., Dong, Y., Liu, S., ... & Zhu, J. (2023, July). Gnot: A general neural operator transformer for operator learning. In International Conference on Machine Learning (pp. 12556-12569). PMLR.

[8] Li, Z., Huang, D. Z., Liu, B., and Anandkumar, A. Fourier neural operator with learned deformations for pdes on general geometries. arXiv preprint arXiv :2207.05209, 2022a.

[9] Tran, A., Mathews, A., Xie, L., & Ong, C. S. (2022, September). Factorized Fourier Neural Operators. In The Eleventh International Conference on Learning Representations.

[10] Kossaifi, J., Kovachki, N. B., Azizzadenesheli, K., & Anandkumar, A. (2022). Multigrid Tensorized Fourier Neural Operator for High Resolution PDEs.

[11] Fanaskov, V. and Oseledets, I. Spectral neural operators. arXiv preprint arXiv:2205.10573, 2022.

[12] Raonic, B., Molinaro, R., Rohner, T., Mishra, S., & de Bezenac, E. (2023, March). Convolutional neural operators. In ICLR 2023 Workshop on Physics for Machine Learning.

[13] Gupta, G., Xiao, X., & Bogdan, P. (2021). Multiwavelet-based operator learning for differential equations. Advances in neural information processing systems, 34, 24048-24062.

[14] Rahman, M. A., Ross, Z. E., & Azizzadenesheli, K. (2023). U-NO: U-shaped Neural Operators. Transactions on Machine Learning Research.

[15] Wu, H., Hu, T., Luo, H., Wang, J., & Long, M. (2023). Solving High-Dimensional PDEs with Latent Spectral Models. arXiv preprint arXiv:2301.12664.

[16] Brandt, A. (1977). Multi-level adaptive solutions to boundary-value problems. Mathematics of computation, 31(138), 333-390.

[17] Takamoto, M., Praditia, T., Leiteritz, R., MacKinlay, D., Alesiani, F., Pflüger, D., & Niepert, M. (2022). PDEBench: An extensive benchmark for scientific machine learning. Advances in Neural Information Processing Systems, 35, 1596-1611.

[18] Zhao, J., George, R. J., Zhang, Y., Li, Z., & Anandkumar, A. (2022). Incremental fourier neural operator. arxiv preprint arxiv:2211.15188.

[19] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[20] Kabri, S., Roith, T., Tenbrinck, D., & Burger, M. (2023, May). Resolution-invariant image classification based on fourier neural operators. In International Conference on Scale Space and Variational Methods in Computer Vision (pp. 236-249). Cham: Springer International Publishing.

[21] Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., ... & Courville, A. (2019, May). On the spectral bias of neural networks. In International conference on machine learning (pp. 5301-5310). PMLR.

[22] Xu, Z. Q. J., Zhang, Y., & Luo, T. (2022). Overview frequency principle/spectral bias in deep learning. arxiv preprint arxiv:2201.07395.

[23] Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In ICLR, 2015.

[24] Zhang, X., Helwig, J., Lin, Y., e, Y., Fu, C., Wojtowytsch, S., & Ji, S. SineNet: Learning Temporal Dynamics in Time-Dependent Partial Differential Equations. In The Twelfth International Conference on Learning Representations.

[25] Deng, H., Xu, Z., Duan, Y., Wu, X., Shu, W., & Deng, L. J. (2024). Exploring the Low-Pass Filtering Behavior in Image Super-Resolution. arxiv preprint arxiv:2405.07919.

[26] Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. IEEE transactions on pattern analysis and machine intelligence, 43(10), 3365-3387.

[27] Viswanath, H., Rahman, M. A., Vyas, A., Shor, A., Medeiros, B., Hernandez, S., ... & Bera, A. (2023). Neural Operator: Is data all you need to model the world? An insight into the impact of Physics Informed Machine Learning. arXiv preprint arXiv:2301.13331.

[28] Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A., and Benson, S. M. U-fno—an enhanced Fourier neural operator-based deep-learning model for multiphase flow.Advances in Water Resources, 163:104180, 2022.

[29] Multi-scale Physical Representations for Approximating PDE Solutions with Graph Neural Operators, Léon Migus, Yuan Yin, Jocelyn Ahmed Mazari, Patrick Gallinari.