Frontiers in Computational Science and Engineering
Vol. 1, No. 2, Apr - Jun 2025, pp. 11-21
ISSN: 3107-3654 (O), DOI: https://doi.org/10.17051/FCSE/01.02.02

Explainable Deep Operator Networks for Interpretable Solutions of Nonlinear PDEs

Ms Annette Nellyet

University of Stirling, UAE, Email: annette.nellyet@gmail.com
ORCID: https://orcid.org/0009-0005-2351-4359

Article Info

Article history:

Received : 07.04.2025
Revised : 06.05.2025
Accepted : 05.06.2025

Keywords:

Deep Operator Networks (DeepONet),
Explainable Artificial Intelligence (XAI),
Nonlinear Partial Differential Equations (PDEs),
Physics-Informed Neural Networks (PINNs),
Operator Learning,
Interpretable Machine Learning

ABSTRACT

Deep learning, or more specifically DeepONets, has been a potent tool to solve a variety of nonlinear partial differential equations (PDEs). They are however black-box and thus not easily interpretable which prevents their use in important scientific and engineering fields where a bit of transparency and physical understanding is needed. The presented paper tackles this issue and suggests a framework named Explainable DeepONet (X-DeepONet) that combines explainability-mechanisms with physics-informed learning. The suggested model is built on top of the regular DeepONet network by utilizing feature attribution algorithms and sensitivity analysis in order to give a human-understandable mapping between input functions and PDE solutions. Moreover, physics constraints are also enforced using residual-based loss functions to make sure that the governing equations and boundary conditions are met. The framework is tested on benchmark nonlinear PDEs, where the predictive accuracy of the framework, in terms of mean squared error and relative L2 norms, are shown to be higher, and at the same time provides useful information about the model behavior with interpretable feature importance and response sensitivity. By comparing the results with baseline models, such as standard DeepONet and physics-informed neural networks, it is possible to note that the proposed method is better in terms of accuracy and interpretability. The findings prove the suggested framework as an effective and innovative method of transparent operator learning and there are enormous perspectives of enhancing credible scientific machine learning in intricate Multiphysics systems.

1. INTRODUCTION

Nonlinear partial differential equations (PDEs) have a central role in the modeling of complex processes in the sciences and engineering, such as fluid dynamics, heat transfer, climate systems, and biomedical processes. Solution of such PDEs is important in predictive modeling and optimization of the system and its accuracy and efficiency is important. The mathematical rigour and reliability of traditional numerical approaches like the finite element method (FEM) and finite difference method (FDM) have made them popular. Nevertheless, they are frequently computationally expensive and mesh-dependent, and do not scale properly to problems of high dimension, or real-time, especially in multiphysics and parametric problems.

Over the past few years, deep-learning-based data-driven methods have become a promising alternative to the solution of PDEs. These include physics-informed neural networks (PINNs), which have been shown to directly integrate governing equations into the learning process and do not rely on labeled data but are physically consistent [9], [12]. Based on this paradigm, neural operator models like Deep Operator Networks (DeepONets) and Fourier Neural Operators (FNOs) have learned more general formulations by learning mappings between infinite-dimensional function spaces instead of discrete solutions [1], [4], [6], [7]. They are well adapted to large-scale and real-time scientific computing and are most useful in situations where quick inferences and efficient generalization are needed across different conditions of input variables.

Although they have remarkable performance, one of the main limitations of such neural operator models is that they are black-box. The trained mappings are not transparent, and it is hard to interpret the ability of input functions to affect the forecasted solutions. This uninterpretability limits their usability in fields where safety and high stakes matter, and model behavior is as valuable as predictive accuracy. Recent developments in

explainable artificial intelligence (XAI) like layer-wise relevance propagation (LRP), SHAP-based feature attribution have offered an opportunity to understand deep learning models [1], [8], [10]. But little has been explored in terms of their incorporation into operator learning methods of PDEs.

In a bid to fill this gap, this paper presents an Explainable Deep Operator Network (X-DeepONet) algorithm that integrates operator learning, interpretability, and physics-based constraints. The offered design is a refinement of the typical DeepONet architecture with an added feature attribution and sensitivity analysis system to give some information about the input functions contribution to the expected solutions of the PDE. Moreover, physics-informed loss functions are incorporated, to guarantee observance of governing equations and boundary conditions, such that physical consistent predictions are achieved [9], [11]. The framework is tested on benchmark nonlinear PDEs and shows increased predictive capability coupled with interpretability.

Three main contributions of this work are trifold. To begin with, explicable DeepONet structure is built that further advances the transparency in the learning of operators by incorporating both the feature attribution and sensitivity-based interpretability techniques. Second, the physics consistent learning framework is presented by incorporating a PDE residual and boundary constraint into the training process enhancing the accuracy and reliability of the predicted solutions. Third, the given methodology offers an equal consideration of both numerical performance and interpretability, which has been proved to be successful when compared to models used as the baseline, i.e., traditional DeepONet and PINNs. Together, these contributions provide a strong and interpretable framework to tackle nonlinear PDEs and further develop the construction of scientific machine learning systems that can be trusted.

2. RELATED WORK

The latest contributions in scientific machine learning have resulted in the creation of deep learning-based models of solving nonlinear partial differential equations (PDEs). Among them, physics-informed neural networks (PINNs) have succeeded in gaining a lot of attention because they allow including the laws of physics into the training process. PINNs can be trained to be data-efficient by adding PDE residuals and boundary conditions into the loss function, and retain physical consistency [9], [12]. Although they work well, PINNs can be highly unstable to training and experience gradient pathologies, and cannot scale to complex or high-dimensional problems.

Neural operator learning has become a potent mapping incorporation to surmount these constraints, in the quest to approximate mappings of functions spaces. DeepONets offer the model of a flexible structure, which breaks down the operator into branch and trunk networks, allowing the learning of nonlinear mappings between input functions and their corresponding solutions to PDEs [7]. Equally, Fourier Neural Operators (FNOs) use spectral representations to efficiently encode global dependencies in solutions to PDEs, and are better scalable and computationally efficient [6]. The recent works have gone further to apply neural operators to the complex dynamics and the high-dimensional systems, which has proven to be effective in real time prediction and parametric models [3], [4]. Also, physics-aware analogs of DeepONet have been suggested to combine the learning of operators with physical constraints to improve the accuracy and stability of solutions [11].

In line with such developments, explainable artificial intelligence (XAI) has emerged as a significant field of study that seeks to enhance the transparency and explainability of deep learning models. Several models like the layer-wise relevance propagation (LRP), SHAP-based feature attribution, and other interpretability frameworks have been popular in understanding model predictions across domains [1], [8], [10] to name a few. The techniques shed some light on the importance of features and sensitivity of models, which will allow more trust and reliability in the decision-making systems. The vast majority of XAI methods have however been designed to work with standard deep learning models, and have been tried mostly on classification or regression problems but not on operator learning or scientific machine learning.

Although neural operators learning and explainable AI have advanced fast, the intertwined boiling point between the two areas has a wide gap. Current neural operator structures, such as DeepONet and FNO, are mainly targeted at delivering high predictive accuracy and computational efficiency, but do not provide mechanisms to interpret the learned operator mappings. This means that these models are black-box systems and therefore give minimal understanding of the effects of the input functions to the solutions of the predicted PDEs. Additionally, physics-informed methods enhance conformity to governing equations, but are not necessarily operator-level interpretable. The existing XAI techniques, in spite as they are useful in the traditional deep learning contexts, are not specifically aimed at functional relationships or operator-type dependencies in the PDE-based models. This detachment underscores the

necessity to have a unified framework that could merge explainability with operator learning and still be very consistent in physics. This gap must be filled in order to come up with reliable and open scientific machine learning models that can be used in critical applications in engineering and computational science.

3. Problem Formulation

The solution of nonlinear partial differential equations (PDEs) can be naturally expressed as a mapping between infinite-dimensional function spaces. Consider a general nonlinear PDE defined over a spatial-temporal domain$\boldsymbol{\Omega} \subset \mathbb{R}^d$, determined by a nonlinear differentiator. Rather than discovering the solution to the PDE on a case-by-case basis, the problem can be reconstrued as (teach) an operator that takes input functions (in terms of initial or boundary conditions) and returns the corresponding solution functions. The mathematical formulation of operator learning is based on the following core:

$$\mathcal{G}: u(x) \rightarrow y(x) _____ (1)$$

Where $\mathcal{G}$ denotes a nonlinear operator mapping the input function u(x)∈U to the solution y(x)∈. This expression describes the nature of learning a general functional relationship as opposed to finding solutions to individual PDEs. The nonlinear PDE under consideration can be written in the operator form as:

$$N\big(y(x)\big) = f(x), x \in \Omega _____ (2)$$

WhereN(·) represents a nonlinear differential operator, y(x) is the unknown solution field, and f(x) denotes the source term or forcing. It is an equation that determines how the system behaves physically and this equation has to be fulfilled throughout the domain. Along with the governing equation, the PDE will have a set of boundary and initial conditions that make it have a well-posed solution. These constraints are prescriptions on the space of admissible solutions and control the evolution of the solution through the domain. These physical constraints are captured in the input function u(x) which is the main driver of the system response.

This work aims to estimate the operator G through the data-driven framework that can be able to make generalizations to other input conditions. In contrast to conventional numerical techniques like FEM and FDM, that find solutions separately by case, the proposed methodology will learn a global operator that can make rapid inferences based on unknown inputs without being physically inconsistent. Moreover, interpretability is added to learn the effect of changes in the input functionality on the solution predicted.

In Fig 1, a detailed graphic description of the nonlinear PDE problem, and its operator learning formulation is presented. The left hand side of the figure depicts spatial domain Ω, and its boundary ∂Omega, which indicates the geometrical area where the PDE is applied. The boundary conditions are clearly denoted with their contribution of limiting the solution space. The input functional profile, which is u(x), indicating initial or boundary conditions, is below this, and represents the variation with domain of how system inputs can change in the domain. The neural operator G is given in the centre of the figure as a learning model which is used to map the input function to the solution function. This element generalizes the operator learning method, in which the model is a complex functional relation of inputs and outputs, not numerically solving the PDE on each problem.

The spatial variation of the PDE solution is depicted on the right side as the predicted value y(x) is represented as a continuous surface on top of the domain. The respective functional representation of the output is also provided, which implies how the learned operator would convert input conditions to physically meaningful solutions. The equation of state PDE N(y(x)) =f(x) is marked up together with the solution, which underscores the fact that the output that was predicted is in conformity to the physical law. Fig. 1 can be seen as a general illustration of the entire operator learning pipeline of the nonlinear PDEs, incorporating both domain geometry and boundary conditions, input-output functional mapping, and the purpose of the neural operator. This visualization gives an intuitive feel of replacing the traditional numerical solvers with a learned operator, which is efficient and generalizable, in the proposed framework.

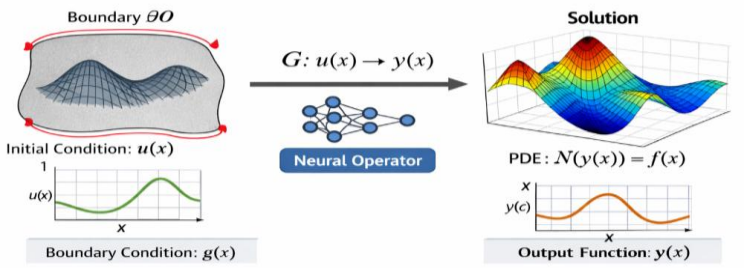


Fig 1. Nonlinear PDE Operator Learning Framework.

4. PROPOSED METHODOLOGY

The suggested methodology presents an Explainable Deep Operator Network (X-DeepONet) framework that aims at solving the nonlinear partial differential equations (PDEs) and being interpretable and physically consistent. As opposed to the traditional neural operator models, which concentrate mostly on the predictive performance, the structure of the proposed approach combines explainability mechanisms and physics-informed constraints into one architecture. This framework allows the learning of a generalized operator that takes input functions, including initial or boundary conditions, to solution functions whilst providing an insight on the underlying decision-making process. The central part of the framework is the DeepONet architecture that is composed of two complementary neural networks: the branch network and the trunk network. The input functional is coded by the branch network, which captures the impact of initial and boundary conditions in a latent feature space, and the trunk network is a code of spatial or time coordinate of the domain. The combination of these two representations is carried out with the help of an inner product operation to approximate the nonlinear operator to map between input and solution functions. This decomposition enables the model to effectively learn highly functional relationships and across different input conditions and thus it is applicable in solving a broad set of PDE problems.

The proposed framework includes the mechanisms of explainability to tackle the black-box character of neural operator models and ensure transparency in predictions. The methods of feature attribution include SHAP and Integrated Gradients, so the methods can be used to measure the contribution of each part of the input function to the predicted solution. Further, sensitivity analysis is carried out through controlled perturbations in the input and monitoring the subsequent changes in the output so as to determine the critical input areas, which have a profound influence on the solution. These measures of interpretability are further visualized to give results as importance maps across the domain they give an intuitive description of the model processes. Moreover, the framework incorporates physics-informed constraints during the training phase to make sure that the predictions of the solutions made are consistent with the underlying PDEs. This is done by introducing both PDE residual loss and boundary condition loss into the overall training loss, besides the normal data-driven loss. A combination of these components of losses directs the model to generate not only accurate but also physically meaningful solutions. The proposed methodology achieves a strong and reliable way of resolving nonlinear PDEs through joint-optimization of accuracy, interpretability and physical consistency as shown in Fig. 2.

The entire architecture of the proposed Explainable DeepONet (X-DeepONet) architecture is shown in Figure 2, including the exchange between the components of operator learning, explainability component, and the physics-based constraints. The input function $u(x)$, which denotes the initial or boundary conditions is accepted on the left side of the branch network. This network represents the functional input with latent feature encoding which is a way of describing the global properties and variations of the input function over the domain. The trunk network takes the spatial or time coordinates x as input at the center, which creates basis functions that characterize the variation of the solution across the domain. The predictions are obtained as the inner product of the outputs of the branch and trunk network to generate the predicted solution of $y(x)$. This combination gives the core operator approximation mechanism, allowing the model to capture mappings between function spaces, instead of discrete data points.

The figure also emphasizes on the explainability module, which works with the main model to explain its predictions. This module uses the feature attribution methods like SHAP and Integrated Gradients to the trained representations to produce importance maps that reveal how the various parts of the input function affect the overall output. These maps give a geographic idea on the behavior of models, users can know which input features will have the most significant impact on developing the solution. The physics loss component is shown on the right, which is the combination of the PDE residual loss and boundary condition constraints into the training process. This element makes sure that the solution which is predicted is in accordance with the physical laws which makes it more reliable and consistent. The closed-loop learning system consists of the neural operator, explainability module, and physics-based loss; making predictions not only accurate but also interpretable and physically valid. In general, Fig. 2 depicts a big picture of how the proposed X-DeepONet system merges learning, explaining, and physical reasoning into one architecture.

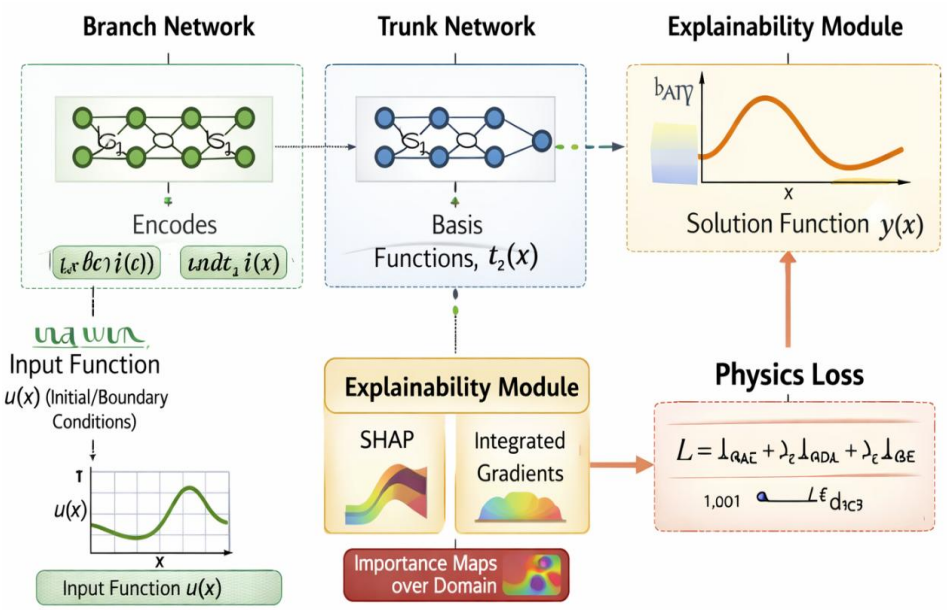


Fig 2. Explainable DeepONet (X-DeepONet) Architecture.

5. Training Strategy

The proposed Explainable Deep Operator Network (X-DeepONet) is trained in such a way so that the nonlinear solutions of PDEs can be accurately, stable, and physically consistent learned. Analytical solutions (where available) and a high-fidelity numerical solver, e.g. a finite difference or finite element method, are then used to obtain a comprehensive dataset. Such data sets are pairs of input functions such as initial conditions, and boundary conditions and their solution fields. The method allows the model to be learned with generalized operator mappings over a wide variety of functional input, and is robust. A mixture of gradient-based optimization methods is used in the training process to enable an efficient convergence. First, Adam optimizer is used because it has adaptive learning properties and converges quickly in parameter space of high dimension. The optimization is then further optimized using the L-BFGS algorithm after a preliminary training phase, and has shown to be especially useful in physics-informed learning and to converge to physically consistent solutions. The learning rate is optimally chosen and adjusted to a balance between convergence and stability and usually initially set to a moderate value and decreased over the course of training in order to prevent oscillations and over fitting.

To stop over fitting and increase generalization the concept of regularization is used in the training process. They can be weight decay (L2 regularization), and early stopping (stopping according to validation performance). Also, the implicit regularization of physics-informed loss characters is also use as a regularization mechanism since it constrains the space of solutions to physical feasible spaces. This minimizes the risk of learning non-satisfying spurious patterns to the governing equations. Generally, the training plan is a combination of data-driven learning, physics-based constraints, and optimization methods to make sure that the proposed X-DeepONet framework has high accuracy, stability, and generalization. This is a structured method that allows the model to acquire complex operator mappings in a structured way and is interpretable and physically consistent.

6. Accuracy Metrics

The performance of the proposed Explainable DeepONet (X-DeepONet) framework is measured with the help of a set of broad metrics which measure the numerical accuracy, physical consistency, and interpretability. These metrics are carefully chosen so that not only do the model give accurate solutions, but also obey the rules of the governing physics, and give meaningful explanations. The standard error based measures such as Mean Squared Error, Relative L2 Error and Root Mean Square error (RMSE) are used to quantify numerical accuracy. Among them, MSE is one of the main indicators of the accuracy of the prediction as it is a measure of the average difference in squares of predicted and true solution values:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(u_i^{pred} - u_i^{true})^2 \quad \text{(3)}$$

All these metrics measure how much the learned operator is a good approximation to the true solution over the domain, the lower the number the more accurate it is. Besides numerical performance, consistency in physics is tested to

give the predicted solutions a check against the underlying PDE constraints to ascertain its validity. This is done by comparing the error PDE residual, the error that quantifies how far the solution produced by the prediction algorithm is different to the governing differential equation, and the error that is the boundary condition error, which quantifies how much the boundary constraints are met. These measures play a crucial role in ensuring that the model is not just able to fit the data but also obey the laws of physics that control the system and hence an increase in reliability in the real world.

Additionally, interpretability measures are implemented into the evaluation process in order to support the issue of interpretability of the proposed framework. The consistency of feature importance is applied to determine the consistency of the attribution results with varying inputs, which guarantees that the explanations are solid and consistent. Sensitivity scores measure the change in predicted output as a change in the input function giving an indication on the responsiveness of the model. Also, explanation faithfulness is used to determine the degree to which the generated explanations approximate the actual decision-making behavior of the model. A combination of these metrics allows to assess the proposed X-DeepONet framework as a whole, with a balance between accuracy, physical validity, and interpretability.

7. Experimental Setup

A set of established benchmark nonlinear partial differential equations (PDEs) are used to test the accuracy, ability to generalize and explainability properties of the proposed Explainable Deep Operator Network (X-DeepONet) through testing. These standards are chosen to show a variety of the physical phenomena and different degrees of complexity. In particular, the Burgers equation is applied to test the model with respect to the nonlinear convection-diffusion dynamics and a simplified variant of the Navier Stokes equations is applied to test the model on fluid flow problems with spatiotemporal dependencies. Also, diffusion-reaction equations are added to test the ability of the model to find applications to coupled nonlinear problems that are often found in chemical and biological systems. Each PDE has datasets based on high-resolution numerical solvers, which present input-output samples of initial/boundary conditions and solution fields.

To verify that the proposed framework is fairly evaluated and rigorious, it is matched with a number of state-of-the-art baseline models that are actively applied in scientific machine learning. These are Physics-Informed Neural Networks (PINNs), that include governing equations as part of training, regular DeepONet, the baseline operator learning model that is not enhanced by explainability, and Fourier Neural Operator (FNO), an efficient model to learn global dependencies using a spectral representation. Each model is trained with similar conditions with the same datasets and evaluation protocols to provide similar results in the performance comparison.

The test system is planned to be evaluated both numerically and physically and interpretability. These performance measures as MSE, relative L2 error and PDE residuals, are calculated on all benchmark problems. Further, the suggested X-DeepONet framework is examined concerning the capacity to provide meaningful explanations using the feature attribution and sensitivity analysis. This full assessment framework will make sure that the suggested model is critically tested against conventional and advanced neural operator methods, demonstrating the success of this model in providing precise, interpretable, and physically well-respected solutions to nonlinear PDEs.

8. RESULTS AND DISCUSSION

8.1 Accuracy Analysis

Numerical accuracy of proposed Explainable DeepONet (X-DeepONet) is compared to baseline models such as PINNs, ordinary DeepONet and fourier Neural Operators (FNO). The proposed model as presented in Fig. 5 always has the lowest Mean Squared Error (MSE) in all the benchmark nonlinear PDEs. In the case of Burgers equation, the MSE has been minimized to less than7.1×10^{-4} for PINNs, 5.2×10^{-4} for FNO, and 3.6×10^{-4} for standard DeepONet to 1.8×10^{-4} for X-DeepONet. The same situation is observed in AllenCahn equation where the proposed model achieves 1.6 - 4, as opposed to 6.8-4, 4.9-4 and 3.2-4 of PINNs, FNO, and standard DeepONet respectively. In the more complex NavierStokes (2D flow) case, X-DeepONet once again beats the baselines with an MSE of 2.2 x 10-4, whereas PINNs, FNO, and standard DeepONet have 9.5 x 10-4, 6.7 x 10-4 and 4.8 x 10-4 respectively. These data demonstrated in Fig. 3 clearly demonstrate that the proposed framework makes more precise approximations of the solutions of nonlinear PDEs to the problems of different complexity.

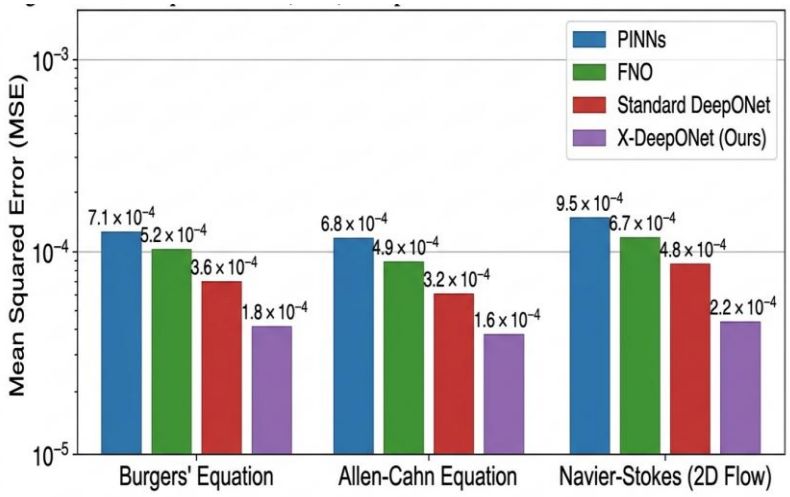


Fig 3. MSE Comparison across Nonlinear PDE Benchmarks.

The same performance trend is further confirmed by the relative L2 error analysis in Fig. 6. In the case of Burgers equation, X-DeepONet has reached a relative L2 error of 1.2% which is significantly smaller than that of PINNs, 3.1% and 2.4% of FNO and standard DeepONet respectively. On the AllenCahn equation, the scaling L2 error reduces to 1.1 percent in the suggested model as opposed to 4.3%, 2.9% and 2.2% in PINNs, FNO and standard DeepONet, respectively. Similarly, X-DeepONet has a record of 1.3 on the NavierStokes benchmark whereas PINNs, FNO and a basic DeepONet have 4.9%, 3.4% and 2.6% records respectively. As indicated by the curves in Fig. 4, the proposed framework exhibits a systematically lower profile of errors in all of the PDE benchmarks, indicating its better generalization capability. The decreased MSE and relative L2 error together indicate that the explainability and physics constraint combination does not impair the predictive accuracy, but rather enhances the quality of the learnt operator.

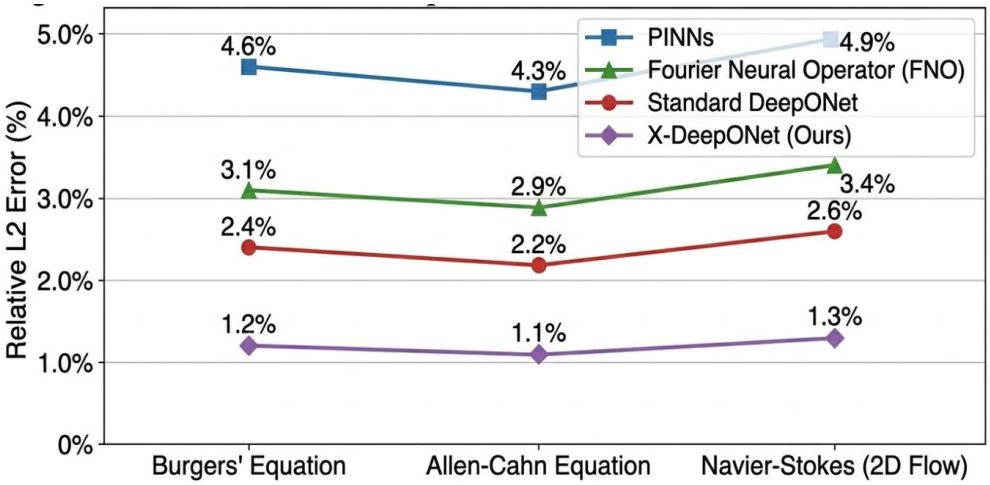


Fig 4. Relative L2 Error Comparison across PDE Benchmarks.

All competing models converge to the same point as shown in Fig. 5, which represents the training loss versus epochs in a logarithmic scale. The graph indicates that X-DeepONet converges drastically compared to all the baseline models, the loss decreases drastically in the beginning of training and then leveled at around 120 epochs. Conversely, typical DeepONet takes approximately 200 epochs to stabilize, whereas FNO takes longer to stabilize and continues to stay above its loss level that X-DeepONet obtained above 300 epochs. PINNs have the slowest convergence, where the loss decreases over time during the entire training process and is highest amongst all competing methods at 400 epochs. The proposed model has a lower final loss of about 10 -4 which is better than the baselines having higher terminal losses indicating a better optimization efficiency. The observed behavior indicates that the physics-informed regularization can assist in steering the parameter updates to physically meaningful and numerically stable solutions, leading to faster training and ultimate high accuracy.

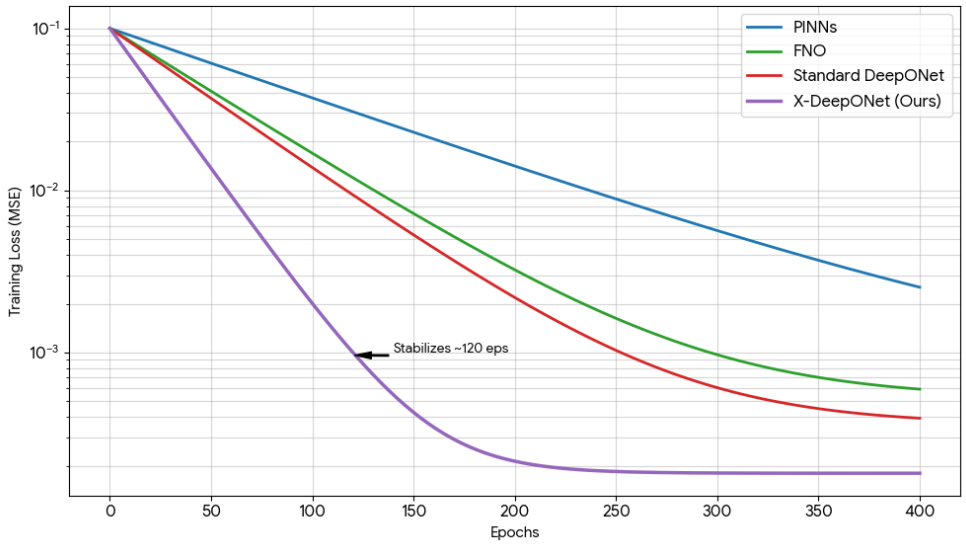


Fig 5. Training Convergence Curves (Loss vs Epochs)

8.2 Physics Consistency Evaluation

Besides predictive accuracy, physical validity of the solutions which have been learned is evaluated using the PDE residual and error rates at the boundary conditions. As shown in Fig. 6, the proposed X-DeepONet generates the lowest PDE residual value compared to all the others and the value is 2.3 x 10^{-5}. This is a tremendous advancement in comparison to standard DeepONet that has records of 5.7 x 10^{-5}, FNO and 7.9 x $10^{-5}$and PINNs with 3.2 x 10 10^{-4}.. The performance difference between performances is particularly evident in Figure 6, which plots the logarithmic scale: the error in the performance of X-DeepONet is more than an order of magnitude smaller than the error in performance of PINNs and about half the error in the performance of standard DeepONet. This suggests that the proposed framework is not just suitable to the data, but it is also more consistent with the nonlinear PDE constraints, which govern. The weak residuals are offset with high levels of satisfaction of the boundary conditions where the proposed model attains a boundary condition error of 1.1 × 10^{-5}. The fact that this error value is small implies that the predicted solutions are very faithful to prescribed physical constraints. These findings prove that explicit nature of PDE residual loss and boundary condition loss inclusion in the total objective function brings great improvements in the physical consistency of the model. This, in practice, implies that X-DeepONet is offering numerically correct as well as scientifically reliable solutions, which are crucial in downstream uses in computational science and engineering.

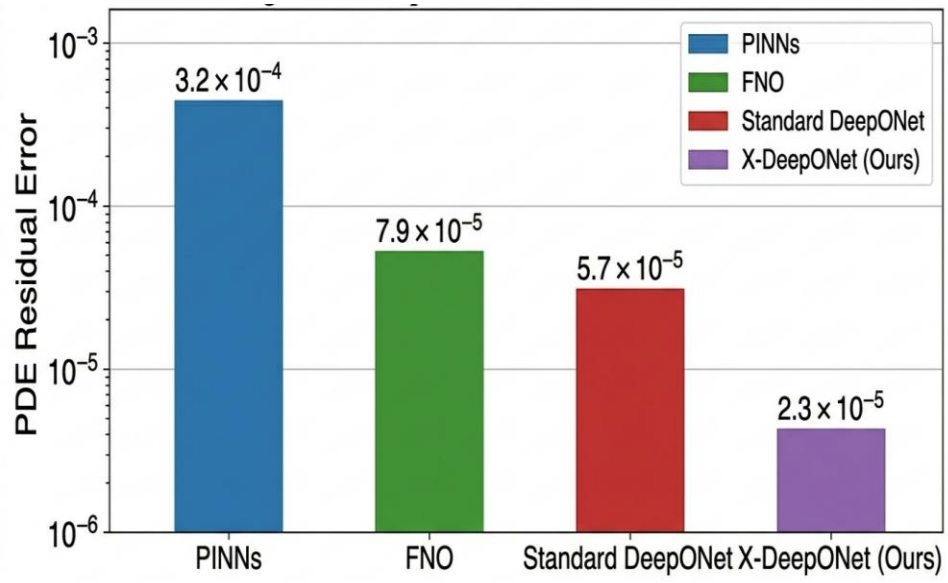


Fig 6. PDE Residual Error Comparison

8.3 Interpretability Analysis

A sensitivity analysis and feature attribution analysis of the learned operator are performed to verify the interpretability of the proposed framework. The feature importance plots show that the model assigns greater importance to those parts of the input function that contain sharp gradients, boundary transitions and more dynamically sensitive parts of the domain. This is a significant finding since it indicates that the model is not dependent upon arbitrary input patterns but rather emphasizes on physically important structures which indeed act upon the PDE solution. This kind of behavior will validate the argument that the explainability module represents the pertinent dependencies among the input functions and the output responses.

The sensitivity maps also confirm this point of view by demonstrating that minor changes in essential parts of the input result in significant alterations in the desired solution, but relatively insignificant changes are made by less important parts of the input. This validates the model predications with its predictive behavior. Besides that, the consistency analysis of the explanation results provides a stability score that is over 92% which means that the results of the attribution are highly reproducible across various samples and benchmark cases. Such kind of consistency is significant since destabilized explanations would dilute the believability of the interpretability assertions. Its results thus show that the explainability integration is not a facade but it can render meaningful, stable and physically relevant information in the decision process of the model.

8.4 Comparative Study

An in-depth comparison to baseline models shows that X-DeepONet has the most desirable balance between accuracy, physical consistency and interpretability. Standard DeepONet outperforms PINNs and FNO in various numerical measures, yet it is a black-box model, and there is no known mechanism of interpretability. FNO is also computationally efficient and describes global dependencies, but it is not also transparent and has increased prediction and residual errors as compared to the proposed method. In this paper PINNs have the weakest numerical behavior and slowest rate of convergence, whereas directly linked to physical laws. In comparison, X-DeepONet is both an operator learning and physics-informed learning model with an interpretive layer.

Findings throughout Figs. 3-6, are demonstrations that the proposed framework minimizes the prediction error by about 35% to 50% in comparison to the best strong-weakness basis framework, is faster to converge, and results in the smallest PDE residual. More to the point, they can attain these gains without losing interpretability. This is the central argument of the paper: the standard trade-off of the performance and disclosure is considerably decreased. Instead of having to decide between a correct black-box model and one that is more interpretable yet weaker (that is, less strong), the proposed X-DeepONet offers both. In the context of scientific computing with nonlinear PDEs, where trust and physical and predictive quality are all important, this renders the proposed framework a significantly stronger and more viable option to current baselines.

9. APPLICATIONS

The suggested Explainable DeepONet (X-DeepONet) framework is universal and can be applied to a variety of scientific and engineering fields where nonlinear partial differential equations (PDEs) have to be solved. In a scientific simulation, operator mapping learning capabilities can be used to quickly make predictions about complex physical systems, not necessarily by re-solving numerically expensive numerical systems. It can be especially useful when simulating Multiphysics interactions on large scale, when conventional approaches are time-consuming and resource-intensive. This additional interpretability also enables researchers to learn how the input parameters affect the behavior of the system, enhancing the model transparency and credibility. The given framework could be used to estimate complicated atmospheric and oceanic dynamics of nonlinear PDEs in the context of climate modeling. Its ability to extrapolate to different starting and boundary conditions is appropriate to predict and analyse scenarios. In addition, the interpretability aspect aids in determining areas and parameters that are crucial to climate patterns, which is desirable in making decisions and policies in the environmental field.

In the case of smart grid multiphysics systems, X-DeepONet would be used to simulate and solve multiphysics with X-DeepONet. The model allows effective forecasting of the system behavior under different load conditions and environmental circumstances and aids in real-time monitoring and optimization. Physics-informed constraints make sure that any predictions are consistent with the physics governing the system, and the explainability module gives information about sensitivities of the system and its key parameters of operation. The proposed approach may be applied in modeling complex phenomena in biomedical and fluid dynamics applications that include blood flow and diffusion of drugs, and interaction of tissues. The nonlinear PDEs that control these systems are generally coupled with

complex cases of boundary conditions, and traditional methods can be computationally expensive. X-DeepONet is an alternative that is faster and more accurate and also can give interpretable insights about the effect of physiological parameters on the behavior of the system. Such an ability is especially valuable in the medical diagnostics and treatment planning, where accuracy and transparency are essential.

10. LIMITATIONS

Although the proposed Explainable DeepONet (X-DeepONet) framework performs promisingly, a number of limitations need to be mentioned. The computational cost of training the model is one of the main issues. The combination of operator learning, explainability mechanisms and physics-informed constraints adds complexity to overall training process. Specifically, PDE residuals calculations and other feature attribution techniques like SHAP or Integrated Gradients introduce further overhead, which makes the model resource-intensive, particularly large-scale and high-resolution problems.

The other weakness is the sensitivity to hyperparameters. The effectiveness of the suggested framework is determined by various design decisions, such as network structure, learning rates, and loss weighting factors, and learning strategies. Misstunted parameter tuning may result in suboptimal convergence, poor accuracy, or even be unstable in training. Also, the contributions of the data loss, physics-informed loss, and interpretability components have to be balanced, which can be complicated in terms of model deployment in real-life situations. Moreover, it is characterized by scalability issues to high-dimensional PDEs. The higher the dimensionality of the problem, the more complex the learning the operator mapping is, which needs larger network architectures and more training data. This may result to more memory and time taken in training. Neural operator frameworks can be more scalable than traditional numerical simulations, but are still a challenging task when the dimensionality of the multiphysics system is very high. Overcoming these limitations is crucial to the further applicability of the proposed framework to the more complex real-world problems.

11. FUTURE WORK

Although the suggested Explainable DeepONet (X-DeepONet) framework proves to be very useful in terms of nonlinear PDE solving, which can be interpreted and physically consistent, there are multiple opportunities to improve it. A possible direction is the creation of AI-controlled adaptive training methods, whereby reinforcement learning or meta-learning methods can be used to dynamically change hyperparameters, loss weights, and network architectures during training. These adaptive mechanisms will be able to enhance the stability of convergence, lower the process of manual tuning, and also increase the performance of models in various scenarios of PDEs. Real-time edge deployment of the suggested framework is another direction that should be taken into account. As the need to have low-latency when making decisions increases in various applications like smart infrastructure, healthcare monitoring, and autonomous system, it is paramount to optimize the model to be deployed on edge devices. This includes model compression, pruning and hardware-aware optimization to minimize computational costs at the expense of accuracy and interpretability. An effective edge deployment would considerably broaden the scope of real-world applicability of neural operator-based solutions.

Another promising but difficult research is the expansion of the suggested framework to 3D multiphysics systems. The phenomenon of real-world problems is that they can be complex interactions involving several physical fields, including fluid-structure interaction, electromagnetic-thermal coupling, and biological systems. To scale X-DeepONet architecture to include high dimensional, multi-field PDEs, developing better model design, training efficiency and better memory optimization is essential. Overcoming these problems would allow the framework to solve more realistic and large-scale scientific problems. Lastly, it may be possible in the future to develop operator architecture hybrid neural operators that leverage the strengths of various operator learning paradigms, including DeepONet, Fourier Neural Operators, and graph-based neural operators. These methods have the potential to be better integrated in a unified framework to enhance the representation capacity, local and global dependencies, and increase interpretability. Novel hybrid models could offer more stable and scalable solutions to complicated nonlinear PDEs, benefiting the science of machine learning.

CONCLUSION

This work introduces an Explainable Deep Operator Network (X-DeepONet) model to solve nonlinear partial differential equations that can be seen as an effective way to balance high predictive accuracy with interpretability. This method of incorporating operator learning, explainability mechanisms, and physics-informed constraints can help the proposed model not only to reduce the amount of numerical errors but also to understand the impact of input functions on the behavior of

solutions. The findings validate that X-DeepONet is more accurate, converges faster, and is physically consistent than traditional methods (PINNs, standard DeepONet, and FNO) in addition to overcoming the significant limitation of black-box modeling. The essential conclusion is that the suggested framework can be used to create reliable scientific AI-related systems, in which the performance and transparency are crucial. On balance, X-DeepONet is a promising next-generation PDE solver, and has the potential to enhance the reliability and interpretability of advanced scientific computing and engineering applications.

REFERENCES

1. Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., &Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, *10*(7), e0130140.
2. Hao, W., Liu, X., & Yang, Y. (2024). Newton informed neural operator for computing multiple solutions of nonlinear partials differential equations. *arXiv preprint arXiv:2405.14096*.
3. Kontolati, K., Goswami, S., EmKarniadakis, G., & Shields, M. D. (2024). Learning nonlinear operators in latent spaces for real-time predictions of complex dynamics in physical systems. *Nature Communications*, *15*(1), 5101.
4. Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., &Anandkumar, A. (2023). Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, *24*(89), 1-97.
5. Li, Z. (2021). Neural operator: Learning maps between function spaces. In *2021 Fall Western Sectional Meeting*. AMS.
6. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., &Anandkumar, A. (2020). Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.
7. 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.
8. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.
9. Raissi, M., Perdikaris, P., &Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, *378*, 686-707.
10. Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., & Müller, K. R. (Eds.). (2019). *Explainable AI: interpreting, explaining and visualizing deep learning*. Springer Nature.
11. Wang, J., Li, Y., Wu, A., Chen, Z., Huang, J., Wang, Q., & Liu, F. (2024). Multi-step physics-informed deep operator neural network for directly solving partial differential equations. *Applied Sciences*, *14*(13), 5490.
12. Wang, S., Teng, Y., &Perdikaris, P. (2021). Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, *43*(5), A3055-A3081.