DSFA-PINN: Deep Spectral Feature Aggregation Physics Informed Neural Network

MUHAMMAD RAFIQ, (Graduate Student Member, IEEE), **GHAZALA RAFIQ**, **AND GYU SANG CHOI**, (Member, IEEE)
Department of Information and Communication Engineering, Yeungnam University, Gyeongsan-si 38541, South Korea
Corresponding author: Gyu Sang Choi (castchoi@ynu.ac.kr)

ABSTRACT Solving parametric partial differential equations using artificial intelligence is taking the pace. It is primarily because conventional numerical solvers are computationally expensive and require significant time to converge a solution. However, physics informed deep learning as an alternate learns functional spaces directly and provides approximation reasonably fast compared to conventional numerical solvers. The Fourier transform approach directly learns the generalized functional space using deep learning among various approaches. This work proposes a novel deep Fourier neural network that employs a Fourier neural operator as a fundamental building block and employs spectral feature aggregation to extrude the extended information. The proposed model offers superior accuracy and lower relative error. We employ one and two-dimensional time-independent as well as two-dimensional time-dependent equations. We employ three benchmark datasets to evaluate our contributions, i.e., Burgers' equation as one dimensional, Darcy Flow equation as two dimensional, and Navier-Stokes as two spatial dimensional with one temporal dimension as benchmark datasets. We further employ a case study of fluid-structure interaction used for the machine component designing process. We employ a computation fluid dynamics simulation dataset generated using the ANSYS-CFX software system to evaluate the regression of the temporal behavior of the fluid. Our proposed method achieves superior performance on all four datasets employed and shows improvements to baseline. We achieve a reduced relative error on the Burgers' equation by approximately 30%, Darcy Flow equation by approximately 35%, and Navier-Stokes equation by approximately 20%.

INDEX TERMS Partial differential equation approximation, physics informed neural network, computational fluid dynamics, neural operators, functional spaces, spectral feature learning.

I. INTRODUCTION

Numerical simulations use a multidimensional discretization mechanism for solving partial differential equations. They are computationally expensive and require significant time to solve numerical problems. It requires large-scale forward numerical runs, often time-consuming and computationally expensive. In general, the approximation process is time-consuming, so one has to wait long enough for results. Therefore, it delays the organizational decision process. Machine learning researchers have made breakthrough progress in providing alternative solutions to numerical simulators or solvers in the past few years. Neural networks that use the data-driven finite-dimensional operator [1]–[3]

The associate editor coordinating the review of this manuscript and approving it for publication was Mingbo Zhao.

and parameterize the solution function are called physics informed, physics constrained, or neural finite-difference learning approaches [4], [5].

DeepONets [6] proposed a solution to the numerical approximations using deep learning. In [7] a multi-pole graphical neural network (MGNO) was proposed. (Li *et al.* 2020) [8] proposed the Fourier Neural Operator (FNO) based on the Fourier transform that solves the problem through the functional parametric dependency and learns directly from the infinite dimension mapping. Then U-FNO [9] improved the FNO to work with the multi-phase flow by adding a U-network inside the FNO. In this context, FNO made significant progress and demonstrated cutting-edge performance.

Solving partial differential equations using deep learning or machine learning is an active research field, and various
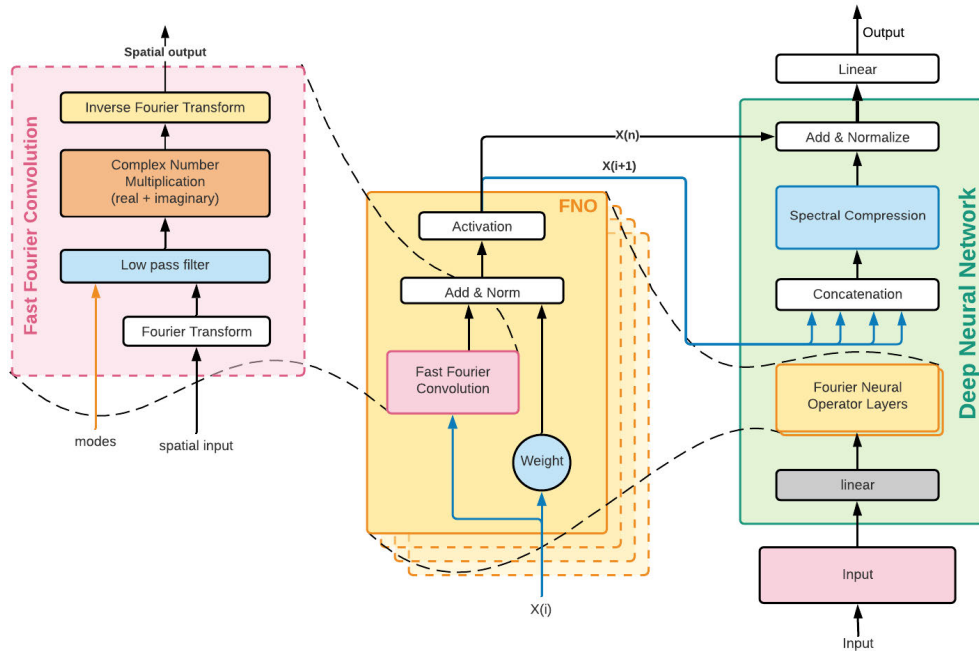
FIGURE 1. System Architecture: Deep spectral feature aggregation model using Fourier neural operator.

solutions with excellent performance exist already, but deep learning and computer vision are not yet perfect enough to produce a strong performance impression. It motivates us to contribute to this field of research. As an active research field, more and more solutions are being proposed in this area.

Fourier neural operators (FNO) have presented a novel method with excellent performance for functional space learning. So, applying the underlying architecture inherits feature space learning to each extended model by itself. However, we analyze that cascading Fourier convolutions with iterative calls cause information loss during reconstruction, the process of transforming spatial domain representations into spectral domains, performing complex multiplications, and inverting transforms into spatial domains. In this process, information loss is observed at the boundary due to spectral convolution's forward and backward transformation process. So it allows us to propose the model to enhance the performance while processing the same information iteratively by saving feature loss at the edge after each layer.

This study aims to fulfill the performance gap and to develop a deep neural network model that takes advantage of the Fourier neural operator to learn partial differential equation (PDE) functional space using the fast Fourier transform domain and improve the feature representation by collecting each layer in a dedicated tensor and fusing this feature-filled tensor with final outputs. We extend the research work by FNO to achieve our extended goals.

Our contributions in this study are as follows.

- We introduce a novel deep spectral-aggregation approach with block-wide feature aggregation consuming the Fourier neural operator.

- We incorporate spectral channel compression to extract most learned information and keep from information loss during layer cascading.
- We introduce spectral feature fusion at the final layer of each aggregation block.
- Finally, we develop a novel deep-spectral-feature-aggregation neural network architecture made of deep-spectral-aggregation blocks and a fully connected layer for the output.

This study proposes a deep neural network model, which fundamentally is a deep layer aggregation model with spectral feature compression. Spatial convolutions learn feature spaces as spatial features, whereas spectral convolutions learn functional spaces directly. The model as shown in Figure 1 is designed to accept Inputs in one, two, or three dimensions that are kept compatible with the base paper; it mainly depends on the composition of the dataset, i.e., Burgers' equation is a one-dimensional dataset; hence, the model for the one-dimensional dataset accepts a compatible one-dimensional input. Similarly, the Darcy Flow equation possesses two-dimensional and Navier-Stokes two-dimensional with additional temporal dimension, which can be used as three-dimensional input. The model is configured in a regression topology with a loss function of mean squared error (MSE); hence the output possesses the same dimension as the input.

We used three benchmark datasets: Burgers' equations, Darcy Flow equations, and Navier-Stokes and Navier-Stokes-based computational fluid problems. A new state-of-the-art performance was achieved by measuring the mean squared

error with the lowest relative error across all benchmark datasets under consideration.

We compare our proposed model with time-independent approaches: Artificial Neural Network (NN), Convolution-based Neural Network Method (FCN) [10], Graph Neural Operator (GNO), Multipole Graph Neural Operator (MGNO), Sub-Neural Operator (LNO), DeepONet, and Principle Component Analysis (PCANN) based neural networks. On the other hand, Time-dependent approaches: 2D ResNet18 with residual connections, image segmentation model U-Net, 2D + temporal turbulence network TF-Net, Fourier neural operator 2D + time (FNO-2D), and 3D model Fourier neuron operator (FNO-3D). The proposed spectral aggregation neural network approach with and without time-dependent problems outperformed and achieved a new state-of-the-art performance.

The primary audience for this work is the deep learning community that promotes and advances computer vision for the physics-informed learning models and estimating numerical problems through computer vision. On the other hand, physical science researchers looking for alternative deep learning computational models can employ our proposed model to estimate partial differential solutions. The Navier-Stokes equation solutions can be approximated quickly and kept from time-consuming simulations for their initial study.

The rest of the paper consists of four sections. Section 2 covers existing approaches and related work in subject areas. Section 3 presents the proposed method. Section 4 describes and discusses the experimental results in detail. Finally, Section 5 concludes the research contribution and presents future aspects of this study.

II. RELATED WORK

A. CONVENTIONAL VS PHYSICS INFORMED MACHINE LEARNING

In general, equation solvers use spatial discretization methods to solve equations. This includes the finite difference method (FDM) or the finite element method (FEM). So there is a trade-off between speed and accuracy due to the size of the grid or mesh. The coarse grid eventually ends with a coarse resolution, whereas the fine grid requires much time to find a solution. Finding solutions to complex problems is difficult and requires patience. [4] has developed a new interface for machine learning that enables synergistic combinations and introduces mechanisms for integrating physical principles into deep neural networks. [11] has developed a new neural operator for Fourier spaces by directly parameterizing the integral kernel. Neural operators can demonstrate zero-shot learning at super-resolution by directly training the feature space. This solution is 100x faster to find a solution. [12] proposed a framework for a multi-grid solver that learns a single mapping to a set of partial differential equations for prolongation operators. This approach integrated unsupervised loss functions and performed the two-dimensional diffusion

problem experiments. [13] introduces an end-to-end deep learning method for improving computational fluid dynamics (CFD) for modeling 2D turbulent flows and performing simulations for turbulence and mass vortex problems. This method showed eight to ten times more precise resolution and 40 to 80 times better spatial dimension performance than the computational speed.

(Shukla, Jagtap and Karniadakis, 2021) [14] introduced a distributed framework for physical information neural networks (PINNs) using domain decomposition. The system uses domain decomposition to distribute distributed tasks across parallel GPUs. This approach mainly focuses on training and inferring models to perform physics information learning. (Jagtap and Karniadakis, 2021) [15] proposed a generalized Spatio-temporal domain decomposition framework for extended physical information neural networks (XPINN). This approach is an evolution of PINN and is designed to solve differential equations. This approach achieved performance mainly by applying residual continuity conditions to adjacent subdomains. This method achieved an excellent L2 error (MSE) of 8.93265e-3 for the Burgers' equation. (K. Li *et al.*, 2020) [16] proposed a deep domain decomposition method (D3M) for solving partial differential equations. The physics-based approach presents a deep subdomain decomposition method and focuses on parallel computing to find solutions to PDEs. This method uses ResNet based by default. In this study, the performance was evaluated using Poisson's equation and the time-independent Schrödinger equation, and it showed good performance at a relative error of 0.0045.

B. FINITE-ELEMENT METHODS (FEM)

A method designed for a specific instance of a PDE trained on a specific problem is called FEM. However, it does not perform as expected for new problems and various functional parameters, requiring retraining by optimization problems similar to neural networks. In this context, [17] proposed a deep learning approach, namely the Deep Ritz method, to solve the numerical variational problem, especially PDE. Deep Ritz is essentially a non-linear adaptive framework. Therefore, it is more likely to solve higher-order problems that fit well with the stochastic gradient descent (SGD) method used in deep learning. This method was evaluated for eigenvalues along with several other numerical problems.

Similarly, [18] proposed an unsupervised DCNN algorithm to solve forward and backward problems for PDE. The network is optimized for a cost function that satisfies the PDE, boundary conditions, and further regularization. This approach is mesh-less, unlike numerical solvers and grid-based solutions. After focusing on 2D second-order elliptical systems of non-constant coefficients and specific applications for electrical impedance tomography (EIT), [19] proposes EikoNet to model seismic ray multi-path, hype-center inversion, and tomographic modeling to solve the Eikonal equation using DCNN. This approach characterizes the first time-to-arrival parameters in heterogeneous 3D velocity structures.

This method utilizes the differentiation of CNNs to compute spatial gradients analytically. This approach provides low memory overhead and avoids lookup tables (LUTs).

C. FINITE-DIMENSIONAL OPERATORS (FDM)

[20] proposed a general and flexible approximation model using CNN for real-time prediction of non-uniform steady laminar flow. This method demonstrated up to 100x faster solutions for CFD solvers using GPU acceleration, and similarly 400x faster than CFD running on CPU at the marginal cost of relative error. A Bayesian approach surrogate model was proposed in [21] for the problem of uncertainty quantification in PDEs using deep encoder-decoder (ED) networks similar to image-to-image regression. The model was trained to consider the standard uncertainty quantification for flow in a heterogeneous medium to realize permeability at corresponding velocities and pressures. This study was evaluated using a probabilistic dimension of 4,225 and produced a good performance basis. Results were compared with Monte Carlo estimates. [22] proposed a partially learned approach to solving the ill-posed inversion problem. This approach advances the ideas of classical regularization theory and incorporates deep learning. Methods were constructed using forward operators, noise models, and regularization functions. This method presented a gradient method similar to deep neural networks. Experiments were performed on the tomographic inversion problem using the Sheep-Logan phantom method and simulated data from computed tomography (CT) of the head. Results were compared with filtered back projection and full transform reconstruction. This method showed a peak signal-to-noise ratio (PSNR) of 5.4 dB with a faster solution at a resolution of 512. [23] proposed an approximate model based on CNN for the flow of fluid prediction. This method predicted velocity and pressure geometries by taking into account the pixelated shape of an object under invisible flow conditions. This method was evaluated against a Reynolds Averaged Navier-Stokes (RANS) flow on airfoil training data. [24] proposed NN to parameterize physical quantities as a function of coefficients and integrated engineering of physics examples in PDE to evaluate the approach.

D. TIME-DEPENDENT PROBLEMS

Various approaches are also proposed based on time dependencies. A spectral pooling-based approach was proposed by [25] and introduced many innovations. A prominent contribution was spectral pooling to reduce dimensionality. Then [26] published computational fluid dynamics using Fourier neural operators for airfoil approximation using deep learning. On the other hand, [27] proposes an approach to PDE-related approximation operators using Fourier neural operators and proves that Fourier neural operators are universal. A sophisticated CNN, Fourier Transform U-Net, is shown in [28] and claims to use Fourier methods to reduce the cost of convolution. The method has identified object information in the given set of images. They also show that this system requires less training time. The

proposed technique was applied to the Broad Bio-image Benchmark Collection (BBBC) data set. Continuing the Fourier approach, [8] proposed a new Fourier neural operator using the Fourier transform function spatial convolution. The approach was evaluated via the Burgers' equation, Darcy Flow, and Navier-Stokes benchmark equations. This model claims to be the first to solve partial differential equations and additionally provides high-quality, one-shot super-resolution. Then, [29] proposed a neural operator for solving partial differential equations (PDEs) using a graph kernel network approach. This method provided a generalization for learning an infinite-dimensional feature space. This approach also used message-passing techniques for kernel integration between objects. They have demonstrated significant contributions in this active field of research for solving partial differential equations. Experiments and data-backed facts validated the performance.

E. AGGREGATION APPROACHES

There are various approaches to feature aggregation. The leading approaches are fusing features hierarchies. Densely connected networks applied the concept of short-connection, or skip-connection [30] achieved a state-of-the-art performance with a significant parameter reduction. It solved the deep network vanishing gradient issue and set a new standard in deep convolution networks. This family of architectures for semantic fusion propagates features and losses through skip connections which are further concatenated over stages. We adapt the skip connection from the DenseNet; however, our approach uses spectral features.

Feature pyramid networks (FPNs) (Vo *et al.*, 2021) [31] is a family of architectures, which focus on spatial fusion, equalize the Resolution and standardize semantic features across various levels of pyramidal hierarchy using a top-down lateral connection approach. Our approach concatenates, compress and fuse the spatial semantic features at the compression block at the end of the stage.

Deep layer aggregation (DLA) (Yu *et al.*, 2018) [32], on the other hand, formulate various hierarchical designs to aggregate the spatial features. This approach handles vanishing gradient problems using skip connection to hierarchical step aggregation. It is further connected to the next level hierarchy, making it easier to see a far layer via skip connections and a small distance due to fewer steps via hierarchical architecture. Our approach aggregate feature similar to deep layer aggregation but different due to no multi-level hierarchy, whereas we employ concatenated aggregation technique and compression as feature reduction approach. Furthermore, our approach aggregates spectral features, whereas deep layer aggregation achieves it over spatial features.

F. FOURIER TRANSFORM

A well-established approach, the Fourier transform, is widely used in image and signal processing and is a well-known transform for solving PDEs because of its mathematical

importance in solving derivatives using frequency-domain multiplication.

A standard multi-layer feed-forward network capable of approximate learning was introduced by [33] and then [34] proposed a method to classify the severity of gear tooth breakage using Fourier transform spectrograms. [35] provided mathematical and empirical evidence suggesting that non-parametric learning, especially kernel methods, can learn complex higher dimensions. A multi-scale neural network for high-dimensional nonlinear maps has been proposed in [36]. This approach approximated discrete nonlinear maps and demonstrated solution maps of nonlinear equations, i.e., radiative transfer equations, Schrödinger equations, and Kohn-Sham map approximations. Meanwhile, [37] proposed a sine-wave representation network (SIREN), a periodic activation function for implicit neural representation. This approach learns complex natural signals derived from 1D, 2D, and 3D functional spaces. Focus on working with images, wave fields, sound, video, and three-dimensional shapes. It also shows that it is suitable for solving boundary value problems such as Ekonal's equation, Poisson's equation, Helmholtz, and wave equations. A new deep learning super-resolution frame using MESHFREEFLOWNET was proposed in [11] to generate Spatio-temporal solutions. This approach has demonstrated empirical studies of the performance of Rayleigh-Bénard convection problems at the super-resolution of turbulence. [38] proposed a frequency convolutional network and [39] developed a deep neural network of multitype signal detection and classification in spectrograms.

G. DEEP LEARNING APPROACHES USING SPATIAL FEATURES

A variety of deep learning Physics informed approaches [40] has been proposed recently. Methods involving deep convolutional neural networks using convolutional deep learning approaches have a long history. These include deep convolutional neural networks (DCNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and generative adversarial networks (GANs). (Mo *et al.*, 2019) [2] (Ren *et al.*, 2021) [41] (Kovachki *et al.*, 2021) [13] (Tang, Liu and Dulofsky, 2020) approach) [3] (Sitzmann *et al.*, 2020) [37] (Guan, Hsu and Chitnis, 2021) [42]–[50] are various approaches. (De Ryck, Lanthaler and Mishra, 2021) [51] Propose a surrogate model for the dynamic subsurface flow problem of wastewater streams by combining spatial and temporal information via LSTM using a cyclic U-Net (RU-Net). The survey (Sun *et al.*, 2021) [52] addressed scientific machine learning based on physics and its application to subsurface flows governing the equations of mass transport in porous media.

III. METHOD

We propose a novel deep spectral feature aggregation Fourier neural network, which employs Fourier neural operator block with feature aggregation and channel compression

mechanism; we control scalability using Fourier modes and network depth. The maximum modes can be as much as one more than half of the width of the input tensor. The core of the model is based on a Fourier neural operator layer, which is further organized into the deep spectral aggregation block (DSAB). A DSAB consists of four FNO layers and one aggregation block.

A. DSFA-NET ARCHITECTURE

The proposed model consumes the Fourier neural operator (FNO) as a basic unit and systematically organizes the layers to make a novel and better deep learning solver for parametric functional learning. The model consists of a series of deep spectral blocks, as shown in Figure 2. A deep spectral block consists of FNO layers, which is an implementation of Fourier neural operators proposed by (Li *et al.*, 2020) [8]. We form a multi-block architecture that performs deep learn-
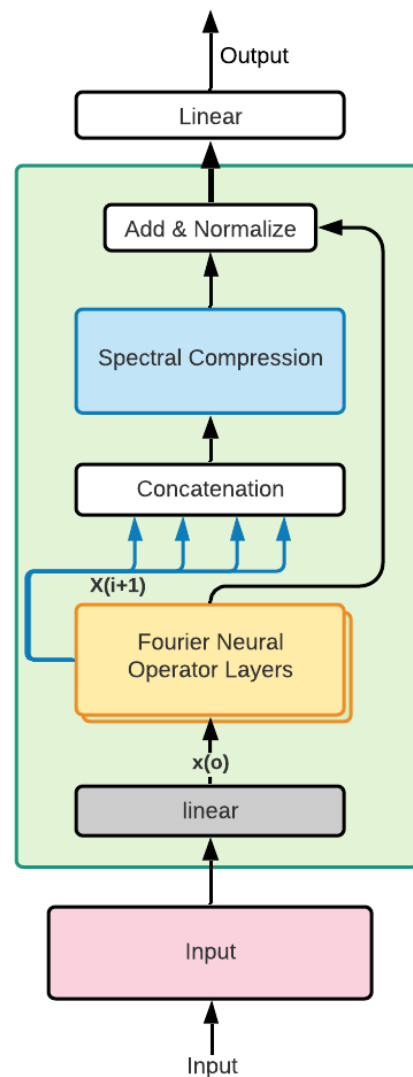

FIGURE 2. Deep spectral feature aggregation neural network.

ing operations staged. The architecture is general-purpose, with a novel organization of deep spectral blocks.

B. FOURIER NEURAL OPERATOR (FNO)

We adapt Fourier neural operator implementation from work presented by (Li *et al.*, 2020) [8]. The Fourier neural operator (FNO) presented a novel scheme with excellent generalized performance. However, there is still room to improve in this approach. We analyze that the Fourier convolution, which primarily transforms the spatial domain representations to spectral-domain, does the complex multiplication, and reverses the transform to the spatial domain, can perform over further depths. In this process, information is lost due to multiple conversions as the nature of the process. Therefore performance degrades on the following layers. Processing the same information with lost features makes it less effective. So it is not efficient to add more layers to increase the depth for further accuracy. We introduce the layer aggregation mechanism explained in the deep spectral aggregation block section to overcome this issue. It enhances the performance of the neural operator significantly.

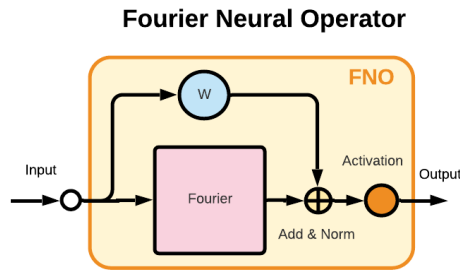Figure 3 depicts the architecture of the Fourier neural operator.

FIGURE 3. Architecture of Fourier neural operator [8].

The Fourier neural operator replaced the kernel integral operator with convolutional operator using Fourier spectral space, hence convert the spatial input to spectral form $\mathcal{F}$ for a function $f : D \to \mathbb{R}^{d_v}$, does complex multiplication $(\mathcal{F}f)_j(k)$ and perform a reverse Fourier transform $\mathcal{F}^{-1}$ for kernel k and input tensor x.

$$(\mathcal{F}f)_j(k) = \int_D f_j(x)e^{-2i\pi\langle x,k\rangle}dx, \quad j = [1, d_v]$$

$$\left(\mathcal{F}^{-1}f\right)_j(x) = \int_D f_j(k)e^{2i\pi\langle x,k\rangle}dk, \quad j = [1, d_v]$$

C. DEEP SPECTRAL FEATURE AGGREGATION

We introduce a concatenation layer to aggregate the features computed at each stage. The last layer is not concatenated to keep final features from compression. The concatenated features are passed to a spectral channel compression layer, which applies Fourier convolution and reduces channels to match the output, which is further added and normalized to both outputs, i.e., one from the last layer and the other from

the spectral compression. Finally, the normalized output is presented for further operations. The Figure 4 depicts the aggregation block.
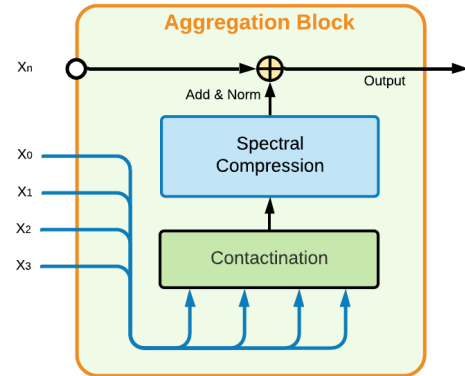

FIGURE 4. Spectral feature aggregation block.

The feature aggregation A performed at each layer i for a tensor x is expressed as

$$A(x_i) = \begin{cases} x_1, & if \ i = 1 \\ A(x_{i-1}, x_i), & otherwise \end{cases}$$

where $A(x_i)$ is obtained aggregation after each layer and final aggregation A_x is then passed through a Fourier based channel compression process to reduce the aggregate channels to same as the input channels and final layer aggregation is activated with activation function σ.

$$A_x = \sigma(A(x_n))$$

the aggregation added and batch normalized as a post process as

$$x_{out} = \sigma(BatchNorm(\sum_i W_i x_i + b) + A_x)$$

where A_x is aggregate for specific input x for layer i, and W represents the layer weights with bias b to obtain the final output x_{out}.

D. DEEP SPECTRAL AGGREGATION BLOCK (DSAB)

A deep spectral aggregation block is an organization of Fourier neural operator layers with an N-layered iterative approach. The mechanism cascades and extracts staged features to a concatenated tensor at the final layer. The concatenated features are then passed through a spectral channel compression which is fundamentally a Fourier spectral convolution layer as shown in Figure 5.

IV. RESULTS AND DISCUSSION

This section shows the performance evaluation of the proposed method and is discussed in more detail later in this section. The relative error for measuring the performance of the proposed method is recorded and compared with the existing method. Performance results were recorded on
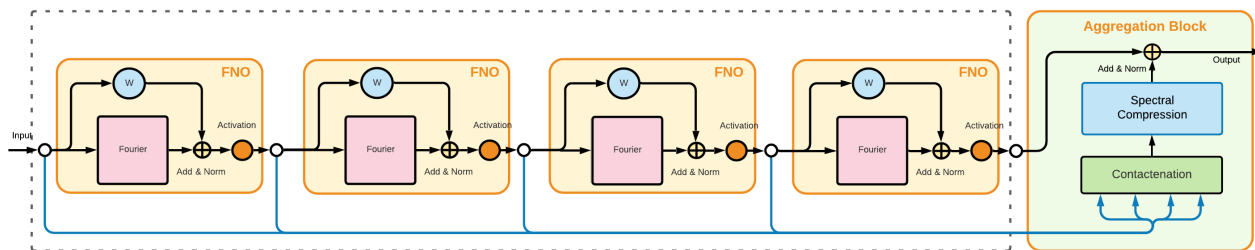
FIGURE 5. End to end deep spectral aggregation block (DSAB).

benchmark data sets such as Burgers' equation, Darcy Flow equation, and Navier-Stokes equation. The data sets used in the experiments remain the same for a fair comparison with existing approaches, especially with current state-of-the-art Fourier neural operators.

A. BENCHMARKS PROBLEMS

We adapt datasets available from the baseline (Li *et al.*, 2020) [8] paper, i.e., Navier-Stokes, Darcy Flow equation, and Burgers' equation. Navier-Stokes consists of temporal information and is a time-dependent problem, whereas Darcy Flow and Burgers' equations are 2D and 1D, respectively.

1) NAVIER-STOKES EQUATION

Navier-Stokes is a famous two-dimensional time-dependent equation. The equation represents the behavior of viscous and incompressible fluids in mathematical form. The equation was first presented by Claude-Louis Navier in 1822 and further modified and extended by George Gabriel Stokes. The equations are recognized among the Clay Mathematics Institute Millennium Problems with a prize offer for a solution.

Navier-Stokes describes the conservation of mass and momentum of Newtonian fluids. The general expression for Navier-Stokes with a momentum and periodic boundary condition with a T^2 unit torus of $[0, 1]^2$ and a fixed viscosity ν, velocity u, forcing function f, and pressure p is:

$$\partial_t u + u.\nabla u + \nabla p = \nu \Delta u + f(x), \quad x \in T^2, \ t \in (0, \infty)$$
$$\nabla .u(x, t) = 0, \quad x \in T^2, \ t \in [0, \infty)$$
$$u(x, 0) = u_0(x), \quad x \in T^2$$

A 256×256 grid is used to generate data with sub-sampling for use at various resolutions. Different datasets are available for N training examples with different times T and viscosities in the baseline paper. The model is evaluated against two additional variants to demonstrate the performance. Table 1 shows the data set distribution for various configurations employed for experiments.

Our proposed model shows promising improvements over the reference paper in terms of relative error. This shows that aggregation of spectral features and the standard layer pipeline improves feature regression. A fundamental

advantage of Deep Spectral Aggregation is that it improves functional learning in the temporal dimension and has fewer spatial features. It is because the Navier-Stokes behavior is transient. We observe a significant reduction in relative error, i.e., L2 loss.

2) DARCY FLOW EQUATION

Henry Darcy's law equations describe the flow of liquids through porous materials, and investigations have been made for water flowing through layers of sand. This is analogous to the Fourier laws of other innate sciences, such as thermal equations, Ohm's law, and Fick's law of the diffusion hypothesis. The basic form of Darcy Flow equation defines a relation between the fluxes $q = Q/A$, where Q is the flow rate (m^3/s) of the liquid flowing through $A(m^2)$ when characterizing the permeability of medium k. Dynamic thickness μ with length L and pressure drop Δp.

$$q = -\frac{k}{\mu L}\Delta p$$

We apply the dataset provided by the baseline paper, where solutions to PDEs were formulated using a grid size of 421×421 employing a second-order finite difference scheme. We keep the same settings and dataset for a fair comparison.

Darcy Flow equations are an active area of research and have a variety of applications, such as popular petroleum engineering and coffee brewing solutions. We use the dataset given in the baseline paper where we formulated a solution to the PDE using a grid size of 421×421 using a second-order finite difference system. Keep the same setup and data set for a fair comparison. All other resolutions were extracted from the highest available resolution (e.g., 421×421).

3) BURGERS' EQUATION

We also apply a one-dimensional fundamental equation for cross-validation, referred to as Burgers' equation with boundary conditions. It is a viscous periodic partial differential equation. The model performance was evaluated, and results are recorded as shown in the table. Burgers' equation can be expressed as below in periodic form [53].

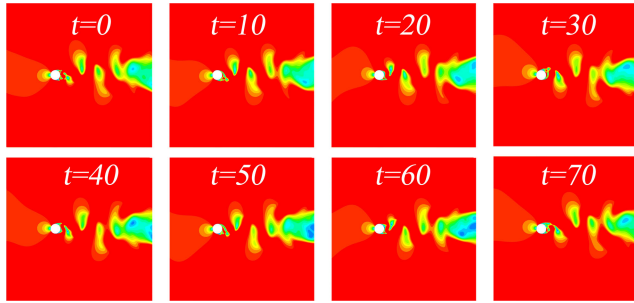$$\frac{dx}{dt} = u, \quad \frac{du}{dt} = 0$$

FIGURE 6. Dataset: samples generated over *t + nT*, where *T* = 10.

With an extended form in periodic form as below

$$\frac{\partial}{\partial t}u(x,t) + \frac{1}{2}\frac{\partial}{\partial x}(u(x,t))^2 = \nu\frac{\partial^2}{\partial x^2}u(x,t), \quad x \in (0, 2\pi),$$
$$t \in (0, \infty))$$
$$u(x, 0) = u_0(x), \quad x \in (0, 2\pi)$$

where fixed viscosity ν is a diffusion term with a value range of 0.1 for time $t > 0$ in our simulations. We adapted the ground truth as available by the baseline paper to compare our results.

4) COMPUTATIONAL FLUID PROBLEM MODEL (CASE STUDY)

We selected case study problem datasets generated for the experiment in a preceding study [26] by a team from Luleo Institute of Technology in Sweden, Yeungnam University in Korea, and Khalifa University Abu Dhabi. The calculation domain size is $30D \times 30D$, where D is the cylinder diameter. The cylinder object is placed $20D$ downstream, and the inlet boundary is at $10D$—the sidewall distance of $15D$ for each cylinder. Create a structural mesh grid using hexahedral elements. An O-grid around the cylinder wall is introduced to maintain a high-quality mesh in the boundary layer thickness region with 50 nodes in this region. The model problem is generated using the Navier-Stokes equation.

A commercial software ANSYS-CFX is utilized to address the transient and incompressible types of the coherence for Navier-Stokes conditions. An average speed at the inlet boundary with normal pressure at the power source is induced and allocated a no-slip condition for sidewalls and create $1,000$ samples for each time t-second in our test dataset.

B. EVALUATION METRICS

The loss function measures the mean squared error (MSE) for 1D, 2D or 3D regression. The MSE is defined as

$$MSE = (\frac{1}{n})\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

where average of n terms is taken from square of the differences of Y_i the observed values and $\hat{Y}_i$ the predicted values.

C. DATASET DISTRIBUTION

Employed dataset distributions for training, validation and test are shown in Table 1.

TABLE 1. Dataset distribution.

Parameter	Train	Validation	Test
Burgers' Equation	1,000	100	100
Darcy Flow 421×421	1,000	100	100
Darcy Flow 211×211	1,000	100	100
Darcy Flow 141×141	1,000	100	100
Darcy Flow 85×85	1,000	100	100
Navier Stokes ν=1e-3	1,000	200	200
Navier Stokes ν=1e-3	5,000	200	200
Navier Stokes ν=1e-4	1,000	200	200
Navier Stokes ν=1e-4	5,000	200	200
Navier Stokes ν=1e-4	10,000	200	200
Computational Fluid Dynamics	1,000	200	200

D. SIMULATION PARAMETERS

Table 2 enlists the parameters employed during our experiments.

TABLE 2. Simulation parameters.

Parameter	Burgers' Equation	Darcy Flow	Navier Stokes
Dimension	1D	2D	2D+T
Epochs	500	500	500
Input Size/Resolution	8192	85 to 421	64
Optimizer	Adam	Adam	Adam
Learning rate	0.001	0.001	0.001
Decay	1×10^{-4}	1×10^{-4}	1×10^{-4}
Scheduler	step scheduling (step size=100)		

E. PERFORMANCE EVALUATION

We evaluate the proposed method using three benchmark datasets and one case study and we measure the MSE score as a performance indicator. We achieved considerable improvements on Burgers' equation dataset where it improved an MSE score by approximately 30% from 0.00081 (baseline) to 0.0005654 (ours). Secondly, the Darcy Flow equation dataset demonstrated an improvement of approximately 37% from 0.0109 (baseline) to 0.0069 (ours). Finally, we observe an improvement of more than 20% for the Navier-Stokes equation dataset with time-dependent data, from 0.0834 (baseline) to 0.0655 (ours).

1) NAVIER-STOKES EQUATION EXPERIMENTS

The deep spectral aggregation approach demonstrated a new state-of-the-art performance in physics-informed deep learning approaches. Our proposed method achieved more than a 20% reduction in relative error for all subject benchmark datasets for the Navier-Stokes equation, i.e., V1e-3-N1000-T50, V1e-3-N5000-T50, V1e-4-N1000-T30, V1e-4-N5000-T30, V1e-4-N10000-T30. The model

TABLE 3. Navier-Stokes benchmarks (Relative error MSE) (fixed resolution 64 × 64 on train and test).

Method	ν=1e-3 T=50 N=1K	ν=1e-3 T=50 N=5K	ν=1e-4 T=30 N=1K	ν=1e-4 T=30 N=5K	ν=1e-4 T=30 N=10K
	Relative error (MSE)				
ResNet	0.0701	-	0.2871	-	0.2311
TF-Net	0.0225	-	0.2253	-	0.1168
U-Net	0.0245	-	0.2051	-	0.1190
FNO-3D	0.0136	0.00586	0.1918	0.1168	0.0820
FNO-2D	0.0128	0.00496	0.1559	0.1010	0.0834
DSFA	**0.0112**	**0.00384**	**0.1458**	**0.0882**	**0.0655**

remained consistent with improved relative error to baseline paper. Navier-Stokes equation dataset is a temporal dataset, and we observe the spectral aggregation notably performed well in temporal datasets, whereas simple 2D Darcy Flow equation dataset did not show such a good response. The baseline paper employed 2D+T and 3D models, from which the 2D+T approach performed better. On the other hand, our proposed model superseded the performance of both of the above variations proposed in the base paper.

We further explored the performance by varying the number of samples (1,000; 5,000; 10,000) per dataset, and the relative error results were observed to be consistently better for the proposed model. Table 3 shows the results recorded from Navier-Stokes experiments. The proposed model outperformed because the model perceives and extracts more temporal features due to the aggregation and compression approaches. Various datasets are shown in columns and networks in rows.

Figure 7 plots the training history of Navier-Stokes over various datasets. The history plot shows that DSFA per-
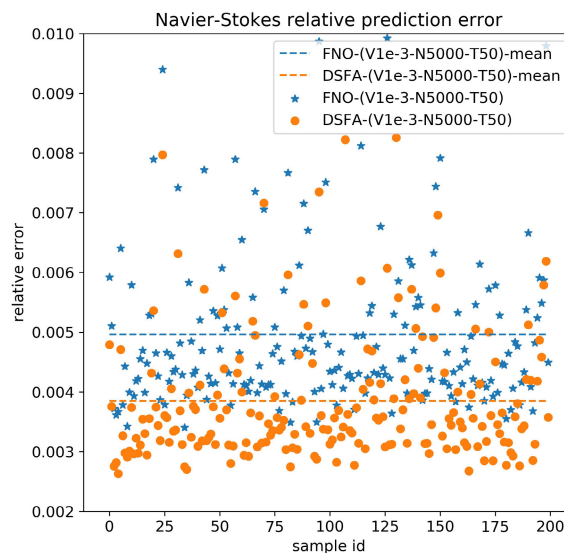

FIGURE 8. Test dataset with 200 samples. Navier-Stokes dataset prediction error with average line. DSFA mean(sample wise error mean) is observed reasonably low, demonstrating the lower error on all 200 samples.
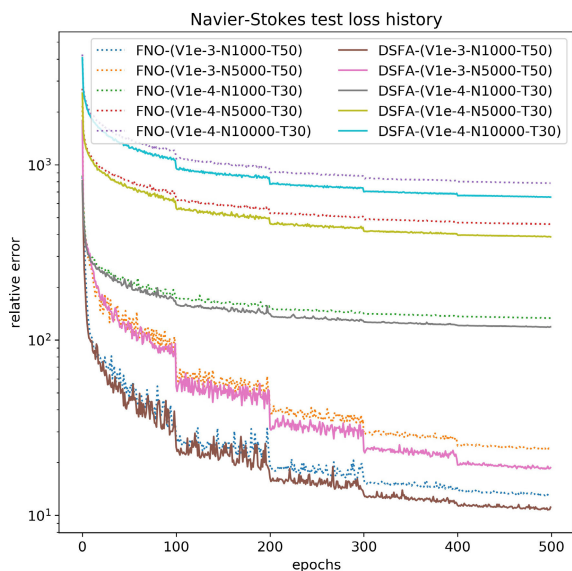
formed much better and demonstrated improvement over the 2D+T/3D volume dataset for Navier-Stokes. We recorded and plotted relative test error in Figure 8. The plot shows orange dots for the proposed model prediction relative error and mean-line for the same. DSFA's relative mean error is lower than the Fourier neural operator model.

We performed experiments on Navier-Stokes equations dataset [8] and recorded the results in Table 3. We employ the dataset with viscosities ν: $1e - 3$ and $1e - 4$ and generated training samples with times $T = 30$ and 50 seconds as discussed in the benchmark problems section. We reproduced baseline results per available code at the GitHub repository and compared them with ours using the same parameters and environment. The results show a performance gain in terms of minimizing the relative error.

2) DARCY FLOW EQUATION EXPERIMENTS
We evaluated our model over the Darcy Flow equation dataset for various resolutions, i.e., 421×421, 211×211, 141×141,

FIGURE 7. Training validation histogram Navier-Stokes datasets.

TABLE 4. Darcy flow equation (Relative error MSE) (resolution size of 421, 211, 141, 85).

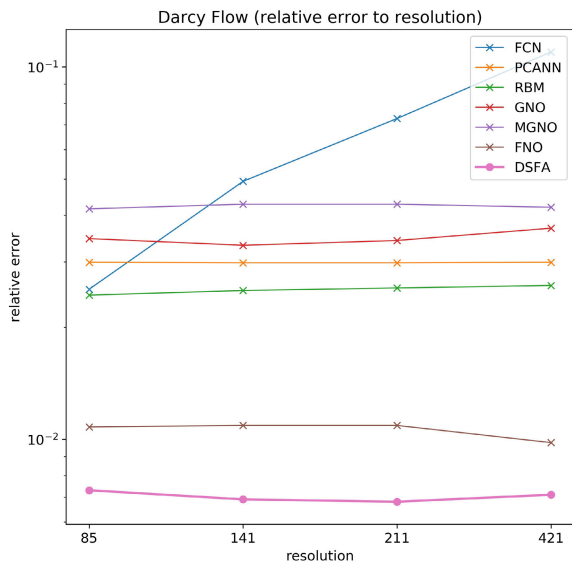| Network	Resolutions/size			
	421×421	211×211	141×141	85×85
	Relative error (MSE)			
LNO	NA	0.0445	0.0461	0.0520
MGNO	0.0420	0.0428	0.0428	0.0416
GNO	0.0369	0.0342	0.0332	0.0346
PCANN	0.0299	0.0298	0.0298	0.0299
FCN	0.1097	0.0727	0.0493	0.0253
RBM	0.0259	0.0255	0.0251	0.0244
FNO	0.0098	0.0109	0.0109	0.0108
DSFA	**0.0073**	**0.0069**	**0.0068**	**0.0071**

FIGURE 9. Darcy Flow relative error over different resolutions, the graph shows DSFA with lowest relative error over various resolutions.

FIGURE 10. Test dataset 100 samples predictions. Orange dots shows overall a lower relative error on the scatter plot. A lower relative error shows a better performance.

and 85×85. The results show that our proposed model outperformed all existing approaches by approximately 20% to the most recent paper, i.e., the baseline paper as shown in Figure 9. The Darcy Flow equation is a two-dimensional dataset and is presented to the model like an image dataset. Therefore, the spectral features are applied to the spatial domain. Due to the nature of the Fourier neural operator, the data is presented in the spatial domain, and then within the layer, it is converted to the spectral domain using the forward operation of the Fourier transform. A complex multiplication takes place, which is transformed back to the spatial domain using a reverse Fourier transform. The process ultimately mimics more like a spatial convolution. Darcy Flow equation dataset at all various resolutions showed that the proposed model is equally performing better in addition to the benchmark dataset of Navier-Stokes, which contains three-dimensional data as two-dimensional spatial and a temporal dimension. We plotted prediction error reported by the model test phase in Figure 10. The plot shows that the proposed model prediction shown as orange dots is reasonably
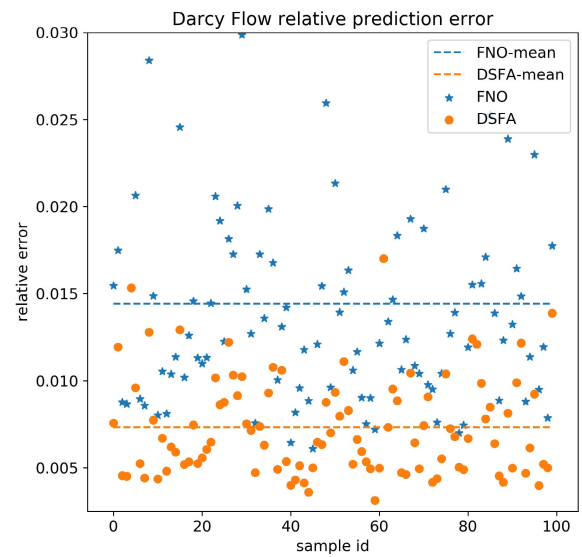
low and is evident from the average line. Table 4 enumerates the results observed over Darcy Flow equation experiments.

### 3) BURGERS' EQUATION EXPERIMENTS
The Burgers' equation is a one-dimensional equation dataset and is a collection of waveforms generated by the equation. The model learns the functional space and regenerates the waveform. The single dimension data is a kind of temporal data and hence shows major improvements in response to the proposed model. Spectral concatenation and feature compression contribute to major performance gain to learn the longer temporal states. We achieve a relative error decrease by approximately 30%. Which shows the superior performance of the proposed model. Recorded results are shown in Table 5 compared with baseline paper.

4) COMPUTATIONAL FLUID DYNAMICS EXPERIMENTS
We performed experiments for a CFD-generated dataset from [26] using ANSYS-CFX software. We compare ground

TABLE 5. Burgers' equation (resolution 8192).

Network	Relative error (MSE)
FNO	0.00081243
DSFA	**0.00056542**
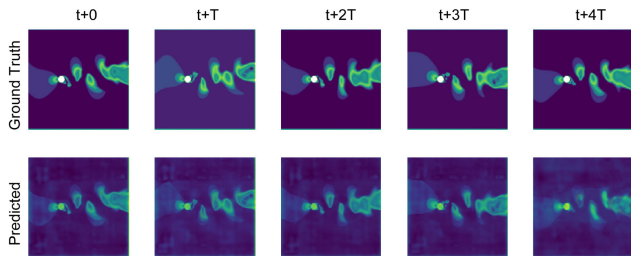
FIGURE 11. CFD - Ground Truth vs. Predictions (T=10).

truth with predictions using our proposed model. The output is encouraging, and reconstruction looks much more meaningful. The experiments exhibit a relative error of 0.00007 for the sequences generated. Although deep learning approaches cannot yet replace or compete the numerical solvers, they are still getting popular over the trade-off on computation cost over accuracy. With this active field of research, we believe further improvement can make it near perfect towards accuracy with a massive benefit of time and computation cost. We plan to optimize further the loss function and a better temporal approach in our future experiments.

5) ANALYSIS WITH NOISY DATA
Partial differential equations are known to be noise-sensitive. We performed experiments over Darcy Flow dataset by adding 10% synthetic generated noise using the random normal distribution to the training data to collect the results. The proposed model demonstrated a minor change in relative error and implicitly canceled the micro noise because of a low pass filter employed with fast Fourier frequency bin selection. The observed results reflect no evident deviation. We obtained a relative error of 0.0071 for the Darcy Flow equation over a resolution of 211, which is 0.0069 without noise experiment. Such a change can occur due to randomness; like training, the model with random seed can reflect slightly different values after every run. Overall the model demonstrates resistance to noise and learns the functional space from noisy inputs with confidence.

V. CONCLUSION AND FUTURE WORK
This study shows that neural operators have started a new direction for physics-informed machine learning. It is a one-shot and straightforward generalized solution for partial differential equations. Various convolutional and recurrent approaches learn and regress the equation-generated data in a conventional sequential method, whereas neural operators

transform the functional space and learn functional features in a single go. Our experiments demonstrate that the aggregation approach followed by spectral compression outperformed the previous state-of-the-art. Although machine learning and deep learning have started to understand the computationally expensive task of numerical solvers, there is a long way to replace the numerical solver. Physics-informed machine learning is still not ready for noisy data; however, Fourier approaches control noisy data using frequency-domain filters. This study is a step to build a model that better understands the parametric partial differential equation functional. However, there is a long way for an artificial intelligence system to approximate perfectly. We plan to try new methods to generalize the models using other mathematical functions like Laplace transform combined with Fourier and conventional convolutional neural networks. The Fourier neural operator-based networks, like a proposed method, can learn computer vision tasks and partial differential equations. Computer vision tasks can employ 2D models like the Darcy Flow dataset and 3D for videos like Navier-Stokes datasets. Therefore, future research on this domain can change the tradition of convolution-only networks.

REFERENCES

[1] Y. Zhu and N. Zabaras, "Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification," *J. Comput. Phys.*, vol. 366, pp. 415–447, Aug. 2018, doi: 10.1016/j.jcp.2018.04.018.

[2] S. Mo, Y. Zhu, N. Zabaras, X. Shi, and J. Wu, "Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media," *Water Resour. Res.*, vol. 55, no. 1, pp. 703–728, Jan. 2019.

[3] M. Tang, Y. Liu, and L. J. Durlofsky, "A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems," *J. Comput. Phys.*, vol. 413, Jul. 2020, Art. no. 109456.

[4] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019, doi: 10.1016/j.jcp.2018.10.045. [Online]. Available: https://ui.adsabs.harvard.edu/abs/2019JCoPh.378..686R/abstract and https://linkinghub.elsevier.com/retrieve/pii/S0021999118307125

[5] Y. Zhu, N. Zabaras, P.-S. Koutsourelakis, and P. Perdikaris, "Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data," *J. Comput. Phys.*, vol. 394, pp. 56–81, Oct. 2019.

[6] 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 Mach. Intell.*, vol. 3, no. 3, pp. 218–229, Mar. 2021. [Online]. Available: http://www.nature.com/articles/s42256-021-00302-5

[7] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Multipole graph neural operator for parametric partial differential equations," in *Proc. Adv. Neural Inf. Process. Syst.*, Jun. 2020, pp. 6755–6766.

[8] Z. Li, N. Kovachki, K. Azizzadenesheli, A. Stuart, and A. Anandkumar, "Fourier neural operator for parametric partial differential equations," in *Proc. ICLR*, Oct. 2020, pp. 1–16.

[9] G. Wen, Z. Li, K. Azizzadenesheli, A. Anandkumar, and S. M. Benson, "U-FNO—An enhanced Fourier neural operator based-deep learning model for multiphase flow," 2021, *arXiv:2109.03697*.

[10] H. Pratt, B. Williams, F. Coenen, and Y. Zheng, "FCNN: Fourier convolutional neural networks," in *Machine Learning and Knowledge Discovery in Databases*, M. Ceci, J. Hollmén, L. Todorovski, C. Vens, and S. Džeroski, Eds. Cham, Switzerland: Springer, 2017, pp. 786–798.

[11] C. L. Jiang, S. Esmaeilzadeh, K. Azizzadenesheli, K. Kashinath, M. Mustafa, H. A. Tchelepi, P. Marcus, M. Prabhat, and A. Anandkumar, "MESHFREEFLOWNET: A physics-constrained deep continuous space-time super-resolution framework," in *Proc. SC, Int. Conf. High Perform. Comput., Netw., Storage Anal.*, Nov. 2020, pp. 1–15. [Online]. Available: https://ieeexplore.ieee.org/document/9355293/

[12] D. Greenfeld, M. Galun, R. Kimmel, I. Yavneh, and R. Basri, "Learning to optimize multigrid PDE solvers," in *Proc. 36th Int. Conf. Mach. Learn. (ICML)*, Jun. 2019, pp. 4305–4316.

[13] D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer, "Machine learning-accelerated computational fluid dynamics," *Proc. Nat. Acad. Sci. USA*, vol. 118, no. 21, May 2021. [Online]. Available: https://www.pnas.org/content/118/21/e2101784118 and https://www.pnas.org/content/118/21/e2101784118.abstract

[14] K. Shukla, A. D. Jagtap, and G. E. Karniadakis, "Parallel physics-informed neural networks via domain decomposition," *J. Comput. Phys.*, vol. 447, Dec. 2021, Art. no. 110683.

[15] A. D. Jagtap and G. E. Karniadakis, "Extended physics-informed neural networks (XPINNs): A generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations," in *Proc. CEUR Workshop*, vol. 2964, 2021, pp. 1–6. [Online]. Available: https://github.com/AmeyaJagtap/XPINNS

[16] K. Li, K. Tang, T. Wu, and Q. Liao, "D3M: A deep domain decomposition method for partial differential equations," *IEEE Access*, vol. 8, pp. 5283–5294, 2020.

[17] E. Weinan and B. Yu, "The deep Ritz method: A deep learning-based numerical algorithm for solving variational problems," *Commun. Math. Statist.*, vol. 6, no. 1, pp. 1–12, 2018, doi: 10.1007/s40304-018-0127-z.

[18] L. Bar and N. A. Sochen, "Unsupervised deep learning algorithm for PDE-based forward and inverse problems," 2019, *arXiv:1904.05417*.

[19] J. D. Smith, K. Azizzadenesheli, and Z. E. Ross, "EikoNet: Solving the eikonal equation with deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10685–10696, Dec. 2021.

[20] X. Guo, W. Li, and F. Iorio, "Convolutional neural networks for steady flow approximation," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 481–490, doi: 10.1145/2939672.2939738.

[21] Y. Zhu and N. Zabaras, "Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification," *J. Comput. Phys.*, vol. 366, pp. 415–447, Aug. 2018.

[22] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," *Inverse Problems*, vol. 33, no. 12, Nov. 2017, Art. no. 124007.

[23] S. Bhatnagar, Y. Afshar, S. Pan, K. Duraisamy, and S. Kaushik, "Prediction of aerodynamic flow fields using convolutional neural networks," *Comput. Mech.*, vol. 64, no. 2, pp. 525–545, Aug. 2019, doi: 10.1007/s00466-019-01740-0.

[24] Y. Khoo, J. Lu, and L. Ying, "Solving parametric PDE problems with artificial neural networks," *Eur. J. Appl. Math.*, vol. 32, no. 3, pp. 421–435, Jun. 2021.

[25] O. Rippel, J. Snoek, and R. P. Adams, "Spectral representations for convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Jan. 2015, pp. 2449–2457.

[26] A. Usman, M. Rafiq, M. Saeed, A. Nauman, A. Almqvist, and M. Liwicki, "Machine learning computational fluid dynamics," in *Proc. Swedish Artif. Intell. Soc. Workshop (SAIS)*, Jun. 2021, pp. 1–4. [Online]. Available: https://ieeexplore.ieee.org/document/9483997/

[27] N. Kovachki, S. Lanthaler, and S. Mishra, "On universal approximation and error bounds for Fourier neural operators," 2021, *arXiv:2107.07562*.

[28] V. Nair, M. Chatterjee, N. Tavakoli, A. Siami Namin, and C. Snoeyink, "Fast Fourier transformation for optimizing convolutional neural networks in object recognition," 2020, *arXiv:2010.04257*.

[29] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, and A. Anandkumar, "Neural operator: Graph kernel network for partial differential equations," 2020, *arXiv:2003.03485*.

[30] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269. [Online]. Available: https://github.com/liuzhuang13/DenseNet

[31] X.-T. Vo, T.-D. Tran, D.-L. Nguyen, and K.-H. Jo, "Stair-step feature pyramid networks for object detection," in *Frontiers of Computer Vision*, H. Jeong and K. Sumi, Eds. Cham, Switzerland: Springer, 2021, pp. 168–175.

[32] F. Yu, D. Wang, E. Shelhamer, and T. Darrell, "Deep layer aggregation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2403–2412.

[33] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Netw.*, vol. 2, no. 5, pp. 359–366, Dec. 1989.

[34] R. Monteiro, C. Bastos-Filho, M. Cerrada, D. Cabrera, and R.-V. Sanchez, "Convolutional neural networks using Fourier transform spectrogram to classify the severity of gear tooth breakage," in *Proc. Int. Conf. Sensing, Diagnostics, Prognostics, Control (SDPC)*, Aug. 2018, pp. 490–496.

[35] Y. Bengio and Y. Lecun, "Scaling learning algorithms toward AI," *Large-Scale Kernel Mach.*, vol. 34, no. 5, pp. 1–41, Aug. 2019.

[36] Y. Fan, J. Feliu-Fabà, L. Lin, L. Ying, and L. Zepeda-Núñez, "A multiscale neural network based on hierarchical nested bases," *Res. Math. Sci.*, vol. 6, no. 2, pp. 1–28, Jun. 2019, doi: 10.1007/s40687-019-0183-3.

[37] V. Sitzmann, J. N. Martel, A. W. Bergman, D. B. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Proc. Adv. Neural Inf. Process. Syst.*, Dec. 2020, pp. 1–12.

[38] X. Zhu, Z. Cui, T. Zhang, Y. Li, and J. Yang, "Going deeper in frequency convolutional neural network: A theoretical perspective," 2021, *arXiv:2108.05690*.

[39] W. Li, K. Wang, and L. You, "A deep convolutional network for multitype signal detection and classification in spectrogram," *Math. Problems Eng.*, vol. 2020, pp. 1–16, Sep. 2020.

[40] M. Cranmer and S. Ho, "Physics-informed inductive biases in deep learning," in *Proc. 35th Conf. Neural Inf. Process. Syst. (NeurIPS)*, 2021. Accessed: Jan. 16, 2022. [Online]. Available: https://neurips.cc/Conferences/2021/ScheduleMultitrack?event=30598

[41] P. Ren, C. Rao, Y. Liu, J. Wang, and H. Sun, "PhyCRNet: Physics-informed convolutional-recurrent network for solving spatiotemporal PDEs," 2021, *arXiv:2106.14103*.

[42] S. Guan, K.-T. Hsu, and P. V. Chitnis, "Fourier neural operator networks: A fast and general solver for the photoacoustic wave equation," 2021, *arXiv:2108.09374*.

[43] L. Zhang, T. Luo, Y. Zhang, W. E, Z.-Q. John Xu, and Z. Ma, "MOD-Net: A machine learning approach via model-operator-data network for solving PDEs," 2021, *arXiv:2107.03673*.

[44] Y. Chen, B. Hosseini, H. Owhadi, and A. M. Stuart, "Solving and learning nonlinear PDEs with Gaussian processes," *J. Comput. Phys.*, vol. 447, 2021. [Online]. Available: https://arxiv.org/abs/2103.12959, doi: 10.1016/j.jcp.2021.110668.

[45] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, and J. Xia, "Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: Evaluation of the diagnostic accuracy," *Radiology*, vol. 296, no. 2, pp. E65–E71, 2020, doi: 10.1148/radiol.2020200905.

[46] J. Blechschmidt and O. G. Ernst, "Three ways to solve partial differential equations with neural networks—A review," *GAMM-Mitteilungen*, vol. 44, no. 2, Jun. 2021, Art. no. e202100006, doi: 10.1002/gamm.202100006.

[47] L. Zhang, T. Luo, Y. Zhang, W. E, Z.-Q. John Xu, and Z. Ma, "MOD-Net: A machine learning approach via model-operator-data network for solving PDEs," 2021, *arXiv:2107.03673*.

[48] N. Roberts, M. Khodak, T. Dao, L. Li, C. Ré, and A. Talwalkar, "Learning operations for neural PDE solvers," in *Proc. ICLR SimDL Workshop*, 2021. Accessed: Jan. 16, 2022.

[49] T. Marwah, Z. C. Lipton, and A. Risteski, "Parametric complexity bounds for approximating PDEs with neural networks," 2021, *arXiv:2103.02138*.

[50] M. Mathieu, M. Henaff, and Y. LeCun, "Fast training of convolutional networks through FFTS," in *Proc. 2nd Int. Conf. Learn. Represent., (ICLR) Conf. Track Proc.*, Dec. 2014, pp. 1–9.

[51] T. D. Ryck, S. Lanthaler, and S. Mishra, "On the approximation of functions by tanh neural networks," *Neural Netw.*, vol. 143, pp. 732–750, Nov. 2021.

[52] A. Y. Sun, H. Yoon, C.-Y. Shih, and Z. Zhong, "Applications of physics-informed scientific machine learning in subsurface science: A survey," 2021, *arXiv:2104.04764*.

[53] J. M. Burgers, "A mathematical model illustrating the theory of turbulence," *Adv. Appl. Mech.*, vol. 1, pp. 171–199, Jan. 1948.

MUHAMMAD RAFIQ (Graduate Student Member, IEEE) received the M.S. degree in electronics engineering from International Islamic University, Pakistan, in 2008. He is currently pursuing the Ph.D. degree with Yeungnam University, South Korea.

He has extensive industry experience with a background in databases, business applications, and industrial technical solutions. His research interests include smart sensors, computer vision, medical image analysis, and video description using deep learning.

GYU SANG CHOI (Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, USA, in 2005.

He was a Research Staff Member of the Samsung Advanced Institute of Technology (SAIT), Samsung Electronics, from 2006 to 2009. Since 2009, he has been a Faculty Member of the Department of Information and Communication, Yeungnam University, South Korea. His research interests include data mining and deep learning.

• • •
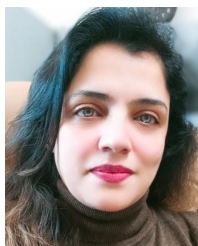
GHAZALA RAFIQ received the B.Sc. degree in mathematics from Punjab University, Lahore, Pakistan, in 2000, and the master's degree in computer science, in 2002. She is currently pursuing the Ph.D. degree with the Data Sciences Laboratory, Department of Information and Communication Engineering, Yeungnam University, Republic of Korea.

She has over 15 years of industry experience. Her research interests include deep learning applications, video description, reinforcement learning, natural language processing, computer vision, and pattern recognition.