



Available online at www.sciencedirect.com

ScienceDirect

Comput. Methods Appl. Mech. Engrg. 407 (2023) 115944

**Computer methods
in applied
mechanics and
engineering**

www.elsevier.com/locate/cma

An unsupervised latent/output physics-informed convolutional-LSTM network for solving partial differential equations using peridynamic differential operator

Arda Mavi^a, Ali Can Bekar^a, Ehsan Haghigat^b, Erdogan Madenci^{a,*}

^a *The University of Arizona, Tucson, AZ 85721, USA*

^b *Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

Received 20 October 2022; received in revised form 4 February 2023; accepted 5 February 2023

Available online 21 February 2023

Abstract

This study presents a novel unsupervised convolutional Neural Network (NN) architecture with nonlocal interactions for solving Partial Differential Equations (PDEs). The nonlocal Peridynamic Differential Operator (PDDO) is employed as a convolutional filter for evaluating derivatives of the field variable. The NN captures the time-dynamics in smaller latent space through encoder-decoder layers with a Convolutional Long-short Term Memory (ConvLSTM) layer between them. The ConvLSTM architecture is modified by employing a novel activation function to improve the predictive capability of the learning architecture for physics with periodic behavior. The physics is invoked in the form of governing equations at the output of the NN and in the latent (reduced) space. By considering a few benchmark PDEs, we demonstrate the training performance and extrapolation capability of this novel NN architecture by comparing against Physics Informed Neural Networks (PINN) type solvers. It is more capable of extrapolating the solution for future timesteps than the other existing architectures.

© 2023 Elsevier B.V. All rights reserved.

Keywords: Deep learning; Peridynamic differential operator; Convolutional-recurrent learning

1. Introduction

Partial Differential Equations (PDEs) are crucial tools to understand and describe phenomena such as sound, diffusion, electrodynamics and fluid dynamics. However, obtaining analytical solutions to most PDEs is challenging and may be not possible. Therefore, variety of numerical techniques for solving different PDEs are developed, including Finite Element Method (FEM), Finite Difference Method (FDM), and Spectral Methods to name a few [1–4]. Although these methods are well established and capable of solving a wide range of PDEs, the use of deep learning-based solution techniques has become attractive in recent years due to their inherent capability in handling data and their advantage as surrogate models.

The deep-learning based PDE solvers rely on the Fourier NN [5] and universal approximation theorem for neural networks [6]. A standard multilayer perceptron approximating a given function with arbitrary error is one of the

* Corresponding author.

E-mail addresses: amavi@arizona.edu (A. Mavi), acbekar@arizona.edu (A.C. Bekar), ehsanh@mit.edu (E. Haghigat), madenci@arizona.edu (E. Madenci).

earliest efforts to solve Ordinary Differential Equations (ODEs) and PDEs using a deep neural network (DNN) [7]. Recently, Berg and Nyström [8] extended this idea to domains with complex geometries.

Han et al. [9] reformulated PDEs as backward stochastic PDEs and approximated the gradient of the solution using DNNs. They were able to solve high dimensional PDEs and overcome the curse of dimensionality. A new paradigm shift for solving PDEs using DNNs emerged with the development of Physics Informed Neural Network (PINN) [10]. PINN benefits from the performance and accuracy of automatic differentiation [11]. The inputs of the NN are space–time independent variables and the partial derivatives of the output with respect to inputs are compiled into a physics-based loss term in the form of a PDE.

If available, known values of the field are included in the loss term, and initial and boundary conditions are enforced by adding corresponding terms to the loss in the form of mean squared error. The parameters of the feed forward NN are optimized by minimizing these losses. PINNs show extraordinary performance for interpolation and surrogate modeling. They have been successfully applied to solve many problems [12–25] with convergence and error analyses [26–28].

However, the training of PINNs still presents challenges despite the extensive improvements by Wang et al. [29–31] and Chen et al. [32]. Also, the convergence of PINNs can be slow for complex problems with multiple governing equations. Another issue is the extrapolation capability of PINNs. Even though PINNs are successful interpolators, their extrapolation performance is rather poor. Predicting a solution outside the training time interval gives unsatisfactory results [33]. A recent work from Lu et al. [34] introduces DeepOnets which is based on learning the underlying operators of the governing equations. Oommen et al. [35] proposed a hybrid autoencoder network to predict the time evolution of two-phase microstructures by learning the dynamics in latent space using a DeepOnet architecture.

Recently, significant attention is concentrated on recurrent learning based PDE solvers because of their robustness for time-extrapolation [33,36,37]. One of the fundamental recurrent learning structures used for solving PDEs is Recurrent Neural Networks (RNNs). They are designed to capture long-time behavior of time-series inputs [38]. However, their training is challenging because of vanishing or exploding gradients [39]. Long–short Term Memory (LSTM) structure is proven to be easier to train compared to RNNs due to their ingenious gated structure [40]. However, the basic LSTM architecture is a fully-connected architecture and it does not conserve spatio-temporal dependencies. On the other hand, NNs with convolutional layers have proven to perform better on data sets with spatial dependencies and easier to train [41]. Recently, Shi et al. [42] introduced Convolutional LSTM (ConvLSTM) architecture and demonstrated its superior performance to LSTMs for capturing data with spatio-temporal correlations. Ren et al. [33] proposed a PDE solver using ConvLSTM by employing FDM to estimate physical losses.

There exist a few studies on the performance of physics informed nonlocal algorithms. Pang et al. [43] extended the idea of PINNs to nonlocal integral equations. Xu et al. [44] and Barros et al. [45] presented methods for learning nonlocal kernels from field observations. Burkova et al. [46] developed a framework for the identification of the fractional kernel parameters. Bekar and Madenci [47] considered learning PDEs from experimental observations by adopting a nonlocal model. However, a recurrent learning based PDE solver leveraging features of a nonlocal method does not exist.

Extending the work of Ren et al. [33], we present a novel unsupervised convolutional NN architecture with nonlocal interactions for solving PDEs. Instead of FDM, we enforce the physical governing equations using Peridynamic Differential Operator (PDDO) [48] as convolutional filters. Unlike FDM, PDDO accounts for nonlocal phenomena with arbitrary horizon size. Therefore, the filters can be generic and of any size.

We employ LSTM because it captures long range time evolution of physical systems [49]. Vanilla RNN architecture can also be used; however, the vanishing or exploding gradients make them unsuitable for learning data for longer time periods [38,50]. In order to prevent overfitting, we eliminate redundancies in input by using an encoder–decoder architecture. The encoder eliminates 50% of the input pixels and decreases the workload of ConvLSTM; thus, reducing the chances of overfitting.

ConvLSTM architecture is shown to be successful at capturing time evolution of spatiotemporal data [42] without suffering the optimization drawbacks affecting the simple RNN architecture. However, LSTMs are more complex models compared to RNNs and their training is more computationally expensive. Also, modifying LSTM architecture does not improve training performance significantly [50] unless there are inductive biases that can be directly imposed. In this study, we decrease the size of the input data by using an encoder architecture to alleviate

the computational burden of the ConvLSTM. We also accelerate its convergence for periodic data by introducing a novel activation function at the state update gate. There are other potential recurrent learners such as transformer neural networks [51] or gated recurrent units (GRU)s [52] which might offer better performance. However, we consider ConvLSTM as our recurrent learner to leverage the already available rich literature about LSTMs.

The current approach leverages the nonlocal nature of PDDO for more efficient convolutional filter design. The partial derivative filters created using finite difference method are sparse and only account for local interactions. On the other hand, partial derivative filters created using PDDO includes nonlocal interactions with more efficiency. The convolution operation using a uniform discretization is Eulerian by default and suitable for PDEs in Eulerian reference frame. Therefore, it requires modification for PDEs in Lagrangian reference frame. There exist approaches combining phase field modeling and deep learning to model fracture such as that of Goswami et al. [53] by employing the Eulerian reference frame. The Lagrangian reference frame enables full advantage of the inherent capability of PD theory. Therefore, the current architecture requires modification for Lagrangian reference frame in order to consider fracture problems.

Our neural network has encoder-decoder layers with a ConvLSTM layer between them for capturing time-dynamics in a smaller latent space. Inspired by reduced order modeling [54,55], we also enforce the physical losses as a constraint in the latent (reduced) space. Furthermore, we modify the ConvLSTM architecture by employing a novel activation function [56] to improve the predictive capability of the present learning architecture for physics with periodic behavior. Finally, we demonstrate its improved training performance and better extrapolation capability by comparing against PINN type solvers.

This study is organized as follows. In Section 2, we explain the specific form of the PDEs and provide the details of the present NN architecture and the solution strategy. In Section 3, we briefly describe the PDDO and explain the structure of the derivative filters. In Section 4, we present numerical experiments and report performance comparison between PINN and the present NN architecture. Finally, we discuss the results and summarize the main conclusions in Section 5.

2. Problem statement and methodology

In this study, we consider the PDE systems in the following form

$$\mathbf{u}_t + \mathcal{N}[\mathbf{u}] = 0, \quad \mathbf{x} \in \Omega, t > 0 \quad (2.1a)$$

$$\mathbf{u}(\mathbf{x}, t) \in \mathbb{R}_{per} \quad \forall t > 0 \quad (2.1b)$$

and

$$\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}) \quad (2.1c)$$

where \circ_t denotes partial derivative with respect to time and $\mathcal{N}[\circ]$ is a nonlinear differential operator acting on \mathbf{u} . The physical domain is denoted by Ω and \mathbb{R}_{per} indicates that the solution to the PDE satisfies periodicity in \mathbb{R}^d , i.e. the solution satisfies periodic boundary conditions and $\mathbf{u}_0(\mathbf{x})$ is the initial condition. In a 2-dimensional domain, the unknown vector includes 2 dependent variables, $\mathbf{u} = \{u, v\}$. The shape of the domain is square with a uniform discretization.

2.1. Network architecture

Motivated by the work of Ren et al. [33], our neural network depicted in Fig. 1 consists of 3 main components: (1) convolutional encoder, (2) ConvLSTM, and (3) convolutional decoder. The encoder part projects a given input to a smaller latent space by extracting the hidden representations. This part removes half of the input pixels to decrease the computational burden of the recurrent learning unit, ConvLSTM. The ConvLSTM part encodes the time dynamics of the process between the time steps while capturing spatial correlations in the data. This unit takes the reduced output of the encoder and passes the temporal evolution information to next layer. The decoder takes the accumulated temporal state of the system from the ConvLSTM and expands it to the same dimension as that of the input of the NN. Subsequently, this output is multiplied with the time step of δt and added to the input which presents the integrated field value at the next time step, i.e., $\mathbf{u}_i = \mathbf{u}_{i-1} + \delta t \mathcal{N}[\mathbf{u}_{i-1}, \lambda]$. Therefore, this neural network functions as a time integrator. The output of every sequence is passed to the next layer as input. Thus, this architecture in essence can be thought of as an autoregressive model.

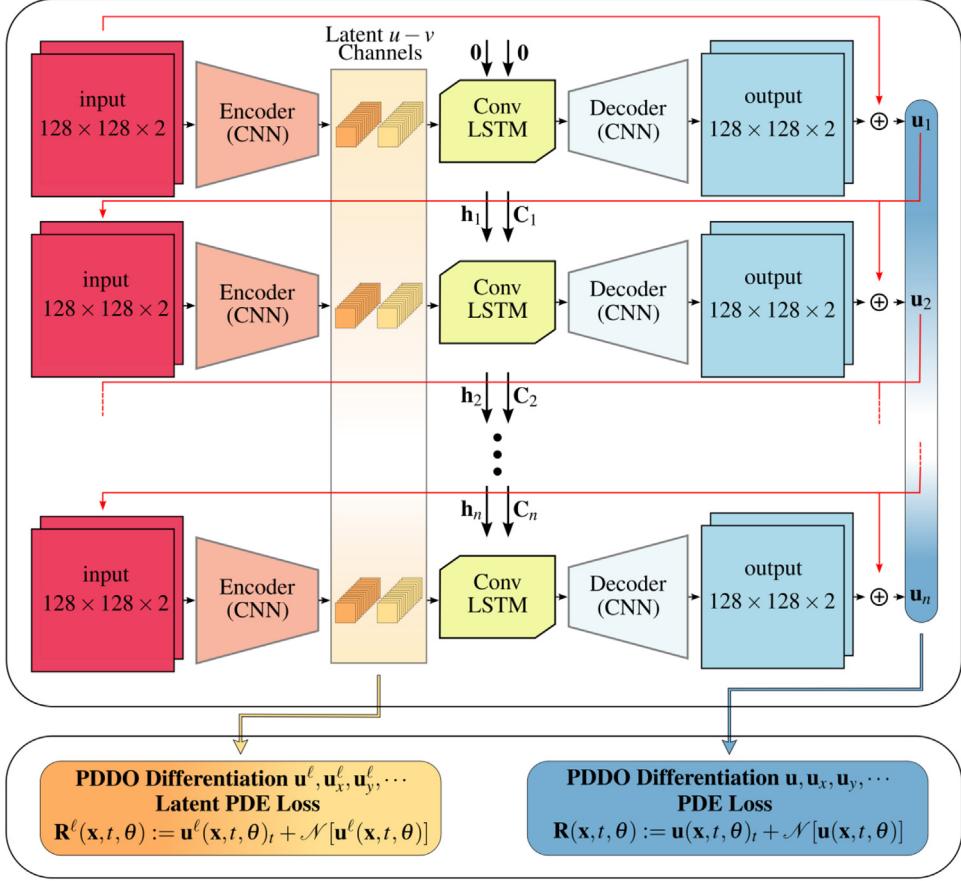
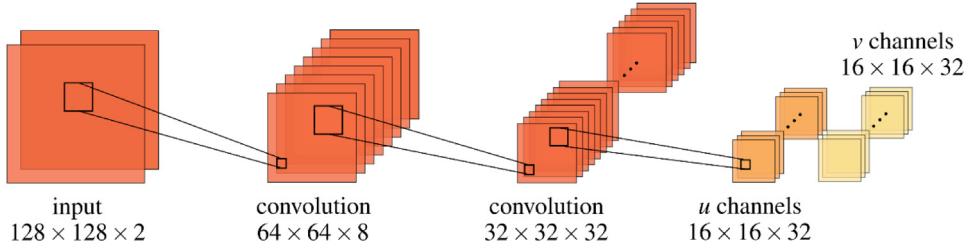
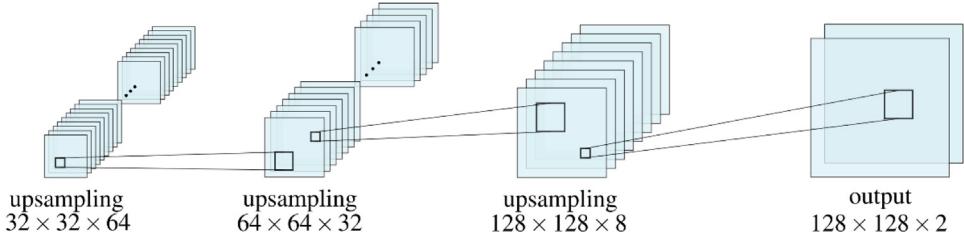


Fig. 1. Proposed network architecture.

Using the discretely integrated field variables, we construct the time step loss by imposing the discretized PDEs. Requiring that physical governing equations are satisfied at any scale, the present architecture also constrains the latent space to satisfy the governing relations. Therefore, the learning process is unsupervised. This unsupervised framework is trained by stacking the layers and minimizing the physics loss at every timestep all at once.

2.2. The encoder-decoder architecture

In this study, the encoder consists of 3 convolutional layers with hyperbolic-tangent activation functions, and 1×1 periodic padding as shown in Fig. 2. First layer uses 8 filters with 4×4 kernels and 2 strides, second layer has 32 filters with 4×4 kernels and 2 strides and the third layer has 64 filters with 4×4 kernels and 2 strides. The output tensor of these three layers has dimensions of $16 \times 16 \times 64$. This tensor is split into two tensors with dimensions of $16 \times 16 \times 32$. These two tensors contain the lower dimensional representations of field variables. The physics informed PDE residual is enforced separately as explained in Section 2.4. Unlike the architecture proposed by Ren et al. [33], the present structure compresses the input using the autoencoder, i.e., extracts smaller features in the latent space. This modification increases the generalization capability of the NN by preventing the NN from memorizing the field data, i.e., overfitting. This modification also shrinks the input data so that the computational burden on ConvLSTM decreases. Fig. 2 shows the structure of the encoder part of the present NN. It has 3 layers with hyperbolic-tangent activation functions and 1×1 periodic padding. First layer uses 64 filters with 3×3 kernels and 1 stride, second layer has 32 filters with 3×3 kernels and 1 stride and the third layer has 8 filters with 3×3 kernels and 1 stride. The output tensor has dimensions of $128 \times 128 \times 2$ so that the input size is recovered. Fig. 3 shows the structure of the encoder part of the NN. Kernel is the applied filter to a channel with constant size

**Fig. 2.** Encoder architecture and tensor sizes.**Fig. 3.** Decoder architecture and tensor sizes.

and it extracts the features of a channel. Padding is the extent of pixels added to the channel edges when the kernel is applied. Stride is the level of shift (movement) of a filter while calculating the convolution between the channel and the kernel.

The common kernel sizes for convolutional neural networks are 3×3 , 4×4 and 5×5 . However, the appropriate size of the kernel is dictated by the characteristics of the input. The function of a convolutional layer is to detect local conjunctions in input. Therefore, smaller local motifs (details) in input requires a smaller kernel size to capture the details and features of the input. Each filter of the convolutional layer captures a different feature of the input. In general, kernel sizes, number of filters and strides are hyperparameters specified through trial and error. The expressional power of the network increases with increasing number of filters. However, number of filters increases the number of trainable parameters and chances of overfitting. Therefore, the purpose of the encoder in this study is to compress the data to a smaller latent space by removing irrelevant detail from the input. It reduces the input size by 50% with filter sizes corresponding to 3×4 or 4×4 filters. Furthermore, the compression of the encoder is designed to be gradual since abrupt changes in sizes between layers can degrade the performance of the NN. The purpose of the decoder is to expand the output of the ConvLSTM to the input size of the neural network for performing time integration using the neural network. The decoder architecture is similar to that of the encoder architecture. The procedure for appropriate values of the hyperparameters is explained in [Appendix E](#).

2.3. Periodic convolutional LSTM

As depicted in [Fig. 4](#), the convolutional LSTM cell architecture is expressed mathematically as follows:

$$\mathbf{i}_t = \sigma(W_{xi} * \mathbf{x}_t + W_{hi} * \mathbf{h}_{t-1} + W_{ci} * \mathbf{C}_{t-1} + b_i) \quad (2.2a)$$

$$\mathbf{f}_t = \sigma(W_{xf} * \mathbf{x}_t + W_{hf} * \mathbf{h}_{t-1} + W_{cf} * \mathbf{C}_{t-1} + b_f) \quad (2.2b)$$

$$\tilde{\mathbf{C}}_t = \Xi_\alpha(W_{xc} * \mathbf{x}_t + W_{hc} * \mathbf{h}_{t-1} + b_c) \quad (2.2c)$$

$$\mathbf{C}_t = \mathbf{f}_t \circ \mathbf{C}_{t-1} + \mathbf{i}_t \circ \tilde{\mathbf{C}}_t \quad (2.2d)$$

$$\mathbf{o}_t = \sigma(W_{xo} * \mathbf{x}_t + W_{ho} * \mathbf{h}_{t-1}) \quad (2.2e)$$

and

$$\mathbf{h}_t = \mathbf{o}_t \circ \tanh(\mathbf{C}_t) \quad (2.2f)$$

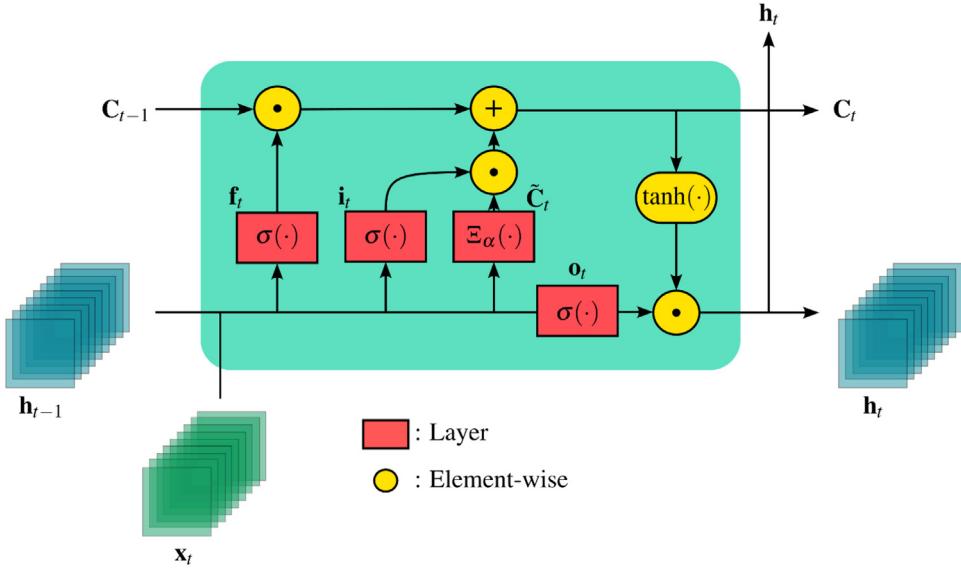


Fig. 4. The modified ConvLSTM cell.

in which, x_t is the input, C_t is the cell state, \tilde{C}_t is the internal cell state, h_t is the hidden state, i_t is the input gate, f_t is the forget gate and o_t is the output gate. The operations, ‘*’ and ‘ \circ ’ denote the convolution operation and element-wise multiplication (Hadamard product), respectively.

The hidden state h_t is computed using the recent input of the ConvLSTM, x_t and previous hidden state, h_{t-1} . It is the output of the ConvLSTM cell and stores the information about the steps up to time, t . Cell state C_t passes the information between steps and chooses what to propagate to the future. It is more local compared to hidden state of the model and its long-term influence is stored in hidden state. Internal cell state \tilde{C}_t stores the information that can be passed to the cell state. The Hadamard product of internal cell state and input gate is added on to the cell state.

Forget gate decides whether to pass the information to the cell state or to “forget” it by filtering that information. Input gate decides on the information added to the cell state from the internal cell state. Output gate decides on the features passed to the hidden state from the updated cell state.

There exist state-of-the-art activation functions [56–60], instead of the commonly accepted hyperbolic-tangent function, we employ the periodic activation function $\Xi_\alpha(x) = x + \frac{1}{\alpha} \sin^2(\alpha x)$ to account for periodicity [56]. Unlike the previous study [56], we do not assign a fixed value of α heuristically. In this study, α is as a trainable parameter whose training interval can be arbitrary for the periodic training data. However, the output of the activation function oscillates fast for large values of α which is not desirable for a nonperiodic input data. In order to consider periodic and nonperiodic data, the range of parameter α is limited between -2π and 2π . Hence, it is initialized randomly between -2π and 2π .

This modification helps the network learn faster when periodic solutions are present. It is also capable of representing nonperiodic and decaying behavior. Appendix B contains further information about the periodic activation function. ConvLSTM has kernel sizes of 3×3 with 1×1 periodic padding. Fig. 4 depicts the structure of the novel ConvLSTM cell.

2.4. Output and latent PDE losses

The learning algorithm is unsupervised, i.e., we do not use any labeled data. However, we use the governing equation of the field to construct the loss. The output (standard) loss has the following form

$$\mathbf{R}(x, t; \boldsymbol{\theta}) := \mathbf{u}(x, t; \boldsymbol{\theta})_t + \mathcal{N}[\mathbf{u}(x, t; \boldsymbol{\theta})] \quad (2.3)$$

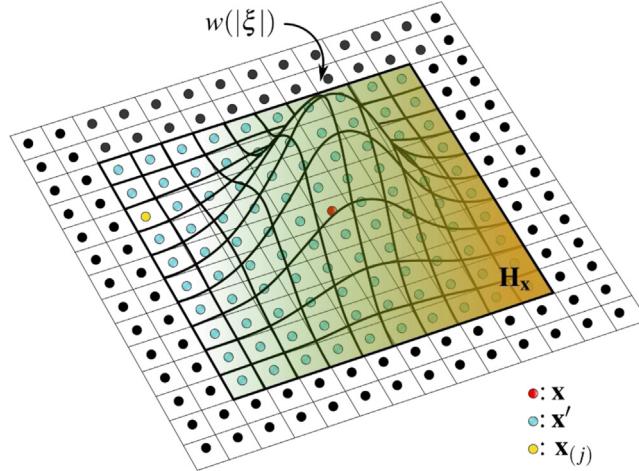


Fig. 5. Interaction domain, \mathbf{H}_x of point \mathbf{x} includes its family members, $\mathbf{x}_{(j)}$.

where $\boldsymbol{\theta}$ represents the NN parameters and $\mathcal{N}[\circ]$ represents the differential operator which often include derivatives and products of derivatives.

Motivated by the fact that physical laws hold true at any scale, we also constrain the latent (reduced) space with a similar PDE-loss as shown in Fig. 1. The loss is similar to Eq. (2.3); however, it acts on a different scale and only affects the encoding part of the NN. The latent loss has the following form

$$\mathbf{R}^\ell(\mathbf{x}, t; \boldsymbol{\theta}) := \mathbf{u}^\ell(\mathbf{x}, t; \boldsymbol{\theta})_t + \mathcal{N}[\mathbf{u}^\ell(\mathbf{x}, t; \boldsymbol{\theta})] \quad (2.4)$$

where \circ^ℓ denotes the quantity in latent space. The remaining challenge is to evaluate a loss function that includes differential operators. Among NN based solvers, PINN uses automatic differentiation to calculate the partial derivatives of the field variable of desired order [10]. PhyCRNet uses convolutional filters based on FDM to evaluate the partial derivatives [17,18,29–33]. As discussed in detail in the next section, we use the Peridynamic Differential Operator (PDDO) [48] to construct the derivative filters. Unlike FDM that can only account for interaction between points in close proximity, PDDO can handle an arbitrary filter size. Hence, it is suitable for both local and nonlocal approximations. Therefore, it can be viewed as a generalization of PhyCRNet.

3. Nonlocal convolutional derivative filters using PDDO

To evaluate derivatives at a point of interest \mathbf{x} , PDDO uses a family of points, \mathbf{H}_x shown in Fig. 5. It is defined as $\mathbf{H}_x = \{ \mathbf{x}' | w(\mathbf{x}' - \mathbf{x}) > 0 \}$. Each point \mathbf{x} has its own unique family in its domain of interaction. The non-dimensional weight function defining the degree of interaction between points is defined as

$$w(|\xi|) = e^{-4|\xi|^2/\delta^2} \quad (3.1)$$

where $\xi = \mathbf{x}' - \mathbf{x}$ is the relative position of a point \mathbf{x}' with the point of interest, \mathbf{x} . The parameter δ , a.k.a, the horizon, defines the extent of the interaction domain (long-range interactions). In discrete form, the family members of point \mathbf{x} are denoted as $\mathbf{H}_x = (\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(N)})$, and their relative positions are defined as $\xi_{(j)} = \mathbf{x}_{(j)} - \mathbf{x}$.

As detailed by Madenci et al. [48], the derivation of PDDO employs Taylor Series Expansion (TSE), weighted integration and orthogonality conditions (see Appendix A). Hence, the nonlocal PD representation of function $f(\mathbf{x})$ and its first and second order derivatives in 2D, using 2nd order Taylor Series Expansion (TSE), can be expressed in continuous and discrete forms as

$$f(\mathbf{x}) = \int_{\mathbf{H}_x} f(\mathbf{x} + \xi) g_2^{00}(\xi) dA \approx \sum_{\mathbf{x}_{(j)} \in \mathbf{H}_x} f(\mathbf{x}_{(j)}) g_2^{00}(\xi_{(j)}) A_{(j)}, \quad (3.2)$$

$$\begin{Bmatrix} f_x \\ f_y \end{Bmatrix} = \int_{\mathbf{H}_x} f(\mathbf{x} + \xi) \begin{Bmatrix} g_2^{10}(\xi) \\ g_2^{01}(\xi) \end{Bmatrix} dA \approx \sum_{\mathbf{x}_{(j)} \in \mathbf{H}_x} f(\mathbf{x}_{(j)}) \begin{Bmatrix} g_2^{10}(\xi_{(j)}) \\ g_2^{01}(\xi_{(j)}) \end{Bmatrix} A_{(j)}, \quad (3.3)$$

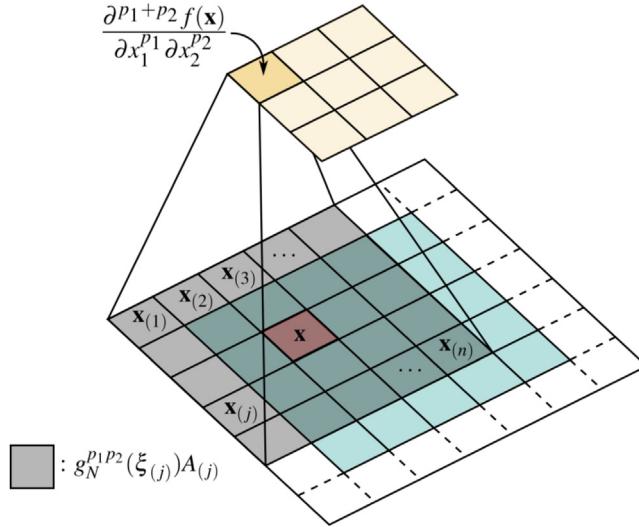


Fig. 6. The general form of Peridynamic convolutional filters.

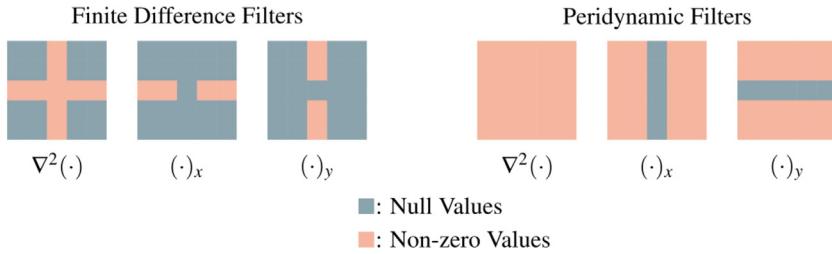


Fig. 7. Comparison of sparsity for PDDO and FDM filters.

and

$$\begin{Bmatrix} f_{xx} \\ f_{yy} \\ f_{xy} \end{Bmatrix} = \int_{\mathbf{H}_x} f(\mathbf{x} + \boldsymbol{\xi}) \begin{Bmatrix} g_2^{20}(\boldsymbol{\xi}) \\ g_2^{02}(\boldsymbol{\xi}) \\ g_2^{11}(\boldsymbol{\xi}) \end{Bmatrix} dA \approx \sum_{\mathbf{x}_{(j)} \in \mathbf{H}_x} f(\mathbf{x}_{(j)}) \begin{Bmatrix} g^{20}(\boldsymbol{\xi}_{(j)}) \\ g^{02}(\boldsymbol{\xi}_{(j)}) \\ g^{11}(\boldsymbol{\xi}_{(j)}) \end{Bmatrix} A_{(j)}, \quad (3.4)$$

where $g_2^{p1 p2}(\boldsymbol{\xi})$ with $(p, q = 0, 1, 2)$ represent the PD functions obtained by enforcing the orthogonality condition of PDDO [48,61], and the integration is performed over the interaction domain.

3.1. PDDO-CNN filters

Discrete summation in Eqs. (3.2)–(3.4) can be considered as a convolution operation. The PD functions, $g_2^{p1 p2}(\boldsymbol{\xi})$ form the basis of the convolutional filters. Fig. 6 depicts the PD integration as a convolutional operation.

The PDDO filters are intrinsically more efficient because of the arbitrary size of spatial area they cover. Fig. 7 shows the difference between PDDO filters and FDM filters. Because of the locality of the FDM filters, most of the values in the convolutional filter are zero. On the other hand, due to nonlocality, PDDO filters have mostly non-zero values; thus, they incorporate more information from the surrounding pixels or family member values. Hence, PDDO derivative filters are genuine convolutional filters.

The spatial derivatives are evaluated using PDDO-Convolutional filters. To determine the time derivative at the output, we use a filter based on a second order finite difference scheme as

$$\circ_t \approx [-1, 0, 1] \times \frac{1}{2\Delta t} \quad (3.5)$$

Consequently, all necessary spatial and temporal filters are established to evaluate the derivatives in the output and latent loss functions.

4. Numerical results

This section consists of the numerical experiments conducted to test the present neural network architecture. It presents solutions to 2D Coupled Burgers' equation, $\lambda - \omega$ Reaction–Diffusion equation and Gray–Scott equation. Training losses for these numerical experiments are shown in [Appendix C](#). We use the same discretization for all cases and compare our interpolation and extrapolation results with those of PINN. The solutions are constructed on (AMD EPYC 7642 48-core) CPU with 2.4 GHz clock speed and 64 GB RAM. The code for our network architecture is written in Python 3.6.13 using TensorFlow version 2.5.2; the remaining library requirements and code are available at: <https://github.com/ardamavi/PI-rCNN>.

4.1. Burgers' equation

Two-dimensional coupled nonlinear Burgers' equation is commonly used to verify numerical solvers. It can model traffic flow, hydrodynamic turbulence, vorticity transport, fluid flow with shock and numerous other phenomena. Burgers' equation in vector form can be expressed as

$$\mathbf{u}_t + \mathbf{u}(\nabla \mathbf{u}) - \nu \nabla^2 \mathbf{u} = 0 \quad (4.1)$$

where $\mathbf{u}^T = \{u, v\}$ is the vector of fluid velocity and ν is the kinematic viscosity. It can be rewritten as

$$u_t = -uu_x - vu_y + \nu \nabla^2 u \quad (4.2a)$$

and

$$v_t = -uv_x - vv_y + \nu \nabla^2 v \quad (4.2b)$$

in which subscript denotes differentiation. We chose the spatial domain for this problem as $\mathbf{x} \in [0, 1] \times [0, 1]$. Initial condition is randomly generated using $\mathbf{u}_0 \sim \mathcal{N}(0, 625(-\Delta + 25I)^{-2})$ [33]. Our network is trained for $t \in [0, 2]$ s using 1000 time steps with learning rate of 0.001; it is updated by using adaptive learning rate. Training time is 28.51 h. [Fig. 8](#) depicts the training outputs and absolute errors for present architecture and PINN. The absolute error for training domain is relatively small for PINN because of the hard constraint of initial conditions; however, the error for PINN gradually increases as time increases. It is worth noting that a decay in accuracy does not occur in our method due to the specification of initial condition as input. [Fig. 9](#) shows the results for extrapolation predictions from present architecture and PINN for $t \in [2, 3]$ s. Pairs of columns in [Figs. 8](#) and [9](#) show u and v fields, respectively. The present model performs better with more accurate extrapolation results compared to PINN. Also, the error in our method decreases as time increases while PINN error consistently increases.

The reason for this superior extrapolation performance compared to PINN is its ability to generalize the dynamics of the field. This capability is due to its architecture. It projects the input to a smaller latent space using an encoder structure and learns the time evolution in the smaller latent space. This learning involves two key features: first one is the convolutional LSTM, which excels at forecasting time dynamics, and the second one is the latent physics loss.

As shown in [Fig. 10](#), the performance of the model is demonstrated by using periodic and nonperiodic activation functions. The novel adaptive activation function does not adversely affect the training in the presence of physics with decaying behavior.

4.2. $\lambda - \omega$ reaction–diffusion equation

$\lambda - \omega$ reaction–diffusion equation is a coupled oscillatory differential equation. It is a special case of Ginzburg–Landau equation [62,63]. It generates stable traveling waves and has a limit cycle for certain λ and ω values. A 2D $\lambda - \omega$ reaction–diffusion equation can be written as

$$u_t = 0.1 \nabla^2 u + \lambda(r)u - \omega(r)v \quad (4.3a)$$

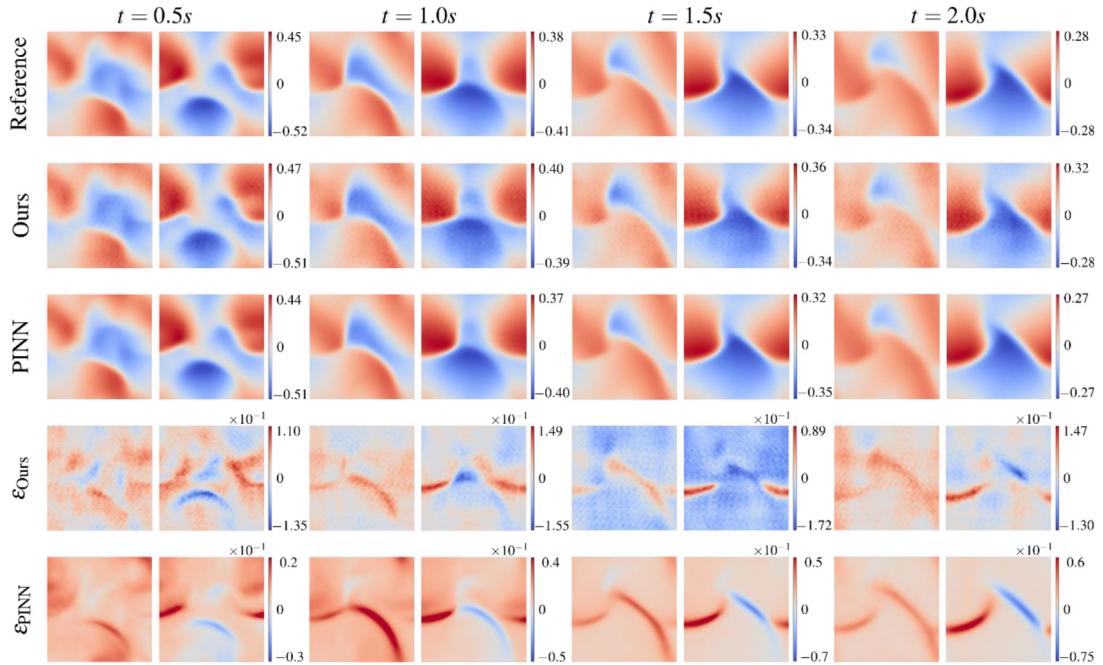


Fig. 8. Training results for Burgers' equation at $t = [0.5, 1.0, 1.5, 2.0]$ s.

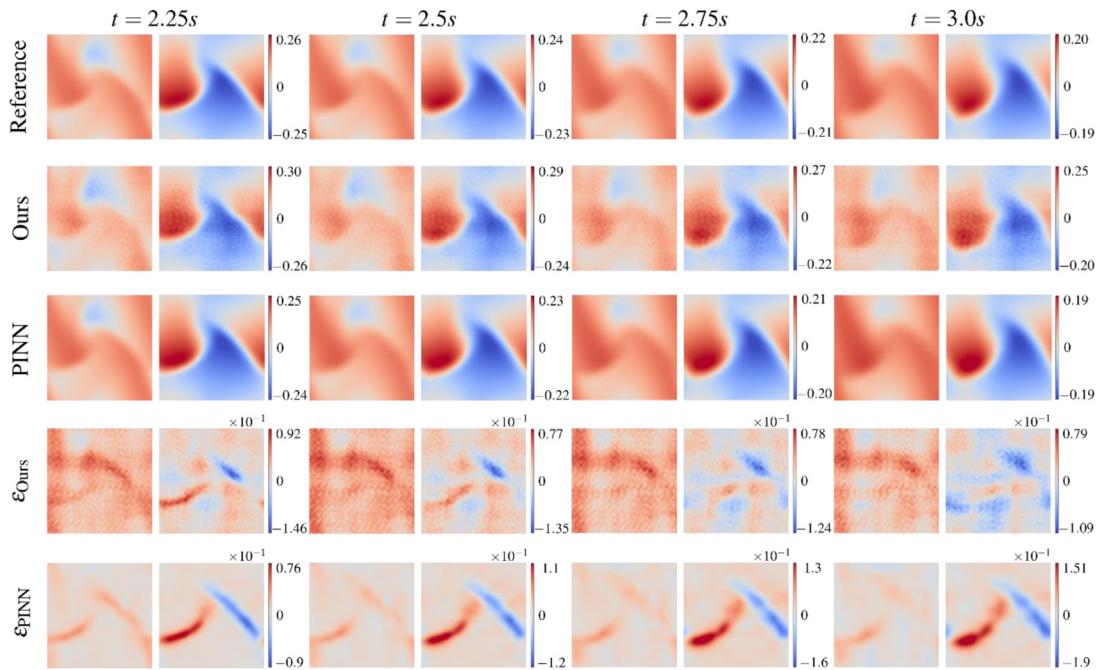


Fig. 9. Testing results for Burgers' equation at $t = [2.25, 2.5, 2.75, 3.0]$ s.

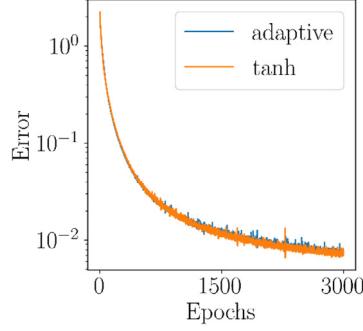


Fig. 10. Training error for Burgers' equations for adaptive periodic and tanh activation function at the state update gate of the ConvLSTM.

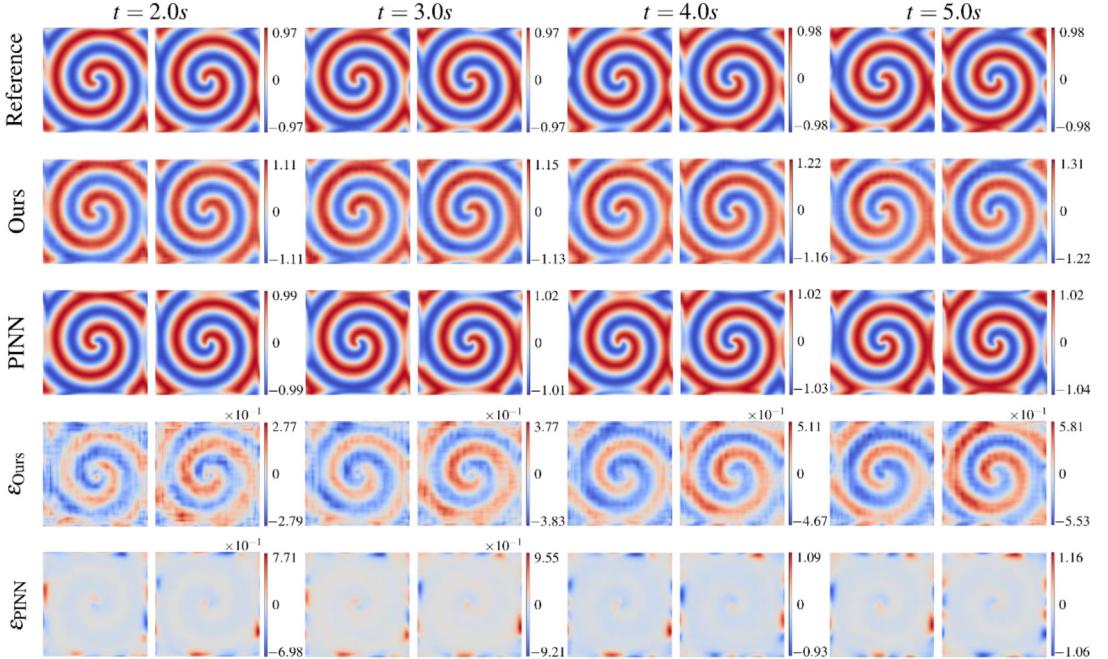


Fig. 11. Training results, $\lambda - \omega$ reaction–diffusion equation $t = [2.0, 3.0, 4.0, 5.0]$ s.

and

$$v_t = 0.1\nabla^2 v + \omega(r)u + \lambda(r)v \quad (4.3b)$$

where $r = u^2 + v^2$, $\lambda(r) = 1 - r^2$ and $\omega(r) = -r^2$. For the solution, we chose $\mathbf{x} \in [-10, 10] \times [-10, 10]$ and employ the dataset provided by Rudy et al. [64]. Their solution is simulated using spectral method. We train our model for $t \in [0, 5]$ s using 400 time steps. Learning rate at the beginning is specified as 0.001 and updated using adaptive learning rate. Training time is specified as 15.6 h. Fig. 11 shows the training results of the current model and PINN for $\lambda - \omega$ reaction–diffusion equation. Our architecture performs again better in interpolation compared to PINN and has a smaller absolute error for $\lambda - \omega$ reaction–diffusion equation. Fig. 12 shows the predictions from the present approach and PINN for $t \in [5, 10]$ s. Pairs of columns in Figs. 11 and 12 show u and v fields, respectively. Our model performs better in extrapolation with a smaller absolute error than that of PINN.

As shown in Fig. 13, the performance of the model is demonstrated by using periodic and nonperiodic activation functions. The adaptive activation function leads to faster convergence to a lower minima than that of tanh in the presence of physics with periodic behavior.

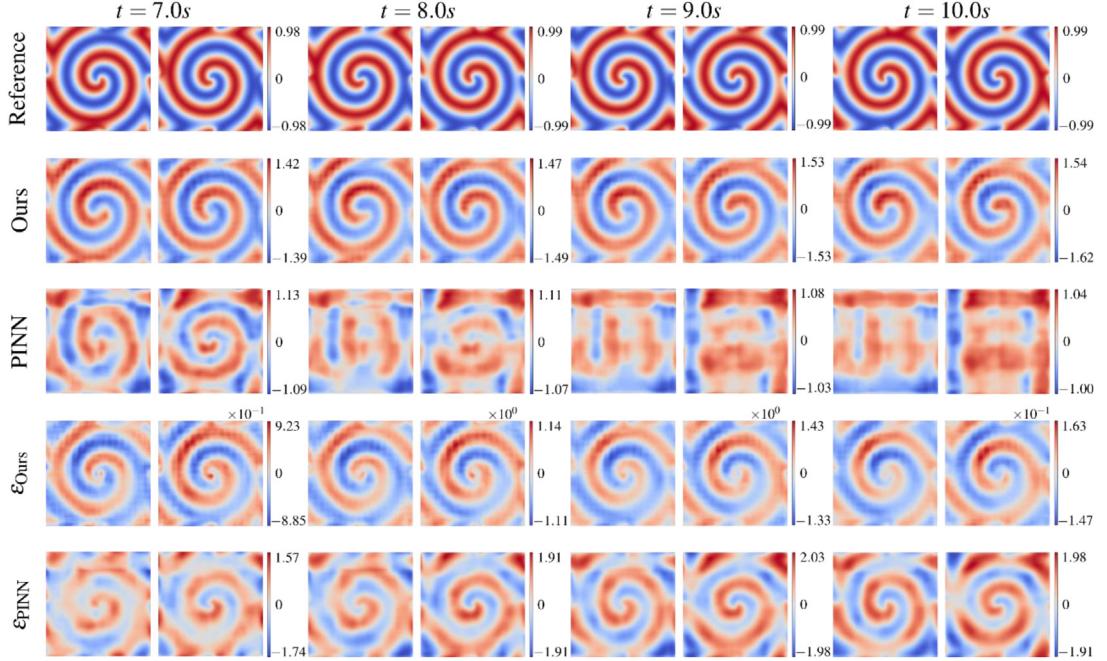


Fig. 12. Testing results, $\lambda - \omega$ reaction–diffusion equation $t = [7.0, 8.0, 9.0, 10.0]$ s.

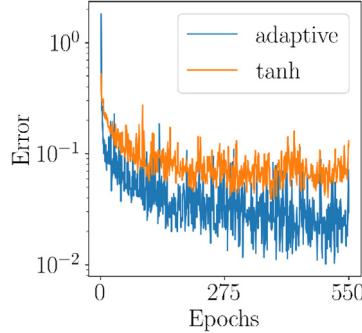


Fig. 13. Training error for R-D equations for adaptive periodic and tanh activation function at the state update gate of the ConvLSTM.

4.3. Gray-Scott equation

Gray–Scott equation is a coupled reaction–diffusion type equation which models the following chemical reactions



and



where \mathcal{V} is a catalyst, \mathcal{P} is an inert product and \mathcal{U} is a chemical species. Gray–Scott equation can be expressed as

$$u_t = \epsilon_1 \nabla^2 u + b(1 - u) - uv^2 \quad (4.5a)$$

and

$$v_t = \epsilon_2 \nabla^2 v - dv + uv^2 \quad (4.5b)$$

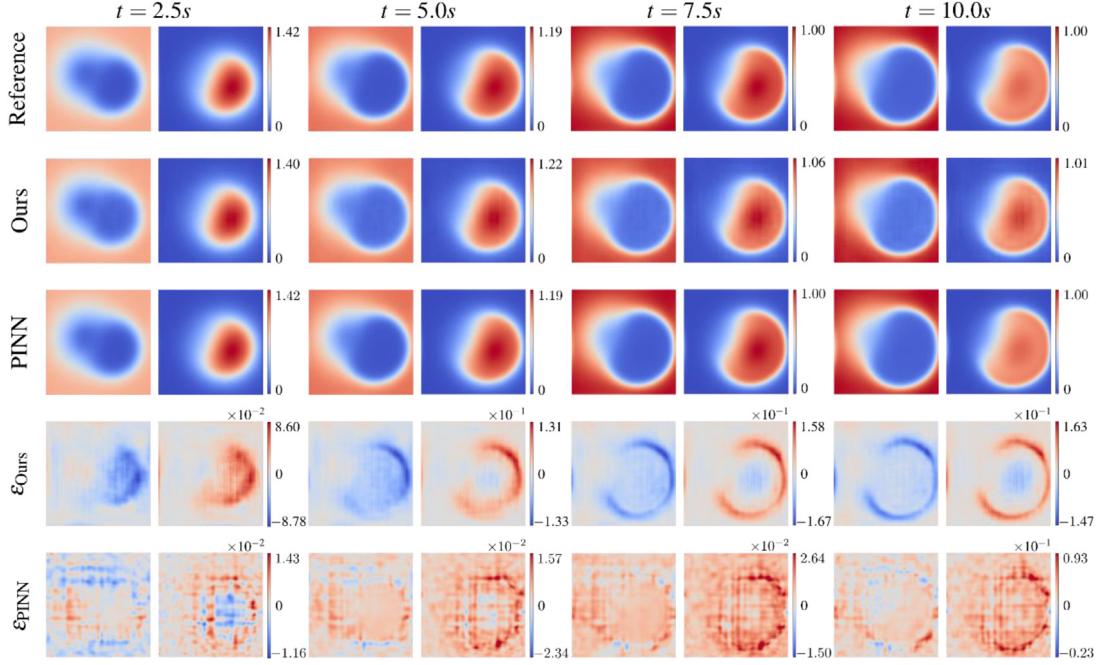


Fig. 14. Training results, Gray–Scott equation for $t = [2.5, 5.0, 7.5, 10.0]$ s.

where $u(x, y, t)$ and $v(x, y, t)$ denote the concentrations for \mathcal{U} and \mathcal{V} , b is the rate of \mathcal{U} feed, ϵ_1 and ϵ_2 are diffusivity coefficients and d is the rate of \mathcal{V} transforming to \mathcal{P} . For this specific case, we chose $\mathbf{x} \in [-0.2, 0.2] \times [-0.2, 0.2]$, $\epsilon_1 = 0.00002$, $\epsilon_2 = 0.00001$, $b = 0.04$, $d = 0.1$ and use **chebfun** package to obtain simulation results for comparison [65]. We train our network for $t \in [0, 10]$ s using 1000 time steps. Learning rate at the beginning is 0.001 and updated using adaptive learning rate. Training time is specified as 23.84 h, and solution is predicted for $t \in [10, 15]$ s. Fig. 14 shows the training and testing results of current method for Gray–Scott equation. Fig. 15 compares the extrapolation predictions of the present model with those of PINN for $t \in [10, 15]$ s. Pairs of columns in Figs. 14 and 15 show u and v fields, respectively. The present model predictions have a smaller absolute error than those of PINN. However, the training errors are smaller for PINN than those of current method because of hard imposition of initial condition.

5. Discussion and conclusions

This study presents a novel deep learning architecture to obtain the forward solution for coupled PDEs by employing only convolutional layers. The main reason for this choice is that the convolutional layers have sparse connections with more robustness to overfitting. The solutions to PDEs contain spatial dependencies; therefore, convolutional layers are more appropriate for capturing such dependencies and provide better generalization performance.

The classical methods for solving PDEs do not offer any generalization property. We use a ConvLSTM to capture the time dynamics and enhance the extrapolation capability of the neural network. Its training is unsupervised with physics informed in the form of governing equation at the output. Also, we impose the conservation relations on the latent space; therefore, we reach higher accuracy compared to other similar architectures.

We use PDDO to obtain the nonlocal convolutional filters for derivative calculation of the field variable. This choice improves the efficiency of the derivative filters and enables us to choose the filter size freely. We compare our results with PINN approach and observe that our method is more capable of extrapolating the solution for future time steps than PINN. One limitation of the present deep learning architecture is the generalization of its performance for different initial conditions. This drawback can be handled by using a generative model at encoding–decoding layers.

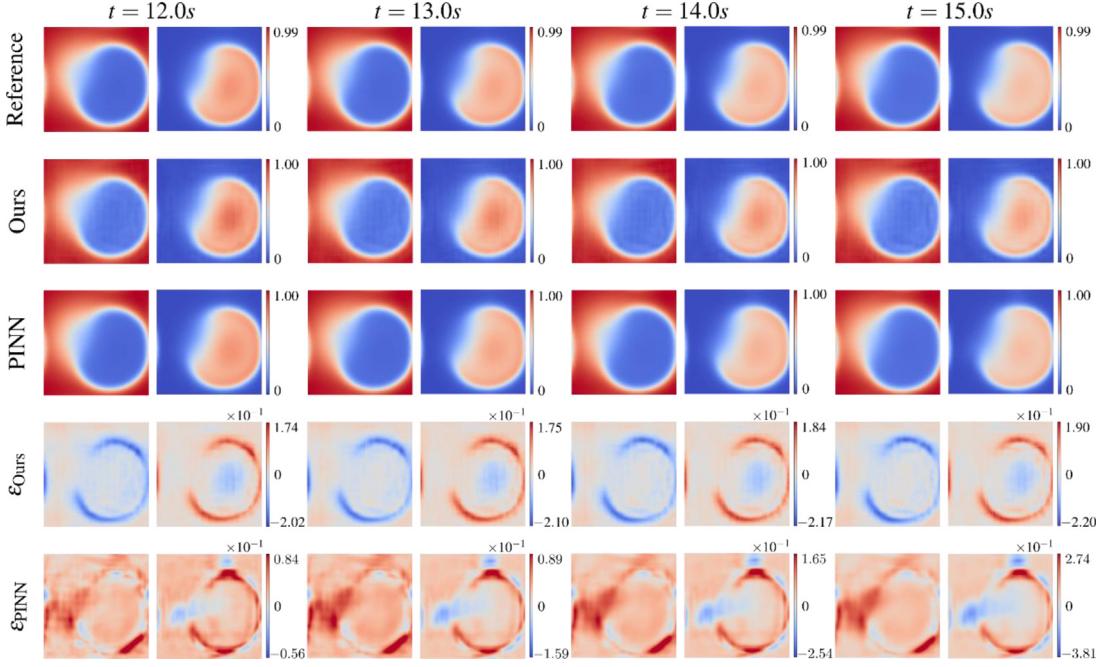


Fig. 15. Testing results, Gray–Scott equation for $t = [12.0, 13.0, 14.0, 15.0]$ s.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

EM, ACB and AM performed this work as part of the ongoing research at the MURI Center for Material Failure Prediction through Peridynamics at the University of Arizona (AFOSR, USA Grant No. FA9550-14-1-0073).

Appendix A. Peridynamic differential operator

In a 2-dimensional space, a function, $f(\mathbf{x} + \xi)$ can be expressed in terms of Taylor Series Expansion (TSE) as

$$f(\mathbf{x} + \xi) = f(\mathbf{x}) + \xi_1 \frac{\partial f(\mathbf{x})}{\partial x_1} + \xi_2 \frac{\partial f(\mathbf{x})}{\partial x_2} + \frac{1}{2!} \xi_1^2 \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} + \frac{1}{2!} \xi_2^2 \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} + \xi_1 \xi_2 \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} + \mathcal{R} \quad (\text{A.1})$$

where \mathcal{R} is the remainder. Multiplying each term with PD functions, $g_2^{p_1 p_2}(\xi)$ and integrating over the domain of interaction (family), \mathbf{H}_x results in

$$\begin{aligned} \int_{\mathbf{H}_x} f(\mathbf{x} + \xi) g_2^{p_1 p_2}(\xi) dV &= \int_{\mathbf{H}_x} f(\mathbf{x}) g_2^{p_1 p_2}(\xi) dV + \frac{\partial f(\mathbf{x})}{\partial x_1} \int_{\mathbf{H}_x} \xi_1 g_2^{p_1 p_2}(\xi) dV \\ &\quad + \frac{\partial f(\mathbf{x})}{\partial x_2} \int_{\mathbf{H}_x} \xi_2 g_2^{p_1 p_2}(\xi) dV + \frac{1}{2} \frac{\partial f^2(\mathbf{x})}{\partial x_1^2} \int_{\mathbf{H}_x} \xi_1^2 g_2^{p_1 p_2}(\xi) dV \\ &\quad + \frac{1}{2} \frac{\partial f^2(\mathbf{x})}{\partial x_2^2} \int_{\mathbf{H}_x} \xi_2^2 g_2^{p_1 p_2}(\xi) dV + \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} \int_{\mathbf{H}_x} \xi_1 \xi_2 g_2^{p_1 p_2}(\xi) dV \end{aligned} \quad (\text{A.2})$$

in which the point \mathbf{x} is not necessarily symmetrically located in the domain of interaction. The initial relative position, ξ , between points \mathbf{x} and \mathbf{x}' can be expressed as $\xi = \mathbf{x}' - \mathbf{x}$. This ability permits each point to have its own unique family with an arbitrary position. Therefore, the size and shape of each family can be different, and they significantly influence the degree of nonlocality. In general, the family of a point can be nonsymmetric due to nonuniform spatial discretization.

The degree of interaction between the material points in each family is specified by a nondimensional weight function, $w(|\xi|)$, which can vary from point to point. The interactions become more local with decreasing family size. Thus, the family size and shape are important parameters. Each point occupies an infinitesimally small entity such as volume, area or a distance.

The PD functions are constructed such that they are orthogonal to each term in the TSE as

$$\frac{1}{n_1!n_2!} \int_{\mathbf{H}_x} \xi_1^{n_1} \xi_2^{n_2} g_2^{p_1 p_2}(\xi) dV = \delta_{n_1 p_1} \delta_{n_2 p_2} \quad (\text{A.3})$$

with $(n_1, n_2, p, q = 0, 1, 2)$ and δ_{ij} is the Kronecker delta symbol. Enforcing the orthogonality conditions in the TSE leads to the nonlocal PD representation of the function itself and its derivatives as

$$f(\mathbf{x}) = \int_{\mathbf{H}_x} f(\mathbf{x} + \xi) g_2^{00}(\xi) dV \quad (\text{A.4a})$$

$$\left\{ \begin{array}{l} \frac{\partial f(\mathbf{x})}{\partial x} \\ \frac{\partial f(\mathbf{x})}{\partial y} \end{array} \right\} = \int_{\mathbf{H}_x} f(\mathbf{x} + \xi) \begin{Bmatrix} g_2^{10}(\xi) \\ g_2^{01}(\xi) \end{Bmatrix} dV \quad (\text{A.4b})$$

$$\left\{ \begin{array}{l} \frac{\partial^2 f(\mathbf{x})}{\partial x^2} \\ \frac{\partial^2 f(\mathbf{x})}{\partial y^2} \\ \frac{\partial^2 f(\mathbf{x})}{\partial x \partial y} \end{array} \right\} = \int_{\mathbf{H}_x} f(\mathbf{x} + \xi) \begin{Bmatrix} g_2^{20}(\xi) \\ g_2^{02}(\xi) \\ g_2^{11}(\xi) \end{Bmatrix} dV \quad (\text{A.4c})$$

The PD functions can be constructed as a linear combination of polynomial basis functions

$$g_2^{p_1 p_2} = a_{00}^{p_1 p_2} w_{00}(|\xi|) + a_{10}^{p_1 p_2} w_{10}(|\xi|)\xi_1 + a_{01}^{p_1 p_2} w_{01}(|\xi|)\xi_2 + a_{20}^{p_1 p_2} w_{20}(|\xi|)\xi_1^2 \\ + a_{02}^{p_1 p_2} w_{02}(|\xi|)\xi_2^2 + a_{11}^{p_1 p_2} w_{11}(|\xi|)\xi_1\xi_2, \quad (\text{A.5})$$

where $a_{q_1 q_2}^{p_1 p_2}$ are the unknown coefficients, $w_{q_1 q_2}(|\xi|)$ are the influence functions, and ξ_1 and ξ_2 are the components of the vector ξ . Assuming $w_{q_1 q_2}(|\xi|) = w(|\xi|)$ and incorporating the PD functions into the orthogonality equation lead to a system of algebraic equations for the determination of the coefficients as

$$\mathbf{A}\mathbf{a} = \mathbf{b}, \quad (\text{A.6})$$

where

$$\mathbf{A} = \int_{\mathbf{H}_x} w(|\xi|) \begin{bmatrix} 1 & \xi_1 & \xi_2 & \xi_1^2 & \xi_2^2 & \xi_1 \xi_2 \\ \xi_1 & \xi_1^2 & \xi_1 \xi_2 & \xi_1^3 & \xi_1 \xi_2^2 & \xi_1^2 \xi_2 \\ \xi_2 & \xi_1 \xi_2 & \xi_2^2 & \xi_1^2 \xi_2 & \xi_2^3 & \xi_1 \xi_2^2 \\ \xi_1^2 & \xi_1^3 & \xi_1^2 \xi_2 & \xi_1^4 & \xi_1^2 \xi_2^2 & \xi_1^3 \xi_2 \\ \xi_2^2 & \xi_1 \xi_2^2 & \xi_2^3 & \xi_1^2 \xi_2^2 & \xi_2^4 & \xi_1 \xi_2^3 \\ \xi_1 \xi_2 & \xi_1^2 \xi_2 & \xi_1 \xi_2^2 & \xi_1^3 \xi_2 & \xi_1 \xi_2^3 & \xi_1^2 \xi_2^2 \end{bmatrix} dV, \quad (\text{A.7a})$$

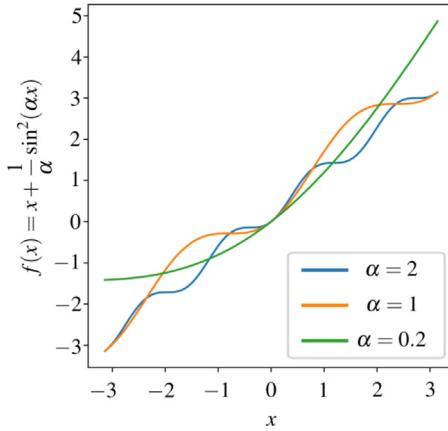


Fig. 16. Novel periodic activation function for different α values.

$$\mathbf{a} = \begin{bmatrix} a_{00}^{00} & a_{10}^{00} & a_{01}^{00} & a_{20}^{00} & a_{02}^{00} & a_{11}^{00} \\ a_{00}^{10} & a_{10}^{10} & a_{01}^{10} & a_{20}^{10} & a_{02}^{10} & a_{11}^{10} \\ a_{00}^{01} & a_{10}^{01} & a_{01}^{01} & a_{20}^{01} & a_{02}^{01} & a_{11}^{01} \\ a_{00}^{20} & a_{10}^{20} & a_{01}^{20} & a_{20}^{20} & a_{02}^{20} & a_{11}^{20} \\ a_{00}^{02} & a_{10}^{02} & a_{01}^{02} & a_{20}^{02} & a_{02}^{02} & a_{11}^{02} \\ a_{00}^{11} & a_{10}^{11} & a_{01}^{11} & a_{20}^{11} & a_{02}^{11} & a_{11}^{11} \end{bmatrix}, \quad (\text{A.7b})$$

and

$$\mathbf{b} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (\text{A.7c})$$

After determining the coefficients $a_{q_1 q_2}^{p_1 p_2}$ via $\mathbf{a} = \mathbf{A}^{-1}\mathbf{b}$, the PD functions $g_2^{p_1 p_2}(\xi)$ can be constructed. The detailed derivations and the associated computer programs can be found in [61]. The PD-DD is nonlocal; however, in the limit as the horizon size approaches zero, it recovers the local differentiation as proven by Silling and Lehoucq [66].

Appendix B. Periodic activation function

Ziyin et al. [56] proposed a new activation function for improving predictive capabilities of a DNN for data with periodic behavior. This activation function is $f_\alpha(x) = x + \frac{1}{\alpha} \sin^2(\alpha x)$, where α is the frequency of the periodic part of the activation function. It is superior for predicting periodic behavior because it has inductive bias for periodicity. Also, it is easier to optimize compared to other periodic activation functions with infinitely many local minima. With this activation function at every layer, they employ a fully connected DNN structure. Ziyin et al. [56] specified the value of parameter α as a fixed parameter. Fig. 16 shows the effect of different α values on the behavior of the activation function. They demonstrated the performance of this activation function for image classification, financial,

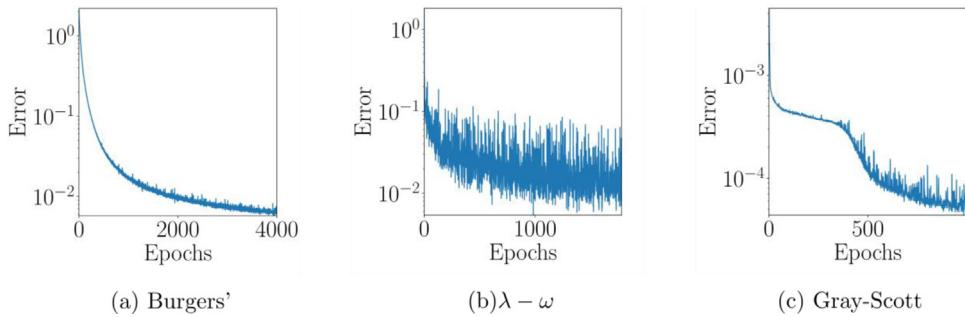


Fig. 17. Training histories of the numerical experiments.

atmospheric and body temperature predictions. It is also capable of predicting time evolution of non-periodic data.

Appendix C. Training losses for numerical experiments

Training losses of the numerical experiments with the present solutions for 2D Coupled Burgers' equation, $\lambda - \omega$ Reaction–Diffusion equation and Gray–Scott equation are shown in Fig. 17 as a semilog plot. Models for Burgers' and Gray–Scott equations are terminated before the plateau to prevent the overfitting. The same approach is followed for $\lambda - \omega$ Reaction–Diffusion equation; however, the periodic behavior of the solution creates oscillation in training error. Therefore, it is more difficult to observe the convergence.

Appendix D. Training of PINN models

We use SciANN library [67] to build and train the PINN models. Uniform sampling is adopted for the imposition of initial conditions and uniform random sampling for the domain collocation points. Number of collocation points for the initial condition and domain points are specified as 16 384 (128×128) and 327 680 ($20 \times 128 \times 128$), respectively. Periodic boundary conditions are imposed by using Fourier Features [68]. Our C^k periodic layer, which comes after the input layer, has 20 frequency components. The PINN setup has 4 hidden layers and 80 neurons with tanh activation function while using Glorot uniform initialization [69] for the network weights. Also, we use Adam optimizer [70] with initial learning rate of 10^{-3} which exponentially decreases to 10^{-4} and loss weights are updated using gradient norm algorithm [32]. We use a batch size of 2000 and trained the PINN for 800 epochs. Training takes 8 h with (AMD EPYC 7642 16-core) CPU with 2.4 GHz clock speed and 32 GB RAM.

Appendix E. Selection of hyperparameters

In order to ensure the right hyperparameters, we trained the proposed architecture with varying hyperparameters of the network. Fig. 18a depicts the training error variation with changing learning rates. We observe that the NN converges to a lower solution for increasing learning rate from 0.0001 to 0.001 with a fixed number of epochs. Further increase in learning rate causes high error gradients and the error growth is unbounded. Therefore, the learning rate is specified as 0.001.

Fig. 18b shows the training error for changing number of layers in autoencoder–decoder architecture. Increase in number of layers provides no advantage for a fixed number of epochs. Also, decreasing the number of layers does not decrease the training error. Finally, changing the convolutional filter dimensions of our architecture results in no significant gain from the increased or decreased filter sizes as shown in Fig. 18c. The hyperparameters are selected the smallest training error occurs in our neural network.

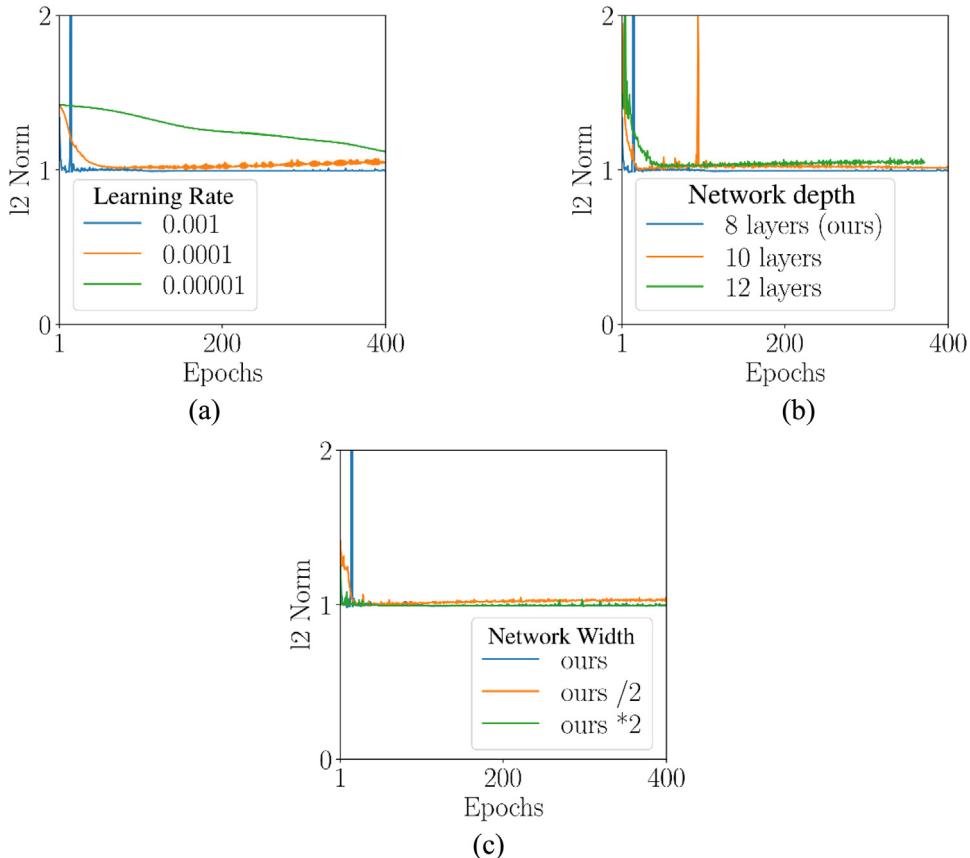


Fig. 18. Training error comparisons for (a) different initial learning rates, (b) different layer sizes, (c) different filter sizes for reaction–diffusion equation.

References

- [1] R. Courant, K. Friedrichs, H. Lewy, On the partial difference equations of mathematical physics, *IBM J. Res. Dev.* 11 (2) (1967) 215–234.
- [2] R.J. LeVeque, *Finite Volume Methods for Hyperbolic Problems*, (Vol. 31), Cambridge University Press, 2002.
- [3] O.C. Zienkiewicz, R.L. Taylor, R.L. Taylor, R.L. Taylor, *The Finite Element Method: Solid Mechanics*, (Vol. 2), Butterworth-heinemann, 2000.
- [4] J.S. Hesthaven, S. Gottlieb, D. Gottlieb, *Spectral Methods for Time-Dependent Problems* (Vol. 21), Cambridge University Press, 2007.
- [5] A.R. Gallant, H. White, There exists a neural network that does not make avoidable mistakes, in: *ICNN*, 1988, pp. 657–664.
- [6] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, *Neural Netw.* 2 (5) (1989) 359–366.
- [7] I.E. Lagaris, A. Likas, D.I. Fotiadis, Artificial neural networks for solving ordinary and partial differential equations, *IEEE Trans. Neural Netw.* 9 (5) (1998) 987–1000.
- [8] J. Berg, K. Nyström, A unified deep artificial neural network approach to partial differential equations in complex geometries, *Neurocomputing* 317 (2018) 28–41.
- [9] J. Han, A. Jentzen, W. E, Solving high-dimensional partial differential equations using deep learning, *Proc. Natl. Acad. Sci.* 115 (34) (2018) 8505–8510.
- [10] M. Raissi, P. Perdikaris, 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.* 378 (2019) 686–707.
- [11] A.G. Baydin, B.A. Pearlmutter, A.A. Radul, J.M. Siskind, Automatic differentiation in machine learning: a survey, *J. Machine Learn. Res.* 18 (2018) 1–43.
- [12] E. Haghighat, M. Raissi, A. Moure, H. Gomez, R. Juanes, A deep learning framework for solution and discovery in solid mechanics, 2020, arXiv preprint [arXiv:2003.02751](https://arxiv.org/abs/2003.02751).
- [13] C. Rao, H. Sun, Y. Liu, Physics-informed deep learning for computational elastodynamics without labeled data, *J. Eng. Mech.* 147 (8) (2021) 04021043.

- [14] E. Haghigat, A.C. Bekar, E. Madenci, R. Juanes, A nonlocal physics-informed deep learning framework using the peridynamic differential operator, *Comput. Methods Appl. Mech. Engrg.* 385 (2021) 114012.
- [15] X. Jin, S. Cai, H. Li, G.E. Karniadakis, Nsfnets (Navier–Stokes flow nets): Physics-informed neural networks for the incompressible Navier–Stokes equations, *J. Comput. Phys.* 426 (2021) 109951.
- [16] Z. Mao, A.D. Jagtap, G.E. Karniadakis, Physics-informed neural networks for high-speed flows, *Comput. Methods Appl. Mech. Engrg.* 360 (2020) 112789.
- [17] H. Gao, L. Sun, J.X. Wang, PhyGeoNet: Physics-informed geometry-adaptive convolutional neural networks for solving parameterized steady-state PDEs on irregular domain, *J. Comput. Phys.* 428 (2021) 110079.
- [18] Y. Chen, L. Lu, G.E. Karniadakis, L. Dal Negro, Physics-informed neural networks for inverse problems in nano-optics and metamaterials, *Opt. Express* 28 (8) (2020) 11618–11633.
- [19] K. Shukla, A.D. Jagtap, J.L. Blackshire, D. Sparkman, G.E. Karniadakis, A physics-informed neural network for quantifying the microstructural properties of polycrystalline nickel using ultrasound data: A promising approach for solving inverse problems, *IEEE Signal Process. Mag.* 39 (1) (2021) 68–77.
- [20] A.D. Jagtap, D. Mitsotakis, G.E. Karniadakis, Deep learning of inverse water waves problems using multi-fidelity data: Application to Serre–Green–Naghdi equations, *Ocean Eng.* 248 (2022) 110775.
- [21] L. McCleny, U. Braga-Neto, Self-adaptive physics-informed neural networks using a soft attention mechanism, 2020, arXiv preprint [arXiv:2009.04544](https://arxiv.org/abs/2009.04544).
- [22] K. Shukla, A.D. Jagtap, G.E. Karniadakis, Parallel physics-informed neural networks via domain decomposition, *J. Comput. Phys.* 447 (2021) 110683.
- [23] A.D. Jagtap, E. Kharazmi, G.E. Karniadakis, Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems, *Comput. Methods Appl. Mech. Engrg.* 365 (2020) 113028.
- [24] A.D. Jagtap, Z. Mao, N. Adams, G.E. Karniadakis, Physics-informed neural networks for inverse problems in supersonic flows, 2022, arXiv preprint [arXiv:2202.11821](https://arxiv.org/abs/2202.11821).
- [25] A.D. Jagtap, 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: AAAI Spring Symposium: MLPS, 2021.
- [26] Z. Hu, A.D. Jagtap, G.E. Karniadakis, K. Kawaguchi, When do extended physics-informed neural networks (XPINNs) improve generalization?, 2021, arXiv preprint [arXiv:2109.09444](https://arxiv.org/abs/2109.09444).
- [27] T. De Ryck, A.D. Jagtap, S. Mishra, Error estimates for physics informed neural networks approximating the Navier–Stokes equations, 2022, arXiv preprint [arXiv:2203.09346](https://arxiv.org/abs/2203.09346).
- [28] S. Mishra, R. Molinaro, Estimates on the generalization error of physics-informed neural networks for approximating a class of inverse problems for PDEs, *IMA J. Numer. Anal.* 42 (2) (2022) 981–1022.
- [29] S. Wang, Y. Teng, P. Perdikaris, Understanding and mitigating gradient flow pathologies in physics-informed neural networks, *SIAM J. Sci. Comput.* 43 (5) (2021) A3055–A3081.
- [30] S. Wang, X. Yu, P. Perdikaris, When and why PINNs fail to train: A neural tangent kernel perspective, *J. Comput. Phys.* 449 (2022) 110768.
- [31] S. Wang, H. Wang, P. Perdikaris, On the eigenvector bias of fourier feature networks: From regression to solving multi-scale pdes with physics-informed neural networks, *Comput. Methods Appl. Mech. Engrg.* 384 (2021) 113938.
- [32] Z. Chen, V. Badrinarayanan, C.Y. Lee, A. Rabinovich, Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks, in: International Conference on Machine Learning, PMLR, 2018, pp. 794–803.
- [33] P. Ren, C. Rao, Y. Liu, J.X. Wang, H. Sun, PhyCRNet: Physics-informed convolutional-recurrent network for solving spatiotemporal PDEs, *Comput. Methods Appl. Mech. Engrg.* 389 (2022) 114399.
- [34] L. Lu, P. Jin, G. Pang, Z. Zhang, G.E. Karniadakis, Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators, *Nat. Mach. Intell.* 3 (3) (2021) 218–229.
- [35] V. Oommen, K. Shukla, S. Goswami, R. Dingreville, G.E. Karniadakis, Learning two-phase microstructure evolution using neural operators and autoencoder architectures, 2022, arXiv preprint [arXiv:2204.07230](https://arxiv.org/abs/2204.07230).
- [36] R. Zhang, Y. Liu, H. Sun, Physics-informed multi-LSTM networks for metamodeling of nonlinear structures, *Comput. Methods Appl. Mech. Engrg.* 369 (2020) 113226.
- [37] N. Geneva, N. Zabaras, Modeling the dynamics of PDE systems with physics-constrained deep auto-regressive networks, *J. Comput. Phys.* 403 (2020) 109056.
- [38] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [39] S. Hochreiter, Y. Bengio, P. Frasconi, J. Schmidhuber, Gradient flow in recurrent nets: the difficulty of learning long-term dependencies, 2001.
- [40] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Comput.* 9 (8) (1997) 1735–1780.
- [41] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Commun. ACM* 60 (6) (2017) 84–90.
- [42] X. Shi, Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, W.C. Woo, Convolutional LSTM network: A machine learning approach for precipitation nowcasting, *Adv. Neural Inf. Process. Syst.* 28 (2015).
- [43] G. Pang, M. D’Elia, M. Parks, G.E. Karniadakis, Npinns: nonlocal physics-informed neural networks for a parametrized nonlocal universal Laplacian operator: algorithms and applications, *J. Comput. Phys.* 422 (2020) 109760.
- [44] X. Xu, M. D’Elia, C. Glusa, J.T. Foster, Machine-learning of nonlocal kernels for anomalous subsurface transport from breakthrough curves, 2022, arXiv preprint [arXiv:2201.11146](https://arxiv.org/abs/2201.11146).
- [45] E.A.B. de Moraes, M. D’Elia, M. Zayernouri, Machine learning of nonlocal micro-structural defect evolutions in crystalline materials, *Comput. Methods Appl. Mech. Engrg.* 403 (2023) 115743.

- [46] O. Burkowska, C. Glusa, M. D'elia, An optimization-based approach to parameter learning for fractional type nonlocal models, *Comput. Math. Appl.* 116 (2022) 229–244.
- [47] A.C. Bekar, E. Madenci, Peridynamics enabled learning partial differential equations, *J. Comput. Phys.* 434 (2021) 110193.
- [48] E. Madenci, A. Barut, M. Futch, Peridynamic differential operator and its applications, *Comput. Methods Appl. Mech. Engrg.* 304 (2016) 408–451.
- [49] A. Graves, A.R. Mohamed, G. Hinton, Speech recognition with deep recurrent neural networks, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2013, pp. 6645–6649.
- [50] K. Greff, R.K. Srivastava, J. Koutný, B.R. Steunebrink, J. Schmidhuber, LSTM: A search space odyssey, *IEEE Trans. Neural Netw. Learn. Syst.* 28 (10) (2016) 2222–2232.
- [51] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez ..., I. Polosukhin, Attention is all you need, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [52] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder–decoder for statistical machine translation, 2014, arXiv preprint [arXiv:1406.1078](https://arxiv.org/abs/1406.1078).
- [53] S. Goswami, M. Yin, Y. Yu, G.E. Karniadakis, A physics-informed variational DeepONet for predicting crack path in quasi-brittle materials, *Comput. Methods Appl. Mech. Engrg.* 391 (2022) 114587.
- [54] P. Conti, M. Guo, A. Manzoni, J.S. Hesthaven, Multi-fidelity surrogate modeling using long short-term memory networks, *Comput. Methods Appl. Mech. Engrg.* 404 (2023) 115811.
- [55] A. Quarteroni, A. Manzoni, F. Negri, Reduced Basis Methods for Partial Differential Equations: An Introduction (Vol. 92), Springer, 2015.
- [56] L. Ziyin, T. Hartwig, M. Ueda, Neural networks fail to learn periodic functions and how to fix it, *Adv. Neural Inf. Process. Syst.* 33 (2020) 1583–1594.
- [57] A.D. Jagtap, K. Kawaguchi, G.E. Karniadakis, Adaptive activation functions accelerate convergence in deep and physics-informed neural networks, *J. Comput. Phys.* 404 (2020) 109136.
- [58] A.D. Jagtap, K. Kawaguchi, G. Karniadakis, Locally adaptive activation functions with slope recovery for deep and physics-informed neural networks, *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 476 (2239) (2020) 20200334.
- [59] A.D. Jagtap, Y. Shin, K. Kawaguchi, G.E. Karniadakis, Deep kronecker neural networks: A general framework for neural networks with adaptive activation functions, *Neurocomputing* 468 (2022) 165–180.
- [60] A.D. Jagtap, G.E. Karniadakis, How important are activation functions in regression and classification? A survey, performance comparison, and future directions, 2022, arXiv preprint [arXiv:2209.02681](https://arxiv.org/abs/2209.02681).
- [61] E. Madenci, A. Barut, M. Dorduncu, *Peridynamic Differential Operator for Numerical Analysis*, Springer International Publishing, Berlin, 2019.
- [62] M.J. Smith, J.D. Rademacher, J.A. Sherratt, Absolute stability of wavetrains can explain spatiotemporal dynamics in reaction–diffusion systems of lambda-omega type, *SIAM J. Appl. Dyn. Syst.* 8 (3) (2009) 1136–1159.
- [63] N. Kopell, L.N. Howard, Plane wave solutions to reaction–diffusion equations, *Stud. Appl. Math.* 52 (4) (1973) 291–328.
- [64] S.H. Rudy, S.L. Brunton, J.L. Proctor, J.N. Kutz, Data-driven discovery of partial differential equations, *Sci. Adv.* 3 (4) (2017) e1602614.
- [65] T.A. Driscoll, N. Hale, L.N. Trefethen, Chebfun guide, 2014.
- [66] S.A. Silling, R.B. Lehoucq, Convergence of peridynamics to classical elasticity theory, *J. Elasticity* 93 (1) (2008) 13–37.
- [67] E. Haghigat, R. Juanes, Sciamn: A keras/tensorflow wrapper for scientific computations and physics-informed deep learning using artificial neural networks, *Comput. Methods Appl. Mech. Engrg.* 373 (2021) 113552.
- [68] S. Dong, N. Ni, A method for representing periodic functions and enforcing exactly periodic boundary conditions with deep neural networks, *J. Comput. Phys.* 435 (2021) 110242.
- [69] X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 249–256.
- [70] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).