

# Symbolic Neural Ordinary Differential Equations

Xin Li<sup>1</sup>, Chengli Zhao<sup>1†</sup>, Xue Zhang<sup>1†</sup>, Xiaojun Duan<sup>1</sup>

<sup>1</sup>College of Science, National University of Defense Technology, Changsha, Hunan 410073, China  
li\_xin@nudt.edu.cn, chenglizhao@nudt.edu.cn, xuezhang@nudt.edu.cn

## Abstract

Differential equations are widely used to describe complex dynamical systems with evolving parameters in nature and engineering. Effectively learning a family of maps from the parameter function to the system dynamics is of great significance. In this study, we propose a novel learning framework of symbolic continuous-depth neural networks, termed Symbolic Neural Ordinary Differential Equations (SNODEs), to effectively and accurately learn the underlying dynamics of complex systems. Specifically, our learning framework comprises three stages: initially, pre-training a predefined symbolic neural network via a gradient flow matching strategy; subsequently, fine-tuning this network using Neural ODEs; and finally, constructing a general neural network to capture residuals. In this process, we apply the SNODEs framework to partial differential equation systems through Fourier analysis, achieving resolution-invariant modeling. Moreover, this framework integrates the strengths of symbolism and connectionism, boasting a universal approximation theorem while significantly enhancing interpretability and extrapolation capabilities relative to state-of-the-art baseline methods. We demonstrate this through experiments on several representative complex systems. Therefore, our framework can be further applied to a wide range of scientific problems, such as system bifurcation and control, reconstruction and forecasting, as well as the discovery of new equations.

## Introduction

Complex dynamical systems, typically expressed as ordinary differential equations (ODEs) or partial differential equations (PDEs), are integral to scientific research and engineering applications across a wide range of domains (Wang, Lai, and Grebogi 2016; Brunton and Kutz 2022; Devaney 2021). However, these systems are often subject to environmental changes during their evolution, which can lead to temporal and spatial variations in internal parameters or external perturbations (Wang, Wang, and Perdikaris 2021). This yields significant challenges when exploring the underlying dynamic mechanisms and predicting the behavior of the system in new scenarios. To address this issue, we aim to construct a surrogate model from observed data generated by parametric dynamic systems and the parameter functions

(Benner, Gugercin, and Willcox 2015; Li et al. 2020b,a). This model should be capable of approximating the operator mapping from the parameter functions to the system states.

In most cases, we can only obtain the experimentally measured data from complex systems without prior knowledge of the underlying dynamical equations. Thus, data-driven methods are becoming increasingly important, including the autoregressive model (Navarro-Moreno 2008), recurrent neural networks (Van Houdt, Mosquera, and Nápoles 2020; Cho et al. 2014; Zhu, Li, and Lin 2023; De Brouwer et al. 2019), graph neural networks (Di Giovanni et al. 2022; Pilva and Zareei 2022; Liu et al. 2022), and neural ODEs (NODEs) (Chen et al. 2018; Holt, Qian, and van der Schaar 2022; Finlay et al. 2020; Biloš et al. 2021). In addition, recent work has shown that introducing physical priors can significantly enhance the learning performance of neural networks (Raissi, Perdikaris, and Karniadakis 2019; Sanchez-Gonzalez et al. 2019; Cranmer et al. 2020). For instance, the PDE-NET proposed by Long *et al.* (Long et al. 2018; Long, Lu, and Dong 2019) utilizes the special convolution kernels to represent potential spatial derivative terms and adopts the idea of sparse identification to learn the nonlinear dynamics (SINDy) (Wu and Zhang 2019; Rudy et al. 2016). The message-passing PDE solvers (MP-PDE) proposed by Brandstetter *et al.* (Brandstetter, Worrall, and Welling 2022) leverages a combination of traditional numerical methods to predict PDE systems. However, these methods often model only a specific system and require retraining when environmental parameters change.

In recent years, neural operators have gained significant attention (Kovachki et al. 2021; Lu et al. 2021; Jin, Meng, and Lu 2022; Chen et al. 2023; Xiong et al. 2023). Among them, the Deep Operator Network (DeepONet, DON) (Lu, Jin, and Karniadakis 2019) and Fourier Neural Operator (FNO) (Li et al. 2020c) are two of the most prominent works. These methods encode the input function at fixed sample points and use a dedicated neural network architecture to directly predict the system state. Compared to traditional methods, these methods solve parametric dynamics systems faster and improve prediction accuracy during testing due to their grid-independent nature. However, they lack interpretability, do not explain dynamics well, and cannot utilize physical priors. Furthermore, they require a large amount of training data and have limited extrapolation capa-

<sup>†</sup>Corresponding author

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

bilities outside the distribution of the training set.

To overcome the aforementioned challenges, we propose a novel learning framework, referred to as Symbolic Neural Ordinary Differential Equations (SNODEs), which establish symbolic continuous-depth neural networks (SCDNNs) to effectively and accurately learn a family of dynamical systems. Our framework has several key advantages:

- SNODEs enable accurate resolution-invariant reconstruction and prediction of the system state at any given time as well as any spatial grid (for PDEs), yielding a highly flexible and adaptable framework.
- SNODEs adopt a gradient flow matching pre-training strategy, which effectively circumvents the challenges associated with high computational loss and susceptibility to local optima inherent in traditional training method for SCDNNs.
- Our framework exhibits strong interpretability and extrapolation capabilities. This can be attributed to the effective implementation of operator learning and residual capture through a three-stage training approach in SNODEs, which leverages the strengths of both symbolism and connectionism.

## Related Work

Our proposed framework builds upon the Neural Ordinary Differential Equations (NODEs) (Chen et al. 2018). This method has gained significant attention in recent years as a continuous approximation of Residual neural Networks (ResNets) (He et al. 2016). In ResNets, the  $n$ -th residual block transforms the hidden layer state from  $\mathbf{h}_n$  to  $\mathbf{h}_{n+1}$  using a trainable parameter vector  $\boldsymbol{\theta}_n$  and a dimension-preserved function  $\mathbf{f}(\mathbf{h}_n, \boldsymbol{\theta}_n)$ , given by the following equation:

$$\mathbf{h}_{n+1} = \mathbf{h}_n + \Delta t \cdot \mathbf{f}(\mathbf{h}_n, \boldsymbol{\theta}_n),$$

where the step size  $\Delta t = 1$ . This discrete dynamical system can be viewed as the Euler discretization of the following ODE  $\dot{\mathbf{h}}(t) = \mathbf{f}(\mathbf{h}(t), \boldsymbol{\theta})$ , when we reuse the parameter vectors  $\boldsymbol{\theta}_n$  as the parameter vector shared  $\boldsymbol{\theta}$ , and the number of residual blocks approaches infinity, and the step size approaches zero. Given an initial value  $\mathbf{h}(t_0)$ , one can obtain the state at any time using an ODE solver:

$$\mathbf{h}(t_1) = \mathbf{h}(t_0) + \int_{t_0}^{t_1} \mathbf{f}(\mathbf{h}(t), \boldsymbol{\theta}) dt,$$

which can be recorded as  $\text{ODESolve}[\mathbf{h}(t_0), t_0, t_1]$ . Although recent works have extended this framework to various types of differential equations, including neural delay differential equations (Zhu, Guo, and Lin 2021), neural controlled differential equations (Kidger et al. 2020), neural integro-differential equations (Zappala et al. 2022), neural stochastic differential equations (Liu et al. 2019), augmented neural ODEs (Dupont, Doucet, and Teh 2019), and stiff neural ODEs (Kim et al. 2021), there has been limited research on applying the framework to parametric differential equations, particularly parametric PDEs.

## Method

Parametric dynamic systems frequently involve a time and/or space-varying parameter function  $\mathbf{u}(\mathbf{x}, t)$ . To model this family of systems, we present a novel symbolic continuous-depth neural network framework, called symbolic NODEs (SNODEs). This framework is capable of effectively learning an operator mapping from the parameter function  $\mathbf{u}(\mathbf{x}, t)$  to the system state  $\mathbf{s}(\mathbf{x}, t)$ .

### The Framework of SNODEs

Consider a dynamical system of the following general form:

$$\begin{aligned} \partial_t \mathbf{s} &= \mathbf{F}[\mathbf{u}, \mathbf{x}, t, \mathbf{s}, \mathbf{s}_{(x_i)}, \mathbf{s}_{(x_i x_j)}, \dots], \\ \mathbf{s}(\mathbf{x}, 0) &= \mathbf{s}_0, \quad \mathbf{x} \in \Omega, \end{aligned} \quad (1)$$

where space domain  $\Omega \subset \mathbb{R}^d$ , system state  $\mathbf{s} \in \mathbb{R}^{d_s}$ , and  $t \in [0, T]$ . When  $d = 0$ , the system is an ODE with respect to the parameter function  $\mathbf{u}(t)$ . When  $d \geq 1$ , it is a PDE system with respect to the parameter function  $\mathbf{u}(\mathbf{x}, t)$ , which satisfies the periodic boundary conditions. Here,  $\mathbf{s}_{(x_i)}$  represents all first-order partial derivatives of state  $\mathbf{s}$  with respect to spatial variables, and  $\mathbf{s}_{(x_i x_j)}$  represents all second-order partial derivatives, extending to higher-order derivatives up to the  $q$ -th order. In this paper, we focus on the scenario  $q \leq 4$ . We note that the proposed framework can be extended to higher-dimensional cases.

Our proposed framework enhances the accuracy of modeling the parametric dynamical systems by leveraging the information in both temporal and spatial domains. For the temporal domain, as illustrated in Figure 1, SNODEs incorporate the parameter function  $\mathbf{u}(\mathbf{x}, t)$  as an additional input into a symbolic neural network (SymNet) and a general neural network (GeNN), and then numerically obtain the system state value at any given time  $t_1$  by an ODE solver,

$$\begin{aligned} \mathbf{s}(\mathbf{x}, t_1) &= \text{ODESolve}[\mathbf{s}(\mathbf{x}, t_0), t_0, t_1, \mathbf{u}(\mathbf{x}, t)] \\ &= \mathbf{s}(\mathbf{x}, t_0) + \int_{t_0}^{t_1} \hat{\mathbf{F}}[\mathbf{s}(\mathbf{x}, t), \mathbf{u}(\mathbf{x}, t), t] dt, \end{aligned} \quad (2)$$

where  $\hat{\mathbf{F}} = \hat{\mathbf{F}}_1 + \hat{\mathbf{F}}_2$  represents the vector field estimated by the sum of two neural networks. During the ODEsolve process, one can select an appropriate numerical integration to explicitly make a trade-off between the numerical precision and the computational cost.

To address the spatial dimensions in PDE systems, we consider a 2-dimensional PDE with periodic boundary conditions. In this case, directly learning the spatial partial derivatives of the system state from the neural network in Eq. (2) is often challenging and requires the incorporation of additional physical priors. To overcome this challenge, we propose a novel strategy for the integration of the spatial derivative terms. Specifically, our approach involves first applying a two-dimensional Fourier transform  $\mathcal{F}$ ,

$$\tilde{\mathbf{s}}(\tilde{x}, \tilde{y}, t) = \mathcal{F}[\mathbf{s}(x, y, t)] = \int \int_{\Omega} \mathbf{s}(x, y, t) e^{-2\pi i(x\tilde{x} + y\tilde{y})}, \quad (3)$$

where  $(x, y) \in \Omega$  and  $t \in [0, T]$ , and  $\mathcal{F}^{-1}$  is the inverse transform of  $\mathcal{F}$ . According to the differential property of the

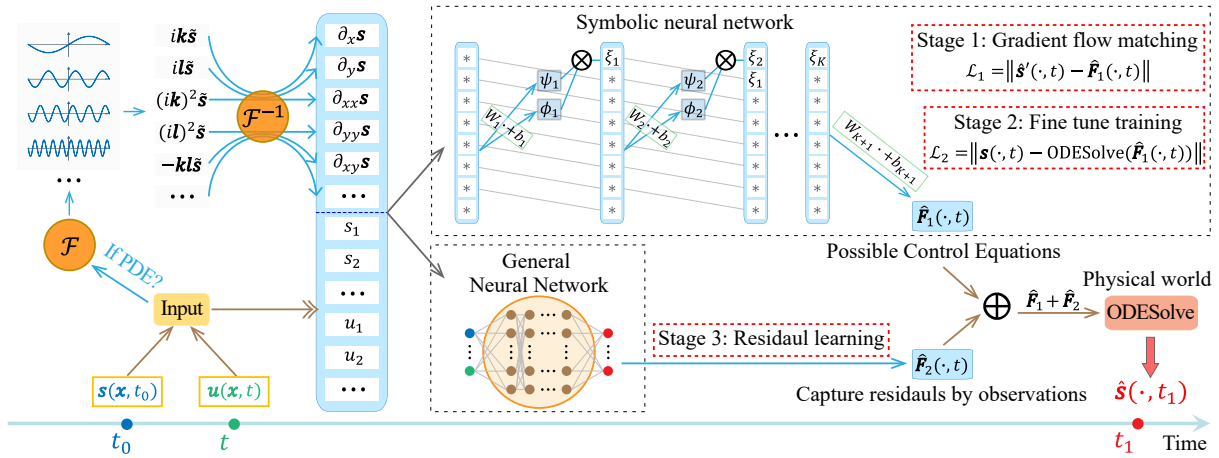


Figure 1: The sketched framework of SNODEs. This framework, which includes the SymNet and GeNN components, takes state variables, parametric functions, and all possible spatial partial derivatives as inputs. It models a family of parametric dynamical systems using the proposed three-stage and adaptive learning strategy.

Fourier transform, the partial derivative of the system state with respect to the space can be calculated as follows:

$$\mathcal{F}\left(\frac{\partial^{n+m}\mathbf{s}}{\partial^n x \partial^m y}\right)(\tilde{x}, \tilde{y}) = (2\pi i \tilde{x})^n (2\pi i \tilde{y})^m \mathcal{F}(\mathbf{s}),$$

$$\frac{\partial^{n+m}\mathbf{s}}{\partial^n x \partial^m y} = \mathcal{F}^{-1}[(2\pi i \tilde{x})^n (2\pi i \tilde{y})^m \mathcal{F}(\mathbf{s})].$$

Given that the right-hand side  $\mathbf{F}$  of most PDEs consists of only a few interacting symbols (see Eq. (2)), incorporating additional physical prior information is possible.

To achieve this, we utilize the concept of the symbolic regression to effectively learn the underlying dynamics. Drawing inspiration from the Symbolic Neural Network (SymNet) (Sahoo, Lampert, and Martius 2018; Long, Lu, and Dong 2019), we propose a continuous-depth multi-layer SymNet. In the first layer, we present the first-order term of all possible symbols, denoted as  $\mathbf{L}_1 = [\mathbf{u}, \mathbf{s}, \partial_x \mathbf{s}, \partial_y \mathbf{s}, \partial_{xx} \mathbf{s}, \dots]^\top$ . The length of  $\mathbf{L}_1$  depends on the spatial dimension and the highest order of partial derivatives. In the next layer, we define

$$\mathbf{L}_{k+1} = [\psi_k \times \phi_k, \mathbf{L}_k^\top]^\top, [\psi_k, \phi_k]^\top = \mathbf{W}_k \mathbf{L}_k + \mathbf{b}_k, \quad (4)$$

where  $k \in \{1, 2, \dots, K\}$ ,  $K$  is the number of hidden layers within SymNet,  $\mathbf{W}_k$  and  $\mathbf{b}_k$  are the trainable parameters. It allows us to use  $\psi_1 \times \phi_1$  to learn all quadratic product terms of elements from the layer  $\mathbf{L}_1$ . Our multi-layer SymNet can be extended to learn arbitrary order polynomials of elements from  $\mathbf{L}_1$ , thereby enabling the learning of the dynamics  $\mathbf{F}$ . The SymNet in Figure 1 illustrates the execution process of a  $K$ -layer SymNet. Here, we need to consider the following discrete form via a proper discretization of the spatial domain,

$$\mathbf{L}_1 = [\mathbf{u}, \mathbf{s}, \mathcal{F}^{-1}(ik\tilde{\mathbf{s}}), \mathcal{F}^{-1}(il\tilde{\mathbf{s}}), \mathcal{F}^{-1}((ik)^2\tilde{\mathbf{s}}), \dots]^\top,$$

$$\mathbf{k} = \frac{2\pi}{N}(e_N, \dots, e_N), \quad \mathbf{l} = \frac{2\pi}{M}(e_M, \dots, e_M)^\top,$$

where  $\tilde{\mathbf{s}} = \mathcal{F}(\mathbf{s})$ ,  $e_X$  denotes the vector  $(0, 1, \dots, X-1)^\top$ ,  $N$  and  $M$  represent the number of sample points along the

$x$  and  $y$  dimensions, respectively. Additionally, the product between  $\mathbf{k}$ ,  $\mathbf{l}$ , and  $\tilde{\mathbf{s}}$  is obtained through the element-wise multiplication, and  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  denote the discrete Fourier transform and inverse transform, respectively.

Furthermore, we also introduce a GeNN for residual learning. This GeNN can embody various prevalent neural network architectures, such as fully connected neural networks, convolutional neural networks. Finally, using Eq. (2) and  $\hat{\mathbf{F}}$ , we can compute the value of  $s(x, y, t_1)$  at any time point  $t_1$ .

### Learning Strategies for SNODEs

Utilizing Eq. (2) directly for training SNODEs encounters several challenges, including the high computational costs and the risk of converging to local optima associated with the classic NODE method, as well as the issue of gradient explosion. Therefore, we propose a three-stage training approach to efficiently model unknown dynamics, as shown in Figure 1.

In stage 1, we employ a strategy of gradient flow matching (Li et al. 2024) to pre-train the SymNet. Here, we estimate the temporal gradient  $\hat{\mathbf{s}}'$  through Fourier analysis

$$\hat{\mathbf{s}}' = \mathcal{F}_t^{-1}(i\mathbf{u}\mathcal{F}_t(\mathbf{s})), \quad \mathbf{u} = \frac{2\pi}{N_t}(e_{N_t}, \dots, e_{N_t}),$$

where  $N_t$  denotes the number of sampling points in the temporal direction,  $\mathcal{F}_t$  is the Fourier transform along the temporal axis. Then the flow matching loss function  $\mathcal{L}_1 = \|\hat{\mathbf{s}}' - \hat{\mathbf{F}}_1\|$ . In stage 2, we fine-tune the SymNet  $\hat{\mathbf{F}}_1$  by minimizing the prediction error of ODESolve, and the loss function  $\mathcal{L}_2 = \|\mathbf{s} - \text{ODESolve}(\hat{\mathbf{F}}_1)\|$ . In stage 3, we keep SymNet fixed and further train GeNN for residual learning with the loss function  $\mathcal{L}_3 = \|\mathbf{s} - \text{ODESolve}(\hat{\mathbf{F}}_1 + \hat{\mathbf{F}}_2)\|$ . Moreover, to enhance the sparsity of the inferred SymNet network, L1 regularization was incorporated into both  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , with the regularization coefficient,  $\alpha$ , serving as a hyperparameter.

During the training process in stages 2 and 3, we employ an adaptive training strategy. To begin, we simply set the

prediction steps to 1 and increase it once the training error falls below a predefined threshold. Secondly, we incorporate an adaptive learning rate, which involves adjusting the learning rate at evenly spaced intervals across batches. Finally, we consider numerical methods for ODE solvers. In the early stages of training, simpler methods such as Euler should be used to expedite training (high-precision methods may cause numerical instability and training failure). In the later stages of training, higher-precision methods such as Runge-Kutta, adaptive-step solvers, can be employed to further train the model and achieve longer-term predictions on the test set.

## Theoretical Results

The proposed framework effectively models the mapping operator from parametric functions to system states. To theoretically validate our method, here we present the Theorem 1 about universal approximation theorem for SNODEs (see Appendix A.1 for specific proof). Additionally, see Appendix C.1 for the extension of the SymNet.

**Theorem 1.** (Universal approximation theorem for SNODEs.) Suppose  $F$  (denoted by  $F(\mathbf{u}, \mathbf{s})$  in Equation (1)) is any nonlinear function acting on  $\mathbf{u}$  and at most  $q$ -th order space derivatives of  $\mathbf{s}$ , the functions  $\mathbf{u}(\mathbf{x}, t)$  and  $\mathbf{s}(\mathbf{x}, t)$  are defined on the domain  $\Omega \subset \mathbb{R}^d$  with periodic boundary conditions with  $0 \leq t \leq T$ . Then for any  $\varepsilon > 0$ , there are positive integers  $K$ , weights  $\mathbf{W}_k$ ,  $\mathbf{b}_k$ ,  $k = 1, \dots, K$  such that

$$\begin{aligned} \mathbf{L}_{k+1} &= [\boldsymbol{\psi}_k \times \boldsymbol{\phi}_k, \mathbf{L}_k^\top]^\top, [\boldsymbol{\psi}_k, \boldsymbol{\phi}_k]^\top = \mathbf{W}_k \mathbf{L}_k + \mathbf{b}_k, \\ \mathbf{L}_1 &= [\mathbf{u}, \mathbf{s}, \partial_{x_1} \mathbf{s}, \dots, \partial_{x_d} \mathbf{s}, \partial_{x_1 x_1} \mathbf{s}, \dots]^\top, \\ |F(\mathbf{u}, \mathbf{s}) - (\mathbf{W}_n \mathbf{L}_n + \mathbf{b}_n)| &< \varepsilon, \end{aligned}$$

holds for all  $\mathbf{x} = (x_1, \dots, x_d) \in \Omega$ ,  $\mathbf{u} \in C(\Omega \times [0, T])$  and the corresponding solution  $\mathbf{s}$  of the Equation (1) with the initial  $\mathbf{s}(\cdot, 0) \in C(\Omega)$ . Here,  $\mathbf{L}_1$  is the dictionary including all possible  $q$ -th order terms,  $C(\Omega)$  is the Banach space of all continuous functions defined on  $\Omega$  with norm  $\|f\|_{C(\Omega)} = \max_{\mathbf{x} \in \Omega} |f(\mathbf{x})|$ .

In fact, the number of elements in the  $\mathbf{L}_1$  layer of SymNet is a crucial parameter, and we can calculate it using Theorem 2 (see Appendix A.2 for specific proof). In most practical scenarios, it is typically observed that  $d \leq 3$  and  $q \leq 4$ . Based on this, one can infer that the dimension of  $\mathbf{L}_1$  is generally less than  $35d_s + d_u$ , and our improved training strategies demonstrate robustness under this condition.

**Theorem 2.** Assuming the system state variable  $\mathbf{s}(\mathbf{x}, t)$  in Eq. (1) has a dimensionality of  $d_s$  and a spatial dimensionality of  $d$ , and  $F$  represents at most  $q$ -th order space derivatives of  $\mathbf{s}$ . Then we can derive a recursive formula for the length  $S_d(d_s, q)$  of  $\mathbf{L}_1$  with respect to  $d$ , and it holds that

$$S_d(d_s, q) = \begin{cases} d_s(q+1) + d_u, & d = 1, \\ \sum_{i=0}^q (S_{d-1}(d_s, i) - d_u) + d_u, & d > 1. \end{cases}$$

In addition, in the NODEs (Chen et al. 2018; Zhuang et al. 2020), the adjoint sensitivity method is used to calculate the gradients with a low cost of memory independent of the depth/time. Similarly, we can obtain the adjoint dynamics of the SNODEs, see Appendix A.3 for details.

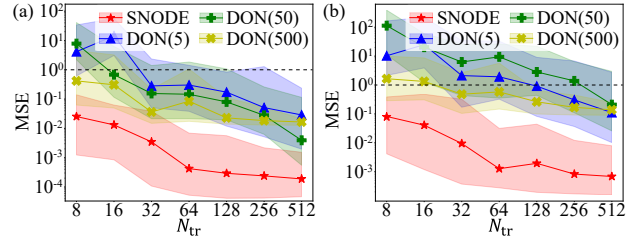


Figure 2: Operator learning for system (6). (a) The mean squared error (MSE) of different methods and different  $N_{tr}$  in the testing set. (b) The MSE in the extrapolation experiment. Here,  $m$  in “DON( $m$ )” represents the number of the uniform sampling points of the parameter function.

## Experiments

In this section, we provide a comprehensive analysis of the performance of our method across experiments with 64GB RAM and NVIDIA Tesla V100 GPU 16GB. To train and test our approach for a given dynamical system  $\mathbf{s}(\mathbf{x}, t, \mathbf{u}(\mathbf{x}, t))$ , we generate  $N_{tr}$  and  $N_{te}$  trajectories, respectively. Specifically, we utilize a Gaussian random field (GRF) to generate a 1-d parameter function for each track. The GRF is characterized by its mean, denoted by  $\mu$ , and a radial basis function (RBF) kernel, given by

$$u \sim \mathcal{G} \left\{ \mu, \exp \left[ -\frac{\|t_1 - t_2\|^2}{(2l)^2} \right] \right\}, \quad (5)$$

where  $l$  represents the length scale that determines the smoothness of the sampling function. We then multiply the sampling function by an output scaling factor  $c$  to obtain our parameter function  $u$ . In practice, we can only sample a discrete set of points instead of a continuous function. In this case, we can obtain a continuous parameter function through interpolation methods, such as the cubic spline interpolation, as described in Appendix B.1. In addition, the parameter settings for all experiments are provided in Appendix B.2.

### A Simple 1-d Parametric ODE

We first consider a simple 1-d ODE, described by

$$\partial_t s(t) = F[s(t), u(t), t], \quad t \in [0, 1], \quad (6)$$

where  $u(t)$  represents a sampling function from a GRF on the interval  $[0, 1]$ . We conduct experiments on the nonlinear example from the DeepONet work (Lu, Jin, and Karniadakis 2019), where  $F[s(t), u(t), t] = -s^2(t) + u(t)$ . To increase the task difficulty, we randomly sample the initial value  $s_0$ , the length scale  $l$ , and the output scale  $c$  from their respective uniform distributions, as shown in the Appendix B.2.

The experimental results are presented in Figure 2. Our approach consistently outperforms the baselines regarding the prediction error, regardless of the number of points sampled from  $u(t)$  by the DeepONet method. Moreover, our method exhibits a more significant advantage in the extrapolation prediction when the output scale  $c$  selected exceeds the training setting ( $c \in [0, 10]$  for training and  $c = 14$  for the extrapolation experiment). This can be attributed to our approach’s ability to fully utilize the temporal information of  $u(t)$  within the framework of NODEs and learn the inherent relationship between the parameter function  $u(t)$  and the

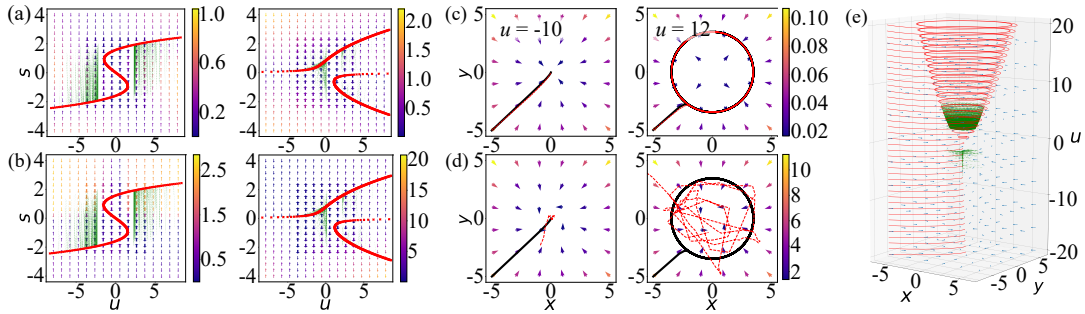


Figure 3: Experimental results in systems with saddle-node, pitchfork and Hopf bifurcations. (a), (c), (e) Predicting bifurcation dynamics using SNODEs. (b), (d) Predicting bifurcation dynamics using DeepONet. Here, the green dashed lines indicate the training data, the red dots represent the true bifurcation diagram, the arrows depict the predicted vector field, the black line represents the true trajectory, and the red line represent predicted trajectory.

system dynamics, thereby enabling highly accurate modeling of the parametric ODEs.

### Learning Dynamical Systems Near the Bifurcation

In this section, we consider a ubiquitous and significant bifurcation scenario that has gained considerable attention across various scientific fields. When a parameter function reaches a bifurcation point, the qualitative or topological nature of the system undergoes a mutation, exhibiting highly complex dynamical behavior. By accurately learning the operator from the parameter function to the system state, SNODEs can facilitate the extrapolation prediction of the bifurcation dynamics of the system. This capability is vital for comprehending and predicting the system’s tipping point.

We select three distinct dynamical systems with varying bifurcation types with respect to the parameter function  $u(t)$ , where  $u(t)$  is sampled from the GRF near the bifurcation value  $u^*$  of the system parameter. The first is the saddle-node bifurcation with  $u^* = \pm 2.28$ , and its dynamical equation is

$$\partial_t s(t) = u + \theta_1 s + \theta_2 s^3, \theta_1 = 2.5, \theta_2 = -1.$$

The second is the pitchfork bifurcation with  $u^* = 1.19$ , which has the form

$$\partial_t s(t) = \theta_1 + us + \theta_2 s^3, \theta_1 = 0.5, \theta_2 = -1.$$

The finally is the Hopf bifurcation with  $u^* = 0$ , which reads

$$\begin{aligned} \partial_t s_1(t) &= us_1 - s_2 - s_1(s_1^2 + s_2^2), \\ \partial_t s_2(t) &= s_1 + us_2 - s_2(s_1^2 + s_2^2). \end{aligned}$$

Subsequently, we assess our SNODEs method by predicting these three bifurcation dynamics.

To begin with, we consider two 1-d dynamical systems that exhibit a saddle-node bifurcation with a hysteresis loop, and a pitchfork bifurcation with a cusp catastrophe (Szép, Dalchau, and Csikász-Nagy 2021), respectively. To increase the task difficulty, we control the parameter function to avoid passing through a neighborhood near the bifurcation point, and then test the extrapolation ability of our method (the results are shown in Figure 3(a)). To verify the superiority of our method, we conduct the same experiments using the DeepONet as well. The results are presented in Figure 3(b). It is evident that the DeepONet method exhibits poor performance in the extrapolation prediction outside the green

area. However, our method, with its powerful extrapolation prediction ability, accurately learns the bifurcation dynamics within a larger region.

Additionally, we consider a 2-d dynamical system that exhibits Hopf bifurcation. To investigate the bifurcation dynamics of the system, we sample the parameter functions on both sides of the bifurcation value and generate the corresponding dynamical trajectories as the training set, as illustrated by the green dashed line in Figure 3(e). To discover the bifurcation dynamics of the system, we perform experiments on the parameter functions with a constant value in a wide range. As depicted in Figures 3(c)-(e), the red dashed line represents the predicted trajectory, the arrows depict the predicted vector field, and the black solid line corresponds to the true trajectory. The experimental results demonstrate that the learned operator mapping using our proposed framework exhibits superior capabilities in reconstructing and extrapolating the bifurcation dynamics.

The SNODEs framework, augmented with its efficient training strategies, has been successfully applied in the aforementioned parametric ODE systems. When furnished with sufficient training data, the SymNet within the SNODEs framework accurately infers the underlying equations of unknown systems. In scenarios where the training data is scarce and the system is complex, SymNet can still precisely identify key components of the underlying dynamics. Following this identification, by integrating with GeNN’s residual learning, it can realize a more accurate modeling of the system. For more training details, please refer to Appendix B.3.

### Applications in Parametric PDEs

We further demonstrate the modeling capability of SNODEs in parametric PDEs with unknown equations, such as the diffusion-reaction (DR), Kuramoto-Sivashinsky (KS), and 2-d Navier-Stokes (NS) systems. To leverage the physical prior knowledge, we incorporate the potential spatial derivatives as additional inputs to our SNODEs method, thereby enhancing the precision in learning the dynamics from parametric functions to system states. In the following experiments, we consider periodic boundary conditions, and the initial conditions and other experimental configurations are presented in Appendix B.2.

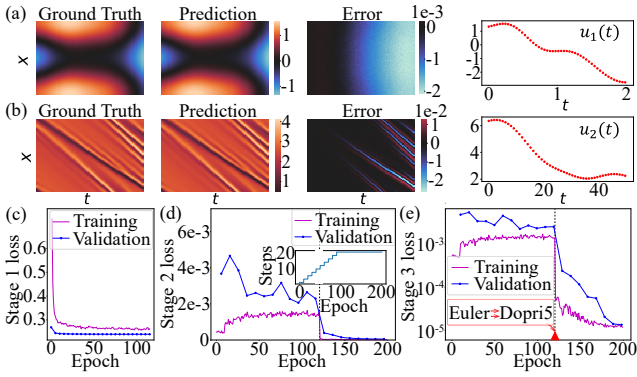


Figure 4: Predicting DR and KS systems using SNODEs. (a) and (b) show the testing examples for the DR and KS systems, respectively. Here, the training set features a spatial resolution of  $N_x = 32$ , whereas the test set has  $N_x = 128$ . (c), (d), and (e) respectively illustrate the variations in training and validation losses across three training stages.

We first consider the DR system. In practice, the source term may vary over time due to changes in environmental factors, yielding the following equation with a time-varying source term  $u(x, t)$ :

$$s_t = Ds_{xx} + Ks^2 + u(x, t), \quad x \in [0, 1], t \in [0, 1], \quad (7)$$

where  $D = 0.01$  is the diffusion coefficient and  $K = 0.01$  is the reaction rate. Here, we use  $u(x, t) = (\pi x)/5 + u_1(t)$ , where  $u_1(t)$  is a sampling function from the GRF. In addition, we consider the KS equation, which exhibits complex chaotic dynamics, of the following form

$$s_t = -ss_x - s_{xx} - u(x, t)s_{xxx}, \quad x \in [0, 32\pi], t \in [0, 20], \quad (8)$$

where  $u(x, t) = x/16 + u_2(t)$ , and  $u_2(t)$  is the sampling function from the GRF. After training, our SNODEs framework exhibits excellent modeling capabilities on the aforementioned PDE systems. Moreover, our framework can be naturally applied to the super-resolution learning by using higher-resolution spatiotemporal data as the test set. In the temporal dimension, our approach is capable of estimating the system state for any given future moment through the ODESolve process. In terms of spatial dimension, our method allows for training with lower resolution data ( $N_x = 32$ ) and testing in higher resolution data ( $N_x = 128$ ). The corresponding outcomes are presented in Figures 4(a)-(b).

Here, we take the DR system as an example and provide the training details for our SNODEs framework. In stage 1, the rapid capture of critical components of unknown dynamics is facilitated through flow matching pre-training, with the training error depicted in Figure 4(c), and we obtain  $\hat{F}_1 = 0.0099s_{xx} + 0.9955u$ . Herein, the regularization parameter  $\alpha$  is set to 0.01, resulting in the training loss exceeding the validation loss. In stage 2,  $\hat{F}_1$  was fine-tuned through the ODESolve prediction, with the corresponding prediction error illustrated in Figure 4(d), culminating in  $\hat{F}_1 = 0.0098s_{xx} + u - 0.0083s + 0.0046$ . Herein, we employ the "Euler" method for the initial 120 epochs, incorporating

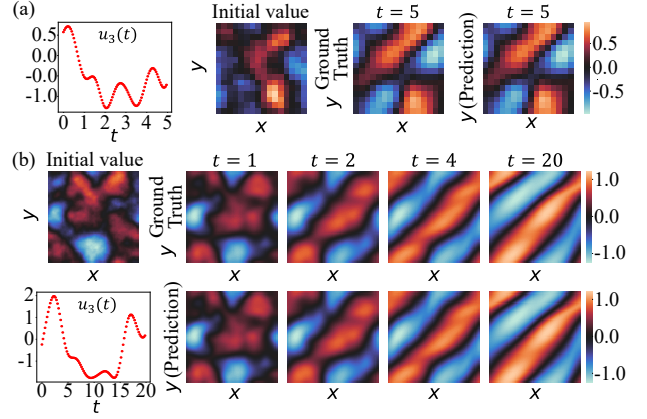


Figure 5: Predicting NS systems using SNODEs. (a) The predicted result for training data with spatial resolution  $N_x = N_y = 16$ . (b) The predicted result for testing data with spatial resolution  $N_x = N_y = 80$ .

progressively increasing prediction steps. Subsequently, for the remaining 80 epochs, the number of prediction steps is maintained at a constant 20, and we switch to the "Dopri5" method for further training. In stage 3, we maintain  $\hat{F}_1$  fixed and employ a GeNN for residual learning, where the optimization objective is  $F - \hat{F}_1$  and the prediction error is presented in Figure 4(e). Herein, the training strategy remains consistent with that of stage 2. In fact, within the temporal-spatial domain of the DR experiment, the influence of  $Ks^2$  on the dynamics is minimal. Consequently, this term was not identified in stage 1. However, accurate predictions were achieved through the employment of simple substitute terms in stage 2. Then in the stage 3, the modeling accuracy was further enhanced through residual learning.

Finally, we consider a 2-d NS system for a viscous, incompressible fluid in vorticity form, which reads

$$s_t = \gamma_x s_y - \gamma_y s_x + \nu \Delta s + u(x, y, t), \quad \Delta \gamma = -s, \quad (x, y) \in [0, 2]^2, t \in [0, 20], \quad (9)$$

where  $\gamma$  represents the stream function,  $\Delta$  is the Laplacian operator, and  $\nu = 0.001$ . Additionally,  $u(x, y, t) = u_3(t) \times \{0.1 \sin[2\pi(x+y)] + \cos[2\pi(x+y)]\}$  is the forcing function, where  $u_3(t)$  is a function obtained from a GRF. The system is defined over a square domain with dimensions  $[0, 2]^2$ , and the time interval is  $[0, 20]$ . Under this condition, we can express the stream function as the vorticity, i.e.,  $\gamma = -\Delta^{-1}s$ . In the discrete scenario within the Fourier domain, this corresponds to  $\tilde{\gamma} = -1/(\mathbf{k}^2 + \mathbf{l}^2)\tilde{s}$ . To facilitate the training, we augment the first layer of SymNet with  $i\mathbf{k}/(\mathbf{k}^2 + \mathbf{l}^2)\tilde{s}$  and  $i\mathbf{l}/(\mathbf{k}^2 + \mathbf{l}^2)\tilde{s}$  in the Fourier domain. Then after training, Figure 5 demonstrates that our framework achieves the accurate operator learning in modeling the underlying dynamics, enabling the precise and stable prediction of system evolutions, even the initial values and parameter functions outside the training set distribution. Additional training details and experimental results for the above parametric PDE systems are provided in Appendix B.4.

Conditions	$N_{tr} = 100, \sigma_n = 0\%$			$N_{tr} = 1000, \sigma_n = 0\%$			$N_{tr} = 1000, \sigma_n = 3\%$		
Experiments	DR	KS	NS	DR	KS	NS	DR	KS	NS
SNODEs	<b>6e-4±5e-3</b>	<b>0.03±0.02</b>	<b>0.11±0.09</b>	<b>3e-6±2e-6</b>	<b>2e-5±3e-4</b>	<b>6e-4±5e-3</b>	<b>0.53±0.28</b>	<b>0.41±0.14</b>	<b>1.37±0.67</b>
DeepONet	27.5±32.4	30.2±24.1	38.3±42.4	0.67±0.53	2.56±1.87	7.30±5.24	3.74±2.64	4.39±4.33	21.6±29.8
FNO	30.1±35.3	72.1±42.2	43.9±35.1	1.70±1.09	2.90±3.08	5.10±3.54	2.45±2.28	3.93±2.68	9.22±7.63
PDE-NET	4.21±3.81	3.13±2.66	8.20±6.68	2.04±1.25	1.31±1.42	3.30±2.87	2.28±1.86	2.01±2.42	3.49±2.38
MP-PDE	2.79±3.24	4.67±3.12	6.65±3.68	0.62±0.28	1.57±0.95	2.26±1.74	1.18±0.61	1.94±1.34	2.85±2.06
E1 (no SymNet)	0.18±0.12	0.32±0.21	0.35±0.28	0.01±0.03	0.11±0.14	0.08±0.13	1.06±0.84	1.39±0.66	1.55±1.27
E2 (no GeNN)	0.41±0.62	0.08±0.04	0.68±0.27	0.17±0.15	5e-5±6e-4	3e-3±8e-3	0.93±1.09	0.82±0.94	1.67±1.52

Table 1: The prediction MSE ( $\pm$  two standard deviations) under different training set sizes  $N_{tr}$  and noise levels  $\sigma_n$ . Here, we consider zero-mean Gaussian noise with a standard deviation of  $\sigma_n$  times the mean absolute value of the training data.

## Comparative Analysis and Ablation Studies

For the fair comparison, we conduct a comparative analysis with several standard baselines, namely DeepONet (Lu, Jin, and Karniadakis 2019), FNO (Li et al. 2020c), PDE-NET (Long, Lu, and Dong 2019), and MP-PDE (Brandstetter, Worrall, and Welling 2022). Additionally, we conducted ablation experiments on SNODEs by separately removing SymNet and GeNN, denoted as E1 (no SymNet) and E2 (no GeNN) respectively. Then we employ the high-precision pseudospectral method to solve parametric PDEs and obtain the dataset. In particular, to increase the task difficulty, we set the output scale of the sampling function to be larger than that of the training set and randomly generated initial values (see Appendix B.2 for details).

The experimental results, which are presented in Table 1, demonstrate that our framework is capable of accurately modeling system dynamics even in scenarios where training data is limited and noisy. Consequently, the SNODEs outperform the DeepONet and FNO methods in terms of interpretability and extrapolation capability, particularly in the NS experiment with random initial values. This is because SNODEs sufficiently leverage additional physical priors by adding the partial derivative terms into  $\mathbf{F}$ . Compared to the PDE-NET and MP-PDE methods, our framework enables efficient symbolic regression over a larger search space that includes parametric functions. This is due to the joint use of three-stage learning and ODEsolve strategies, enabling the more precise modeling of spatiotemporal dynamics from limited noisy data. Furthermore, results from ablation experiments E1 and E2 clearly demonstrate that these two components mutually enhance each other, with SymNet enhancing the model’s interpretability and generalization abilities, while GeNN strives for greater modeling precision.

Config	SNODEs	NODE	DeepONet	FNO	PDE-NET	MP-PDE
DR	221	1580	372	557	430	943
KD	489	2920	407	649	513	1017
NS	824	12476	694	1120	902	2389

Table 2: Comparison of training times using different methods. Here, the unit of measurement is seconds.

As an advancement of the NODEs, our SNODEs framework employs the flow matching pre-training strategy, effectively optimizing the high computational cost associated with classical NODE methods. To assess the efficiency of our proposed method, we conduct the runtime comparisons with the baselines in the DR, KS, and NS experiments. The

results are displayed in the Table 2 of the attached file. It demonstrates that our framework not only achieves the higher precision but also maintains the good training efficiency. However, it is worth noting that the computation cost during the inference process is contingent upon the choice of the ODE solver, with higher precision numerical solvers potentially requiring longer computation time.

## Concluding Remarks

In this work, we present the SNODEs framework for learning the operator mapping from the parameter functions to system states in parametric dynamical systems. Specifically, this framework establishes the SCDNNs which effectively integrate the strengths of symbolism and connectionism. It initially employs SymNet to learn potential underlying equations, followed by utilizing GeNN for residual learning, thereby achieving a more accurate model of the physical world. To enhance the robustness of our method, we also propose the three-stage and adaptive training strategy that simultaneously improves the stability and efficiency of modeling unknown dynamics.

Moreover, by leveraging the Fourier transform to incorporate potential partial derivative terms into the first layer of SymNet, SNODEs enable accurate and resolution-invariant modeling of the PDE systems at any given time and on any spatial grid. We evaluate SNODEs method across multiple representative ODE and PDE systems. Experimental results demonstrate that, compared to the standard baselines, our framework not only achieves higher prediction accuracy but also exhibits superior extrapolation capabilities, interpretability, and robustness. Consequently, our framework can be applied to a broader range of scientific problems and real-world systems.

However, our approach has certain limitations that warrant further investigation in the future. For example, the Fourier transform within the SNODEs framework is primarily suited for uniformly spaced spatial data. Therefore, for PDE systems encompassing complex geometric domains, integration with advanced strategies (such as non-uniform Fourier transform or other numerical methods) is necessary. In addition, when the underlying dynamics become increasingly complex, deep SymNet may exhibit instability during the training process. Therefore, incorporating more sophisticated symbolic regression approaches, such as KAN (Liu et al. 2024), into the SNODEs framework represents a potential and significant direction for future exploration.

## References

- Benner, P.; Gugercin, S.; and Willcox, K. 2015. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM review*, 57(4): 483–531.
- Biloš, M.; Sommer, J.; Rangapuram, S. S.; Januschowski, T.; and Günnemann, S. 2021. Neural flows: Efficient alternative to neural ODEs. *Advances in Neural Information Processing Systems*, 34: 21325–21337.
- Brandstetter, J.; Worrall, D. E.; and Welling, M. 2022. Message Passing Neural PDE Solvers. *ArXiv*, abs/2202.03376.
- Brunton, S. L.; and Kutz, J. N. 2022. *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.
- Chen, G.; Liu, X.; Li, Y.; Meng, Q.; and Chen, L. 2023. Laplace neural operator for complex geometries. *ArXiv*, abs/2302.08166.
- Chen, R. T.; Rubanova, Y.; Bettencourt, J.; and Duvenaud, D. K. 2018. Neural ordinary differential equations. *Advances in neural information processing systems*, 31.
- Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- Cranmer, M.; Greydanus, S.; Hoyer, S.; Battaglia, P.; Spergel, D.; and Ho, S. 2020. Lagrangian neural networks. *arXiv preprint arXiv:2003.04630*.
- De Brouwer, E.; Simm, J.; Arany, A.; and Moreau, Y. 2019. GRU-ODE-Bayes: Continuous modeling of sporadically-observed time series. *Advances in neural information processing systems*, 32.
- Devaney, R. L. 2021. *An introduction to chaotic dynamical systems*. CRC press.
- Di Giovanni, F.; Rowbottom, J.; Chamberlain, B. P.; Markovich, T.; and Bronstein, M. M. 2022. Graph neural networks as gradient flows. *arXiv preprint arXiv:2206.10991*.
- Dupont, E.; Doucet, A.; and Teh, Y. W. 2019. Augmented Neural ODEs. *ArXiv*, abs/1904.01681.
- Finlay, C.; Jacobsen, J.-H.; Nurbekyan, L.; and Oberman, A. 2020. How to train your neural ODE: the world of Jacobian and kinetic regularization. In *International conference on machine learning*, 3154–3164. PMLR.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
- Holt, S. I.; Qian, Z.; and van der Schaar, M. 2022. Neural Laplace: Learning diverse classes of differential equations in the Laplace domain. In *International Conference on Machine Learning*, 8811–8832. PMLR.
- Jin, P.; Meng, S.; and Lu, L. 2022. MIONet: Learning multiple-input operators via tensor product. *SIAM J. Sci. Comput.*, 44: 3490–.
- Kidger, P.; Morrill, J.; Foster, J.; and Lyons, T. 2020. Neural Controlled Differential Equations for Irregular Time Series. *ArXiv*, abs/2005.08926.
- Kim, S.; Ji, W.; Deng, S.; and Rackauckas, C. 2021. Stiff Neural Ordinary Differential Equations. *Chaos*, 31 9: 093122.
- Kovachki, N. B.; Li, Z.-Y.; Liu, B.; Azizzadenesheli, K.; Bhattacharya, K.; Stuart, A.; and Anandkumar, A. 2021. Neural Operator: Learning Maps Between Function Spaces. *ArXiv*, abs/2108.08481.
- Li, X.; Zhang, J.; Zhu, Q.; Zhao, C.; Zhang, X.; Duan, X.; and Lin, W. 2024. From Fourier to Neural ODEs: Flow matching for modeling complex systems. *arXiv preprint arXiv:2405.11542*.
- Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; and Anandkumar, A. 2020a. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*.
- Li, Z.; Kovachki, N.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; and Anandkumar, A. 2020b. Neural operator: Graph kernel network for partial differential equations. *arXiv preprint arXiv:2003.03485*.
- Li, Z.-Y.; Kovachki, N. B.; Azizzadenesheli, K.; Liu, B.; Bhattacharya, K.; Stuart, A.; and Anandkumar, A. 2020c. Fourier Neural Operator for Parametric Partial Differential Equations. *ArXiv*, abs/2010.08895.
- Liu, X.; Xiao, T.; Si, S.; Cao, Q.; Kumar, S.; and Hsieh, C.-J. 2019. Neural SDE: Stabilizing Neural ODE Networks with Stochastic Noise. *ArXiv*, abs/1906.02355.
- Liu, Y.; Liu, Q.; Zhang, J.-W.; Feng, H.; Wang, Z.; Zhou, Z.; and Chen, W. 2022. Multivariate Time-Series Forecasting with Temporal Polynomial Graph Neural Networks. In *Advances in Neural Information Processing Systems*.
- Liu, Z.; Wang, Y.; Vaidya, S.; Ruehle, F.; Halverson, J.; Soljačić, M.; Hou, T. Y.; and Tegmark, M. 2024. Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*.
- Long, Z.; Lu, Y.; and Dong, B. 2019. PDE-Net 2.0: Learning PDEs from data with a numeric-symbolic hybrid deep network. *Journal of Computational Physics*, 399: 108925.
- Long, Z.; Lu, Y.; Ma, X.; and Dong, B. 2018. Pde-net: Learning pdes from data. In *International conference on machine learning*, 3208–3216. PMLR.
- Lu, L.; Jin, P.; and Karniadakis, G. E. 2019. DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. *ArXiv*, abs/1910.03193.
- Lu, L.; Meng, X.; Cai, S.; Mao, Z.; Goswami, S.; Zhang, Z.; and Karniadakis, G. E. 2021. A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data. *Computer Methods in Applied Mechanics and Engineering*.
- Navarro-Moreno, J. 2008. ARMA prediction of widely linear systems by using the innovations algorithm. *IEEE Transactions on Signal Processing*, 56(7): 3061–3068.

Pilva, P.; and Zareei, A. 2022. Learning time-dependent PDE solver using Message Passing Graph Neural Networks. *arXiv preprint arXiv:2204.07651*.

Raissi, M.; Perdikaris, P.; and Karniadakis, G. E. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378: 686–707.

Rudy, S. H.; Brunton, S. L.; Proctor, J. L.; and Kutz, J. N. 2016. Data-driven discovery of partial differential equations. *Science Advances*, 3.

Sahoo, S.; Lampert, C.; and Martius, G. 2018. Learning equations for extrapolation and control. In *International Conference on Machine Learning*, 4442–4450. PMLR.

Sanchez-Gonzalez, A.; Bapst, V.; Cranmer, K.; and Battaglia, P. 2019. Hamiltonian graph networks with ode integrators. *arXiv preprint arXiv:1909.12790*.

Szép, G.; Dalchau, N.; and Csikász-Nagy, A. 2021. Parameter Inference with Bifurcation Diagrams. *ArXiv*, abs/2106.04243.

Van Houdt, G.; Mosquera, C.; and Nápoles, G. 2020. A review on the long short-term memory model. *Artificial Intelligence Review*, 53: 5929–5955.

Wang, S.; Wang, H.; and Perdikaris, P. 2021. Learning the solution operator of parametric partial differential equations with physics-informed deepnets. *Science advances*, 7(40): eabi8605.

Wang, W.-X.; Lai, Y.-C.; and Grebogi, C. 2016. Data based identification and prediction of nonlinear and complex dynamical systems. *Physics Reports*, 644: 1–76.

Wu, Z.; and Zhang, R. 2019. Learning physics by data for the motion of a sphere falling in a non-Newtonian fluid. *Communications in Nonlinear Science and Numerical Simulation*, 67: 577–593.

Xiong, W.; Huang, X.; Zhang, Z.; Deng, R.; Sun, P.; and Tian, Y. 2023. Koopman neural operator as a mesh-free solver of non-linear partial differential equations. *ArXiv*, abs/2301.10022.

Zappala, E.; de Oliveira Fonseca, A. H.; Moberly, A. H.; Higley, M. J.; Abdallah, C. G.; Cardin, J. A.; and van Dijk, D. 2022. Neural Integro-Differential Equations. *ArXiv*, abs/2206.14282.

Zhu, Q.; Guo, Y.; and Lin, W. 2021. Neural delay differential equations. *arXiv preprint arXiv:2102.10801*.

Zhu, Q.; Li, X.; and Lin, W. 2023. Leveraging neural differential equations and adaptive delayed feedback to detect unstable periodic orbits based on irregularly sampled time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(3): 031101.

Zhuang, J.; Dvornik, N.; Li, X.; Tatikonda, S.; Papademetris, X.; and Duncan, J. 2020. Adaptive checkpoint adjoint method for gradient estimation in neural ode. In *International Conference on Machine Learning*, 11639–11649. PMLR.