Operator Learning and Deep Operator Networks

Simone Venturi, Tiernan Casey

Extreme-Scale Data Science & Analytics (8739)

Part of the Code Documentation for
Neural Networks for Reduced Order Modeling (ROMNet)

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Operator Learning

- Machine Learning (ML)-Based Surrogates [1-3]

“PINNs penalize the residual of governing physics-based relations for the system in the loss function, where the partial derivatives are computed through automatic differentiation.” [10]

- Physics-Informed ML [4] and Physics-Informed Neural Networks [5-9]

[1] S. L. Brunton et al - Machine Learning for Fluid Mechanics - 2020
[2] K. Lee and K. T. Carlberg (SANDIA) - Model reduction of dynamical systems on nonlinear manifolds using … - 2020
[3] E. Quin et al - Lift & Learn: Physics-informed machine learning for large-scale nonlinear dynamical systems - 2020

[4] G. Karniadakis et al - Physics-informed machine learning – 2021
[5] M. W. M. G. Dissanayake and N. Phan-Thien - Neural-network-based approximations for solving partial differential equations - 1994
[6] I.E. Lagaris et al - Artificial neural networks for solving ordinary and partial differential equations - 1998
[7] M. Raissi et al - Physics Informed Deep Learning (Part I)- Data-driven Solutions of Nonlinear Partial Differential Equations – 2017
[8] M. Raissi et al - Physics Informed Deep Learning (Part II)- Data-driven Discovery of Nonlinear Partial Differential Equations – 2017
[9] M. Raissi et al - Physics-informed neural networks- A deep learning framework for solving forward … - 2019

[10] M. Yin et al - Simulating progressive intramural damage leading to aortic dissection using DeepONet: … - 2022

Operator Learning

- Machine Learning (ML)-Based Surrogates [1,2,3]

"PINNs penalize the residual of governing physics-based relations for the system in the loss function, where the partial derivatives are computed through automatic differentiation." [10]

- Physics Informed ML [4] and Neural Networks [5-9]

"Yet, for a system with different boundary/initial conditions, one has to retrain the network for each case, making the algorithm time-consuming.
Hence, there is a pressing need to develop models that can learn the operator level of mapping between functions; that is, predicting the physical system under diverse boundary/initial conditions." [10]

- **Operator Learning**

[10] M. Yin et al - Simulating progressive intramural damage leading to aortic dissection using DeepONet: … - 2022

Operator Learning

"The goal is to learn the functional response of a system from a functional input, such as an initial/boundary condition or forcing term. In the context of learning the response of systems governed by differential equations, these learned models can function as fast surrogates of traditional numerical solvers.

…

While architectures such as convolutional neural networks may perform well under this setting, this approach can be somewhat limited. For instance, if we desired the value of the output at a query location outside of the training grid, an entirely new model would need to be built and tuned from scratch. This is a consequence of choosing to discretize the regression problem before building a model to solve it. If instead we formulate the problem and model at the level of the (infinite-dimensional) input and output function spaces, and then make a choice of discretization, we can obtain methods that are more flexible with respect to the locations of the point-wise measurements." [1]

[1] G. Kissas et al - Learning operators with coupled attention – 2022

Operator Learning

1. “Parametric approaches (Part I):

   - Chen et al. [2]: The authors proposed a method for learning non-linear operators based on a one-layer feed-forward neural network architecture. Moreover, the authors presented a universal approximation theorem which ensures that their architecture can approximate any continuous operator with arbitrary accuracy.
   - Lu et al. [3,4]: Deep Operator Network (DeepONet): An extension of the architecture above, built with multiple layer feed-forward neural networks

   - Li et al. [5]: Graph Neural Operator (GKN), motivated by the solution form of linear partial differential equations (PDEs) and their Greens’ functions.
   - Li et al. [6]: As an extension of this work, the authors also proposed a Graph Neural Operator architecture where a multi-pole method is used sample the spatial grid allowing the kernel to learn in a non-local manner.
   - Li et al. [7]: Fourier Neural Operator (FNO): In later published work, this framework has been extended to the case where the integral kernel is stationary, enabling one to efficiently compute the integral operator in the Fourier domain.” [1]

[1] G. Kissas et al - Learning operators with coupled attention – 2022
[2] T. Chen and H. Chen - Universal Approximation to Nonlinear Operators by Neural Networks … - 1995
[3] L. Lu et al - DeepONet: Learning nonlinear operators for identifying differential equations based on the universal … - 2019
[4] L. Lu et al - Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators - 2021
[5] Z. Li et al - Neural Operator: Graph Kernel Network for Partial Differential Equations - 2020
[6] Z. Li et al - Multipole Graph Neural Operator for Parametric Partial Differential Equations - 2020
[7] Z. Li et al - Fourier Neural Operator for Parametric Partial Differential Equations – 2020

Operator Learning

1. "Parametric approaches (Part II):

   - R. Patel et al [8] (SANDIA): Applies a neural network parameterization of governing physics in modal space, allowing a characterization of differential operators while providing structure which may be used to impose biases related to symmetry, isotropy, and conservation form.

   - H. You et al. [9] (SANDIA): Nonlocal Kernel Network (NKN): NKN stems from the interpretation of the neural network as a discrete nonlocal diffusion reaction equation that, in the limit of infinite layers, is equivalent to a parabolic nonlocal equation.

   - K. Bhattacharya et al. [10]: General framework for data-driven approximation of input-output maps between infinite-dimensional spaces. The proposed approach is motivated by the recent successes of neural networks and deep learning, in combination with ideas from model reduction.

[1] G. Kissas et al - Learning operators with coupled attention – 2022
[8] R. Patel et al - A physics-informed operator regression framework for extracting data-driven continuum models - 2020
[9] H. You et al - Nonlocal Kernel Network (NKN): a Stable and Resolution-Independent Deep Neural Network - 2022
[10] K. Bhattacharya et al - Model Reduction and Neural Networks for Parametric PDEs - 2020

Operator Learning

1. “Parametric approaches (Part II):

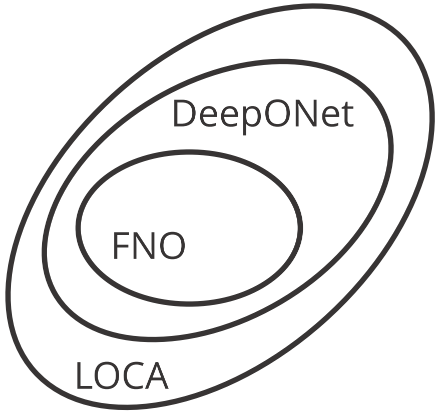
    - C. R. Gin et al. [11]: DeepGreen: A deep learning approach for directly approximating the Green’s function of differential equations [2]

    - N. H. Nelsen et al. [12]: Multi-wavelet approach for learning projections of an integral kernel operator to approximate the true operator and a random feature approach for learning the solution map of PDEs [3]

Note: While the Fourier Neural Operator and the DeepONet methods come with theoretical guarantees of universal approximation, meaning that under some assumptions these classes of models can approximate any continuous operator to arbitrary accuracy, no theoretical guarantees of the approximation power of these last approaches are presented.” [1]

[1] G. Kissas et al - Learning operators with coupled attention – 2022
[11] C. R. Gin et al - DeepGreen: deep learning of Green’s functions for nonlinear boundary value problems - 2021
[12] N. H. Nelsen et al - The Random Feature Model for Input-Output Maps between Banach Spaces - 2020

Operator Learning



2. “Non-parametric approaches:

   - Operator-Valued Kernel methods:

     - …

     - H. Kadri et al. [13]: Focused on the use of reproducing kernel Hilbert space theory to learn from such functional data.

     - G. Kissas et al. [1]: Learning Operators with Coupled Attention (LOCA), motivated from the recent success of the attention mechanism. The input functions are mapped to a finite set of features which are then averaged with attention weights that depend on the output query locations.

     - N. Trask et al [14] (SANDIA): Generalized Moving Least Squares (GMLS) - Generalizes CNNs by introducing methods for data on unstructured point clouds based on GMLS. GMLS is a non-parametric technique for estimating linear bounded functionals from scattered data, and has recently been used in the literature for solving partial differential equations.

[1] G. Kissas et al - Learning operators with coupled attention – 2022
[13] H. Kadri et al. - Operator-valued Kernels for Learning from Functional Response Data – 2016
[14] N. Trask et al - GMLS-Nets- A framework for learning from unstructured data - 2019

Operator Learning

3. Functional Data Analysis (FDA) [15,16]: Formulating models with functional data"[1]

[1] G. Kissas et al - Learning operators with coupled attention – 2022
[14] J. O. Ramsay - When the data are functions - 1982
[15] J. O. Ramsay et al - Some Tools for Functional Data Analysis - 1991

Critiques to Fourier Neural Operator – Part I

"Another parallel effort on operator regression started in 2020 with a paper on a graph kernel network (GKN) for PDEs [2]. The authors represented the infinite-dimensional mapping by composing nonlinear activation functions and a class of integral operators with the kernel integration computed by message passing on graph networks.

Unfortunately, GKN was of limited use as it was shown to be unstable with the increase of the number of hidden layers [3].

A different architecture was then proposed by the same group in [4], where they formulated the operator regression by parameterizing the integral kernel directly in Fourier space.

…

According to an independent work by [5], FNO in its continuous form can be viewed as a DeepONet with a specific architecture of the branch and a trunk represented by a trigonometric basis.

(… In other words, FNO can be thought of as a special form of branch network in DeepONet. …)

(… FNO in its continuous form can be thought of as a subcase of DeepONet with a specially-designed branch network and a discrete trigonometric basis to replace the trunk net. …)

…

On this point, it is worth noted that DeepONet was based from the onset on the theorem of Chen & Chen, whereas the formulation of FNO was not theoretically justified originally, and the recent theoretical work covers only invariant kernels." [1]

[1] L. Lu et al - A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data - 2021
[2] Z. Li et al - Neural Operator: Graph Kernel Network for Partial Differential Equations – 2020
[3] H. You et al - Nonlocal Kernel Network (NKN): a Stable and Resolution-Independent Deep Neural Network - 2022
[4] Z. Li et al - Fourier Neural Operator for Parametric Partial Differential Equations – 2020
[5] N. Kovachi et al - On universal approximation and error bounds for Fourier neural operators - 2021

Critiques to Fourier Neural Operator – Part II

“Several examples are demonstrated in Fig. 6, where we find that FNO is extremely sensitive to noise, i.e., it cannot handle 0.1% Gaussian noise of the input.

…

However, one major difference we found is the robustness to noise. In the examples of simulating the advection equation and instability waves, FNO was shown to be extremely sensitive to noise and totally failing to predict the solution even only when 0.1% Gaussian noise was imposed on the testing data.

…

This extreme sensitivity of FNO to noise implies that it may learn an unstable operator, which may not be able to generalize well even within the distribution of inputs.

…

The main difference between DeepONet and FNO is that DeepONet does not discretize the output, but FNO does. Moreover, DeepONet can employ any type of neural network architectures in the branch net whereas FNO has a fixed architecture, and hence DeepONet is more flexible than FNO in terms of problem settings and datasets (see comparison details in Table 1).” [1]

[1] L. Lu et al - A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data - 2021
[2] Z. Li et al - Neural Operator: Graph Kernel Network for Partial Differential Equations – 2020
[3] H. You et al - Nonlocal Kernel Network (NKN): a Stable and Resolution-Independent Deep Neural Network - 2022
[4] Z. Li et al - Fourier Neural Operator for Parametric Partial Differential Equations – 2020
[5] N. Kovachi et al - On universal approximation and error bounds for Fourier neural operators - 2021

Deep Operator Networks (DeepONets)

ARTICLES
https://doi.org/10.1038/s42256-021-00302-5
nature machine intelligence

Check for updates

Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators

Lu Lu [1], Pengzhan Jin [2,3], Guofei Pang [2], Zhongqiang Zhang [4] and George Em Karniadakis [2] ✉

It is widely known that neural networks (NNs) are universal approximators of continuous functions. However, a less known but powerful result is that a NN with a single hidden layer can accurately approximate any nonlinear continuous operator. This universal approximation theorem of operators is suggestive of the structure and potential of deep neural networks (DNNs) in learning continuous operators or complex systems from streams of scattered data. Here, we thus extend this theorem to DNNs. We design a new network with small generalization error, the deep operator network (DeepONet), which consists of a DNN for encoding the discrete input function space (branch net) and another DNN for encoding the domain of the output functions (trunk net). We demonstrate that DeepONet can learn various explicit operators, such as integrals and fractional Laplacians, as well as implicit operators that represent deterministic and stochastic differential equations. We study different formulations of the input function space and its effect on the generalization error for 16 different diverse applications.

[1] L. Lu et al - Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators - 2021

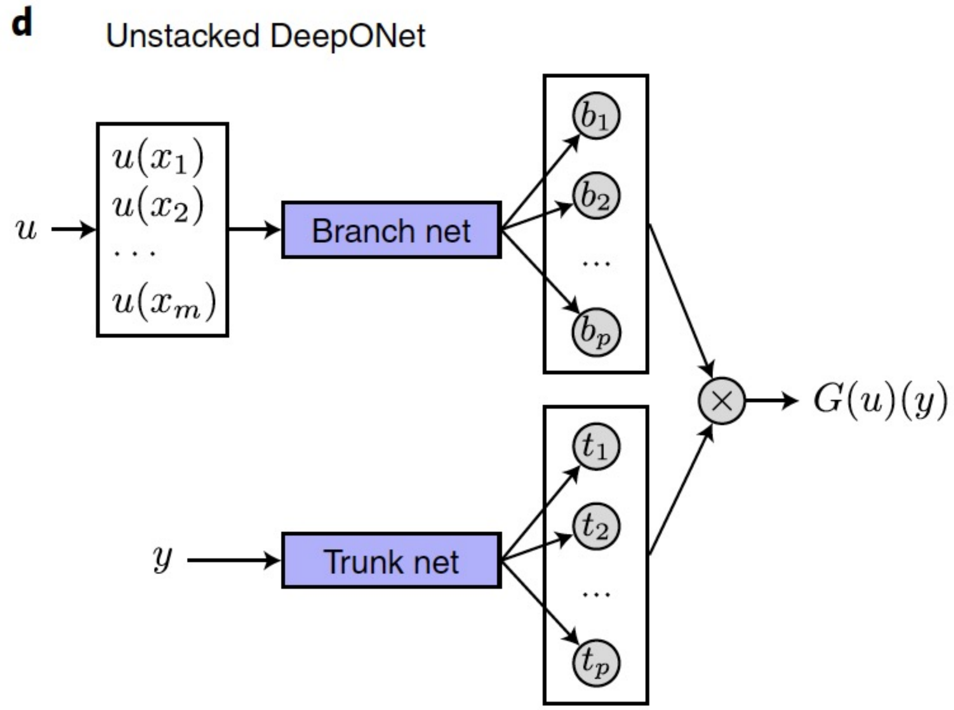
Deep Operator Networks (DeepONets)

Let G be an operator taking an input function u, with $G(u)$ being the corresponding output function. For any point y in the domain of $G(u)$, the output $G(u)(y)$ is a real number.

Theorem 2 (Generalized Universal Approximation Theorem for Operator). *Suppose that X is a Banach space, $K_1 \subset X$, $K_2 \subset \mathbb{R}^d$ are two compact sets in X and $\mathbb{R}^d$, respectively, V is a compact set in $C(K_1)$. Assume that $G: V \to C(K_2)$ is a nonlinear continuous operator. Then, for any $\epsilon > 0$, there exist positive integers m, p, continuous vector functions $\mathbf{g}: \mathbb{R}^m \to \mathbb{R}^p$, $\mathbf{f}: \mathbb{R}^d \to \mathbb{R}^p$, and $x_1, x_2, \ldots, x_m \in K_1$, such that*

$$\left| G(u)(y) - \langle \underbrace{\mathbf{g}(u(x_1), u(x_2), \cdots, u(x_m))}_{\text{branch}}, \underbrace{\mathbf{f}(y)}_{\text{trunk}} \rangle \right| < \epsilon$$

holds for all $u \in V$ and $y \in K_2$, where $\langle \cdot, \cdot \rangle$ denotes the dot product in $\mathbb{R}^p$. Furthermore, the functions $\mathbf{g}$ and $\mathbf{f}$ can be chosen as diverse classes of neural networks, which satisfy the classical universal approximation theorem of functions, for example, (stacked/unstacked) fully connected neural networks, residual neural networks and convolutional neural networks.



[1] L. Lu et al - Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators - 2021

Deep Operator Networks (DeepONets)

Universal Approximation Theorems for Operators (“Theory of DeepONets” [5]):
”Are statements that establish the density of a class of NNs within a space of mappings. Thus, UATs imply that NNs represent a wide variety of mappings when given appropriate weights and biases” [3]

1. [1]
2. [2]
3. [3]

“Both DeepONet and FNO suffer from the curse of dimensionality if one uses ReLU or tanh networks for Lipschitz continuous operators, due to the approximation capacity of these networks for high dimensional. However, rates of convergence for the DeepONet are obtained for some solution operators from PDE [2,4]” [5]

“In addition, complementary error analyses [2, 4] provide upper bounds for the approximation error in terms of network size, operator type, and data regularity, while practical performance demonstrates the low generalization and optimization errors associated with this architecture.” [6]

[1] T. Chen and H. Chen - Universal Approximation to Nonlinear Operators by Neural Networks … - 1995
[2] S. Lanthaler et al. - Error estimates for DeepONets- A deep learning framework in infinite dimensions - 2021
[3] A. Yu et al - Arbitrary-depth universal approximation theorems for operator neural networks – 2021
[4] B. Deng et al - Convergence rate of DeepONets for learning operators arising from advection-diffusion equations - 2021
[5] L. Lu et al - A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data – 2021
[6] B. Meuris et al - Machine-learning custom-made basis functions for partial differential equations - 2021

Deep Operator Networks (DeepONets)

"Once DeepONet learns a given operator, it has better generalization to new input functions compared to simpler feedforward or recurrent deep neural network baselines.
...
Once DeepONet is trained, it can be applied to new input functions, thus producing new results substantially faster than numerical solvers.
...
Another benefit of DeepONet is that it can be applied
to simulation data, experimental data or both, and the experimental data may span multiple orders of magnitude in spatio-temporal scales, thus allowing scientists to estimate dynamics better by pooling the existing data.
...
it opens up exciting opportunities, like modelling the dynamics of complex systems where no analytical descriptions exist, for example social dynamics." [1]

[1] I. Higgins - Generalizing universal function approximators – 2021 - Nature Machine Intelligence