Journal of Machine Learning ISSN: 2790-2048(e), 2790-203X(p)

Stochastic Operator Network: A Stochastic Maximum Principle Based Approach to Operator Learning

Ryan Bausback[1], Jingqiao Tang[1], Lu Lu[2], Feng Bao[1], and Phuoc-Toan Huynh* [1]

[1]Department of Mathematics, Florida State University, Florida 32304, Tallahassee, USA.
[2]Department of Statistics and Data Science, Yale University, New Haven, CT 06511, USA.

Abstract. We develop a novel framework for uncertainty quantification in operator learning, namely the stochastic operator network (SON). SON combines the stochastic optimal control concepts of the stochastic neural network (SNN) with the deep operator network. By formulating the branch net as an SDE and backpropagating through the adjoint BSDE, we replace the gradient of the loss function with the gradient of the Hamiltonian from the stochastic maximum principle in the stochastic gradient descent (SGD) update. This allows SON to learn the uncertainty present in operators through its diffusion parameters. We then demonstrate the effectiveness of SON when replicating several noisy operators in 2D and 3D.

Keywords:
Stochastic operator learning,
Deep operator network,
Stochastic maximum principle,
Backward stochastic differential equation,
Uncertainty quantification.

Article Info.:
Volume: 5
Number: 1
Pages: 71 - 96
Date: March/2026
doi.org/10.4208/jml.250709

Article History:
Received: 09/07/2025
Accepted: 16/12/2025

Communicated by:
Zhi-Qin Xu

1 Introduction

Over the past five years, operator learning has emerged as a promising alternative to traditional numerical solvers for a wide range of differential equations [5, 9, 14, 18, 21, 22, 26, 30, 38]. Unlike classical machine learning models that approximate functions at discrete spatial or temporal points, neural operators take input as an entire function and produce a corresponding output function [25]. The two most influential architectures, deep operator network (DeepONet) and Fourier neural operator (FNO), have driven the evolution of deep operator learning. DeepONet, which is based on the universal approximation theorem for operators [9], comprises two parallel subnetworks: the "branch" net, which learns coefficients, and the "trunk" net, which learns a data-driven basis for the output function. FNO, in contrast, encodes and decodes using the Fourier basis (via fast Fourier transform (FFT) and inverse FFT) within successive Fourier layers, relying on a single feed-forward backbone and assuming that input and output domains coincide [22]. The authors in [26] later showed that FNO was simply a special parameterization of DeepONet under the Fourier basis, which means that all variants of the FNO benefits from the same universal approximation feature as the DeepONet [26]. Crucially, neither DeepONet nor FNO relies on a fixed mesh discretization, a key limitation of many classical PDE solvers in solving large scale problems, which paves the way for rapid, mesh-free inference across a broad class of parameterized PDEs.

*Corresponding author. th24ba@fsu.edu.

While DeepONet has been applied to certain stochastic problems, such as stochastic ordinary differential equations (ODEs) and stochastic PDEs in [27], the randomness appears only in the input function, not as intrinsic stochasticity in the operator itself. The network still learns a deterministic map from noisy inputs to outputs, rather than an operator that injects noise at each point along a trajectory. Beyond these stochastic input applications, several advanced operator learning frameworks have emerged. For example, GANO (generative adversarial neural operator) uses an FNO backbone to train a generator-discriminator pair for richer output sampling, and PCAnet replaces the Fourier basis with a data-driven principal component analysis (PCA) basis to reduce dimensionality [5, 30]. A handful of architectures explicitly target uncertainty quantification, for example, the information bottleneck-uncertainty quantification (IB-UQ) replaces the branch network with an encoder-decoder that learns a latent representation of input noise to produce predictive distributions [14]. However, most operator-learning work still focuses on deterministic accuracy, and even UQ-oriented variants have not yet tackled SDE-style operators with pointwise noise along trajectories.

In this work, we propose a novel training strategy for stochastic operators by merging probabilistic learning, namely, stochastic neural networks, with the DeepONet framework. The SNN is an extension of the so-called "neural ODE" network architecture where the evolution of hidden layers in the deep neural network (DNN) is formulated as a discretized ordinary differential equation system [8, 10, 13, 15, 34]. More specifically, an additive Brownian motion noise, which characterizes the randomness caused by the uncertainty of models and noises of data, is added to the ODE system corresponding to the hidden layers to transform the ODE into an stochastic differential equation (SDE) [16, 17, 23, 24, 33]. In the SNN model, the prediction of the network is represented through the drift parameters, and the stochastic diffusion governs the randomness of network output, which serves to quantify the uncertainty of deep learning. Compared to other probabilistic learning methods, such as the Bayesian neural network (BNN) [11,12,19,29,31,35–37], the SNN approach takes less computational cost to evaluate the diffusion coefficient, while it is still able to characterize sufficient probabilistic behavior of the neural network by stacking (controlled) diffusion terms together through the multilayer structure [2]. The main challenge in implementing the SNN is to construct an efficient numerical solver for the backpropagation process, since the standard backpropagation approach used in the neural ODE is deemed computationally inefficient [2, 3]. We address this by adopting the sample-wise backward-SDE solver of [2, 3], which treats backward samples as "pseudo data" and solves only a small, randomly selected subset of the backward SDE per iteration, thus drastically reducing computational burden.

The main contributions of our paper are as follows. We first construct the general framework of our training strategy, namely the stochastic operator network, which effectively combines the methodology of the DeepONet and the training process of the SNN. The stochastic maximum principle (SMP) [8] is adopted to formulate the loss function of the new backpropagation process and transforms the training algorithm of the neural operator into a stochastic optimal control problem. We then consider several numerical experiments to validate our method on a range of stochastic operators. Finally, we also compare the performance our method with the standard DeepONet approach.