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

A Multimodal PDE Foundation Model for Prediction and Scientific Text Descriptions

Elisa Negrini * [1], Yuxuan Liu[1], Liu Yang[2], Stanley J. Osher[1], and Hayden Schaeffer[1]

[1]Department of Mathematics, University of California Los Angeles, Los Angeles, CA 90095, USA.
[2]Department of Mathematics, National University of Singapore, Singapore 119077.

Abstract. Neural networks are one tool for approximating non-linear differential equations used in scientific computing tasks such as surrogate modeling, real-time predictions, and optimal control. PDE foundation models utilize neural networks to train approximations to multiple differential equations simultaneously and are thus a general purpose solver that can be adapted to downstream tasks. Current PDE foundation models focus on either learning general solution operators and/or the governing system of equations, and thus only handle numerical or symbolic modalities. However, real-world applications may require more flexible data modalities, e.g. text analysis or descriptive outputs. To address this gap, we propose a novel multimodal deep learning approach that leverages a transformer-based architecture to approximate solution operators for a wide variety of ODEs and PDEs. Our method integrates numerical inputs, such as equation parameters and initial conditions, with text descriptions of physical processes or system dynamics. This enables our model to handle settings where symbolic representations may be incomplete or unavailable. In addition to providing accurate numerical predictions, our approach generates interpretable scientific text descriptions, offering deeper insights into the underlying dynamics and solution properties. The numerical experiments show that our model provides accurate solutions for in-distribution data (with average relative error less than 3.3%) and out-of-distribution data (average relative error less than 7.8%) together with precise text descriptions (with correct descriptions generated 100% of times). In certain tests, the model is also shown to be capable of extrapolating solutions in time.

Keywords:
Multimodal AI,
Generative AI,
PDE foundation model,
Transformer-based operator learning,
Scientific text generation.

Article Info.:
Volume: 4
Number: 4
Pages: 290 - 317
Date: December/2025
doi.org/10.4208/jml.250214

Article History:
Received: 14/02/2025
Accepted: 03/10/2025

Communicated by:
Lu Lu

1 Introduction

Neural networks have become increasingly important for solving non-linear differential equations, with applications in climate modeling, financial forecasting, biological systems analysis, and structural optimization (see for instance [8, 32, 36]). Their ability to model complex, non-linear relationships allows for efficient and accurate predictions for various scientific computing tasks such as surrogate modeling, real-time predictions, and optimal control. Previous work in deep learning for partial differential equations (PDE) have focused on learning either the solution operator, which maps input functions to their solutions, or the governing system of equations, which describes the constitutive model based on observations of state variables [6, 21, 22, 29, 35, 51]. These approaches, however, tackle one task at a time and are limited to the use of numerical data.

*Corresponding author. enegrini@math.ucla.edu

Building on the observation that families of differential equations frequently share fundamental characteristics, recent work has introduced transformer-based architectures to enable simultaneous encoding of various parametric differential equations [5,18,23,25,38, 46–48]. Although effective, these methods require structured input and output data, with vanilla in-context operator network (ICON) [46] focusing on numerical data and predicting operators and symbolic expressions using multimodal transformers (PROSE) utilizing numerical and symbolic data [25, 38]. In this work, we consider additional modalities as both inputs and outputs to the model. Often, one has access or would like to produce heuristic descriptions of the observed dynamics that are neither in symbolic nor numerical form, but instead come as text descriptions. For example, consider modeling the dynamics of a complex ecological system: the numerical inputs can include measured population levels, while text inputs could describe key processes such as predator-prey interactions or migration patterns. Similarly, in material science, numerical data could represent experimental results, while text inputs provide the governing equation or describe the experimental setup. By combining these modalities, the model may better capture the underlying rules and provide more accurate and contextually informed predictions. The use of mathematical formulae, text descriptions, and numerical values can provide a more robust development toward a PDE foundation model. Note that in fine-tune language models as multi-modal differential equation solvers (ICON-LM) [47] both textual and numerical prompts were provided as inputs; however, the model does not generate text descriptions since the outputs of ICON-LM are the numerical predictions.

Consider the following parametrized differential equation:

$$\begin{cases} \mathcal{F}\big(u(x,t;c)\big) = 0, & (x,t) \in \Omega \times [0,T], \\ \mathcal{B}\big(u(x,t;c)\big) = 0, & (x,t) \in \partial\Omega \times [0,T], \\ u(x,0;c) = \mathcal{G}(x;c), & x \in \Omega, \quad c \sim \mathcal{D}, \end{cases} \tag{1.1}$$

where $\mathcal{F}$ denotes the governing equation, $\mathcal{B}$ denotes boundary conditions. The initial condition $\mathcal{G}$ is a generating function, and c denotes the parameters that determine the initial conditions from distribution $\mathcal{D}$. The objective is to develop a single neural network model capable of approximating numerically and describing in text the solution operators for a range of governing equations $\mathcal{F}$ in (1.1), which can include ordinary differential equations (ODEs) and PDEs. For simplicity in experiments and data design, for the PDE dataset, we restrict our attention to periodic boundary conditions. However, we note that the model would be able to handle different boundary conditions simply by augmenting or extending the input by either providing these conditions as numerical values or analytic formulae. We leave the interesting and potentially more complex case of higher-dimensional PDEs for future work, as we expect this to also require a novel data processing as in [23].

The text input may contain either the equation to solve or a text description of the system. This is important in applications where one may only have a partial model or a description of the underlying process. The numerical data includes the equation's parameters and initial conditions. For instance, suppose we want to solve the heat equation with parameter $c = 0.003$ and initial condition $u(x,0) = u_0(x)$. The input to our model could be: