

Simulating Fluid Flows with Continuous Vision Transformers

¹Paris Perdikaris *; ²Sifan Wang; ¹Shyam Sankaran; ¹Hanwen Wang

¹*University of Pennsylvania*

²*Yale University*

Abstract

Operator learning, which aims to approximate maps between infinite-dimensional function spaces, is an important area in scientific machine learning with applications across various physical domains. Here we introduce the Continuous Vision Transformer (CViT), a novel neural operator architecture that leverages advances in computer vision to address challenges in learning complex physical systems. CViT combines a vision transformer encoder, a novel grid-based coordinate embedding, and a query-wise cross-attention mechanism to effectively capture multi-scale dependencies. This design allows for flexible output representations and consistent evaluation at arbitrary resolutions. We demonstrate CViT's effectiveness across a diverse range of partial differential equation (PDE) systems, including fluid dynamics, climate modeling, and reaction-diffusion processes. Our comprehensive experiments show that CViT achieves state-of-the-art performance on multiple benchmarks, often surpassing larger foundation models, even without extensive pretraining and roll-out fine-tuning. Taken together, CViT exhibits robust handling of discontinuous solutions, multi-scale features, and intricate spatio-temporal dynamics. Our contributions can be viewed as a significant step towards adapting advanced computer vision architectures for building more flexible and accurate machine learning models in the physical sciences.

Keywords:

deep learning; neural operators; fluid dynamics