Less Data, More Physics?

A Data-Efficiency Study for PDE Surrogate Modelling

To what extent can physics-informed neural operator training improve data efficiency compared with purely data-driven neural operator training for PDE surrogate modelling?

TU Delft

Samuel Campos Vilar - scamposvilar@tudelft.nl

Supervisors: Jing Sun & Tieixing Wang

1. MOTIVATION

Neural operators can learn fast Partial Differential Equations (PDE) surrogate models, but standard Fourier Neural Operator (FNO) training needs labelled solver data.

Since generating those labels can be expensive, this work asks whether adding the known PDE as a training constraint can reduce the amount of labelled data needed.

2. NEURAL OPERATORS

Neural operators are reusable PDE surrogates that:

- Map an **input field** to a **solution field**
- **Generalise** to new inputs without retraining

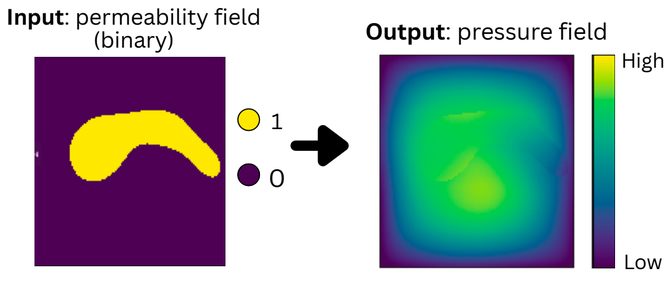
FNO learns from **labelled data only**
PINO also penalises **PDE residuals**

3. PROBLEM SETUP

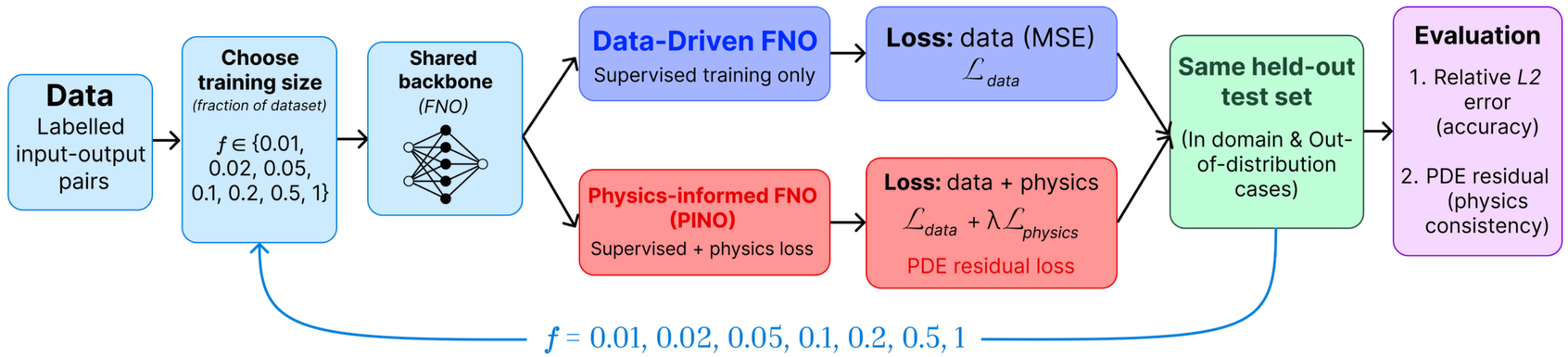
We evaluate on two benchmarks from PDEBench:

- **Darcy flow** (2D steady elliptic)
- **Burgers' equation** (1D nonlinear time-dependent)

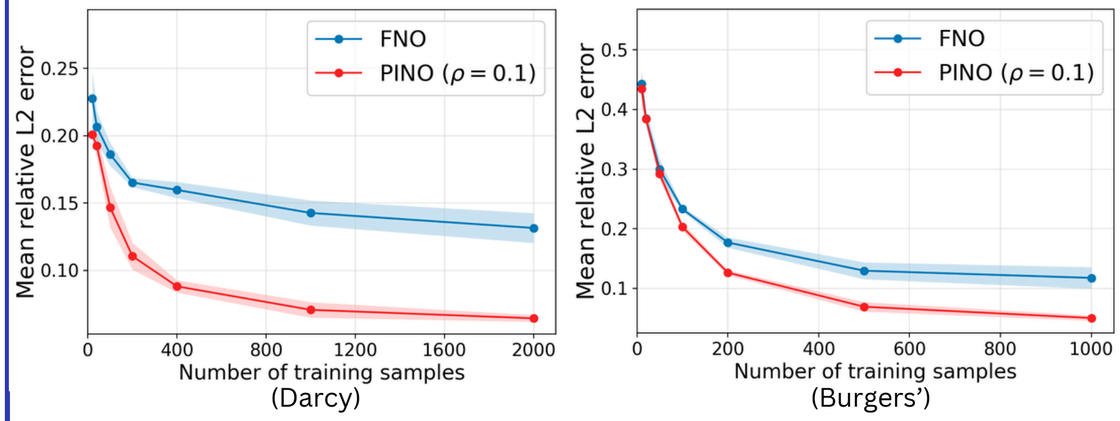
Example shown: Darcy Flow



4. METHODOLOGY



5. RESULTS



Across both Darcy flow and Burgers' equation, PINO is consistently equal or below FNO, meaning that adding the PDE residual loss improves in-distribution accuracy.

The improvement is also data-efficient. On Darcy flow, PINO with only 10% of the labels already beats the FNO trained with all labels. On Burgers' equation, PINO beats the full-data FNO using 50% of the labels.

PDE residual MSE on held-out predictions

Benchmark	FNO	PINO	Effect
Darcy, full-data	2841	20	142x lower
Burgers, full-data	4.8	1.2	4x lower

6. INTERPRETATION

PINO was not strongest in the tiniest-data regime. The gain increased once enough labels taught the solution scale and structure.

Physics-only training failed, and too much PDE weight hurt accuracy.

Darcy OOD remained poor because the shifted parameter was not given as an input.

7. CONCLUSION

PINO improves data-efficient PDE surrogate modelling when the physics loss is properly balanced with labelled data.

The PDE residual acts as a useful regulariser, improving physical consistency as well as accuracy, but it does not replace labelled simulations or guarantee extrapolation.

Future work: test adaptive physics-loss weights, include shifted PDE parameters as model inputs, and extend the comparison to full-resolution datasets and harder PDE families.