

DPOT: Auto-Regressive Denoising Operator Transformer for Large-Scale PDE Pre-Training

Zhongkai Hao, advisor: Hang Su, Jun Zhu, Anima Anandkumar

Dept. of Computer Science, Tsinghua University; Contact: hjz21@mails.tsinghua.edu.cn

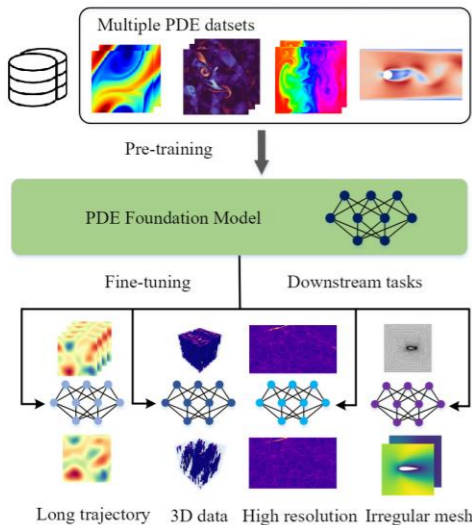


Introduction

Learning operators of PDEs is an essential problem with applications in numerous domains of science and engineering. However, collecting PDE data using simulation or experiments are costly (thousands of paired data or trajectories).

Can we use pre-training to build a foundational PDE model for downstream tasks?

Yes! we propose **DPOT**, an auto-regressive Denoising Pre-training Operator Transformer for diverse PDE data.



Contributions

The first unified transformer model up to **1B** pre-trained on 10+ different PDE datasets.

Methodology

- Objective for pre-training: predicting next step using noisy data for robust rollouts

$$\min_w \mathcal{L} = \mathbb{E}_{u \sim p(\mathcal{D})} \sum_{1 \leq t \leq T} \|\mathcal{G}_w(u^{<t} + \varepsilon) - u^t\|_2^2$$

- Learning the integral transform of PDE solution map using Fourier Mixer

$$(\mathcal{K}_\phi z^l)(x) = \int_\Omega \kappa(x, y; \phi) z^l(y) dy$$

Approximate it in frequency space using FFT/IFFT + MLP

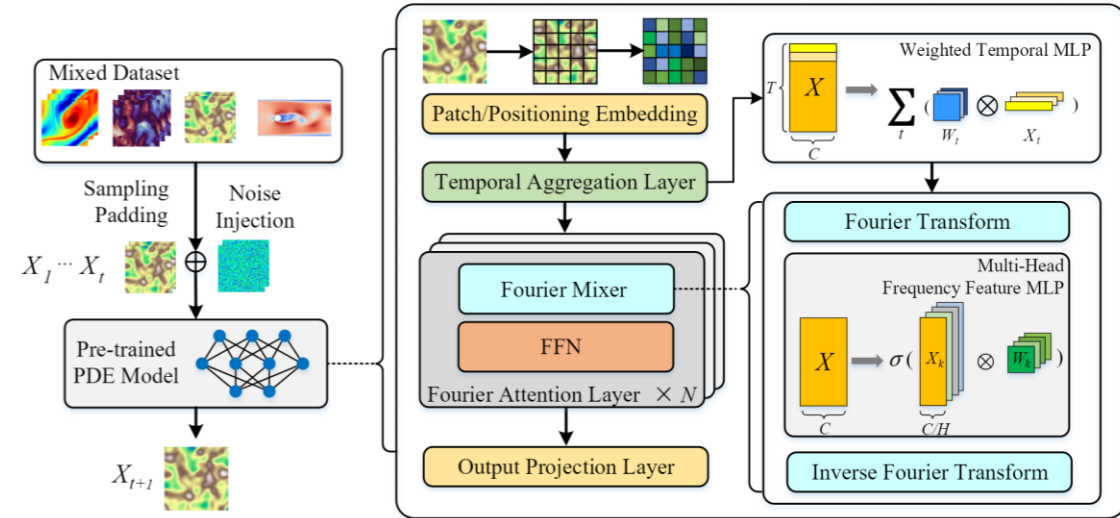
$$z^{l+1}(x) = \mathcal{F}^{-1}[W_2 \cdot \sigma(W_1 \cdot \mathcal{F}[z^l] + b_1) + b_2](x)$$

- Temporal aggregation to extract temporal features

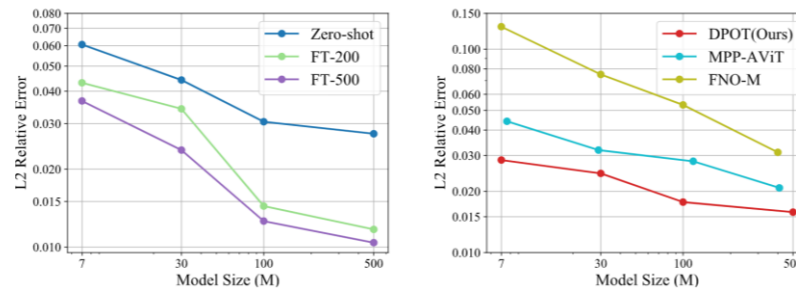
$$z_{\text{agg}} = \sum_t W_t \cdot z_p^t e^{-i\gamma t}$$

- Balanced data sampling by choosing dataset with equal probability

Illustration of DPOT



Scaling Law Results of Neural PDE Solver



Model Parameters

Size	Attention dim	MLP dim	Layers	Heads	Model size
Tiny	512	512	4	4	7M
Small	1024	1024	6	8	30M
Medium	1024	4096	12	8	122M
Large	1536	6144	24	16	509M
Huge	2048	8092	27	8	1.03B

Downstream Tasks

L2RE/L1-Medium*	Turbulence PDB-nurb	3D dataset PDB	Steady* CNO
(Geo)-FNO	0.193	0.410	0.0357
MPP-FT	0.152	-	-
DPOT-Vanilla	0.167	0.262	0.0331
DPOT-FT	0.135	0.226	0.0230

Main Results

L2RE Subset	Params	FNO- ν			PDEBench CNS-(η, ζ), DR, SWE				PDEArena		CFDBench				
		1e-5	1e-4	1e-3	1.0,1	1.0,1	MI	M0,1	DR	SWE		NS	NS-cond		
Small Model															
FNO	0.5M	0.156	0.0834	0.0128	0.098	0.096	0.097	0.360	0.170	0.265	0.12	0.0044	0.0912	0.319	0.00761
UNet	25M	0.198	0.119	0.0245	0.334	0.291	0.313	0.569	0.357	0.463	0.0971	0.0521	0.102	0.337	0.209
FFNO	1.3M	0.121	0.0503	0.0099	0.0212	0.052	0.0366	0.162	0.0452	0.104	0.0571	0.0116	0.0839	0.602	0.00714
GK-T	1.6M	0.134	0.0792	0.0098	0.0341	0.0377	0.0359	0.0274	0.0366	0.0320	0.0359	0.00692	0.0952	0.423	0.0105
GNOT	1.8M	0.157	0.0443	0.0125	0.0325	0.0420	0.0373	0.0228	0.0341	0.0285	0.0311	0.00678	0.172	0.325	0.00877
Ofortner	1.9M	0.1705	0.0645	0.0104	0.0417	0.0625	0.0521	0.0254	0.0205	0.0229	0.0192	0.00717	0.135	0.332	0.0102
FNO-am	7M	0.116	0.0922	0.0156	0.151	0.108	0.130	0.230	0.076	0.153	0.0321	0.00912	0.210	0.384	0.0274
MPP-Ti	7M	-	-	-	-	-	0.0442	-	-	-	0.153	0.0312	0.0066	-	-
MPP-S	30M	-	-	-	-	-	0.0319	-	-	-	0.1213	0.0112	0.0024	-	-
Ours-Ti	7M	0.0976	0.0606	0.00954	0.0173	0.0397	0.0285	0.0132	0.0220	0.0176	0.0321	0.00560	0.125	0.384	0.00952
Ours-S	30M	0.0553	0.0442	0.0131	0.0153	0.0337	0.0245	0.0119	0.0187	0.0153	0.0379	0.00657	0.0991	0.316	0.00696
Pre-trained															
MPP-L	400M	-	-	-	-	-	0.0208	-	-	-	0.0147	0.0098	0.00220	-	-
Ours-S	30M	0.0553	0.0442	0.0131	0.0153	0.0337	0.0245	0.0118	0.0188	0.0153	0.0379	0.00657	0.0999	0.316	0.00696
Ours-M	122M	0.0409	0.0285	0.00474	0.0116	0.0238	0.0177	0.00866	0.0129	0.0108	0.0292	0.00290	0.0812	0.276	0.00752
Ours-L	500M	0.0550	0.0274	0.00528	0.0100	0.0216	0.0158	0.00872	0.0115	0.0101	0.0232	0.00233	0.0798	0.240	0.00650
Ours-H	1.03B	0.0174	0.0131	0.00229	0.00961	0.0180	0.0138	0.00847	0.0105	0.00948	0.0191	0.00199	0.0379	0.213	0.00749
DPOT-FT															
T-200	7M	0.0511	0.0431	0.00729	0.0136	0.0238	0.0187	0.0168	0.0145	0.0157	0.0194	0.00280	0.103	0.313	0.00537
S-200	30M	0.0449	0.0342	0.00680	0.0152	0.0211	0.0182	0.0150	0.0151	0.0151	0.0171	0.00224	0.0892	0.290	0.00442
M-200	100M	0.0255	0.0144	0.00427	0.0123	0.0179	0.0151	0.0182	0.0117	0.0149	0.0142	0.00218	0.0329	0.191	0.00452
L-200	500M	0.0235	0.0117	0.00383	0.0114	0.0153	0.0133	0.0171	0.0108	0.0140	0.0158	0.00197	0.0307	0.182	0.00480
T-500	7M	0.0520	0.0367	0.00580	0.0112	0.0195	0.0153	0.0174	0.0138	0.0156	0.0148	0.00244	0.0910	0.280	0.00391
S-500	30M	0.0322	0.0237	0.00437	0.0129	0.0167	0.0148	0.0152	0.0126	0.0139	0.0129	0.00235	0.0867	0.268	0.00382
M-500	100M	0.0229	0.0126	0.00335	0.00998	0.0146	0.0123	0.0161	0.00947	0.0128	0.0103	0.00227	0.0294	0.172	0.00373
L-500	500M	0.0213	0.0104	0.00323	0.0108	0.0131	0.0119	0.0160	0.00905	0.0125	0.00739	0.00170	0.0278	0.170	0.00322