A EXPERIMENTAL DETAILS

Table 4: Architecture optimizer settings on PDE tasks with CNN and FNO initializations. Note that the learning rate is updated using the same schedule as the backbone.

task	XD init	optimizer	initial lr	warmup
1d Burgers' equation	CNN	Adam	1E-3	0
1d Burgers' equation	FNO	Momentum(0.5)	1E-4	250
2d Darcy Flow	CNN	Momentum(0.5)	1E-1	0
2d Darcy Flow	FNO	Momentum(0.5)	1E-1	0
2d Navier Stokes ($\nu = 10^{-4}, T = 30$)	CNN	Momentum(0.5)	5E-3	0
2d Navier Stokes ($\nu = 10^{-5,} T = 20$)	CNN	Momentum(0.5)	1E-3	0

For our experiments, we use the code and setup from Li et al. (2021b) provided here: `https://github.com/zongyi-li/fourier_neural_operator`. We use the same training routine and settings as the backbone architecture for each task and only tune the architecture optimizer. We consider the following hyperparameters for the architecture optimizer: Adam vs. SGD (with or without momentum), initial learning rate, and number of warmup epochs. Our CNN backbone is analogous to the FNO architecture used for each problem. In particular, the CNN backbone architecture used for each task is simply the FNO architecture where FNO layers of dimension N with m modes are replaced by N-dimensional convolutional layers with filters of size $(m+1)^N$ and circular padding to match the dimensionality of FNO.

Table 5: Test relative errors on the 1d Burgers' equation. We were not able to match the FNO-1d results reported by Li et al. (2021b) using their published codebase, however, our proposed XD operations outperform our reproduction of their results at every resolution. Furthermore, we outperform their reported test relative errors on every resolution except $s = 4096$, where we roughly match their performance.

Method (source)	$s = 256$	512	1024	2048	4096	8192
NN (Li et al., 2021b)	0.4714	0.4561	0.4803	0.4645	0.4779	0.4452
GCN (Li et al., 2021b)	0.3999	0.4138	0.4176	0.4157	0.4191	0.4198
FCN (Li et al., 2021b)	0.0958	0.1407	0.1877	0.2313	0.2855	0.3238
PCANN (Li et al., 2021b)	0.0398	0.0395	0.0391	0.0383	0.0392	0.0393
GNO (Li et al., 2021b)	0.0555	0.0594	0.0651	0.0663	0.0666	0.0699
LNO (Li et al., 2021b)	0.0212	0.0221	0.0217	0.0219	0.0200	0.0189
MGNO (Li et al., 2021b)	0.0243	0.0355	0.0374	0.0360	0.0364	0.0364
FNO-1d (Li et al., 2021b)	0.0149	0.0158	0.0160	0.0146	**0.0142**	0.0139
CNN backbone (ours)	0.0518	0.1220	0.1830	0.2280	0.2730	0.2970
FNO-1d (reproduced)	0.0181	0.0191	0.0188	0.0184	0.0183	0.0183
CNN backbone XD (ours)	**0.0141**	**0.0079**	**0.0154**	**0.0099**	0.0145	**0.0123**
FNO-1d XD (ours)	0.0153	0.0154	0.0154	0.0167	0.0160	0.0155

Table 6: Test relative errors on 2d Darcy Flow. Our reproduction of the FNO-2d results outperform those reported by Li et al. (2021b). Nonetheless, our proposed XD operations outperform both our reproduction and the reported results of Li et al. (2021b) at every resolution.

Method (source)	$s = 85$	$s = 106$	$s = 141$	$s = 211$	$s = 421$
NN (Li et al., 2021b)	0.1716	-	0.1716	0.1716	0.1716
GCN (Li et al., 2021b)	0.0253	-	0.0493	0.0727	0.1097
FCN (Li et al., 2021b)	0.0299	-	0.0298	0.0298	0.0299
PCANN (Li et al., 2021b)	0.0244	-	0.0251	0.0255	0.0259
GNO (Li et al., 2021b)	0.0346	-	0.0332	0.0342	0.0369
LNO (Li et al., 2021b)	0.0520	-	0.0461	0.0445	-
MGNO (Li et al., 2021b)	0.0416	-	0.0428	0.0428	0.0420
FNO-2d (Li et al., 2021b)	0.0108	-	0.0109	0.0109	0.0098
CNN backbone (ours)	0.0404	0.0495	0.0613	0.0813	0.1150
FNO-2d (reproduced)	0.0096	0.0092	0.0091	0.0091	0.0091
CNN backbone XD (ours)	**0.0065**	**0.0065**	**0.0065**	**0.0071**	**0.0066**
FNO-2d XD (ours)	0.0082	0.0079	0.0077	0.0076	-