LOGLO-FNO

Efficient Learning of Local and Global Features in Fourier Neural Operators

Marimuthu Kalimuthu*, David Holzmüller, Mathias Niepert

*(firstname.lastname@ki.uni-stuttgart.de)

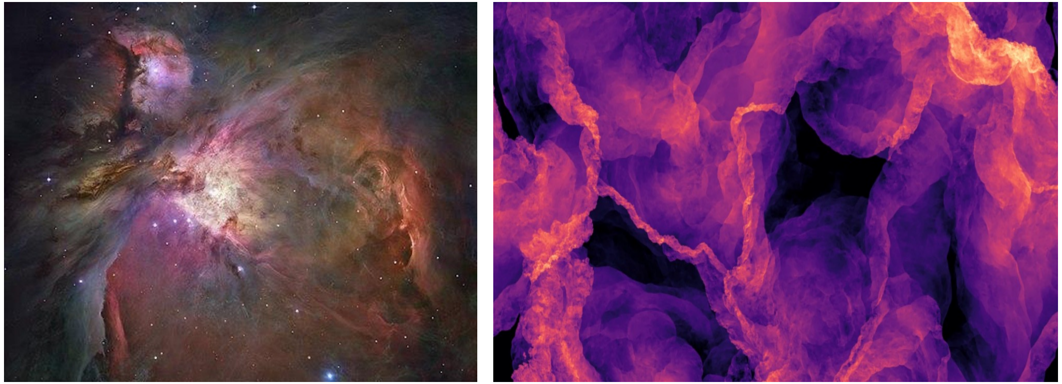
Presenter: Daniel Musekamp

University of Stuttgart, Stutgart Center for Simulation Science (SimTech),
International Max Planck Research School for Intelligent Systems (IMPRS-IS),
Inria Paris, École Normale Supérieure PSL University, NEC Labs Europe

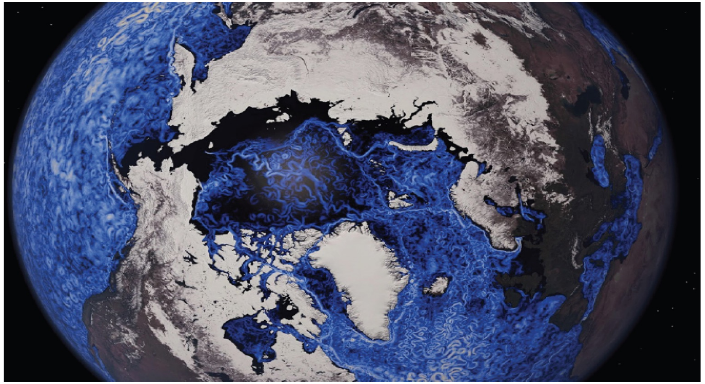
Machine Learning Multiscale Processes

University of Stuttgart Germany

Real-World Physical Phenomena

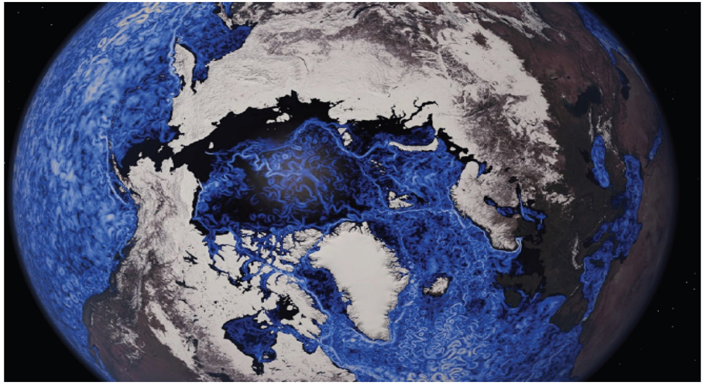


Astrophysical Processes – Star and Cold Gas Formation



Ocean and Atmospheric Circulation

Real-World Physical Phenomena



Astrophysical Processes – Star and Cold Gas Formation

Ocean and Atmospheric Circulation

Governed by Partial Differential Equations such as Navier-Stokes & Shallow-Water equations.

Partial Differential Equations (PDEs)

Fluid Mechanics

*Turbulent **R**adiative **L**ayer*

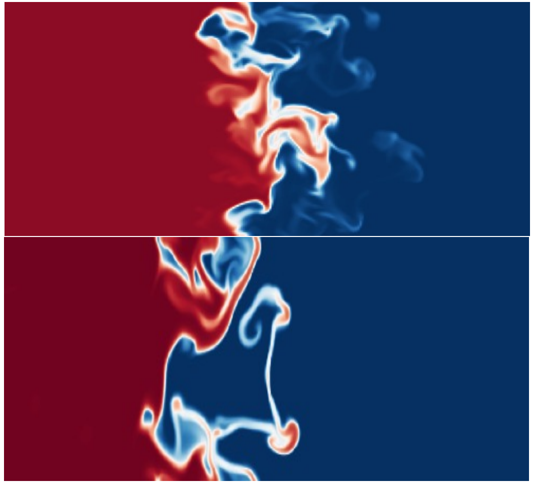
$$\frac{\partial \rho}{\partial t} + \nabla \cdot \big(\overbrace{\rho}^{\text{mass density}} \, \underbrace{\mathbf{v}}_{\text{velocity field}} \big) = 0$$

$$\frac{\partial \rho \mathbf{v}}{\partial t} + \nabla \cdot \big(\overbrace{\rho \mathbf{v} \otimes \mathbf{v}}^{\text{momentum flux tensor}} + \underbrace{P}_{\text{pressure tensor}} \big) = 0$$

$$\frac{\partial E}{\partial t} + \nabla \cdot \big((E + P)\mathbf{v} \big) = -\frac{E}{\underbrace{t_{\text{cool}}}_{\text{cooling timescale}}}$$

$$E = P/(\gamma - 1) \quad \text{where} \quad \gamma = \frac{5}{3}$$

adiabatic index ($\frac{5}{3}$ for monatomic gas)



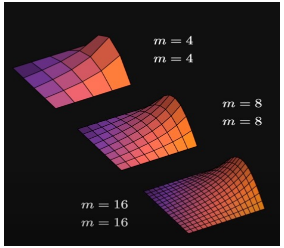
TheWell Ohana et al., (2024)

considered on a given,

- domain **D**
- time interval **[0, T)**

Solving Partial Differential Equations

(using neural networks)



One popular approach involves

(i) Discretizing the given domain **D** and time interval [**0**, **T**)

(ii) Obtaining trajectories using classical numerical solvers

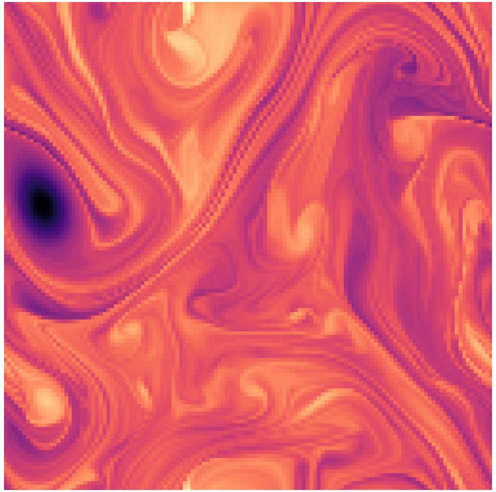
(iii) Training neural networks to predict future states given past state(s)

$$\theta^* = \arg\min_{\theta} \sum_{n=1}^{N} \sum_{t=1}^{T-1} \mathcal{C}(\mathcal{N}_\theta(u^t), u^{t+1})$$

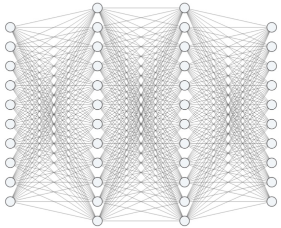
Dynamics Modeling ✅

Forward Problem: Dynamics Modeling

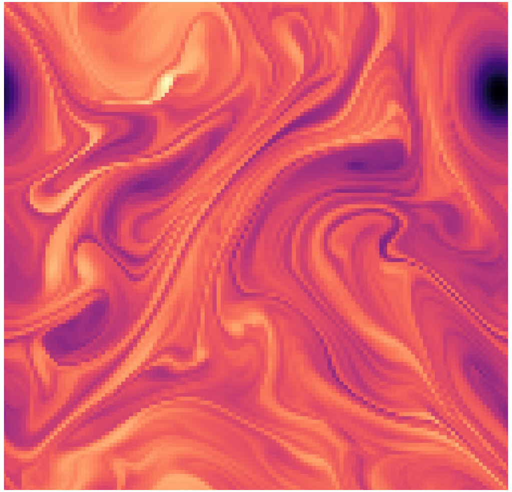
u(., t)



$\mathcal{N}_\theta$



u(., t+Δt)

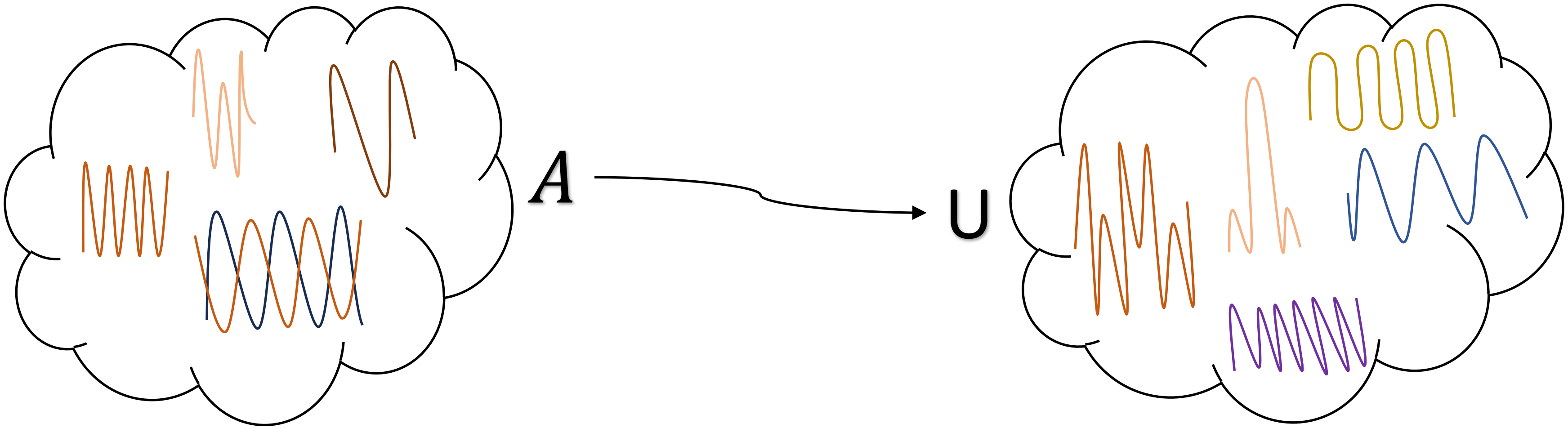


$$\frac{\partial u}{\partial t} + u \cdot \nabla u - \frac{1}{Re}\Delta u = -\nabla p + \sin(ny)\hat{x}, \qquad \nabla \cdot u = 0,$$

Dynamics Modeling

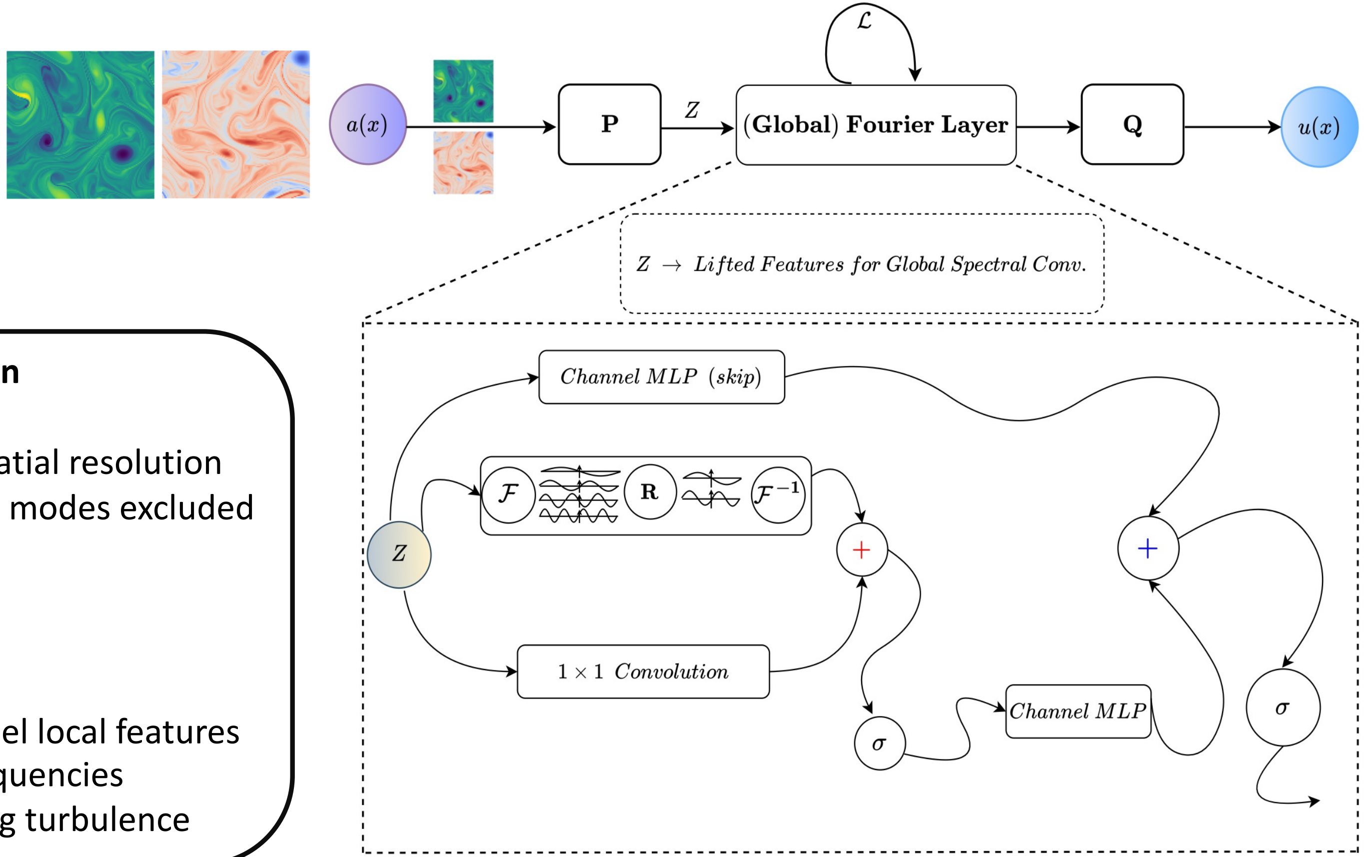
Solving PDEs: Neural Operators

$\mathcal{G}^{\dagger} : \mathcal{A} \to \mathcal{U}$ => (Non-linear map between function spaces)



$\mathcal{G}_{\theta} : \mathcal{A} \to \mathcal{U}, \quad \theta \in \mathbb{R}^p$ => Learn an approximation so that $\mathcal{G}_{\theta^{\dagger}} \approx \mathcal{G}^{\dagger}$

FNO: Fourier Neural Operator



Global Spectral Convolution

- Operates on the full spatial resolution
- High-frequency Fourier modes excluded

Limitations

- Cannot effectively model local features
- Cannot model high frequencies
- Struggles with modeling turbulence

Research Question

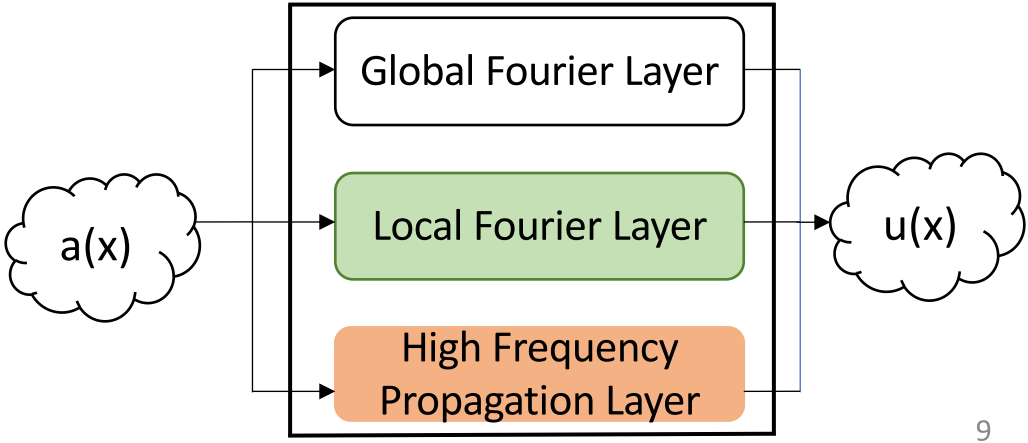
How can we model local features and high-frequencies *without* major changes to FNO architecture?
(i.e., a plug-and-play module, parameter-efficient)

Research Question

How can we model local features and high-frequencies *without* major changes to FNO architecture?
(i.e., a plug-and-play module, parameter-efficient)

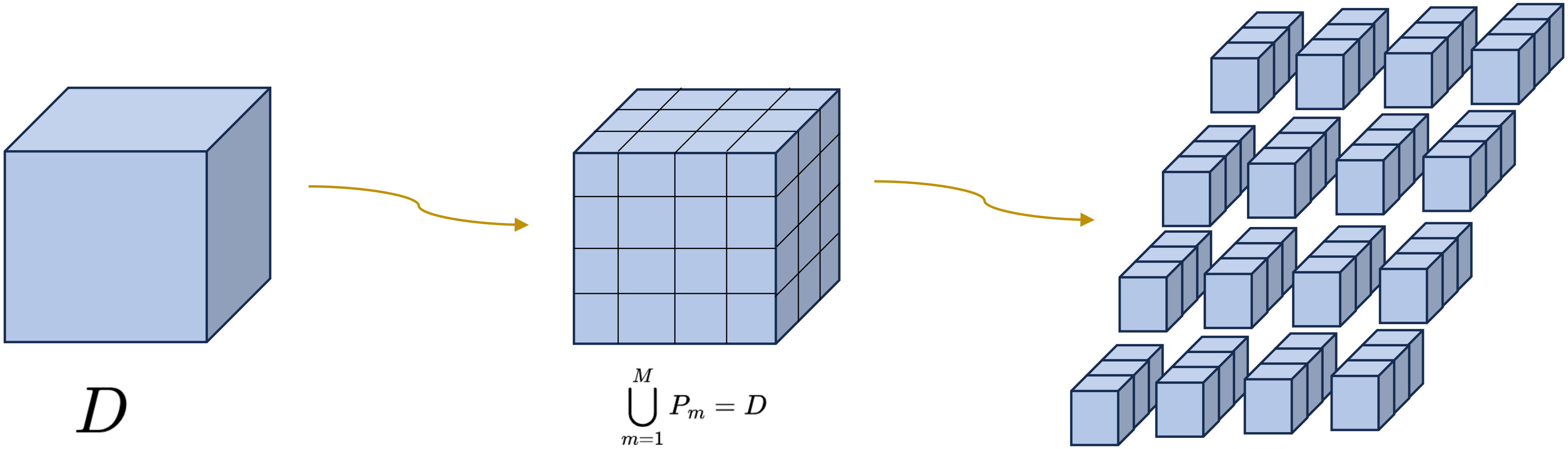
Ideas:
Introduce additional parallel branches to

- ❖ Achieve local convolution
- ❖ Extract & propagate high-frequencies



IDEA (1): Enhance FNO with Local Spectral Convolution

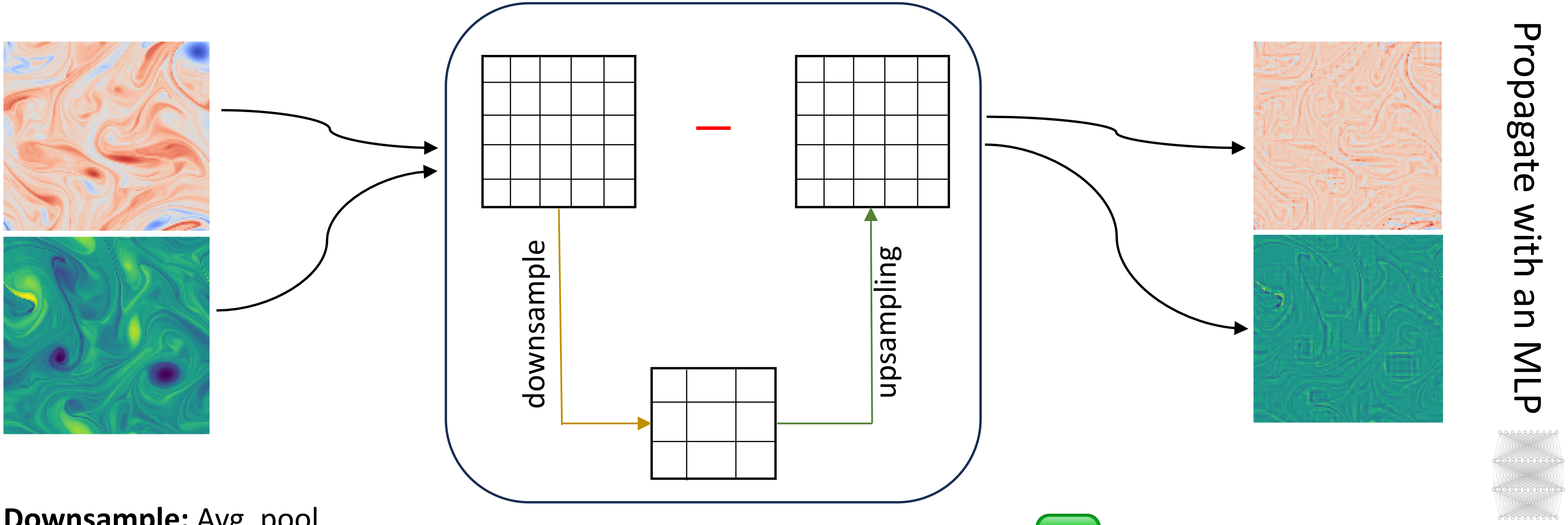
Step 1: Decompose the domain into subdomains (e.g., patches: 8x8, 16x16)
Step 2: Perform spectral convolution on these patches



Local convolution ✅

IDEA (2): Enhance FNO with High Frequency Propagation

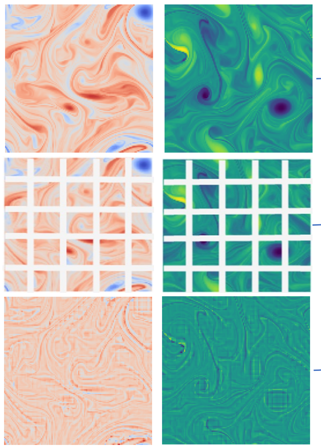
Step 1: Extract high-frequency features
Step 2: Propagate them in a parallel branch

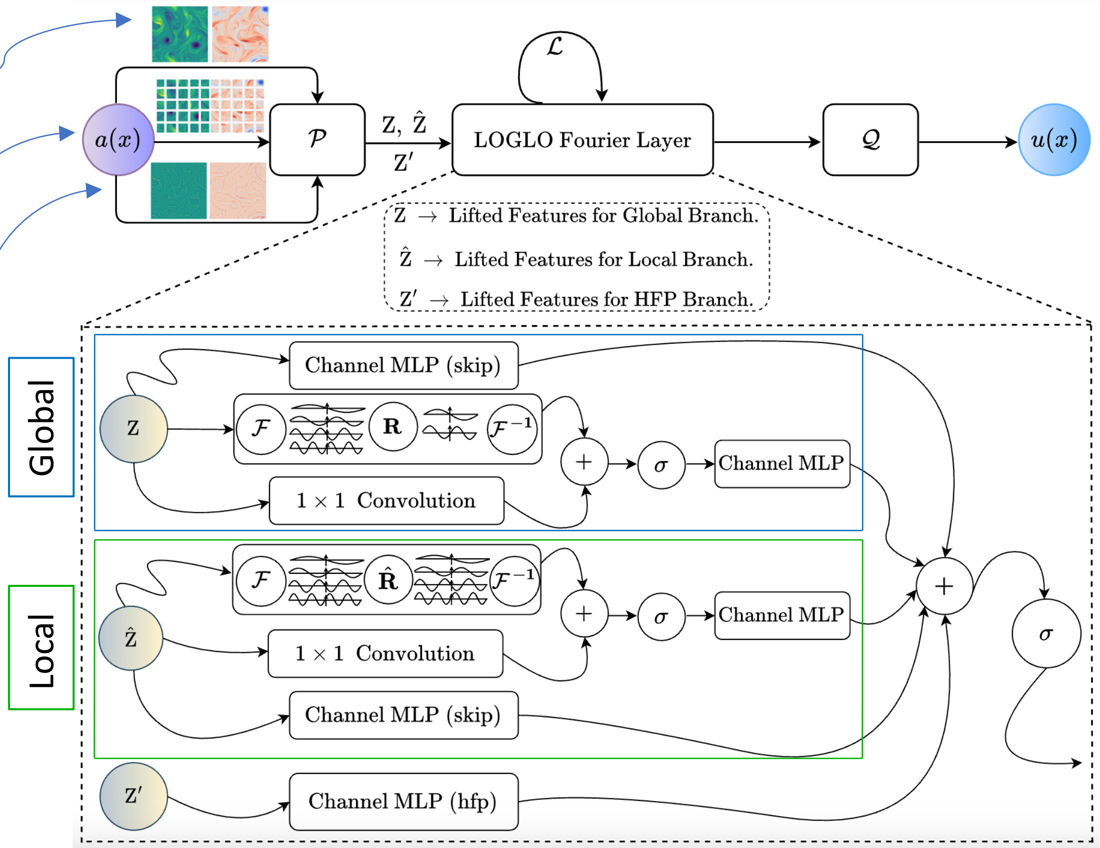


- ❑ **Downsample:** Avg. pool
- ❑ **Upsample:** Interpolation

High-frequency features modeled ✅

Proposed: LOGLO-FNO Architecture





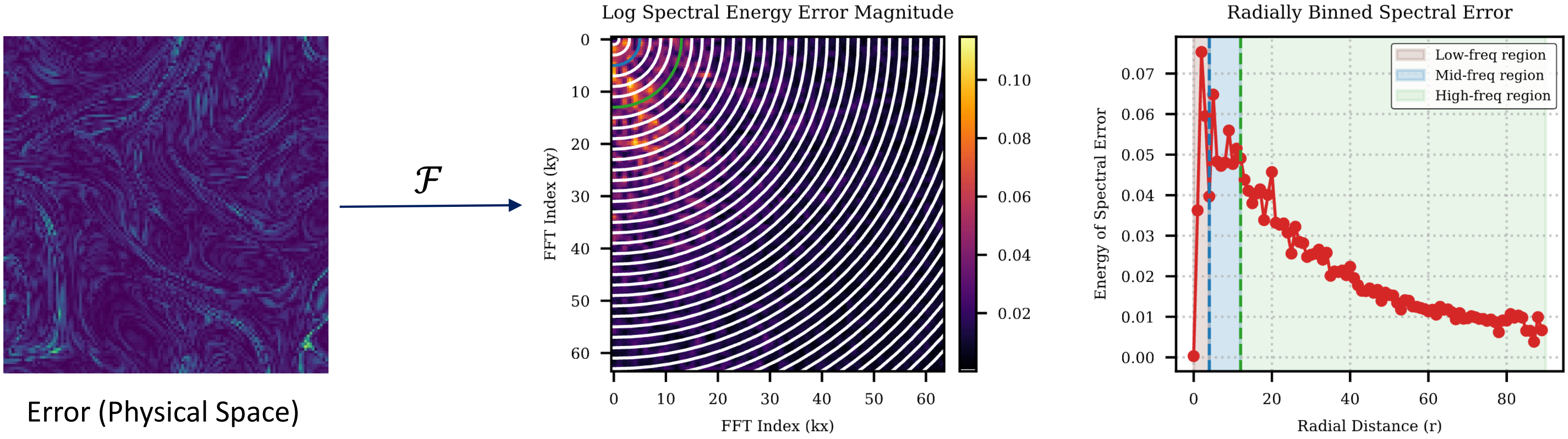
1) Local Spectral Convolution

- Domain decomposition with patching
- Spectral convolution on patches
- Retains all Fourier modes

2) High-Frequency Propagation

- High-pass filter – e.g., Average Pooling

IDEA (3): Spectral Loss with Radially Binned Errors



Combined Objective: $\theta^* = \arg\min_\theta \sum_{n=1}^{N} \sum_{t=1}^{T-1} \mathcal{C}(\mathcal{N}_\theta(u^t), u^{t+1}), \qquad \mathcal{C} = \mathcal{C}_{\text{MSE}} + \lambda \cdot \mathcal{C}_{\text{freq}}, \quad 0 \le \lambda \le 1$

PDEs: 2D and 3D Time-dependent Problems

Kolmogorov flow 2D:

$$\frac{\partial u}{\partial t} + u \cdot \nabla u - \frac{1}{Re}\Delta u = -\nabla p + \sin(ny)\hat{x}, \qquad \nabla \cdot u = 0, \qquad \text{on } [0, 2\pi]^2 \times (0, \infty)$$

Turbulent Radiative Layer 3D:

TheWell

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}) = 0$$

$$\frac{\partial \rho \boldsymbol{v}}{\partial t} + \nabla \cdot (\rho \boldsymbol{v}\boldsymbol{v} + P) = 0$$

$$\frac{\partial E}{\partial t} + \nabla \cdot ((E + P)\boldsymbol{v}) = -\frac{E}{t_{\text{cool}}}$$

on x, y ∈ [-0.5, 0.5], z ∈ [-1, 1]

t ∈ [0, 266.172178]

Diffusion-Reaction 2D:
(PDEBench)

$$\frac{\partial u}{\partial t} = D_u \partial_{xx} u + D_u \partial_{yy} u + R_u(u, v), \quad \frac{\partial v}{\partial t} = D_v \partial_{xx} v + D_v \partial_{yy} v + R_v(u, v), \quad (-1, 1)^2 \times (0, 5]$$

Experiments and Results

2D INS: Kolmogorov flow*

Model	RMSE (↓)	nRMSE	bRMSE	cRMSE	fRMSE(L)	fRMSE(M)	fRMSE(H)	MaxError (↓)	MELR (↓)	WLR (↓)
1-step Evaluation										
U-Net	$7.17 \cdot 10^{-1}$	$1.3 \cdot 10^{-1}$	$1.47 \cdot 10^{0}$	$1.74 \cdot 10^{-2}$	$2.24 \cdot 10^{-2}$	$3.57 \cdot 10^{-2}$	$4.39 \cdot 10^{-2}$	$2.01 \cdot 10^{1}$	$1.64 \cdot 10^{-1}$	$1.48 \cdot 10^{-2}$
FNO	$8.08 \cdot 10^{-1}$	$1.47 \cdot 10^{-1}$	$7.94 \cdot 10^{-1}$	$1.1 \cdot 10^{-2}$	$1.36 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	$4.7 \cdot 10^{-2}$	$1.46 \cdot 10^{1}$	$5.2 \cdot 10^{-1}$	$2.83 \cdot 10^{-2}$
F-FNO	$7.53 \cdot 10^{-1}$	$1.37 \cdot 10^{-1}$	$7.41 \cdot 10^{-1}$	$1.5 \cdot 10^{-2}$	$1.49 \cdot 10^{-2}$	$2.15 \cdot 10^{-2}$	$4.36 \cdot 10^{-2}$	$1.42 \cdot 10^{1}$	$4.74 \cdot 10^{-1}$	$2.28 \cdot 10^{-2}$
LSM	$7.49 \cdot 10^{-1}$	$1.36 \cdot 10^{-1}$	$1.47 \cdot 10^{0}$	$1.36 \cdot 10^{-2}$	$2.59 \cdot 10^{-2}$	$4.42 \cdot 10^{-2}$	$4.64 \cdot 10^{-2}$	$2.05 \cdot 10^{1}$	$1.43 \cdot 10^{-1}$	$1.68 \cdot 10^{-2}$
U-FNO	$6.13 \cdot 10^{-1}$	$1.12 \cdot 10^{-1}$	$1.0 \cdot 10^{0}$	$\mathbf{7.09} \cdot 10^{-3}$	$1.27 \cdot 10^{-2}$	$2.22 \cdot 10^{-2}$	$3.71 \cdot 10^{-2}$	$1.64 \cdot 10^{1}$	$1.38 \cdot 10^{-1}$	$1.09 \cdot 10^{-2}$
NO-LIDK*	$7.25 \cdot 10^{-1}$	$1.33 \cdot 10^{-1}$	$1.12 \cdot 10^{0}$	$9.05 \cdot 10^{-3}$	$1.45 \cdot 10^{-2}$	$2.69 \cdot 10^{-2}$	$4.55 \cdot 10^{-2}$	$1.65 \cdot 10^{1}$	$1.85 \cdot 10^{-1}$	$1.54 \cdot 10^{-2}$
NO-LIDK◇	$6.13 \cdot 10^{-1}$	$1.11 \cdot 10^{-1}$	$5.95 \cdot 10^{-1}$	$1.46 \cdot 10^{-2}$	$1.65 \cdot 10^{-2}$	$2.29 \cdot 10^{-2}$	$3.91 \cdot 10^{-2}$	$1.5 \cdot 10^{1}$	$9.57 \cdot 10^{-2}$	$1.1 \cdot 10^{-2}$
NO-LIDK†	$5.86 \cdot 10^{-1}$	$1.07 \cdot 10^{-1}$	$\mathbf{5.64} \cdot 10^{-1}$	$1.05 \cdot 10^{-2}$	$1.44 \cdot 10^{-2}$	$2.46 \cdot 10^{-2}$	$3.82 \cdot 10^{-2}$	$1.47 \cdot 10^{1}$	$\mathbf{7.11} \cdot 10^{-2}$	$\mathbf{1.0} \cdot 10^{-2}$
LoGLo-FNO	$5.89 \cdot 10^{-1}$	$\mathbf{1.07} \cdot 10^{-1}$	$6.74 \cdot 10^{-1}$	$7.23 \cdot 10^{-3}$	$\mathbf{1.21} \cdot 10^{-2}$	$\mathbf{1.81} \cdot 10^{-2}$	$\mathbf{3.54} \cdot 10^{-2}$	$\mathbf{1.33} \cdot 10^{1}$	$1.29 \cdot 10^{-1}$	$1.06 \cdot 10^{-2}$
Rel. % Diff	-27.1 %	-27.21 %	-15.12 %	-34.31 %	-11.22 %	-11.88 %	-24.64 %	-8.99 %	-75.12 %	-62.63 %
5-step Autoregressive Evaluation										
U-Net	$1.51 \cdot 10^{0}$	$2.65 \cdot 10^{-1}$	$2.31 \cdot 10^{0}$	$5.39 \cdot 10^{-2}$	$6.53 \cdot 10^{-2}$	$1.04 \cdot 10^{-1}$	$9.38 \cdot 10^{-2}$	$1.81 \cdot 10^{1}$	$2.13 \cdot 10^{-1}$	$3.52 \cdot 10^{-2}$
FNO	$1.33 \cdot 10^{0}$	$2.35 \cdot 10^{-1}$	$1.34 \cdot 10^{0}$	$1.46 \cdot 10^{-2}$	$3.37 \cdot 10^{-2}$	$5.80 \cdot 10^{-2}$	$8.37 \cdot 10^{-2}$	$1.60 \cdot 10^{1}$	$6.18 \cdot 10^{-1}$	$4.93 \cdot 10^{-2}$
F-FNO	$1.29 \cdot 10^{0}$	$2.28 \cdot 10^{-1}$	$1.27 \cdot 10^{0}$	$2.28 \cdot 10^{-2}$	$3.60 \cdot 10^{-2}$	$5.52 \cdot 10^{-2}$	$8.12 \cdot 10^{-2}$	$1.50 \cdot 10^{1}$	$5.37 \cdot 10^{-1}$	$3.99 \cdot 10^{-2}$
LSM	$1.81 \cdot 10^{0}$	$3.18 \cdot 10^{-1}$	$2.76 \cdot 10^{0}$	$3.87 \cdot 10^{-2}$	$7.6 \cdot 10^{-2}$	$1.44 \cdot 10^{-1}$	$1.12 \cdot 10^{-1}$	$2.08 \cdot 10^{1}$	$2.04 \cdot 10^{-1}$	$4.39 \cdot 10^{-2}$
U-FNO	$1.15 \cdot 10^{0}$	$2.03 \cdot 10^{-1}$	$1.45 \cdot 10^{0}$	$\mathbf{6.3} \cdot 10^{-3}$	$\mathbf{2.69} \cdot 10^{-2}$	$5.33 \cdot 10^{-2}$	$7.35 \cdot 10^{-2}$	$1.56 \cdot 10^{1}$	$2.01 \cdot 10^{-1}$	$2.51 \cdot 10^{-2}$
NO-LIDK*	$1.36 \cdot 10^{0}$	$2.39 \cdot 10^{-1}$	$1.8 \cdot 10^{0}$	$1.1 \cdot 10^{-2}$	$3.14 \cdot 10^{-2}$	$6.86 \cdot 10^{-2}$	$8.91 \cdot 10^{-2}$	$1.59 \cdot 10^{1}$	$2.23 \cdot 10^{-1}$	$2.93 \cdot 10^{-2}$
NO-LIDK◇	$1.17 \cdot 10^{0}$	$2.04 \cdot 10^{-1}$	$1.12 \cdot 10^{0}$	$1.92 \cdot 10^{-2}$	$3.71 \cdot 10^{-2}$	$5.96 \cdot 10^{-2}$	$7.6 \cdot 10^{-2}$	$1.61 \cdot 10^{1}$	$1.42 \cdot 10^{-1}$	$2.12 \cdot 10^{-2}$
NO-LIDK†	$1.17 \cdot 10^{0}$	$2.05 \cdot 10^{-1}$	$1.14 \cdot 10^{0}$	$1.23 \cdot 10^{-2}$	$3.31 \cdot 10^{-2}$	$5.94 \cdot 10^{-2}$	$7.74 \cdot 10^{-2}$	$1.56 \cdot 10^{1}$	$\mathbf{1.03} \cdot 10^{-1}$	$2.18 \cdot 10^{-2}$
LoGLo-FNO	$\mathbf{1.09} \cdot 10^{0}$	$\mathbf{1.92} \cdot 10^{-1}$	$\mathbf{1.12} \cdot 10^{0}$	$8.99 \cdot 10^{-3}$	$2.80 \cdot 10^{-2}$	$\mathbf{4.55} \cdot 10^{-2}$	$\mathbf{6.93} \cdot 10^{-2}$	$\mathbf{1.26} \cdot 10^{1}$	$1.67 \cdot 10^{-1}$	$\mathbf{2.07} \cdot 10^{-2}$
Rel. % Diff	-18.26 %	-18.39 %	-16.12 %	-38.21 %	-16.86 %	-21.62 %	-17.26 %	-21.31 %	-73.04 %	-58.02 %

*Li et al., (2022) "Learning chaotic dynamics in dissipative systems"; (Re=5000)

Experiments and Results

3D Turbulent Radiative Layer*

Model	RMSE (↓)	nRMSE	bRMSE	cRMSE	fRMSE(L)	fRMSE(M)	fRMSE(H)	vRMSE (↓)	MaxError (↓)
1-step Evaluation									
U-Net	✗	✗	✗	✗	✗	✗	✗	$3.73 \cdot 10^{-1}$	✗
CNextU-Net	✗	✗	✗	✗	✗	✗	✗	$3.67 \cdot 10^{-1}$	✗
FNO	$2.76 \cdot 10^{-1}$	$2.97 \cdot 10^{-1}$	$6.83 \cdot 10^{-1}$	$2.3 \cdot 10^{-2}$	$5.17 \cdot 10^{-2}$	$4.45 \cdot 10^{-2}$	$2.76 \cdot 10^{-2}$	$3.09 \cdot 10^{-1}$	$1.11 \cdot 10^{1}$
LoGlo-FNO	$2.48 \cdot 10^{-1}$	$2.66 \cdot 10^{-1}$	$6.3 \cdot 10^{-1}$	$2.02 \cdot 10^{-2}$	$4.46 \cdot 10^{-2}$	$3.74 \cdot 10^{-2}$	$2.48 \cdot 10^{-2}$	$2.77 \cdot 10^{-1}$	$1.07 \cdot 10^{1}$
Rel. % Diff	-10.08%	-10.4%	-7.71%	-12.11%	-13.85%	-15.91%	-10.34%	-10.46%	-3.68%

*Ohana et al., (2024) "The Well: a Large-Scale Collection of Diverse Physics Simulations for Machine Learning"

Experiments and Results

2D Diffusion-Reaction (PDEBench)*

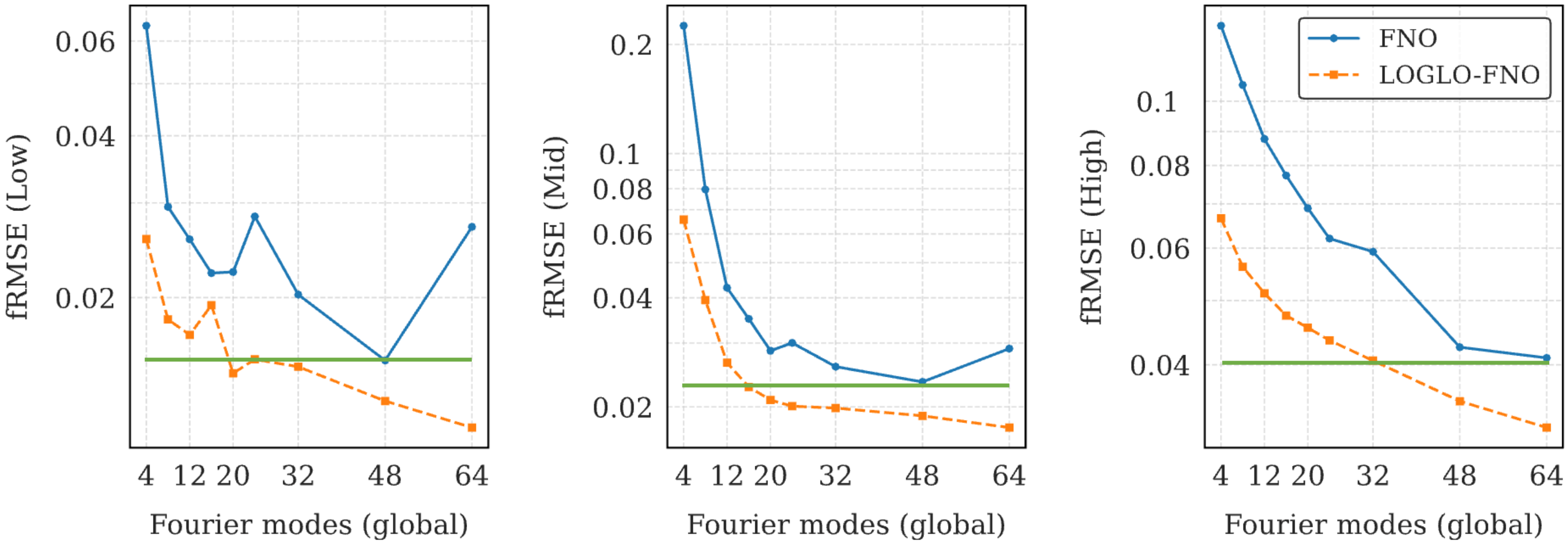
Model	RMSE (↓)	nRMSE	bRMSE	cRMSE	fRMSE(L)	fRMSE(M)	fRMSE(H)	MaxError (↓)	MELR (↓)	WLR (↓)
U-Net	$6.1 \cdot 10^{-2}$	$8.4 \cdot 10^{-1}$	$7.8 \cdot 10^{-2}$	$3.9 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$8.2 \cdot 10^{-4}$	$5.7 \cdot 10^{-2}$	$1.9 \cdot 10^{-1}$	✗	✗
U-FNO	$1.4 \cdot 10^{-2}$	$2.6 \cdot 10^{-1}$	$2.0 \cdot 10^{-2}$	$4.3 \cdot 10^{-3}$	$3.4 \cdot 10^{-3}$	$1.6 \cdot 10^{-3}$	$2.6 \cdot 10^{-4}$	$7.8 \cdot 10^{-2}$	$4.5 \cdot 10^{-1}$	$7.9 \cdot 10^{-2}$
FNO	$5.2 \cdot 10^{-3}$	$8.3 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$1.2 \cdot 10^{-3}$	$6.2 \cdot 10^{-4}$	$5.6 \cdot 10^{-4}$	$2.4 \cdot 10^{-4}$	$7.3 \cdot 10^{-2}$	$2.96 \cdot 10^{-1}$	$1.3 \cdot 10^{-2}$
F-FNO	$4.3 \cdot 10^{-3}$	$7.0 \cdot 10^{-2}$	$7.9 \cdot 10^{-3}$	$2.8 \cdot 10^{-3}$	$9.6 \cdot 10^{-4}$	$4.7 \cdot 10^{-4}$	$\mathbf{1.3} \cdot 10^{-4}$	$5.3 \cdot 10^{-2}$	$2.0 \cdot 10^{-1}$	$1.3 \cdot 10^{-2}$
LSM	$2.81 \cdot 10^{-2}$	$4.47 \cdot 10^{-1}$	$3.45 \cdot 10^{-2}$	$5.92 \cdot 10^{-3}$	$7.17 \cdot 10^{-3}$	$2.4 \cdot 10^{-3}$	$3.67 \cdot 10^{-4}$	$1.32 \cdot 10^{-1}$	$3.43 \cdot 10^{-1}$	$2.08 \cdot 10^{-1}$
NO-LIDK (loc. int)	$\mathbf{3.6} \cdot 10^{-3}$	$\mathbf{6.3} \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$4.8 \cdot 10^{-4}$	$4.0 \cdot 10^{-4}$	$4.6 \cdot 10^{-4}$	$1.5 \cdot 10^{-4}$	$5.0 \cdot 10^{-2}$	✗	✗
LoGlo-FNO	$\underline{3.89} \cdot 10^{-3}$	$\underline{6.4} \cdot 10^{-2}$	$\mathbf{5.2} \cdot 10^{-3}$	$\mathbf{4.6} \cdot 10^{-4}$	$\mathbf{2.8} \cdot 10^{-4}$	$\mathbf{3.2} \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$	$\mathbf{2.2} \cdot 10^{-2}$	$\mathbf{1.6} \cdot 10^{-1}$	$\mathbf{7.9} \cdot 10^{-3}$
Rel. % Diff	-25.19 %	-22.75 %	-65.13 %	-61.67 %	-54.84 %	-42.86 %	-20.83 %	-69.97 %	-44.09 %	-36.98 %

Fully autoregressive training and evaluation.

*Takamoto et al., (2022) “PDEBench: An Extensive Benchmark for Scientific Machine Learning”

Experiments and Results

2D INS: Kolmogorov flow*



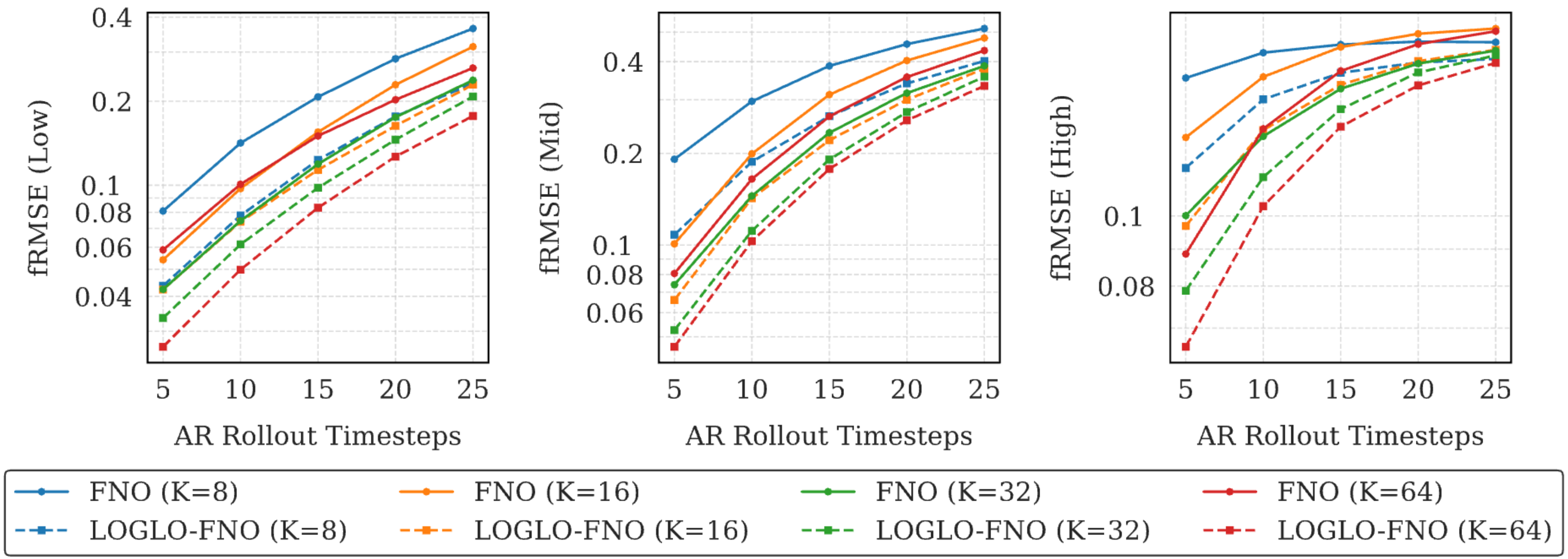
1-step Error (fRMSE):

- Global branch: varying number of modes
- Local branch : patch size (16 x 16)

*Li et al., (2022) "Learning chaotic dynamics in dissipative systems"; (Re=5000)

Experiments and Results

2D INS: Kolmogorov flow*



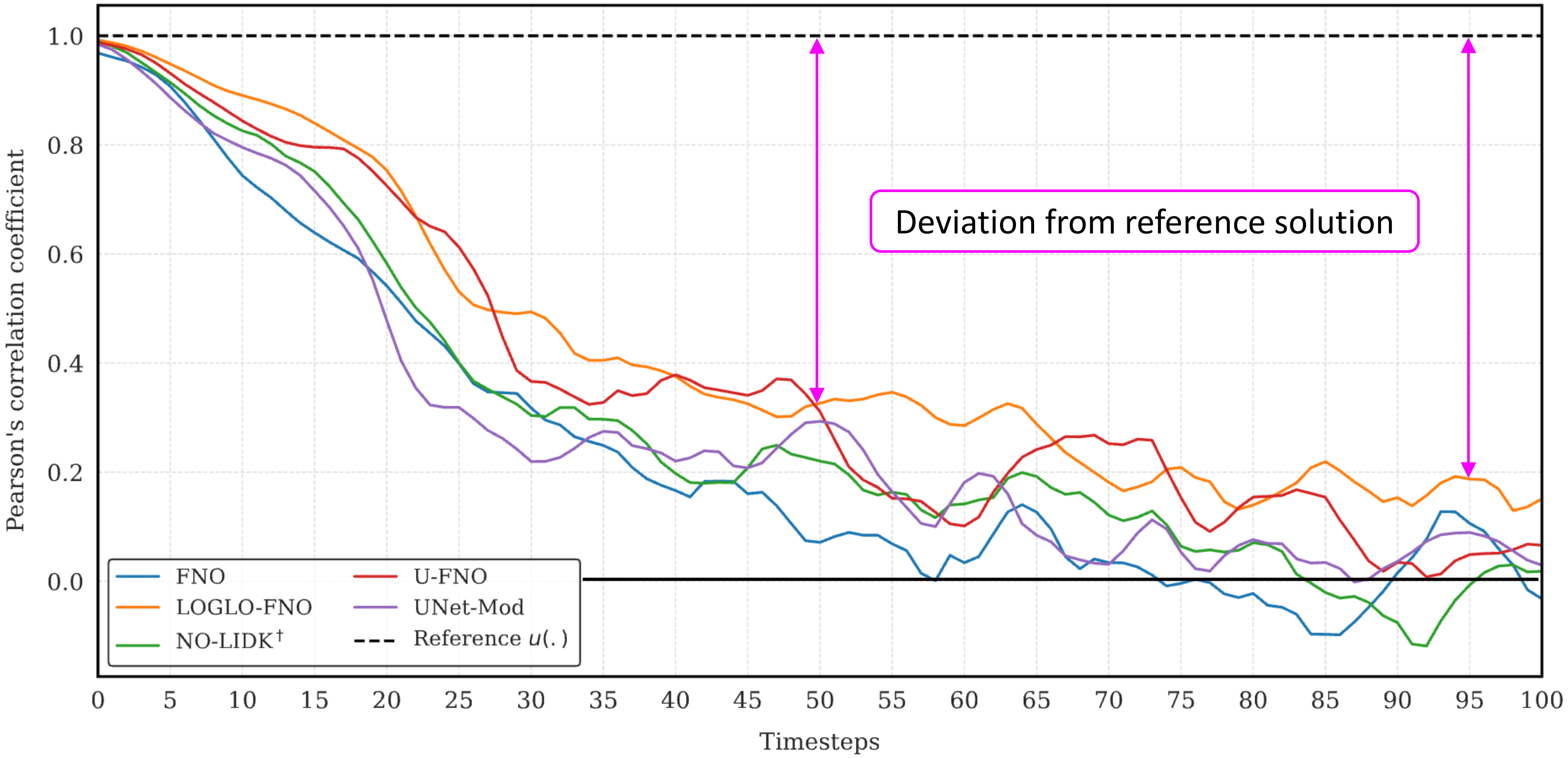
Autoregressive Rollout (fRMSE):

- Global branch: varying number of modes
- Local branch : patch size (16 x 16)

*Li et al., (2022) "Learning chaotic dynamics in dissipative systems"; (Re=5000)

Correlation with Reference Solution

(on a random trajectory from the Kolmogorov flow 2D test set)



Conclusions

- ✓ Proposed LOGLO-FNO model – for modeling **local features** and **high frequencies**.
- ✓ **Local spectral convolutions** significantly boosts FNO accuracy.
- ✓ **50% reduction** in trainable parameters to reach base FNO scores.
- ✓ **Radially binned spectral loss** complements the training objective.
- ✓ Rollout errors have also been improved, but not entirely mitigated.
- ✓ AR error accumulation is a **fundamental problem** that needs tackling.

Acknowledgements

SimTech NATIONALES HOCHLEISTUNGS RECHNEN impres-is

Project Page

https://kmario23.github.io/loglo-fno/