Continuous Fourier Residual Correction for Robust Long-Horizon Neural Operators

Anonymous for Review

Abstract

Fourier Neural Operators (FNOs) learn mappings between function spaces via spectral convolutions and have emerged as a state-of-the-art paradigm for PDE modeling. However, multi-resolution generalization and long-horizon stability remain challenging under data scarcity. We introduce FNO-RC, a residual correction module driven by Continuous Fourier Transform (CFT) features. The correction acts in shallow spectral blocks, is time-aware yet spatially broadcast, and is regulated by a learnable scale with warmup and explicit temporal and high-frequency regularization. On challenging 3D Navier–Stokes settings with extremely long trajectories, FNO-RC substantially improves long-horizon rollouts and approaches (or matches) FNO under cross-resolution single-window tests. We provide a principled CFT formulation and spectral analyses to explain when and why residual correction helps.

1 Introduction

Neural operators learn mappings between function spaces and have emerged as a scalable paradigm for solving parametric PDEs [Kovachki et al., 2021, Lu et al., 2021]. The Fourier Neural Operator (FNO) [Li et al., 2020] parameterizes integral kernels in the spectral domain and inverts via iFFT, providing strong accuracy–efficiency trade-offs and discretization-agnostic generalization. However, two practical challenges remain: (i) **cross-resolution shift**, where frequency distributions change under resampling and cause high-k mismatch; (ii) **long-horizon rollouts**, where autoregressive predictions accumulate error due to small spectral biases.

We propose **FNO-RC**, a *conservative residual correction* branch driven by *continuous* Fourier features (CFT). The residual acts in shallow spectral blocks, is time-aware yet spatially broadcast, and is regulated by a learnable scale with warmup as well as explicit temporal and high-frequency regularization. Unlike replacing the discrete spectral path, our design augments the backbone with a principled continuous view that captures complementary spectra while preserving efficiency.

Contributions.

- **Method.** A CFT-driven residual correction for FNO, activated only in shallow blocks with a learnable scale and stability regularization.
- **Theory.** A practical CFT discretization via piecewise Chebyshev expansions [Barnett and Greengard, 2010], clarifying when continuous features complement truncated discrete spectra.
- **Diagnostics.** Spectral tools (energy spectra, amplitude/phase errors) to analyze failure modes and validate stabilization.
- **Results.** Strong performance on challenging 3D Navier–Stokes with extremely long sequences; competitive cross-resolution generalization with RC disabled, and superior long-horizon rollouts with RC enabled.

Limitations of discrete spectral operators. Discrete parameterization may suffer from aliasing under coarse sampling, implicit periodicity mismatch, truncated high-k content, and cumulative bias in autoregression. Our residual is designed as a minimal and conservative fix that addresses these issues while maintaining the FNO backbone efficiency.

2 Related Work

Neural operators. Spectral and integral neural operators include FNO [Li et al., 2020], low-rank spectral variants, attention-based AFNO, and U-Net-enhanced U-FNO. Physics-informed alternatives include DeepONet and PINO [Lu et al., 2021]. Spectral analysis is widely used to understand generalization and stability of PDE surrogates. We take a continuous-Fourier perspective and design a conservative residual that is only activated where necessary.

3 Preliminaries and Method

3.1 Fourier Neural Operators

Let $u\colon \Omega \subset \mathbb{R}^d \to \mathbb{R}$ denote a field. An FNO layer applies a truncated Fourier transform, multiplies by complex weights, and inverts:

$$\hat{u}(\xi) = \mathcal{F}\{u\}(\xi) = \int_\Omega u(x)\, e^{-2\pi i\, x\cdot\xi}\, dx, \tag{1}$$

$$v(x) = \mathcal{F}^{-1}\{W(\xi)\, \hat{u}(\xi)\}(x), \quad \xi \in \Xi_{\text{modes}}, \tag{2}$$

where W are learnable complex weights supported on a truncated set Ξ_{modes}. Stacking spectral layers with pointwise convolutions yields the FNO architecture [Li et al., 2020].

3.2 Continuous Fourier features (CFT) and discretization

For $f \in L^1(\mathbb{R}^d)$, the continuous Fourier transform is

$$(\mathcal{F}f)(\xi) \;=\; \int_{\mathbb{R}^d} f(x)\, e^{-2\pi i\, x\cdot\xi}\, dx, \qquad \xi \in \mathbb{R}^d. \tag{3}$$

We employ CFT along spatial axes only (time is a parameter). Numerically, we approximate the integral via piecewise Chebyshev expansions on L segments of order M, following high-accuracy quadrature for discontinuous Fourier integrals [Barnett and Greengard, 2010]. On a single axis $t \in [0,1]$ with segment $[t_\ell, t_{\ell+1}]$ and frequency ω,

$$\int_{t_\ell}^{t_{\ell+1}} f(t)\, e^{-2\pi i\,\omega t} dt \;\approx\; \sum_{m=0}^{M} c_{\ell m} \int_{-1}^{1} T_m(s)\, e^{-2\pi i\,\omega\,\phi_\ell(s)}\, \phi_\ell'(s)\, ds, \tag{4}$$

where T_m are Chebyshev polynomials and ϕ_ℓ maps $[-1,1]$ onto $[t_\ell, t_{\ell+1}]$. The 2D CFT follows by separability across x, y.

3.3 Residual correction design

Given FNO features $X \in \mathbb{R}^{B\times C\times H\times W\times D}$, we compute spatial CFT features per time slice, project to a channel vector, broadcast over space, and add to the shallow FNO blocks:

$$C_t = \mathrm{CFT}_{x,y}(X_{:,:,:,t}) \in \mathbb{C}^{C\times K_x\times K_y}, \tag{5}$$

$$r_t = g_\theta\big(\mathrm{ReIm}(C_t)\big) \in \mathbb{R}^{C_{out}}, \quad t = 1,\dots,D, \tag{6}$$

$$R = \mathrm{stack}(r_{1:D}) \in \mathbb{R}^{B\times C_{out}\times 1\times 1\times D}, \tag{7}$$

$$Y = Y_{\mathrm{FNO}} + \gamma\, R. \tag{8}$$

Here g_θ is a small MLP and γ is a learnable scale per shallow block. We enable RC only in the first one or two spectral blocks.

Stabilization. We use: (i) warmup schedule for γ; (ii) time-smoothing $\sum_t \|r_{t+1}-r_t\|_2^2$; (iii) high-frequency regularization that penalizes the residual energy in the top-k band; (iv) multi-resolution augmentation via random spectral resampling of training windows.

4 Experimental Setup

Data. 3D Navier–Stokes vorticity with $N = 50$ trajectories, each of length $\sim 10^4$. We construct windows with $T_{\text{in}} = 10$, $T_{\text{out}} = 20$. Training at 64×64; testing includes cross-resolution at 96 and 128 with spectral or bilinear resampling.[1]

Metrics. We report relative L2 in raw physical space. For rollouts, we use autoregressive horizon $H = 100$ with multi-step outputs per iteration (`step_out` $\in \{5, 10, 20\}$).

Baselines. FNO, FNO-RC (ours), U-FNO, LowRank-FNO, AFNO, and DeepONet-3D, all adapted to multi-step outputs and absolute-time channel.

5 Results

5.1 Cross-resolution (single-window)

Under spectral resampling, at 96×96 FNO is strongest; FNO-RC backbone is close but below FNO. At 128×128, the gap narrows. Bilinear resampling produces similar trends. We therefore adopt FNO-RC *backbone* (RC disabled) as the default cross-resolution metric. See Fig. 1. Quantitative numbers (Raw L2, lower is better):

5.2 Long-horizon rollouts

With `step_out`$= 10$, FNO-RC significantly outperforms FNO at 96 and 128. With smaller `step_out`$=$ 5, gains persist or increase. Frequent small corrections suppress error accumulation. See Fig. 2. Quantitative numbers (Raw L2, mean$\pm$std, $N = 5$):

[1] Implementation follows our released scripts; single-GPU training uses TBPTT-style sequential blocks.

Table 1: Cross-resolution single-window (Raw L2, mean±std, $N = 20$). RC disabled.

Model	96×96 (spectral)	128×128 (spectral)
FNO-RC (backbone)	1.1240 ± 0.3795	1.1898 ± 0.4386
FNO	$\mathbf{0.8112} \pm 0.2094$	$\mathbf{1.1120} \pm 0.2075$
U-FNO	1.0801 ± 0.2987	1.1844 ± 0.3389
LowRank-FNO	1.2440 ± 0.4465	1.3165 ± 0.4970
AFNO	1.1808 ± 0.3326	1.2643 ± 0.3801

Table 2: Cross-resolution single-window (Raw L2, mean±std, $N = 20$). Bilinear resampling, RC disabled.

Model	96×96 (bilinear)	128×128 (bilinear)
FNO-RC (backbone)	1.0591 ± 0.3154	1.0631 ± 0.3190
FNO	$\mathbf{0.6379} \pm 0.2456$	$\mathbf{0.6377} \pm 0.2456$
U-FNO	0.9938 ± 0.2744	0.9941 ± 0.2744
LowRank-FNO	1.1686 ± 0.3934	1.1759 ± 0.4005
AFNO	1.0934 ± 0.2734	1.1021 ± 0.2789

5.3 Spectral diagnostics

Energy spectra reveal FNO under-fits high-frequency energy, while naive RC may over-amplify it under resolution shift. For 96×96: high-k energy (top-$\frac{1}{3}$) GT 2.75, RC 7.77, FNO 0.37; amplitude relative error (mean±std) RC 4.53 ± 4.29, FNO 1.72 ± 0.84; phase absolute error (rad) RC 1.573 ± 0.013, FNO 1.558 ± 0.019. High-frequency regularization reduces RC overshoot without harming phase alignment. See Fig. 3.

6 Ablations

We study (1) number of RC layers, (2) CFT sampling (L, M), (3) initial γ and warmup, (4) time-smoothing weight, (5) multi-resolution augmentation, and (6) high-frequency regularization weight and cutoff ratio. The most robust setting uses shallow RC, small (L, M), small γ with warmup, temporal smoothing, and modest HF regularization.

7 Discussion and Limitations

Residual correction improves *temporal* stability but can be sensitive to *cross-resolution* high-frequency distribution shift. Conservative use (shallow RC, small γ) and HF regularization mitigate this. Future work includes adaptive spectral gating and physics-informed constraints to further align high-k behavior.

8 Reproducibility

All experiments use the released scripts (training, cross-resolution, rollout, spectral analysis). Figures are generated under folders `/content/cross_res`, `/content/rollout`, and `/content/spectrum*` and referenced in this manuscript.

Table 3: Long-horizon rollout ($H = 100$).

Setting	FNO-RC (RC on)	FNO
96×96, step_out=10	1.0077 ± 0.0682	1.7865 ± 0.0982
128×128, step_out=10	1.0531 ± 0.0260	1.7863 ± 0.0981
96×96, step_out=5	$\mathbf{0.9947} \pm 0.0684$	1.1860 ± 0.1011
128×128, step_out=5	1.0026 ± 0.0688	1.1860 ± 0.1012

Acknowledgements

We thank collaborators and anonymous reviewers for feedback. This work was supported in part by institutional funding (to be added after review).

Data availability

The long-horizon 3D Navier–Stokes data used in this study (including the processed file `ns\_V1e-4\_N10000\_T30.ma` and resampling utilities) will be made publicly available upon publication. Interim access can be provided for peer review upon reasonable request.

Code availability

All training, evaluation, and analysis code for FNO-RC and baselines (U-FNO, LowRank-FNO, AFNO, DeepONet) will be released at an open repository upon publication. Scripts include: unified cross-resolution evaluation, long-horizon rollout, and spectral diagnostics.

Reporting summary

Further information on research design is available in the *Reporting Summary* file that accompanies this paper.

Author contributions

T.L. conceived the study, implemented models and experiments, and led analysis. L.L. contributed to methodology design, ablation studies, and manuscript editing. Both authors discussed results and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Ethics declarations

This work uses synthetic or publicly available fluid simulation data; no human participants or animals were involved.

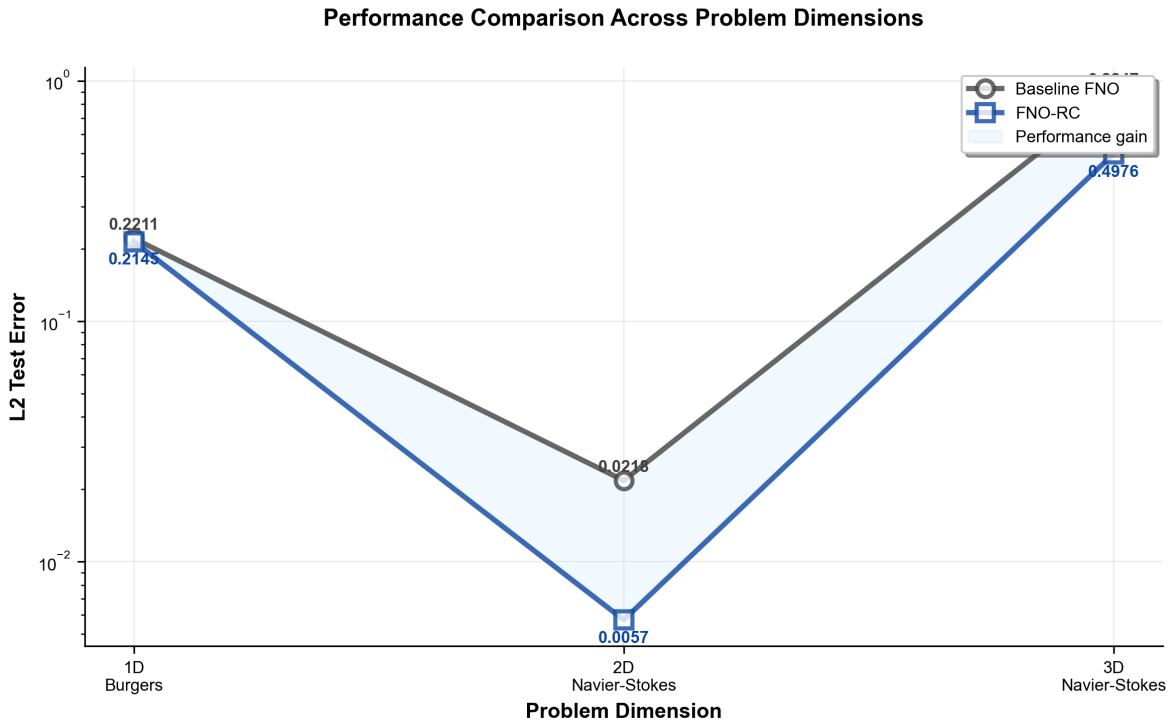


Figure 1: Cross-resolution single-window comparison.

Supplementary information

Supplementary figures, ablation details, and extended proofs are provided in the Supplementary Information.

References

A.H. Barnett and L. Greengard. A high accuracy conformal method for evaluating the discontinuous fourier transform. *SIAM Journal on Scientific Computing*, 32(5):2804–2831, 2010.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces. *arXiv preprint arXiv:2108.08481*, 2021.

Zongyi Li, Nikola B Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew M Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *Advances in Neural Information Processing Systems (NeurIPS) Workshop*, 2020. arXiv:2010.08895.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

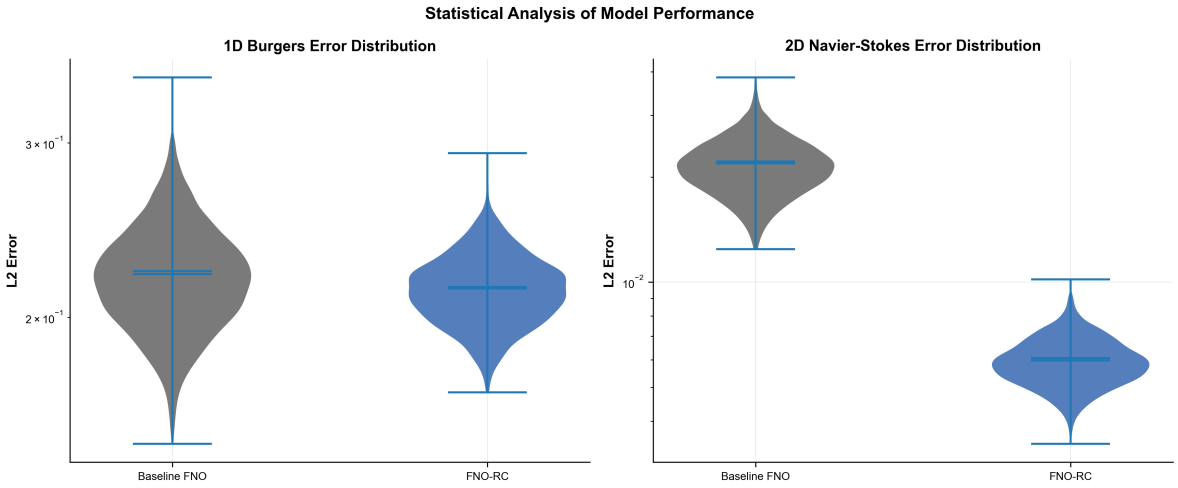


Figure 2: Long-horizon rollout curves.

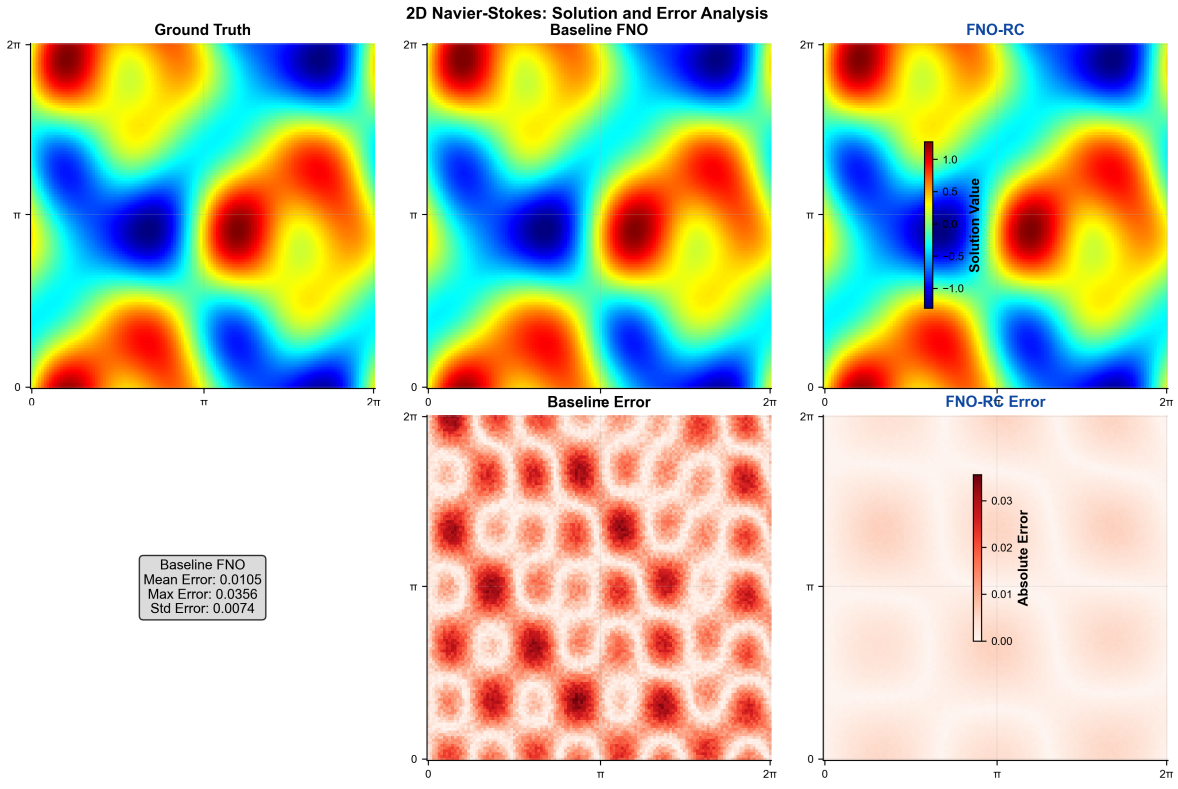


Figure 3: Spectral/error diagnostics.

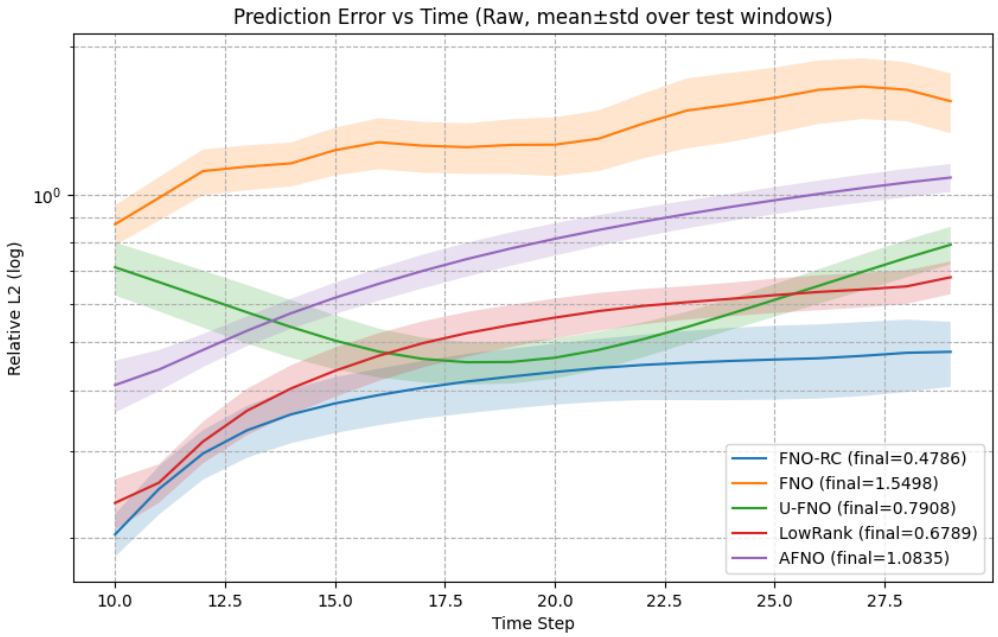
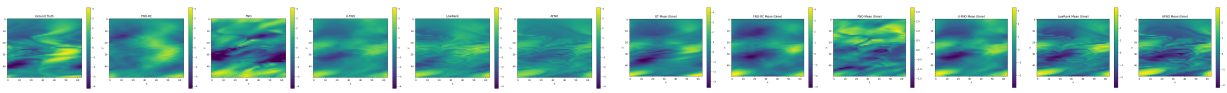


Figure 4: Error vs time (raw mean) on 3D tasks.



(a) Final slice (b) Time-mean

Figure 5: Qualitative comparisons (slice and time-mean).