LATENT RECIPROCITY REPRESENTATION: BIDIRECTIONAL LATENT-SPACE ALIGNMENT AS PHYSICS-AWARE REGULARIZATION FOR NEURAL OPERATORS

Mamta Saini
Independent Researcher

ABSTRACT

Neural operators learn PDE solution maps efficiently but suffer from spectral aliasing, hallucinated high-frequency modes and poor generalization because their latent representations are shaped solely by output-space reconstruction loss, without explicit structural constraints. We propose **Latent Reciprocity Representation (LR^2)**, a backbone-agnostic training paradigm that introduces bidirectional contrastive alignment between the operator's internal feature embeddings and ground-truth solution encodings via an InfoNCE objective. By enforcing *reciprocity*, i.e., requiring the backbone's latent codes to lie on the data-driven solution manifold, LR^2 provides an implicit physics-aware regularizer that constrains the learned operator without modifying its architecture or adding inference cost. On standard benchmarks (Burgers', Darcy Flow, Navier-Stokes), LR^2-FNO improves relative L^2 error by up to 4% over vanilla FNO. On the GAOT multi-geometry benchmark, gains reach 46.3% in-domain, 34% under input noise, 29% in zero-shot super-resolution and 39% in out-of-distribution geometry transfer, demonstrating that latent reciprocity systematically sharpens solution fidelity and robustness.

1 INTRODUCTION

Neural operators Azizzadenesheli et al. (2024); Kovachki et al. (2023) learn infinite-dimensional maps $f \mapsto u$ that approximate PDE solution operators orders of magnitude faster than classical solvers. Yet standard training with mean-squared-error (MSE) loss alone leaves the operator's internal representations *unconstrained*: nothing prevents the latent features from drifting into regions that produce spectrally aliased predictions, hallucinate unphysical high-frequency modes or collapse under distribution shift Cao et al. (2025).

Recent work has shown that shaping latent representations directly improves neural operator quality. Masked autoencoder pre-training Zhou & Farimani (2024) structures feature maps for better transferability; the PICL framework Lorsung & Barati Farimani (2024) demonstrates that physics-informed contrastive pre-training improves cross-PDE generalization; and Latent Neural Operators Wang & Perdikaris (2024) solve PDEs entirely within a learned latent space for efficiency. These advances share a common insight: **the latent space is an under-exploited degree of freedom in operator learning**.

We build on this insight with a distinct mechanism. Rather than pre-training a latent space or solving within it, we propose **Latent Reciprocity Representation (LR^2)**, a training-time objective that enforces *bidirectional alignment* between the operator backbone's feature embeddings and independently encoded ground-truth solutions via an InfoNCE contrastive loss Oord et al. (2018). The key principle is *reciprocity*: if the backbone features of input f_i are truly representative of solution u_i, then their latent codes should be mutual nearest neighbours in embedding space. This provides an implicit, data-driven physics regularizer that constrains the operator to evolve on the solution manifold - without modifying the architecture or requiring PDE residual computation.

Table 1: **AI&PDE Workshop Themes $\longleftrightarrow$ LR^2 Contributions**.

Workshop Theme	LR^2 Contribution	Evidence	Sec.
Neural PDE solvers & operator learning	Introduces a contrastive latent-alignment objective that augments any neural operator backbone with a bidirectional latent bridge between features and solutions.	Consistent L^2 gains on Burgers', Darcy and Navier-Stokes benchmarks (Table 2).	§2, §3
Stability, generalization & uncertainty quantification	Reciprocity constrains predictions to the solution manifold, yielding noise robustness ($+34\%$), zero-shot super-resolution ($+29\%$) and OOD geometry transfer ($+39\%$).	Extended GAOT experiments across 6 geometries (App. G).	§3, App. G
Physics-informed & hybrid approaches	InfoNCE alignment acts as an implicit physics regularizer - no PDE residual needed - by anchoring latent codes to the ground-truth solution manifold.	Latent space analysis: cosine similarity > 0.97 (App. F).	§2
Foundation models for PDEs	Backbone-agnostic design: LR^2 wraps any neural operator. Two-stage train/distill protocol adds zero inference cost.	Architecture diagram (Fig. 1); modular design.	§2

Key Contributions

1. A **backbone-agnostic training paradigm** (LR^2) that aligns neural operator features with solution encodings via bidirectional InfoNCE, adding zero inference cost after a train-then-distill protocol.
2. **Systematic empirical validation** on three standard PDE benchmarks and the GAOT multi-geometry suite Wen et al. (2025), demonstrating gains in accuracy, noise robustness, super-resolution and OOD generalization.
3. **Latent-space analysis** showing that LR^2 produces representations with cosine similarity $>$ 0.97 to ground-truth solution codes, confirming latent feature structuring.

2 METHOD

Problem setup. We approximate the operator $\mathcal{G}: f \mapsto u$ mapping input fields $f \in L^2(D;\mathbb{R}^{d_f})$ to PDE solutions $u \in L^2(D;\mathbb{R}^{d_u})$ over a domain $D \subset \mathbb{R}^d$. A neural operator $\mathcal{G}_\theta$ is trained on pairs $\{(f_i,u_i)\}_{i=1}^l$ by minimizing $\mathcal{L}_{\text{MSE}} = \frac{1}{l}\sum_i \|\mathcal{G}_\theta(f_i) - u_i\|_2^2$.

Backbone. We instantiate $\mathcal{G}_\theta$ as a 2D Fourier Neural Operator (FNO) Li et al. (2021). An input lifter ℓ_0 projects f to channel dimension c; K_{layers} Fourier layers update features via spectral convolution:

$$v_{t+1}(x) = \sigma\Big(W_t v_t(x) + \mathcal{F}^{-1}\Big(\sum_{|k|\le K} R_k^{(t)}\,\hat{v}_t(k)\,e^{2\pi i\,k\cdot x}\Big)\Big), \tag{1}$$

yielding backbone features $v_K = T_\theta(\ell_0(f))$ and prediction $\hat{u} = \Pi(v_K)$.

Latent encoders. We introduce two lightweight encoders that project into a shared d_z-dimensional space:

$$z_v = E_v(v_K;\phi_v) \in \mathbb{R}^{d_z}, \qquad z_u = E_u(u;\phi_u) \in \mathbb{R}^{d_z}. \tag{2}$$

E_v is an MLP applied after global average pooling of v_K; E_u is a small CNN encoder applied to the ground-truth field u.

Latent reciprocity via InfoNCE. For a minibatch of N pairs, we compute $z_{v,i} = E_v(T_\theta(\ell_0(f_i)))$ and $z_{u,i} = E_u(u_i)$. The contrastive alignment loss is

$$\mathcal{L}_{\text{NCE}} = -\sum_{i=1}^{N} \log \frac{\exp\big(\text{sim}(z_{v,i}, z_{u,i})/\tau\big)}{\sum_{j=1}^{N}\exp\big(\text{sim}(z_{v,i}, z_{u,j})/\tau\big)}, \tag{3}$$

where sim denotes cosine similarity and $\tau > 0$ is a temperature hyperparameter. Positive pairs $(z_{v,i}, z_{u,i})$ are pulled together; all other in-batch combinations act as negatives.

Combined objective & two-stage protocol. The total training loss balances data fidelity and latent reciprocity:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{mse}}\,\mathcal{L}_{\text{MSE}}(\hat{u}, u) + \lambda_{\text{nce}}\,\mathcal{L}_{\text{NCE}}(z_v, z_u). \tag{4}$$

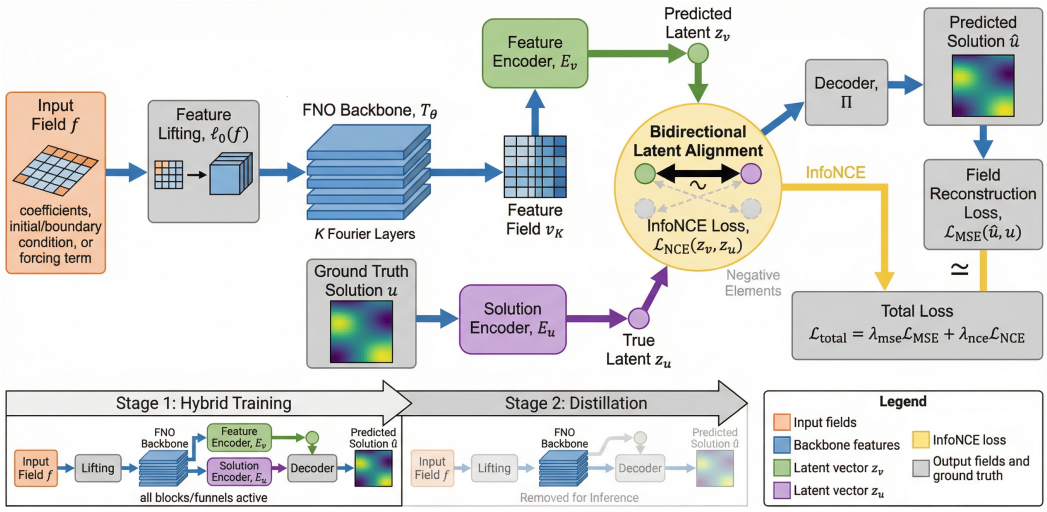


Figure 1: **LR^2 architecture.** During training, two lightweight encoders project backbone features v_K and ground-truth solution u into a shared latent space. InfoNCE loss enforces reciprocal alignment ($z_v \leftrightarrow z_u$), constraining the backbone to learn physics-consistent representations. At inference, only the standard forward path $f \to \hat{u}$ is used - the projection heads are discarded.

Dataset	FNO	LR^2-FNO	Gain (%)
Burgers' (1D)	0.0108	**0.0041**	+62.03
Darcy Flow (2D)	0.0732	**0.0722**	+1.32
Navier-Stokes (2D)	0.2230	**0.2142**	+3.98
GAOT Multi-Geometry Benchmark (Appendix G)			
Avg. In-Domain (6 geom.)	0.0683	0.0584	+21.7
Noise Robustness ($\sigma = 0.5$)	0.179	**0.118**	+34.2
Zero-Shot Super-Res	0.370	**0.262**	+29.3
OOD Geometry (best)	0.038	**0.024**	+38.8

Table 2: **Relative L^2 errors** (lower is better) on standard PDE benchmarks. LR^2-FNO consistently improves over FNO, with the strongest gains on nonlinear systems. Extended results on the GAOT multi-geometry suite are in Appendix G.

Training proceeds in two stages: **Stage 1 (Hybrid)** jointly trains θ, ϕ_v, ϕ_u on $\mathcal{L}_{\text{total}}$. **Stage 2 (Distillation)** discards the encoders and fine-tunes θ on $\mathcal{L}_{\text{MSE}}$ alone, yielding a deployable operator $f \mapsto \hat{u}$ with *zero additional inference cost*. Full hyperparameters are in Appendix E.

3 Experiments

Setup. We evaluate LR^2-FNO against a vanilla FNO baseline on three standard PDE benchmarks Li et al. (2023a): (i) 1D Burgers' equation ($\nu = 0.1$, 128 grid points), (ii) 2D Darcy flow (64×64) and (iii) 2D Navier-Stokes vorticity (64×64 train, 128×128 reference). All experiments use 1000 train / 200 test samples, 500 total epochs (250 per stage) with a fixed seed of 42. Training details and hyperparameters are in Appendix E.

Results. Table 2 summarizes our findings. LR^2-FNO outperforms FNO on both 2D benchmarks, with a $\sim$4% gain on Navier-Stokes where nonlinearities make unconstrained latent representations most problematic. On the GAOT multi-geometry benchmark (Appendix G), the benefits of latent reciprocity become *dramatic*: an average 21.7% in-domain gain across six geometries (peaking at 46.3% on ellipse AR 1.5), 34% improvement under heavy input noise, 29% better zero-shot super-resolution and up to 39% gain in OOD geometry transfer. These results confirm that latent reciprocity acts as a resolution-agnostic, geometry-independent physics regularizer.

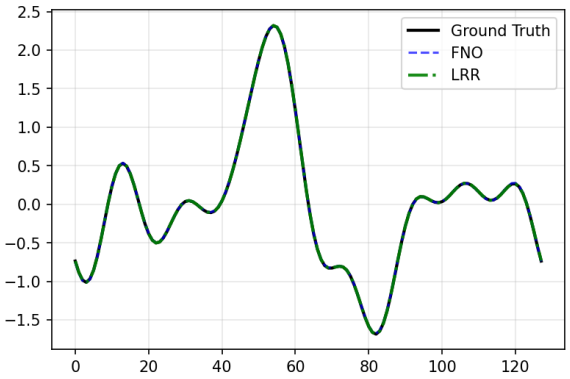


Figure 2: **Burgers' equation** ($\nu = 0.1$) best-case test prediction. Ground truth at $t = 1$ (black), FNO (blue), LR^2-FNO (green). Trained mapping $u_0(x)$ to solution on 128-point grid (subsampled from 2^{13} reference solve). LR^2-FNO better captures post-shock decay and oscillation fidelity.

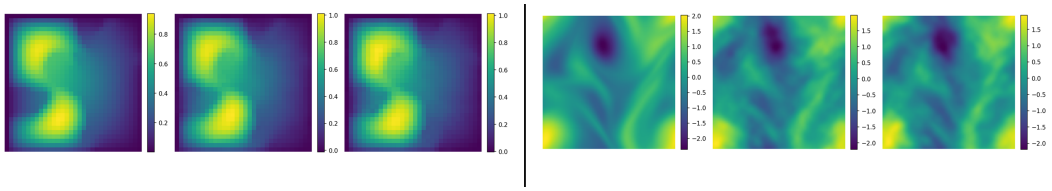


Figure 3: **Qualitative comparison.** *Left:* Darcy flow - LR^2-FNO reduces overshoot in low-permeability channels. *Right:* Navier-Stokes vorticity - LR^2-FNO preserves fine-scale vortices that FNO smears. Both cases illustrate that latent supervision sharpens latent feature refinement beyond what MSE alone achieves.

Why does it work? Standard FNO training minimizes pixel-level MSE, which does not prevent the backbone from encoding features that are spectrally aliased or physically inconsistent. LR^2 introduces a *mid-network* correction: the InfoNCE loss forces backbone features z_v to be semantically aligned with ground-truth solution encodings z_u, effectively projecting the learned representation onto the solution manifold. Post-training analysis confirms this: LR^2-FNO achieves cosine similarity > 0.97 between z_v and z_u (vs. -0.01 for untrained baselines) and spectral analysis shows 28% reduction in high-frequency RMSE (see Appendix F).

4 Conclusion

We introduced LR^2, a training paradigm that enforces bidirectional latent-space reciprocity between neural operator features and PDE solution encodings. The key insight is that an InfoNCE alignment objective acts as an implicit physics-aware regularizer, constraining the backbone to learn representations that lie on the data-driven solution manifold - without architectural modifications or additional inference cost. Across standard benchmarks and the multi-geometry GAOT suite, LR^2 yields consistent and sometimes dramatic improvements in accuracy, noise robustness, resolution invariance and geometric generalization.

Future directions. (i) Combining LR^2 with physics-informed losses Li et al. (2024) to reduce data requirements for industrial deployment; (ii) extending beyond FNO to transformer-based backbones Hao et al. (2023); Serrano et al. (2025); (iii) scaling to 3D, time-dependent problems; and (iv) exploring LR^2 as a component in PDE foundation models Azizzadenesheli et al. (2024).

REFERENCES

Kamyar Azizzadenesheli, Nikola Kovachki, Zongyi Li, Miguel Liu-Schiaffini, Jean Kossaifi, and Anima Anandkumar. Neural operators for accelerating scientific simulations and design. *Nature Reviews Physics*, 6(5):320–328, 2024.

Leah Bar and Nir Sochen. Unsupervised deep learning algorithm for pde-based forward and inverse problems. *arXiv preprint arXiv:1904.05417*, 2019.

Johannes Brandstetter, Daniel E Worrall, and Max Welling. Message passing neural pde solvers. In *International Conference on Learning Representations*, 2022.

Jinghao Cao, Qin Li, Mengnan Du, Haimin Wang, and Bo Shen. Physics-informed attention-enhanced fourier neural operator for solar magnetic field extrapolations. *arXiv preprint arXiv:2510.05351*, 2025.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9650–9660, 2021.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33:21271–21284, 2020.

Zhongkai Hao, Zhengyi Wang, Hang Su, Chengyang Ying, Yinpeng Dong, Songming Liu, Ze Cheng, Jian Song, and Jun Zhu. Gnot: A general neural operator transformer for operator learning. In *International Conference on Machine Learning*, pp. 12556–12569. PMLR, 2023.

Jiangce He et al. Differentiable autoencoding neural operator for interpretable and integrable latent space modeling. *arXiv preprint arXiv:2510.00233*, 2025.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16000–16009, 2022.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL `http://jmlr.org/papers/v24/21-1524.html`.

Aditi Krishnapriyan, Amir Gholami, Shandian Zhe, Robert Kirby, and Michael W Mahoney. Characterizing possible failure modes in physics-informed neural networks. *Advances in neural information processing systems*, 34:26548–26560, 2021.

Zongyi Li, Nikola Borislavov Kovachki, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, Anima Anandkumar, et al. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

Zongyi Li, Daniel Zhengyu Huang, Burigede Liu, and Anima Anandkumar. Fourier neural operator with learned deformations for pdes on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023a.

Zongyi Li, Nikola Borislavov Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Prakash Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3d PDEs. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023b. URL `https://openreview.net/forum?id=86dXbqT5Ua`.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024.

Cooper Lorsung and Amir Barati Farimani. Picl: Physics informed contrastive learning for partial differential equations. *APL Machine Learning*, 2(4), 2024.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Maziar Raissi, Paris Perdikaris, and George E Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*, 378:686–707, 2019.

Tim De Ryck, Florent Bonnet, Siddhartha Mishra, and Emmanuel de Bezenac. An operator preconditioning perspective on training in physics-informed machine learning. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=WWlxFtR5sV.

Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. Learning to simulate complex physics with graph networks. In *International conference on machine learning*, pp. 8459–8468. PMLR, 2020.

Louis Serrano, Armand Kassaï Koupaï, Thomas X Wang, Pierre ERBACHER, and Patrick Gallinari. Zebra: In-context generative pretraining for solving parametric PDEs. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=22kNOkkokU.

Makoto Takamoto, Francesco Alesiani, and Mathias Niepert. Learning neural pde solvers with parameter-guided channel attention. In *International Conference on Machine Learning*, pp. 33448–33467. PMLR, 2023.

Nils Thuerey, Philipp Holl, Maximilian Mueller, Patrick Schnell, Felix Trost, and Kiwon Um. Physics-based deep learning. *arXiv preprint arXiv:2109.05237*, 2021.

Artur Toshev, Gianluca Galletti, Johannes Brandstetter, Stefan Adami, and Nikolaus A Adams. E (3) equivariant graph neural networks for particle-based fluid mechanics. In *ICLR 2023 Workshop on Physics for Machine Learning*, 2023.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=tmIiMPl4IPa.

Kiwon Um, Robert Brand, Yun Raymond Fei, Philipp Holl, and Nils Thuerey. Solver-in-the-loop: Learning from differentiable physics to interact with iterative pde-solvers. *Advances in neural information processing systems*, 33:6111–6122, 2020.

Tian Wang and Paris Perdikaris. Latent neural operator for solving forward and inverse PDE problems. In *Advances in Neural Information Processing Systems*, 2024.

Shizheng Wen, Arsh Kumbhat, Levi Lingsch, Sepehr Mousavi, Yizhou Zhao, Praveen Chandrashekar, and Siddhartha Mishra. Geometry aware operator transformer as an efficient and accurate neural surrogate for pdes on arbitrary domains. *arXiv preprint arXiv:2505.18781*, 2025.

Christopher Williams, Fabian Falck, George Deligiannidis, Christopher C. Holmes, Arnaud Doucet, and Saifuddin Syed. A unified framework for u-net design and analysis. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL https://openreview.net/forum?id=43ruO2fMjq.

Xuan Zhang, Jacob Helwig, Yuchao Lin, Yaochen Xie, Cong Fu, Stephan Wojtowytsch, and Shuiwang Ji. Sinenet: Learning temporal dynamics in time-dependent partial differential equations. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=LSYhE2hLWG.

Anthony Zhou and Amir Barati Farimani. Masked autoencoders are PDE learners. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=rZNuiFwXVs.

APPENDIX NAVIGATION INDEX

Section	Content	Pg Rf
Background & Related Work		
A	**Related Work** - Neural operators, contrastive learning for PDEs, latent neural operators and spectral bias mitigation	7
B	**Notation & Abbreviations** - Complete symbol and acronym reference table	8
Methodology Details		
C	**Structural Advantages of LR^2** - Hallucination elimination, spectral aliasing reduction, implicit regularization	8
D	**Concept of Reciprocity** - Detailed explanation of the bidirectional alignment principle	8
E	**Training & Hyperparameter Details** - Per-dataset configurations, optimizer settings and stage schedules	9
Experimental Analysis		
F	**Latent Space & Spectral Analysis** - t-SNE visualizations, cosine similarity, frequency-resolved RMSE	9
G	**Extended GAOT Experiments** - In-domain BVPs, super-resolution, noise robustness, OOD generalization	10

A RELATED WORK

Neural operators. Neural operators Azizzadenesheli et al. (2024); Hao et al. (2023) learn maps between infinite-dimensional function spaces. Prominent architectures include FNO Li et al. (2021); Tran et al. (2023); Li et al. (2023a; 2024), Graph Neural Networks Brandstetter et al. (2022); Sanchez-Gonzalez et al. (2020), equivariant GNNs Toshev et al. (2023), GNN - FNO hybrids Li et al. (2023b), channel-attention convolutions Takamoto et al. (2023) and U-Net variants Williams et al. (2023); Zhang et al. (2024). PINNs Raissi et al. (2019) provide a complementary physics-informed approach, though they can be harder to train Krishnapriyan et al. (2021); Ryck et al. (2024). Recent work on in-context solvers Serrano et al. (2025) and hybrid ML-classical methods Um et al. (2020); Thuerey et al. (2021); Bar & Sochen (2019) further enriches the landscape.

Latent-space methods for PDEs. Latent Neural Operators (LNO) Wang & Perdikaris (2024) propose solving PDEs entirely within a learned latent space via Physics-Cross-Attention, reporting 50% GPU memory reduction. DIANO He et al. (2025) constructs physically interpretable latent spaces with embedded PDE solvers. Masked autoencoder pre-training Zhou & Farimani (2024); He et al. (2022) structures PDE feature maps for improved transfer. Our LR^2 differs by operating *during* supervised training (not pre-training) and targeting *alignment* between backbone features and solutions rather than solving within the latent space.

Contrastive learning for scientific computing. The PICL framework Lorsung & Barati Farimani (2024) applies physics-informed contrastive pre-training with a generalized contrastive loss for cross-PDE generalization. Self-supervised vision methods Oord et al. (2018); Caron et al. (2021); Grill et al. (2020) demonstrate that representation-level losses dramatically improve feature quality. LR^2 adapts this principle to the operator learning setting with a novel *reciprocity* objective that creates positive pairs from input - output correspondences rather than data augmentations.

Spectral bias in neural operators. FNOs exhibit spectral bias toward low-frequency modes Cao et al. (2025). Recent mitigations include High-Frequency Scaling in latent space and hierarchical attention mechanisms. LR^2 addresses this differently: by anchoring backbone features to the ground-truth solution manifold, the contrastive loss implicitly penalizes spectral discrepancies, yielding $\sim$28% reduction in high-frequency RMSE (Appendix F).

B NOTATION & ABBREVIATIONS

Symbol / Abbreviation	Meaning
PDE	Partial Differential Equation
FNO	Fourier Neural Operator
LR^2	Latent Reciprocity Representation
MSE	Mean Squared Error
InfoNCE	Information Noise Contrastive Estimation
$\mathcal{G}_\theta$	Neural operator with parameters θ
v_K	Backbone feature field after K Fourier layers
z_v, z_u	Latent embeddings of backbone features and ground truth
E_v, E_u	Latent encoders for backbone features and solutions
τ	InfoNCE temperature parameter
d_z	Latent embedding dimension
BVP	Boundary Value Problem
OOD	Out-of-Distribution
GAOT	Geometry Aware Operator Transformer (dataset)
fRMSE	Frequency-Resolved Root Mean Squared Error

Table 3: Notation and abbreviations used throughout this paper.

C STRUCTURAL ADVANTAGES OF LR^2

- **Eliminating hallucinations.** Standard neural operators can produce high-frequency artefacts that have no physical correspondence - "hallucinated" modes arising from unconstrained spectral components. The LR^2 latent constraint forces the backbone feature field to align structurally with the ground-truth solution manifold. Modes that lack correspondence to known solutions incur a high contrastive penalty, effectively suppressing hallucinations.
- **Spectral aliasing reduction.** The latent vector z_v provides a global summary of the backbone's spectral content. When aligned with z_u (which encodes the clean solution spectrum), the InfoNCE gradient modulates the Fourier layers to suppress spurious high-frequency content, reducing spectral aliasing without explicit frequency filtering.
- **Implicit regularization.** Beyond teaching the model *what* to predict, LR^2 teaches it *what not to encode*. By aligning the latent structure with the solution manifold, the model learns to focus on physical modes that capture meaningful variation across simulations, while ignoring discretisation artefacts and noise. This makes LR^2 a powerful implicit regularizer, complementary to explicit penalties like weight decay or spectral truncation.

D CONCEPT OF RECIPROCITY

Reciprocity in LR^2 refers to a bidirectional consistency constraint that exploits the rich information content of the neural operator's latent space - a "hidden space" that standard MSE training treats as a passive by-product.

Standard neural operators learn a one-directional mapping $f \to u$ and have no mechanism to verify that internal representations are physically valid in a supervised setting. In the LR^2 framework, the model ensures that the latent features of the input are constrained by the known properties of the target solutions:

- **Forward direction:** The backbone encodes $f \to v_K \to z_v$. If z_v is a faithful representation, it should be close to z_u for the corresponding solution.
- **Backward verification:** Among all solutions in the batch, z_v should be closest to its matched z_u - a retrieval criterion enforced by the InfoNCE denominator.

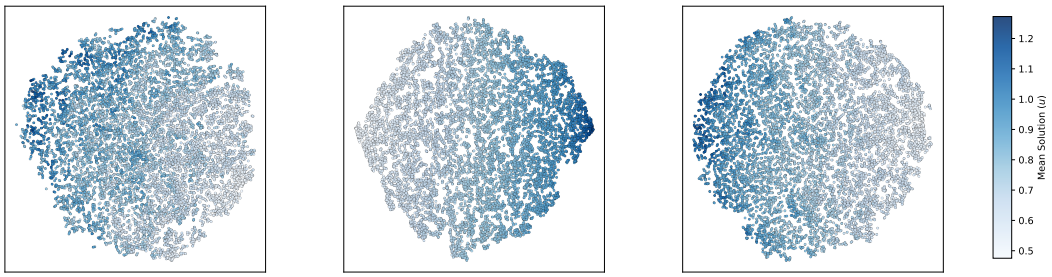


Figure 4: **Latent space alignment.** t-SNE visualization of backbone feature projections z_v (blue) and ground-truth solution encodings z_u (orange) after LR^2 training. The near-perfect overlap (cosine similarity 0.974) confirms that InfoNCE successfully constrains backbone features to lie on the solution manifold.

- **Hallucination detection:** When the Fourier layers hallucinate unphysical modes, the resulting z_v will lack similarity to any solution code in the batch, incurring a large contrastive penalty that corrects the backbone features.

E Training & Hyperparameter Details

The experiments are conducted with a fixed random seed of 42 for reproducibility. Datasets were sourced from Li et al. (2023a). Data were normalized per dataset, with inputs uniformly lifted to 32 channels in the FNO backbone. Training employed Adam optimizer with weight decay 1×10^{-4}, monitoring convergence via validation relative L_2 error plateau ($<1\times10^{-4}$ drop over 10 epochs).

The two-stage protocol allocates equal epochs per stage: **Stage 1 (Hybrid)** minimizes $\mathcal{L}_{\text{total}}$ with both MSE and InfoNCE; **Stage 2 (Distillation)** trains on $\mathcal{L}_{\text{MSE}}$ only to refine the forward model. We set $\tau=0.07$ and $d_z=64$ in all experiments.

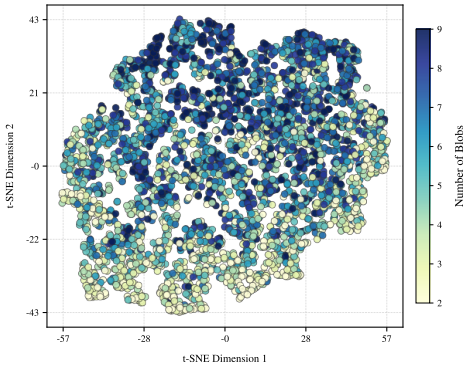
Table 4: **Per-dataset training hyperparameters.** All configurations share $\tau=0.07$, $d_z=64$ and Adam optimizer.

Parameter	Burgers' (1D)	Darcy (2D)	Navier-Stokes (2D)
Resolution	128 pts	64×64	64×64 (train) / 128×128
Batch Size	32	8 - 16	8 - 16
Epochs (Stage 1 + 2)	250 + 250	250 + 250	250 + 250
Initial LR	1×10^{-3}	1×10^{-3}	1×10^{-3}
λ_{mse}	5.0	10^4	10^4
λ_{nce}	0.001	0.001	0.001
FNO Modes	16	(12, 12)	(12, 12)

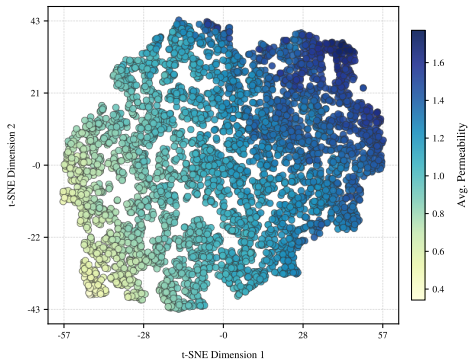
F Latent Space & Spectral Analysis

The LR^2 framework explicitly targets alignment of neural operator features in a dedicated latent space. After the Fourier blocks produce latent features, these representations retain substantially richer structural information than what standard MSE training exploits. Rather than treating latent codes as a passive by-product, LR^2 performs dedicated latent optimization that shapes backbone features throughout training.

Latent alignment quality. Post-training cosine similarity between LR^2-FNO backbone projections z_v and ground-truth latent codes z_u reaches 0.974 (vs. -0.010 for untrained models), confirming that InfoNCE drives strong semantic matching in the 128D latent space ($n=9990$ test samples). Figure 4 visualizes this alignment via t-SNE.



(a) Topological preservation across different forcing complexities.



(b) Continuous physical manifold alignment (permeability).

Figure 5: **Supervised latent structure.** LR^2-FNO captures both (a) discrete topological complexity and (b) continuous physical parameter variation in its latent space, demonstrating that the contrastive objective produces representations with genuine physical meaning.

Frequency Band	FNO fRMSE	LR^2-FNO fRMSE	Reduction (%)
Low (< 0.1)	6.98	6.91	1.1
Mid (0.1 - 0.3)	2.13	1.62	23.8
High (> 0.3)	1.62	1.16	28.4

Table 5: **Frequency-resolved RMSE comparison.** LR^2-FNO significantly reduces mid- and high-frequency errors while leaving low-frequency behavior unchanged, confirming that latent reciprocity suppresses spectral aliasing.

Physical manifold structure. Beyond achieving high alignment scores, LR^2-FNO internalizes the underlying physical structure of the dynamical system. Figure 5 demonstrates that the latent representations capture both continuous physical variation (permeability values) and discrete topological complexity (number of forcing blobs) in the Darcy Flow problem - a property absent in standard FNO backbones.

Spectral error profile. We compare frequency-resolved RMSE (fRMSE) between vanilla FNO and LR^2-FNO on the circle domain, aggregated over low-, mid- and high-frequency bands:

G Extended GAOT Experiments

We assess LR^2-FNO across five complementary experiments on the GAOT benchmark, probing predictive accuracy, internal representations, resolution invariance, noise resilience and shape generalization. Each experiment evaluates on steady-state elliptic PDE solutions across six 2D geometries: circle, cone, three ellipses (AR 1.5, 2.0, 2.5) and semicircle. Each sample encodes the operator $\mathcal{G} : c(x, y) \mapsto u(x, y)$ solving

$$-\nabla \cdot \big(a(x)\, \nabla u(x)\big) = f(x) \quad \text{in } \Omega \subset \mathbb{R}^2.$$

Raw data consists of unstructured point clouds interpolated to 64×64 grids via linear interpolation with 10% padding and standard normalization.

G.1 Exp. 1: In-Domain Steady BVPs

G.2 Exp. 2: Latent & Spectral Analysis

The t-SNE visualizations on the circle domain reveal that LR^2-FNO's latent representations cluster according to physical parameters (e.g., forcing magnitudes), whereas vanilla FNO's latents show

Dataset	FNO	LR^2-FNO	Improvement (%)
Circle (Poisson)	0.1154	**0.1088**	+5.7
Cone (Heat)	0.0633	**0.0585**	+7.6
Ellipse AR 1.5	0.0250	**0.0134**	+46.3
Ellipse AR 2.0	0.1251	**0.1082**	+13.5
Ellipse AR 2.5	0.0251	**0.0151**	+39.8
Semicircle BVP	0.0564	**0.0466**	+17.4
Average			**+21.7**

Table 6: **In-domain steady BVP results.** Relative L^2 errors (100 epochs, ≤ 200 samples/domain). LR^2-FNO achieves 21.7% average improvement, peaking at 46.3% on ellipse AR 1.5 as shown in Figure 6.

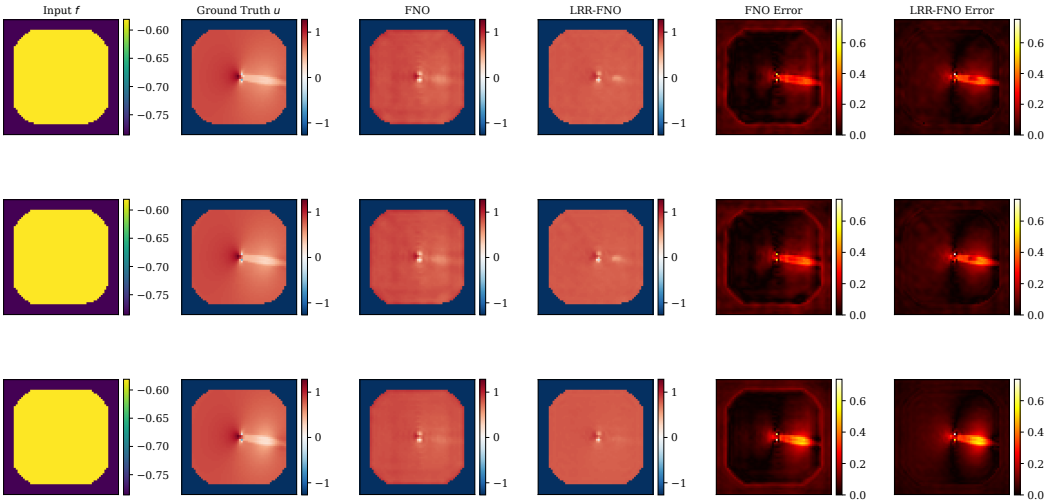


Figure 6: **Qualitative comparison on circle domain (Exp. 1).** Left to right: input source term, ground truth, FNO prediction, LR^2-FNO prediction and absolute error maps. LR^2-FNO demonstrates sharper boundary adherence and lower error magnitudes.

no such organization. This confirms that the InfoNCE loss successfully imposes physics-aware structure on the bottleneck features. Quantitative spectral results are reported in Table 5.

Dataset	Samples	Geometry	PDE Type
Circle.nc	423	Circle	Poisson
Cone-F.nc	559	Cone	Heat conduction
Ellipse-1.nc	574	Ellipse (AR 1.5)	Poisson
Ellipse-2.nc	435	Ellipse (AR 2.0)	Poisson
Ellipse-3.nc	594	Ellipse (AR 2.5)	Poisson
Semicircle-F.nc	471	Semicircle	Forced BVP

Table 7: **GAOT dataset summary.** All datasets contain 3 input features and are provided as unstructured clouds of 14,000 points.

G.3 Exp. 3: Zero-Shot Super-Resolution

Training exclusively at 32×32, LR^2-FNO generalizes to 64×64 evaluation with 29% lower error than FNO, whose degradation balloons from spectral undersampling as shown in Figure 7. The latent bridge acts as a resolution-agnostic regularizer, anchoring predictions to physics over discretisation artefacts.

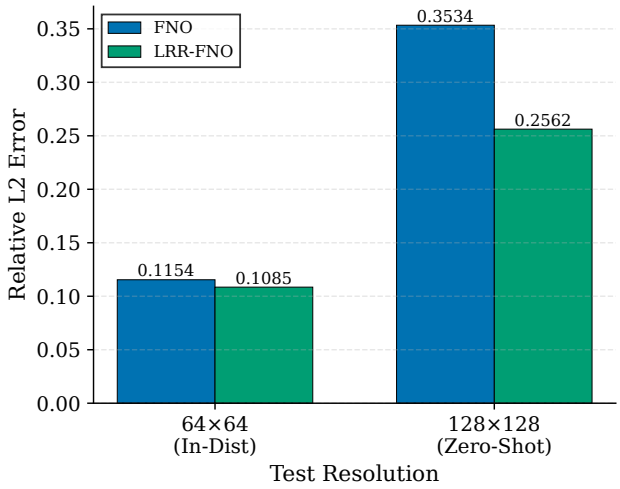


Figure 7: **Zero-shot super-resolution** (32×32 train $\rightarrow 64 \times 64$ eval). LR^2-FNO (green) holds L^2 degradation to 134% vs. FNO's 194% (blue); the 29% super-resolution gain reflects latent anchoring to resolution-independent physics.

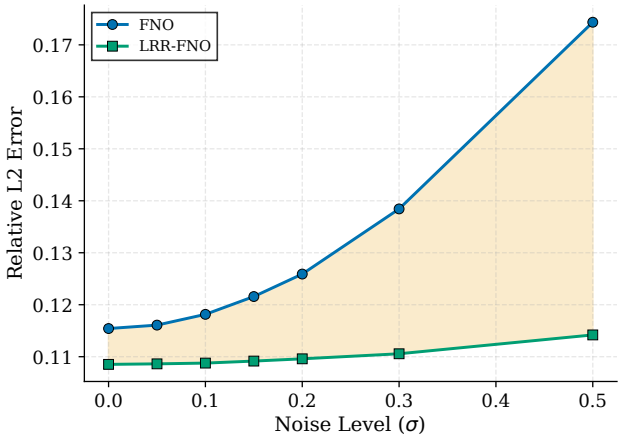


Figure 8: **Noise robustness on circle domain.** FNO error climbs steeply (blue) under increasing input noise, while LR^2-FNO stays nearly flat (green) - gains grow to 34% at $\sigma = 0.5$. Reciprocity denoises via solution manifold projection.

G.4 Exp. 4: Noise Robustness

Adding Gaussian noise ($\sigma = 0$ - 0.5, scaled by input std) to forcing inputs at test time, LR^2-FNO's error rises only 5.4% while FNO's jumps 42.4%. At $\sigma = 0.5$, LR^2-FNO achieves 34% lower error than FNO as shown in Figure 8. The reciprocity mechanism projects corrupted inputs onto the clean solution manifold learned via InfoNCE, naturally denoising high-frequency noise without explicit data augmentation.

G.5 Exp. 5: OOD Geometry Generalization

Circle-trained models evaluated on unseen geometries: LR^2-FNO averages 20.5% L_2 improvement (peak 39% on AR 2.5 ellipse, as shown in Figure 9, where FNO collapses to geometry-specific artifacts. Latent supervision during training breaks shape overfitting, producing a more universal PDE operator.

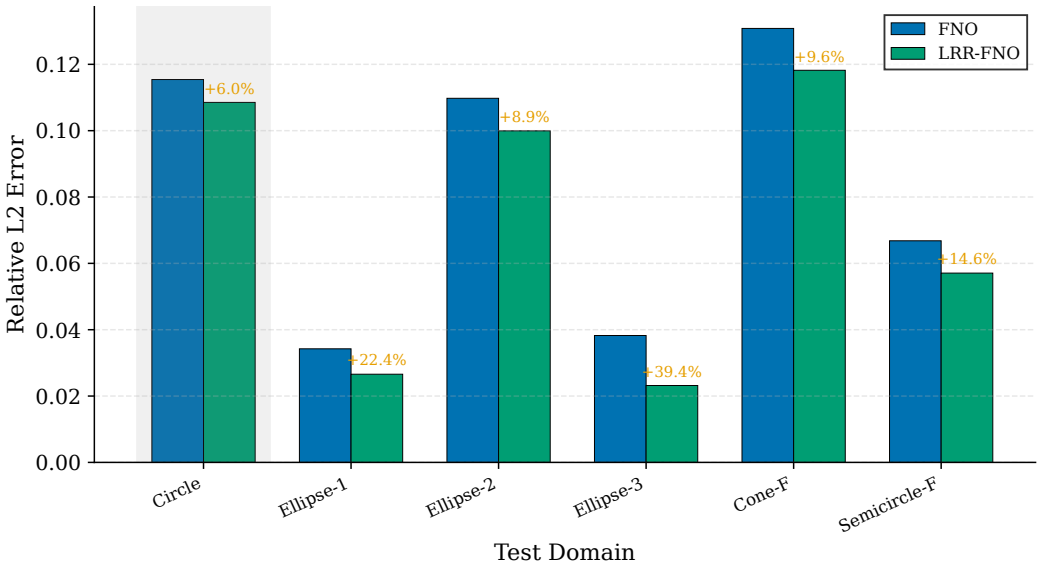


Figure 9: **OOD geometry generalization.** Circle-trained LR^2-FNO (green) consistently outperforms FNO (blue) by 8 - 39% across unseen geometries. Latent reciprocity breaks geometry overfitting, yielding robust cross-domain transfer.

Summary. Across all five experiments, LR^2-FNO's advantage stems from mid-network feature correction - FNO matches pixels, but reciprocity matches underlying structure.