The Operator Atlas: Learning the Geometry of PDE Solution Manifolds

Anonymous authors
Paper under double-blind review

Abstract

Existing PDE learning methods treat each equation in isolation and lack a representation that captures relations between equations. We introduce STAR, a contrastive encoder trained on the local differential features of solution fields. Three objectives jointly shape the latent space: a contrastive loss recovers the topological relations of the solution manifold, a reconstruction loss makes the coordinates physically decodable, and a coefficient loss anchors the shared representation to physical parameters. Trained on only three equation families, the model produces an Operator Atlas over twelve families in a train-free manner. The atlas organizes families by solution morphology, unfolds within-family parameters into continuous axes, and reveals a differential core shared by all twelve families (p ¡ 0.001) together with two cross-family relations: Poisson and Biharmonic, which differ in order, are mutual nearest neighbours under both metrics, and Helmholtz and mKdV are mutual nearest neighbours under centroid distance. Once frozen, the atlas solves four forward and inverse problems without per-query training: geodesic interpolation reconstructs unseen fields (median L2 of 1.6×10^{-2}), patch voting identifies the equation family (0.889), manifold position infers parameters (absolute error 0.081), and departure from the manifold detects anomalies (AUROC 0.811). The first three tasks are further paired with ground-truth-free uncertainty metrics, enabling reliable deployment without labelled validation data.

1 Introduction

Deep learning for partial differential equations (PDEs) follows two main paradigms, and both generalize poorly across equation families: physics-informed neural networks (PINNs) handle one equation instance at a time (Raissi et al., 2019), and neural operators, represented by FNO, fit a single equation family (Li et al., 2021); neither provides a representation that captures relations between equations. A neglected theoretical line closes this gap: Hadamard's well-posedness theory shows that a parametrized PDE family carries a low-dimensional solution manifold in function space (Hadamard, 1923), spectral graph theory shows that the graph Laplacian converges to the Laplace–Beltrami operator on a Riemannian manifold (Belkin & Niyogi, 2006), and HaoChen et al. (HaoChen et al., 2021) further prove that minimizing InfoNCE is equivalent to the spectral decomposition of the graph Laplacian. Together they give a previously unexplored route: if the solutions of a PDE form a low-dimensional manifold and InfoNCE performs its spectral embedding, then contrastive learning can in principle recover the intrinsic geometry of the solution space.

The representations learned by InfoNCE, however, are uniformly distributed on a highly symmetric hypersphere: their coordinates carry no physical meaning and the representation is uninterpretable. We therefore introduce STAR: a contrastive loss decides whether local differential features come from the same equation family and thereby learns the topological relations between equations; a reconstruction loss binds the latent coordinates to the solution field and makes the representation readable; and a coefficient loss anchors the shared representation to the physical parameters. Trained on only three equation families, STAR produces an Operator Atlas over twelve families in a train-free manner: families are organized by solution morphology, within-family parameters unfold into continuous axes, and

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

cross-family correspondences are preserved, rather than scattering randomly (all 66 pairwise angles stay below 11.23°, against a random baseline of 62.74°).

The atlas contains phenomena that agree with classical theory — the degeneration intersection of Laplace and Poisson, the three energy regions of Allen–Cahn, and the continuous manifold that Helmholtz forms beyond $k > 2\pi$ while staying spectrally aligned out of distribution — and phenomena beyond it: Poisson and Biharmonic are mutual nearest neighbours, Helmholtz and mKdV are adjacent, and all twelve families share a three-dimensional differential core ($p < 0.001$).

The frozen atlas reconstructs solutions under unseen coefficients by geodesic interpolation with no additional training; by contrast, after 3000 epochs on the test samples, PINN and FNO still fail on 16 of 36 and 22 of 36 instances respectively to reach STAR's train-free accuracy. The atlas also supports family identification (0.889), parameter inference (median absolute error 0.081) and anomaly detection (AUROC 0.811), and it gives prior uncertainty for the forward and inverse problems (up to $\rho = 0.954$).

The central contribution of this paper is to propose and establish that the geometry of PDE solution manifolds can be learned, verified, and used.

1. **STAR and the Operator Atlas.** A contrastive encoder trained on only three equation families produces a latent space that organizes twelve families by solution morphology and physical parameters, and a shared differential core and cross-family correspondences emerge.
2. **How the three objectives divide their work.** Each loss constrains only the quantity it reads: the contrastive loss reads the coordinates and shapes the topology, the reconstruction loss reads the coordinates and field values and makes them decodable, and the coefficient loss reads the shared representation and anchors the physical parameters (Appendix E.2).
3. **A train-free forward and inverse problem framework.** The frozen atlas solves forward and inverse problems without retraining, with an uncertainty metric for the first three tasks (anomaly detection has none; see Section 6).

2 Related Work

Generalization failure and its mitigation strategies. PINN surveys (Cuomo et al., 2022) and neural-operator reviews (Azizzadenesheli et al., 2024) document recurring generalization failures. We trace these failures to a shared limitation: existing models presuppose the identity of the equation and do not represent relations between solutions. For a specialized model, it is therefore unclear what transfers; for a general model, it is unclear which component fails when generalization collapses.

Solution manifolds and spectral graph theory. Hadamard's well-posedness theory (Hadamard, 1923) establishes the continuous dependence of the solution on the parameters, so a parametrized PDE family carries a low-dimensional solution manifold; G"uneysu and Pflaum (Güneysu & Pflaum, 2017) give a rigorous geometric formulation, and classical reduced-order modelling (ROM, POD, RBM) rests on this hypothesis but relies on explicit basis functions. A second line comes from spectral graph theory: Belkin and Niyogi (Belkin & Niyogi, 2006) prove that the graph Laplacian converges to the Laplace–Beltrami operator and thus encodes the local geometry of the manifold; Burago et al. (Burago et al., 2015) and Lu (Lu, 2022) show that the intrinsic geometry of a manifold embedded in a high-dimensional space can in principle be recovered; Cheng and Wu (Cheng & Wu, 2022) give convergence rates; and HaoChen et al. (HaoChen et al., 2021) prove the equivalence between InfoNCE (van den Oord et al., 2018) and the graph Laplacian spectrum.

Contrastive learning in scientific machine learning. PICL (Lorsung & Barati Farimani, 2024) was the first to bring contrastive learning into PINNs, but its loss is prescribed a priori and lacks reconstruction and coefficient constraints, so its representation clusters like a point cloud; Lyu et al. (Lyu et al., 2023) use InfoNCE to learn Koopman representations of nonlinear control systems and observe a ring-shaped manifold in low-dimensional prob-

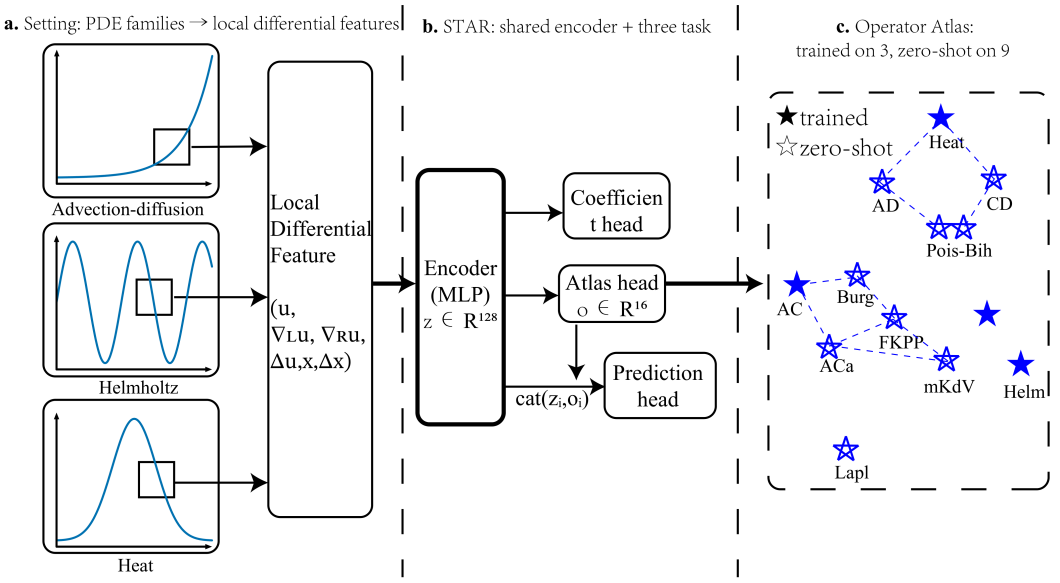

Figure 1: Overall architecture of STAR. A six-dimensional local differential feature is mapped by a shared encoder to a 128-dimensional representation, which then branches into three task heads. The atlas head produces the 16-dimensional atlas coordinate, the only output read by downstream tasks. The coefficient head predicts the sweep parameters. The prediction head reconstructs field values from the concatenated vector.

lems, which however degenerates into a scattered cloud in higher dimensions. Other work (Kumawat et al., 2024; Yu et al., 2025; Germain et al., 2026) neither examines nor demonstrates the geometry of the learned latent space, so it cannot verify that manifold learning occurs; methods that map PDE solutions to a finite-dimensional latent space (Bhattacharya et al., 2021; Fresca et al., 2021) impose structure through architectural inductive bias rather than learning manifold geometry from data.

3 Method

3.1 Overall Architecture

STAR is a contrastive encoder that maps the local differential features of PDE solution fields into a latent space. Three complementary objectives constrain that space: its topological structure, its alignment with physical parameters, and its physical decodability. The resulting latent space is the Operator Atlas. For a steady PDE solution field $u(x)$ discretized on a grid, the model constructs a six-dimensional local differential feature $\mathbf{x}_i \in \mathbb{R}^6$ at every interior point i (§3.2). A shared encoder E_ϕ maps it to a 128-dimensional representation:

$$\mathbf{z}_i = E_\phi(\mathbf{x}_i) \in \mathbb{R}^{128}. \tag{1}$$

The representation $\mathbf{z}_i$ is passed to three parallel task heads: the **atlas head** $g_{\text{atlas}} : \mathbb{R}^{128} \to \mathbb{R}^{16}$ outputs the atlas coordinate $\mathbf{o}_i$ (the only output the downstream tasks read), the **coefficient head** $g_{\text{coeff}} : \mathbb{R}^{128} \to \mathbb{R}^4$ predicts the sweep parameter $\hat{\boldsymbol{\theta}}_i$ of that patch, and the **prediction head** $g_{\text{pred}} : \mathbb{R}^{128} \times \mathbb{R}^{16} \to \mathbb{R}$ reconstructs the field value $\hat{u}_i$ from the concatenated vector $[\mathbf{z}_i; \mathbf{o}_i] \in \mathbb{R}^{144}$; the overall architecture is shown in Figure 1.

3.2 Local Differential Features

For an interior point $i \in \{1, \ldots, N-2\}$, the feature vector consists of the field value, the left first difference, the right first difference, the second difference, the spatial coordinate, and the grid spacing:

$$\mathbf{x}_i = \left[u_i,\ \frac{u_i - u_{i-1}}{\Delta x},\ \frac{u_{i+1} - u_i}{\Delta x},\ \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2},\ x_i,\ \Delta x \right]^\top. \tag{2}$$

3.3 Training Objectives

Coefficient regression supervises the coefficient head with the ground-truth parameters $\boldsymbol{\theta}$:

$$\mathcal{L}_{\text{coeff}} = \left\| \hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right\|_2^2. \tag{3}$$

The contrastive objective is an InfoNCE loss acting on the 16-dimensional atlas coordinates, with temperature $\tau = 0.1$. Positive samples come from fields of the same equation family, and negatives from other families.

$$\mathcal{L}_{\text{cont}} = -\log \frac{\exp(\text{sim}(\mathbf{q}, \mathbf{k}^+)/\tau)}{\sum_{j=1}^{K} \exp(\text{sim}(\mathbf{q}, \mathbf{k}_j)/\tau)}, \tag{4}$$

The prediction head reconstructs local field values, which keeps the coordinates physically decodable:

$$\mathcal{L}_{\text{pred}} = (\hat{u} - u)^2. \tag{5}$$

The total objective is the weighted sum of the three:

$$\mathcal{L} = \lambda_{\text{coeff}} \mathcal{L}_{\text{coeff}} + \lambda_{\text{cont}} \mathcal{L}_{\text{cont}} + \lambda_{\text{pred}} \mathcal{L}_{\text{pred}}. \tag{6}$$

3.4 Data and Training Setup

The training set consists of numerical solutions of three steady equation families: Helmholtz ($k \sim U[0.5, 8]$), the steady heat equation with random source configurations (the equation parameter is fixed; this family sweeps the source term), and advection–diffusion ($v \in [0.5, 3]$).The twelve equation families used for testing are listed in Appendix A.

Every instance yields its differential features from the solution on a grid, and 30 local patches are sampled per instance; each training family contributes 1,000 instances. Further implementation details (network structure, training hyperparameters, the composition of the test families, and the selection of solutions) are given in Appendix A.

4 Results

The Operator Atlas organizes the twelve PDE families into a markedly non-random geometric structure: the family subspaces are nearly parallel and together form a cone-like structure shared by all families, and each family moves smoothly on the cone as its parameter is swept. The geometry of this cone is the parametrization of the structure shared across families.

4.1 Expected Findings

The Poisson–Laplace degeneration intersection. As the source amplitude A of the Poisson equation tends to zero, its solutions degenerate to solutions of the Laplace equation.

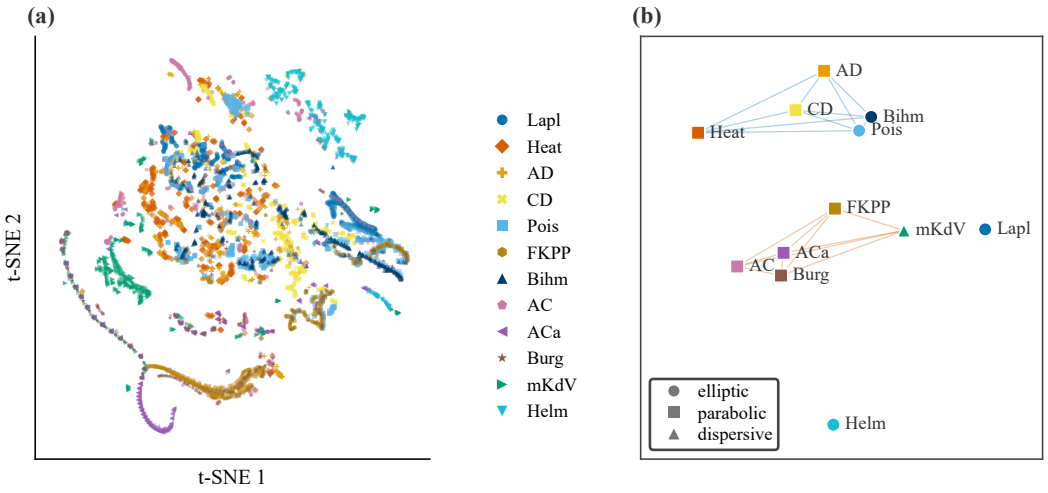

Figure 2: Geometry of the Operator Atlas over twelve equation families. a: joint t-SNE reduction of the latent coordinates of all families; b: MDS embedding derived from the de-cored residual angles between families.

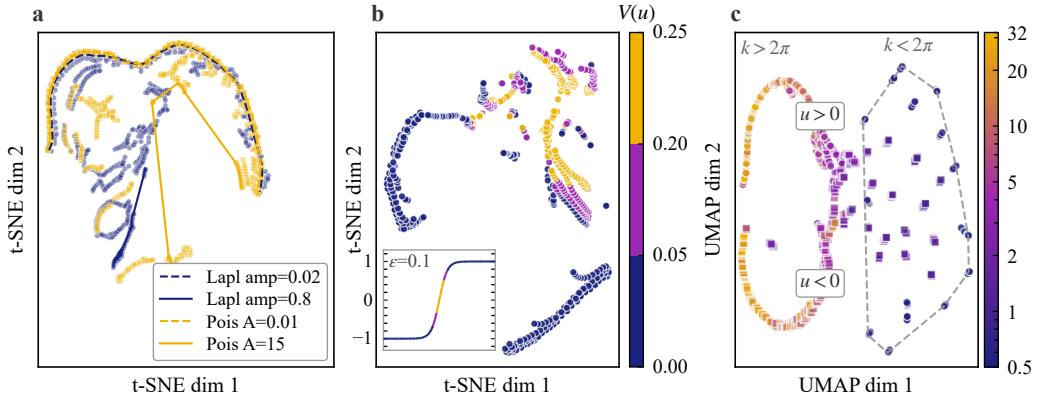

Figure 3: The three expected findings. (a) Poisson→Laplace degeneration intersection; (b) the energy partition of Allen–Cahn; (c) the periodic threshold of Helmholtz.

In the atlas this limit appears as a geometric intersection of the two manifolds: for small A the two manifolds nearly coincide (geodesic distance $d_g < 0.05$) and they separate gradually as A grows (Appendix B.13). **Energy organization of the Allen–Cahn manifold.** Within the Allen–Cahn family, high-energy and low-energy solutions each form an internally connected region. The spatial autocorrelation statistics are Moran's I of 0.93–0.95 and Geary's C of 0.05–0.08 ($z \approx -30$), against a permutation null whose expectations are 0 for Moran's I and 1 for Geary's C ($p \leq 0.0005$; the statistic and the permutation-test protocol are in Appendix B.2). **Periodic threshold of the Helmholtz manifold.** The Helmholtz equation forms a scattered point cloud for $k < 2\pi$ and two smooth symmetric manifolds for $k > 2\pi$; for k values outside the training distribution (OOD), the solutions still lie on these smooth manifolds (Appendix B.10).

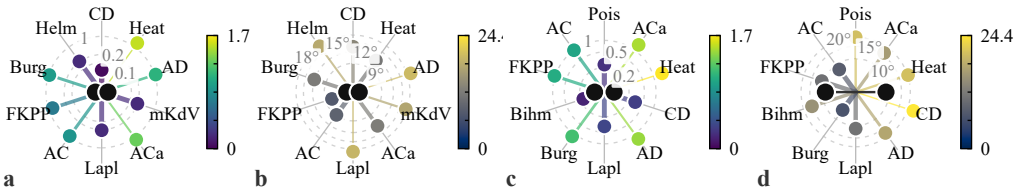

Figure 4: The four cross-family relations under two metrics (centroid distance / de-cored residual angle).

4.2 Unexpected Findings

4.2.1 A Shared Differential Core

In the atlas coordinates of the twelve equation families, we average the projection operators onto the dominant atlas subspaces to characterize **how many families share a direction on average**. The first three directions are shared by an average of **11.8, 11.6, and 10.8** families, and the spectrum drops sharply from the fourth, so **a three-dimensional subspace shared by nearly all families exists**. Its three-dimensionality is judged by a null distribution: substituting subspaces drawn at random in the same input space into the same statistic over 1,000 draws leaves the observed value outside the whole sample, giving $p < 0.001$, while an untrained encoder has only one marginal one-dimensional weak core. The three coordinates are mutually fully coherent ($|\rho| = 1.00$), with identical feature loadings $\pm(0.57, -0.58, -0.98)$ on $\{g_L, g_R, \text{lap}\}$: they read **local differential structure** rather than equation identity or parameters, which is why we name it the differential core. Trained on only 3 equation families, this core is shared by all 12 (Appendix B.9).

4.2.2 Two Cross-Equation Relations

We characterize cross-family relations with two complementary metrics: **centroid distance** for the similarity of solutions (position) and the **de-cored residual angle** for parameter alignment (direction). Beyond the relations expected from degeneration limits (Laplace–Poisson, advection–diffusion–heat), two further relations appear (Figure 4). **Poisson and Biharmonic:** a second- and a fourth-order elliptic operator should occupy different solution subspaces, yet the atlas places each as the other's nearest neighbour under both metrics. **Helmholtz and mKdV:** mutual nearest neighbours under centroid distance, but not close under the de-cored residual angle. Both relations are stable across seeds, and a POD basis of the input features and the Green's-function kernel cannot reproduce them (Appendix B.13).

5 Ablations and Robustness

The metrics used in this section are as follows. **Two decision criteria.** First, **consistency of the degeneration intersection** (d_{g1}): whether the degeneration intersection of the Poisson and Laplace families at the low-parameter end is jointly represented, with the criterion $0 < d_{g1} < 0.05$. Second, **integrity of the shared structure** (λ_{top}): whether the shared compressed structure is destroyed. If either criterion fails, the variant is considered not to preserve the atlas structure. **Three continuous metrics.** Together these characterize the overall quality of the atlas; all lie in $[0, 1]$, are better when higher, and each has a healthy band: **spread** H_d, how widely the families spread in the atlas; **inter-family separation** S_{inter}; and **family purity**, whether points of the same family are not confused with one another, higher being better. The exact definitions and computation of these metrics are given in the appendix.

Loss	d_{g1} (pass if $0 < d_{g1} < 0.05$)	λ_{top}
InfoNCE (reference)	0.0091 ✓	0.977
VICReg	**0.0713** ×	**0.726**
margin	0.0028 ✓	0.927
triplet	**0.0000**	1.000
CosFace	0.0011 ✓	0.952
ProtoNCE	0.0117 ✓	0.923

Table 1: Comparison of the six contrastive losses. Values at or near $d_{g1} = 0$ indicate complete overlap of the two families and are treated as failures.

Variant	Change relative to the reference	d_{g1} (pass if $0 < d_{g1} < 0.05$)	λ_{top}
Baseline	three heads (reference)	0.009 ✓	0.977
noPred	remove the prediction head and its loss	0.024 ✓	0.987
atlasOnly	contrastive objective only	0.002 ✓	0.987
predDetach	block the atlas→prediction gradient	0.014 ✓	0.980
coeffFromAtlas	decode coefficients from the atlas	0.009 ✓	0.956
catFuse	feed the coefficient estimate into the prediction head	**0.058** ×	0.972

Table 2: The three-head loss ablation.

5.1 Ablating the Contrastive Loss: Does InfoNCE Approach the Manifold Structure?

The theoretical basis of STAR is the spectral equivalence between InfoNCE and the graph Laplacian (Section 2). If this equivalence is the source of the atlas geometry, then replacing the contrastive loss should directly change the final geometry. We therefore replace InfoNCE with five alternative contrastive losses and compare the geometric metrics of the resulting atlases; the results are given in Table 2, (Appendix C.1). VICReg loses the intersection structure and its core spectrum degenerates; triplet shows a global structural degeneration; margin, CosFace and ProtoNCE retain part of the structure. Losses without the spectral equivalence do not reproduce the same atlas geometry and show different forms of structural degradation.

5.2 Division of Work among the Three Head Losses

Table 3 gives the variants of the three-head ablation. Three variants (predOnly, coeffOnly, noCont) have a placeholder vector as their atlas output and are therefore not listed in the table (Appendix C.2). The division of work among the three heads is indispensable.

We compare the atlases obtained from different training seeds using the **normalized Procrustes residual after optimal rotational alignment** (res_rot $= \|Z_1 - Z_2 R^*\| / \|Z_1\|$): a residual of 0 means the two solutions differ by exactly a rotation, and larger values mean more difference remains after alignment (Appendix E.2). Averaging over all pairs of four seeds, six pairs per variant, pure contrastive learning (atlasOnly) gives 1.05 ± 0.82; adding the reconstruction loss raises it to 1.73 ± 1.11; adding the coefficient loss brings it back to 0.71 ± 0.11; and the randomly permuted-label control (P_{rand}) gives 0.84 ± 0.28. The rotation residual exceeds 0.7 for every variant, so no uniform rotation reconciles the seeds' final states (Appendix E.2).

5.3 Hyperparameter Stability

This section examines twelve hyperparameter axes with a further metric, **orderedness**: how continuously the solutions within a family change with the parameter while different families remain separable; higher is better (Appendix C.4.2, C.0).

Conclusion. Among the twelve axes, **batch size** has the largest impact: it is the only axis that breaks the structure clearly, in two variants, and it also shows the largest drop in orderedness. Next come **temperature** and the **coefficient loss weight** — the two axes

with the most criterion failures, the latter breaking clearly in its zero-weight setting as well — together with the **learning rate**, which does not break the structure but noticeably weakens intra-family continuity and inter-family separability. The smallest impact comes from a group of data- and scale-level axes (random seed, number of patches, instances per family, sampling density): their criteria all pass and their orderedness is essentially unchanged. We stress that even on the most affected axes λ_{top} remains far above the criterion for collapse of the shared structure: what is affected is always a **secondary property** such as intra-family continuity or inter-family separability, while the shared compressed structure itself is stable on every axis.

5.4 Extension to Higher Dimensions and Time

To test whether the atlas structure transfers across settings, we extend STAR in two directions: a time dimension and a second spatial dimension. Neither changes the input descriptor convention or the contrastive task (Appendix F). The two-dimensional extension runs on 8 equation families (3 training, 5 never seen). Leave-one-out identification reaches 85.0%, far above the random baseline of 12.5%, with no indistinguishable pair; a model-free control shows the solution fields themselves are already 92.5% separable, so this is not the model certifying itself. The median inter-family angle is 6.5° (within-family 1.5°), far below the 31.2° of a random cloud. The shared core persists: its **family-mean membership** drops from 0.55–0.59 in 1D to 0.29–0.40, about two thirds, and removing it raises the within-to-between contrast from 4.28× to 5.26× and the identification rate to 87.5%. The atlas is ordered along time ($\text{isom}_t = 0.894$, above 0.856 without the time input), and equation separability forms an arrow: early snapshots are dominated by the initial condition and the families are similar, then separate, so time supervision redistributes separability from early to late (crossover at $t \approx 0.27$). The negative result is equally clear: the cross-equation geometry does not transfer fully in time.

6 Train-Free Utility

The frozen Operator Atlas answers four kinds of questions. **Reconstruction**: on 36 test instances (3 per family, parameters absent from the test instances) the train-free median L_2 error of STAR is 1.6×10^{-2}; PINN, trained instance by instance, and FNO, trained on the test equation instances for at most 3000 epochs, fail to reach the target on 16 and 22 of them respectively, and need on average about 1.5×10^3 and 1.9–2.1×10^3 epochs to match the same-instance accuracy; the uncertainty metric is the perpendicular distance to the interpolation chord ($\rho = 0.954$). **Identification**: atlas voting reaches 0.889 at the full budget and 0.750 with only 4 patches, against the strongest baseline on the same features, KNN at 0.694; the uncertainty metric is the top-1 cosine similarity ($\rho = 0.289$). **Inference**: the parameter is read from the position geometry in the atlas, with a median absolute error of 0.081 and 47% of instances within 5% of the ground truth, better than the Gaussian-process baseline (8.58); the uncertainty metric is the dispersion ($\rho = 0.771$). **Anomaly detection**: the fraction of patches outside the parameter segment gives an AUROC of 0.811 over all anomaly types (0.978 noise, 0.901 degeneration, 0.682 extrapolation) against the strongest baseline, Isolation Forest (Liu et al., 2008), at 0.622, and the conclusion holds under heavy contamination (0.764 ± 0.221, 0.803 ± 0.095, 0.809 ± 0.048 at $p = 5\%/20\%/50\%$); this task has no uncertainty metric. Methods and per-task baselines are in Appendix D.1–D.4.

7 Discussion

7.1 Conclusion

This paper introduces STAR: a contrastive encoder trained on only three equation families that embeds the local differential features of PDE solution fields into a 16-dimensional latent space and produces an Operator Atlas covering twelve equation families. The spectral equivalence between InfoNCE and the graph Laplacian provides the theoretical basis, and the ablations show that each of the three losses constrains only the quantity it reads (Section 5.2). The atlas shows a markedly non-random geometric organization: within a family the

parameters unfold into continuous axes, and between families the relations both reproduce structures expected from classical theory — the Poisson–Laplace degeneration intersection, the Allen–Cahn energy partition, the Helmholtz periodic threshold — and reveal unexpected ones, such as a differential core shared by all twelve families ($p < 0.001$) and the mutual proximity of Poisson with Biharmonic and of Helmholtz with mKdV. Its limitations are that the local differential features rely on a regular grid and that the evaluation partly relies on geometric metrics defined in this paper (Section 7.3).

7.2 Open Questions

Open question 1: relation-preserving finite-dimensional embedding. The solution manifold of each PDE family is an infinite-dimensional object, yet the relations between families are faithfully represented in a 16-dimensional Operator Atlas. Under what conditions does a family of PDE solution manifolds admit an embedding into a finite-dimensional space that preserves the relations between families? Classical embedding theory guarantees only the finite-dimensional embedding of a single manifold (Whitney, 1936; Nash, 1956); no result guarantees that a *family* of manifolds can be embedded as a whole together with its relations.

Open question 2: an intrinsic classification principle for operators. The classical elliptic–parabolic–hyperbolic trichotomy is not reproduced by the measured geometry ($R^2 \approx 0.14$). Does an intrinsic classification principle for operators exist in the geometry of the solution manifold?

7.3 Limitations

First, the local differential features come from finite differences on a regular grid: the resolution is limited by grid density, the feature count grows with the spatial dimension, and non-uniform grids are not covered. Second, the evaluation partly relies on geometric metrics defined in this paper; model-free controls (the solution fields alone separate at 92.5%), Monte-Carlo baselines and permutation tests reinforce them, but broader third-party benchmarks are missing. Third, relation values do not transfer: the core's family-mean membership drops to about two thirds in two dimensions, and the cross-equation geometry does not transfer fully in time. Future work will extend the scope to non-uniform and adaptive grids, higher-dimensional families, calibration of the anomaly-detection confidence, standard cross-family metrics, combining condition numbers with PINNs, and aligning equation forms with CLIP.

AI use statement

In this work, we used generative AI tools for the following required-disclosure tasks: writing and debugging analysis and plotting code, assisting with the LaTeX typesetting of the manuscript, and language editing and translation of the text. We have not used generative AI tools to generate research ideas, hypotheses, experimental designs, or results; no data, measurements, or numerical results in this paper were produced by AI tools, and all experiments were run and verified by the authors. Additionally, we used generative AI tools for literature-search assistance. We have reviewed all AI-assisted work: every claim, number, and citation in this paper was checked against the underlying experiment records by the authors, AI-assisted code was tested for correctness, and AI-assisted text was edited and approved by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI. See the ICLR 2027 AI Policy for Authors for more details.

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*, 6:320–328, 2024. doi: 10.1038/s42254-024-00712-5. URL `https://doi.org/10.1038/s42254-024-00712-5`.

Mikhail Belkin and Partha Niyogi. Convergence of Laplacian eigenmaps. In *Advances in Neural Information Processing Systems*, volume 19, pp. 129–136, 2006.

Kaushik Bhattacharya, Bamdad Hosseini, Nikola B. Kovachki, and Andrew M. Stuart. Model reduction and neural networks for parametric PDEs. *The SMAI Journal of Computational Mathematics*, 7:121–157, 2021. arXiv:2005.03180.

Dmitri Burago, Sergei Ivanov, and Yaroslav Kurylev. A graph discretization of the Laplace–Beltrami operator. *Journal of Spectral Theory*, 4(4):675–714, 2015. arXiv:1301.2222.

Xiuyuan Cheng and Nan Wu. Eigen-convergence of Gaussian kernelized graph Laplacian by manifold heat interpolation. *Applied and Computational Harmonic Analysis*, 61:132–190, 2022. arXiv:2010.06391.

Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics-informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92:88, 2022. doi: 10.1007/s10915-022-01939-z. URL https://doi.org/10.1007/s10915-022-01939-z.

Stefania Fresca, Ludovica Dede', and Andrea Manzoni. A comprehensive deep learning-based approach to reduced order modeling of nonlinear time-dependent parametrized PDEs. *Journal of Scientific Computing*, 87(2):Article 61, 36 pages, 2021. arXiv:2001.04001.

Thibaut Germain, Sami Chemlal, Rémi Flamary, Vladimir R. Kostic, and Karim Lounici. Geometric dictionary learning of dynamical systems with optimal transport. *arXiv preprint arXiv:2605.18276*, 2026. URL https://arxiv.org/abs/2605.18276.

Batu Güneysu and Markus J. Pflaum. The profinite dimensional manifold structure of formal solution spaces of formally integrable PDEs. *Symmetry, Integrability and Geometry: Methods and Applications*, 13:Article 003, 44 pages, 2017. doi: 10.3842/SIGMA.2017.003.

Jacques Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Silliman Memorial Lectures. Yale University Press, New Haven, CT, 1923.

Jeff Z. HaoChen, Colin Wei, Adrien Gaidon, and Tengyu Ma. Provable guarantees for self-supervised deep learning with spectral contrastive loss. In *Advances in Neural Information Processing Systems*, volume 34, pp. 5000–5011, 2021. arXiv:2106.04156.

Hemant Kumawat, Biswadeep Chakraborty, and Saibal Mukhopadhyay. RoboKoop: Efficient control conditioned representations from visual input in robotics using Koopman operator. *arXiv preprint arXiv:2409.03107*, 2024. URL https://arxiv.org/abs/2409.03107.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO. arXiv:2010.08895.

Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *Proceedings of the 2008 Eighth IEEE International Conference on Data Mining*, pp. 413–422, 2008. doi: 10.1109/ICDM.2008.17. URL https://doi.org/10.1109/ICDM.2008.17.

Cooper Lorsung and Amir Barati Farimani. PICL: Physics informed contrastive learning for partial differential equations. *APL Machine Learning*, 2(4):046107, 2024. doi: 10.1063/5.0223651. arXiv:2401.16327.

Jinpeng Lu. Graph approximations to the Laplacian spectra. *Journal of Topology and Analysis*, 14(1):111–145, 2022. doi: 10.1142/S1793525320500442. arXiv:1910.09224.

Xubo Lyu, Hanyang Hu, Seth Siriya, Ye Pu, and Mo Chen. Task-oriented Koopman-based control with contrastive encoder. In *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pp. 93–105, 2023. arXiv:2309.16077.

John Nash. The imbedding problem for Riemannian manifolds. *Annals of Mathematics*, 63 (1):20–63, 1956. doi: 10.2307/1969989.

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. doi: 10.1016/j.jcp.2018.10.045. URL https://doi.org/10.1016/j.jcp.2018.10.045.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. URL https://arxiv.org/abs/1807.03748.

Hassler Whitney. Differentiable manifolds. *Annals of Mathematics*, 37(3):645–680, 1936. doi: 10.2307/1968482.

Xiaohang Yu, Hui Miao, Chen Pang, and Lei Lyu. Adaptive Koopman contrastive learning for skeleton-based action recognition. *Neurocomputing*, 658:131750, 2025. doi: 10.1016/j.neucom.2025.131750. URL https://doi.org/10.1016/j.neucom.2025.131750.