Independent Latents, Robust Neural Operators

Anonymous Author(s)
Affiliation
Address
email

Abstract

Neural operators have emerged as powerful surrogates for scientific computing, enabling rapid reconstruction of complex physical fields from sparse sensor observations. Despite their strong predictive capability, their reliability often degrades under noisy or corrupted measurements, limiting practical deployment in real-world sensing environments. We show that enforcing statistical independence in latent representations provides a simple yet effective mechanism for improving neural operator robustness. To this end, we introduce Neural Operator Independence Regularization (NOIR), an end-to-end training framework that promotes statistically independent latent features during operator learning. Unlike structured basis approaches such as POD-DeepONet and PCA-Net, which impose orthogonality primarily as an offline compression constraint, NOIR directly shapes latent representations during training to enhance resilience against input perturbations. Across three canonical PDE benchmarks and five additive sensor corruption settings, NOIR consistently improves reconstruction accuracy while preserving clean-data performance. Beyond predictive gains, the induced latent representations exhibit distinctive statistical signatures that reveal architecture-specific robustness characteristics under corruption. These findings identify latent statistical independence as a principled inductive bias for robust neural operator design and provide new insight into the internal structure of operator learning systems. All models, datasets, checkpoints, and training code are publicly available at `https://github.com/t55176853-cmyk/NOIR/`

1 Introduction

Neural operators have emerged as a powerful paradigm for scientific machine learning, enabling direct approximation of mappings between infinite-dimensional function spaces and providing scalable surrogates for complex physical systems. Grounded in operator-theoretic foundations, these models offer principled architectures for learning discretization-invariant representations of parametric partial differential equations. Their ability to reconstruct high-dimensional physical fields from sparse sensor measurements has enabled transformative advances across fluid dynamics, structural mechanics, climate modeling, and other data-intensive scientific applications where conventional numerical solvers remain computationally prohibitive. Yet this promise comes with a critical challenge: practical sensing environments are inherently imperfect, with measurements often corrupted by noise, drift, dropout, or instrumentation degradation. Under such conditions, neural operators may produce physically plausible yet erroneous predictions, exposing a fundamental limitation of current black-box operator learning systems and raising important questions about their reliability in scientific and safety-critical deployment.

Despite rapid progress in operator learning, existing neural operators remain largely optimized and evaluated as end-to-end input-output mappings, offering limited visibility into the internal representations that govern their predictions. This opacity becomes especially problematic under

Submitted to 40th Conference on Neural Information Processing Systems (NeurIPS 2026). Do not distribute.

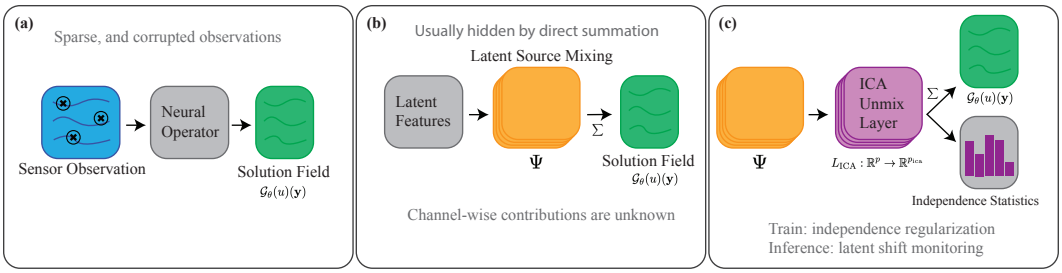


Figure 1: **Schematic of the NOIR framework for robust operator learning and latent monitoring. (a) Operator learning under corrupted observations:** The neural operator maps sparse, potentially corrupted sensor observations to a predicted solution field $G_\theta(u)(y)$. **(b) Hidden latent channel contributions:** In standard implementations, the latent contribution tensor Ψ is directly aggregated (summed) to produce the output, effectively entangling the individual feature contributions. **(c) NOIR unmixing, independence, and monitoring:** Our proposed method introduces a learnable Independent Component Analysis (ICA) unmixing layer to disentangle Ψ into independent source channels S. During training, an independence regularization term is applied. At inference, shifts in the statistics of these independent sources are utilized to generate a corruption score, enabling real-time monitoring of latent distribution shifts and data integrity.

non-ideal sensing conditions, where corrupted measurements may induce subtle degradation in latent physics encoding long before predictive errors become obvious at the output level. Existing approaches to robustness typically focus either on improving reconstruction accuracy under noise or on post hoc anomaly detection mechanisms that operate outside the operator-learning process. Such strategies provide limited insight into how corruption propagates through learned function-space representations and often introduce a tradeoff between task performance and detectability of sensor degradation. These limitations raise a fundamental question: *can neural operators be endowed with internal structural organization that not only improves robustness, but also provides intrinsic signals of when their inputs deviate from the conditions under which they were trained?*

We posit that statistical independence provides a principled structural prior for robust operator learning. In conventional neural operators, latent representations are learned solely to optimize predictive accuracy, often producing entangled feature spaces in which physically meaningful dynamics and sensor-induced perturbations are inseparably mixed. Such entanglement obscures the internal effects of corrupted measurements and limits the model's ability to distinguish genuine changes in system behavior from sensing failures. By encouraging statistical independence among latent features, the operator is forced to learn a more structured decomposition of the underlying physics, where distinct latent components encode complementary information. Under ideal sensing conditions, this latent organization remains stable; when inputs are corrupted, measurable violations of the learned independence emerge as intrinsic signatures of representation degradation. To operationalize this principle, we introduce Neural Operator Independence Regularization (NOIR), an end-to-end training framework that promotes statistically independent latent representations directly during operator learning. Inspired by negentropy-based independence estimation, NOIR integrates covariance regularization and non-Gaussianity promotion into a unified differentiable objective, eliminating the need for offline basis construction or post hoc anomaly detection modules. Unlike structured basis methods that impose orthogonality primarily for compression, NOIR learns task-adaptive independent latent features jointly with the operator itself, enabling both improved robustness and internal awareness of sensing corruption.

In this work, we evaluate NOIR across three canonical PDE benchmarks and multiple additive sensor corruption regimes spanning Gaussian perturbations, spectral distortions, and localized sensor failures. Our results show that enforcing latent statistical independence consistently improves reconstruction robustness while preserving, and in several cases enhancing, clean-data predictive performance. Beyond output-level accuracy, we demonstrate that NOIR induces distinctive latent statistical signatures that expose architecture-specific sensitivity to sensing corruption, offering a new lens for interpreting neural operator behavior under non-ideal inputs. These findings position latent independence not merely as a regularization strategy, but as a foundational design principle for

robust operator learning. Collectively, this work advances neural operators from purely predictive surrogates toward self-monitoring scientific models capable of internal awareness under uncertain sensing environments.

2 Related work

Operator learning. Neural operator learning has emerged as a powerful paradigm for learning mappings between infinite-dimensional function spaces, enabling discretization-invariant approximation of complex physical systems. Foundational architectures such as DeepONet and the Fourier Neural Operator (FNO) established operator learning as an effective framework for solving parametric partial differential equations, while subsequent developments including NOMAD, Wavelet Neural Operators, and physics-informed hybrid variants have further improved expressivity, scalability, and generalization across resolutions [3, 14, 27, 28, 18, 6]. Despite these advances, existing work has largely focused on predictive accuracy and computational efficiency, with comparatively limited attention given to the internal structure of learned latent representations and their sensitivity to corrupted sensing inputs.

Robustness under imperfect sensing. Reliable deployment of scientific machine learning systems requires resilience to imperfect sensing conditions, including measurement noise, drift, dropout, and instrumentation degradation. Prior work on sensor fault detection and mitigation has explored numerical compensation schemes, redundant sensing strategies, and explicit fault-detection algorithms [25, 7, 21, 29]. While effective in specific operational settings, these approaches often introduce substantial computational or hardware overhead and typically operate externally to the predictive model itself. In scientific machine learning, robustness is more commonly addressed through data augmentation, uncertainty quantification, or residual correction strategies, which improve output reliability but provide limited visibility into how corruption propagates through learned internal representations

Statistical independence and structured representation learning. Statistical independence has long played a central role in representation learning, particularly in independent component analysis and related latent-variable methods designed to recover disentangled explanatory factors from mixed observations [8, 20, 13, 19]. More recent work has shown that independence-promoting objectives can improve invariance, interpretability, and robustness in deep latent representations. However, these ideas have been explored primarily in unsupervised learning, generative modeling, and general-purpose representation learning. Their role as an inductive bias for operator learning – and specifically as a mechanism for enhancing robustness to corrupted scientific measurements – remains largely unexplored.

Structured basis methods in operator learning. Several operator-learning approaches impose latent structure through explicit basis decomposition. Methods such as POD-DeepONet, PCA-Net, and resolution-aware operator decompositions leverage orthogonal basis representations to improve compression and computational efficiency [18, 2, 1]. These methods demonstrate the value of structured latent spaces, but their primary objective is compact representation rather than robustness or corruption awareness. Moreover, they typically rely on offline basis construction stages that decouple latent structure learning from end-to-end operator optimization. In contrast, our approach enforces statistical independence directly during training, enabling task-adaptive latent organization that simultaneously improves perturbation resilience and exposes intrinsic indicators of sensing degradation.

Positioning of this work. This work bridges operator learning, robust scientific machine learning, and independence-based representation learning by introducing latent statistical independence as a unifying structural prior for neural operators. Unlike existing approaches that use latent structure primarily for compression or external anomaly detection, NOIR integrates independence regularization directly into end-to-end operator training, enabling robust prediction and internal monitoring within a single unified framework.

3 Methodology

3.1 Problem formulation

We consider operator learning from sparse and potentially corrupted sensor measurements. Let, $\mathcal{G} : \mathcal{U} \to \mathcal{S}$ denote the mapping from an input function $u \in \mathcal{U}$ to the corresponding solution field $s = \mathcal{G}(u) \in \mathcal{S}$. In practice, u is observed only at a finite set of sensor locations $X_s = \{\mathbf{x}_i\}_{i=1}^m$, yielding, $u_{X_s} = \mathcal{P}_{X_s}(u) = \big[u(\mathbf{x}_1), \ldots, u(\mathbf{x}_m)\big] \in \mathbb{R}^m$. At test time, these measurements are often corrupted. The objective is to learn a neural operator $\mathcal{G}_\theta$ such that

$$\mathcal{G}_\theta(\tilde{u}_{X_s})(\mathbf{y}) \approx \mathcal{G}(u)(\mathbf{y}), \quad \mathbf{y} \in \mathcal{Y}, \tag{1}$$

where $\tilde{u}_{X_s}$ represents the corrupted input. To this end, we seek to design an inductive bias, with the goal of improving robustness to input corruption without sacrificing predictive performance.

3.2 Latent entanglement in neural operators

Many neural operators admit a representation consistent with the linear representation hypothesis [22], where predictions arise from a nonlinear feature map followed by a linear readout:

$$\mathcal{G}_\theta(u)(\mathbf{y})_c = \sum_{k=1}^{p} \xi_{k,c}(u, \mathbf{y}), \tag{2}$$

i.e., an additive aggregation of latent channel contributions. For example, in DeepONet [17], with branch output $B \in \mathbb{R}^{N_b \times p}$ and trunk output $T \in \mathbb{R}^{P \times p \times C}$, the prediction is

$$\mathcal{G}_\theta(u_j)(\mathbf{y}_q)_c = \sum_{k=1}^{p} B_{jk} T_{qkc}. \tag{3}$$

Defining the source-mixture tensor, $\Psi_{jqkc} = B_{jk} T_{qkc}$ and $\mathcal{G}_\theta(u_j)(\mathbf{y}_q)_c = \sum_{k=1}^{p} \Psi_{jqkc}$ makes explicit the channel-wise contributions over the joint domain $\mathcal{U} \times \mathcal{Y}$. However, these latent channels are inherently non-identifiable. For any invertible $\mathbf{R} \in \mathbb{R}^{p \times p}$, $b'(u) = \mathbf{R} b(u)$, $t'(\mathbf{y}) = \mathbf{R}^{-\top} t(\mathbf{y})$, yields the same operator output [11]. This rotational ambiguity induces latent entanglement, whereby perturbations in corrupted inputs propagate across channels, degrading interpretability and robustness.

3.3 Statistical independence as a structural prior

To reduce cross-channel interference, we impose statistical independence as a structural prior on the latent representation. Let $\mathbf{s} = [s_1, \ldots, s_p]^\top$ denote latent variables. Statistical independence is defined as

$$\mathbb{P}(\mathbf{s}) = \prod_{k=1}^{p} \mathbb{P}_k(s_k). \tag{4}$$

This condition eliminates both second-order and higher-order dependencies, preventing latent channels from encoding redundant or entangled information. This is particularly important in operator learning, where input corruption may preserve low covariance structure while still inducing nonlinear coupling across latent dimensions. This induces a *diagonally dominant latent sensitivity structure*, which suppresses cross-channel propagation of perturbations in the operator's Jacobian. As a result, the effective sensitivity of the learned operator becomes approximately separable across latent channels. On the theoretical side, statistical independence yields a tighter upper bound on the Lipschitz constant of the operator and directly improves certified robustness under input perturbations. Theoretical analysis of the same are proved in Section 4.

3.4 Neural Operator Independence Regularization (NOIR)

NOIR enforces independence directly on the exposed latent component tensor before final aggregation. Figure 1 summarizes the framework. The key idea is to (i) expose the source-mixture tensor that is typically hidden inside the final summation, (ii) apply a learnable ICA-based unmixing layer to this tensor, and (iii) regularize the resulting source channels to be statistically independent.

Source-mixture tensor. Instead of collapsing Eq. (3) immediately, we retain the intermediate tensor Ψ defined previously. This tensor provides a channel-wise decomposition of the predicted solution across all input-query pairs. As a result, NOIR not only regularizes the latent representation used to construct the solution, but also enables direct attribution of each channel's contribution to the final operator output. **ICA-based unmixing layer.** We apply a learnable unmixing map, $L_{\mathrm{ICA}} : \mathbb{R}^p \to \mathbb{R}^{p_{\mathrm{ica}}}$ to the latent channel dimension of Ψ. After permuting Ψ to shape (N_b, P, C, p), the transformed source tensor is $\mathbf{S}_{4\mathrm{D}} = L_{\mathrm{ICA}}(\Psi) \in \mathbb{R}^{N_b \times P \times C \times p_{\mathrm{ica}}}$. The final operator output is then obtained by summing over the transformed source dimension:

$$\mathcal{G}_\theta(u_j)(\mathbf{y}_q)_c = \sum_{\ell=1}^{p_{\mathrm{ica}}} [\mathbf{S}_{4\mathrm{D}}]_{jqc\ell}. \tag{5}$$

Approximate whitening. For each output channel c, we extract the slice $\mathbf{S}^{(c)} \in \mathbb{R}^{N_b \times P \times p_{\mathrm{ica}}}$ and reshape it into $\mathbf{S}_{2\mathrm{D}}^{(c)} = \mathrm{reshape}\big(\mathbf{S}^{(c)}, (N, p_{\mathrm{ica}})\big)\,,\ N = N_b P$. To satisfy the assumptions of ICA during mini-batch training, we apply per-source standardization, $\hat{S}_{n\ell}^{(c)} = \frac{S_{n\ell}^{(c)} - \mu_\ell^{(c)}}{\sigma_\ell^{(c)} + \epsilon}$, where $\mu_\ell^{(c)}$ and $\sigma_\ell^{(c)}$ are the empirical mean and standard deviation of source ℓ in channel c, and ϵ is a small numerical constant. This approximates whitening under a diagonal covariance assumption, which is justified by the regularization itself and is substantially cheaper than eigendecomposition-based whitening.

Negentropy-based independence loss. Following FastICA [8, 9], we maximize non-Gaussianity of the whitened sources using the LogCosh contrast, $\mathcal{G}_{\mathrm{ICA}}(x) = \log(\cosh(x))$. For output channel c, the NOIR regularizer is

$$\mathcal{L}_{\mathrm{ICA}}^{(c)} = -\sum_{\ell=1}^{p_{\mathrm{ica}}} \left(\frac{1}{N} \sum_{n=1}^{N} \mathcal{G}_{\mathrm{ICA}}\left(\hat{S}_{n\ell}^{(c)}\right) - \mathbb{E}_\nu[\mathcal{G}_{\mathrm{ICA}}(\nu)] \right)^2, \tag{6}$$

where $\nu \sim \mathcal{N}(0,1)$. The total NOIR loss is $\mathcal{L}_{\mathrm{ICA}} = \frac{1}{C} \sum_{c=1}^{C} \mathcal{L}_{\mathrm{ICA}}^{(c)}$, and the complete training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{task}} + \lambda_{\mathrm{ICA}} \mathcal{L}_{\mathrm{ICA}}. \tag{7}$$

The full derivation, together with the relation to orthogonality-based formulations, is given in Appendix A.1 and Appendix A.2.

Latent corruption monitoring. At inference time, we monitor the learned independence structure of the latent sources. Let $m(\cdot)$ denote a summary statistic of the source distribution, such as mean negentropy across channels. We compare the test-time statistic against clean validation statistics using $Z = \frac{|m_{\mathrm{test}} - \mu_{\mathrm{clean}}|}{\sigma_{\mathrm{clean}}}$, where μ_{clean} and σ_{clean} are computed on clean validation data. High Z-scores indicate that corruption has disrupted the learned independence structure, providing an intrinsic confidence signal without requiring labeled failure examples. Further effect on this is discussed in Appendix A.3.

NOIR requires no offline basis construction, no precomputed decomposition, and no auxiliary post hoc detector. Its only requirement is access to the unsummed latent component tensor before the final aggregation. Considering DeepONet, the standardization and negentropy evaluation add $\mathcal{O}(N_b P C\, p_{\mathrm{ica}})$. If latent monitoring is disabled at test time, the extra linear mixing can be absorbed into the final aggregation with no change in asymptotic complexity. Importantly, *NOIR is broadly applicable to operator architectures with exposed latent component tensors*. Any neural operator with a final linear aggregation of latent features can expose an analogous component tensor and apply the same unmixing-and-regularization strategy. Full generalization is provided in Appendix A.4.

Hyperparameter selection. We selected p_{ica} and λ_{ICA} based on empirical results on a held-out training set (See Table 4 for detailed hyperparameter sensitivity and performance).

4 Theoretical analysis

We show that statistical independence in latent representations induces tighter Lipschitz bounds and improved certified robustness for neural operators (see Appendix A.5 for the full formal proof).

Assumption 4.1. [Decorrelation of Latent Sensitivities] Latent perturbations satisfy

$$\mathbb{E}[\delta_i^\top \delta_j] = 0 \quad (i \neq j), \qquad \mathbb{E}[\delta_k] = 0. \tag{8}$$

This assumption is induced in practice by enforcing statistical independence of latent variables.

Proposition 4.2 (Lipschitz Continuity of Latent Maps)**.** *Let each latent coefficient function $s_k(u_{X_s})$ be parameterized by a depth-L neural network of the form*

$$s_k(u) = W_L \sigma_{L-1}(W_{L-1} \cdots \sigma_1(W_1 u)), \tag{9}$$

where each activation function σ_ℓ is L_σ-Lipschitz, and each weight matrix satisfies $\|W_\ell\|_2 \leq \rho_\ell$.

Then s_k is globally Lipschitz with constant

$$L_k \leq \left(\prod_{\ell=1}^{L} \rho_\ell \right) L_\sigma^{L-1}. \tag{10}$$

Lemma 4.3 (Perturbation Energy Decomposition)**.** *Let $\mathcal{G}_\theta(u)(\mathbf{y}) = \sum_{k=1}^{p} s_k(u_{X_s})\psi_k(\mathbf{y})$ and define $\Delta := \mathcal{G}_\theta(u+\epsilon)(\mathbf{y}) - \mathcal{G}_\theta(u)(\mathbf{y})$, $\delta_k := s_k(u_{X_s} + \epsilon) - s_k(u_{X_s})$. Then,*

$$\mathbb{E}\|\Delta\|_2^2 = \sum_{k=1}^{p} \|\psi_k(\mathbf{y})\|_2^2 \mathbb{E}\|\delta_k\|_2^2 + \sum_{i \neq j} \langle \psi_i(\mathbf{y}), \psi_j(\mathbf{y}) \rangle \mathbb{E}[\delta_i^\top \delta_j],$$

where expectation is over the input perturbation ϵ.

Proof sketch. The result follows by substituting $\Delta = \sum_{k=1}^{p} \delta_k \psi_k(\mathbf{y})$, expanding the squared norm, and applying linearity of expectation to separate diagonal and cross-covariance terms. □

Theorem 4.4 (Robustness via Latent Decorrelation)**.** *Let Assumption 4.1 and Proposition 4.2 hold. Then the operator satisfies*

$$\mathbb{E}\|\Delta\|_2^2 \leq \sum_{k=1}^{p} \|\psi_k(y)\|_2^2 L_k^2 \|\epsilon\|_2^2.$$

Consequently, for $\|\epsilon\|_2 \leq \eta$,

$$\|\mathcal{G}_\theta(u+\epsilon) - \mathcal{G}_\theta(u)\|_2 \leq \eta L, \quad L^2 = \sum_{k=1}^{p} \|\psi_k(y)\|_2^2 L_k^2.$$

Proof sketch. The result follows by applying Lemma 4.3, using Assumption 4.1 to eliminate cross-covariance terms, and Proposition 4.2 to bound each latent perturbation by $\|\delta_k\|_2 \leq L_k \|\epsilon\|_2$. □

Remark 4.5 (Mechanism of Robustness via Latent Decorrelation). The robustness improvement induced by NOIR can be understood directly from the perturbation energy decomposition in Lemma 4.3 and Theorem 4.4. Input noise ϵ propagates to the output only through latent perturbations $\delta_k = s_k(u_{X_s} + \epsilon) - s_k(u_{X_s})$, which are aggregated as $\Delta = \sum_{k=1}^{p} \delta_k \psi_k(\mathbf{y})$. In the absence of structural constraints, cross-covariance terms $\mathbb{E}[\delta_i^\top \delta_j]$ introduce coupling between latent channels, causing noise to amplify through interaction effects in function space. NOIR mitigates this effect by enforcing latent decorrelation, which eliminates cross terms and yields an additive decomposition of perturbation energy across channels. Consequently, noise propagation becomes diagonally structured, preventing inter-channel amplification and resulting in a tighter control of the operator's sensitivity to input perturbations.

Remark 4.6 (NOIR as Variance-Reducing Robustness Mechanism). Beyond eliminating cross-channel interactions (Lemma 4.3), NOIR induces a variance reduction effect in the output sensitivity to noise. Under latent decorrelation, perturbation energy decomposes additively as $\mathbb{E}\|\Delta\|^2 = \sum_{k=1}^{p} \|\psi_k(\mathbf{y})\|_2^2 \mathbb{E}\|\delta_k\|_2^2$. If latent channels have comparable variance and normalized contributions, i.e., $\mathbb{E}\|\delta_k\|_2^2 \approx \sigma^2$ and $\|\psi_k(\mathbf{y})\|_2^2 = \mathcal{O}(1/p)$, then the overall output sensitivity satisfies $\mathbb{E}\|\Delta\|_2^2 = \mathcal{O}(p^{-1})$. This implies that increasing the number of independent latent components reduces the effect of input noise through averaging, thereby yielding a tighter Lipschitz constant and improved robustness. In contrast, entangled representations prevent such averaging due to non-vanishing cross-covariance terms, leading to persistent noise amplification.

5 Experiments

5.1 Experimental setup

Benchmarks. We evaluate NOIR on three operator-learning tasks spanning distinct PDE regimes: 1D Burgers' equation and 2D Darcy flow [18], and transonic flow over NACA airfoils [26]. **Test-time contamination.** Inputs are corrupted at test time with noise levels $\delta \in \{0, 0.1, 0.2, 0.5, 1\}$, defined relative to data variability σ_{data}, while targets remain clean. We consider five additive perturbations—high-pass (HP), low-pass (LP), spatial harmonic (SH), unstructured Gaussian (UGR), and localized block (LGB)—covering spectral and localized failure modes. This yields a zero-shot robustness evaluation (details in Appendix B). **Models.** We compare DeepONet, FNO, WNO, NOMAD, and their NOIR variants (DON+N, FNO+N) on Darcy and Burgers; on Airfoil, WNO is omitted due to convergence issues. Architecture details are in Appendix C.1. **Metrics.** Accuracy is measured via relative L_2 error. Robustness is summarized by sensitivity (error ratio at δ vs. $\delta = 0$). For NOIR models, we also report the latent Z-score, $Z = \frac{|m_{\text{test}} - \mu_{\text{clean}}|}{\sigma_{\text{clean}}}$, capturing representation-level deviation under distribution shift.

5.2 Cross-benchmark predictive robustness

Table 1 reports performance across all datasets for clean inputs ($\delta = 0$) and extreme corruption ($\delta = 1$). We first observe that NOIR largely preserves clean-data performance and often improves accuracy in the noiseless regime. For example, on Darcy flow, DeepONet improves from $8.19 \to 6.42$, and FNO from $2.70 \to 2.44$ with NOIR. Similarly, on Burgers' equation, DeepONet improves from $1.86 \to 1.28$, while FNO remains comparable ($0.08 \to 0.12$). On the airfoil dataset, performance is essentially preserved (e.g., DeepONet: $7.25 \to 7.15$, FNO: $1.06 \to 1.04$). This confirms that enforcing latent decorrelation does not degrade, and can even slightly enhance, baseline accuracy. Under severe corruption ($\delta = 1$), the robustness gains become much more pronounced. The most striking improvements occur for FNO, which is otherwise highly sensitive to noise. On the airfoil benchmark under high-pass (HP) noise, FNO error increases to 121.89, whereas FNO+N reduces this to 69.42 (approximately 43% reduction). Under low-pass (LP) noise, the error decreases from $491.77 \to 417.09$, and under spatial harmonic (SH) noise from $123.45 \to 94.89$. A similar trend holds for Darcy flow: under HP noise, FNO improves from $17.02 \to 9.47$, and under SH noise from $13.12 \to 9.65$. These reductions indicate that NOIR mitigates spectral amplification of perturbations. For Burgers' equation, where FNO is already relatively stable, improvements are smaller but still consistent. In contrast, DeepONet – already more robust – shows steady gains. For instance, on Burgers under HP noise, the error reduces from $4.27 \to 2.77$, and under unstructured Gaussian random (UGR) noise from $4.98 \to 3.83$. On Darcy, DeepONet+N improves worst-case error from $9.50 \to 8.73$ (HP) and $9.71 \to 8.94$ (UGR). Across all three benchmarks, the trend is consistent: *disentangled latent representations yield improved robustness, with the largest gains observed in architectures such as FNO that are otherwise highly sensitive to input perturbations*. Detailed results across intermediate noise levels and all corruption types are provided in Appendix D.1, further corroborating these observations.

5.3 Latent statistical signatures

The latent-independence Z-scores (Fig. 2) provide an architecture-dependent negentropy statistic that correlates with predictive robustness. When noise corrupts an architecture's inductive bias, Z-scores increase sharply and track prediction error, yielding a quantitative confidence signal; when the model filters the perturbation, Z remains low despite visible corruption. Low-pass (LP) noise highlights this behavior. On Airfoil, FNO+N reaches $Z = 5127.84$, consistent with its severe degradation ($399\times$), whereas DON+N shows a moderate rise ($Z = 15.42$) aligned with its milder $2.35\times$ sensitivity. On Darcy, DON+N attains $Z = 3.41$ at $\delta = 1.0$, reflecting corruption of its smooth latent manifold. High-pass (HP) noise exhibits the complementary pattern: FNO+N again shows large Z (e.g., 4434.06 on Airfoil) tracking collapse, while DON+N maintains low values ($Z \approx 15$) consistent with stable predictions. For localized block noise (LGB), both models maintain low Z (< 2 across benchmarks), correctly indicating confident reconstruction. The Z-score trajectories are also interpretable: monotonic growth indicates progressive latent corruption, while flat low-magnitude curves indicate robustness. Finally, the Airfoil task (geometry-to-solution mapping) produces significantly larger Z-scores than Darcy/Burgers, reflecting the stronger impact of

Table 1: Relative L_2 error ($\times 10^{-2}$). Top-left: clean inputs ($\delta = 0$). Others: corrupted inputs ($\delta = 1.0$) across noise types.

$\delta = 0$ **(Clean)**	Air	Burg	Dar
DON	7.25	1.86	8.19
DON+N	7.15	1.28	6.42
FNO	1.06	0.08	2.70
FNO+N	1.04	0.12	2.47
WNO	–	0.31	4.22
NOMAD	7.47	24.42	6.31

LP ($\delta = 1$)	Air	Burg	Dar
DON	17.18	8.99	28.87
DON+N	16.92	8.74	25.98
FNO	491.77	8.45	28.64
FNO+N	417.09	8.47	29.13
WNO	–	8.72	251.77
NOMAD	355.75	24.87	21.52

HP ($\delta = 1$)	Air	Burg	Dar
DON	17.15	4.27	9.50
DON+N	16.90	2.77	8.73
FNO	121.89	0.26	17.02
FNO+N	69.42	0.45	9.20
WNO	–	1.22	284.69
NOMAD	344.83	24.42	8.59

SH ($\delta = 1$)	Air	Burg	Dar
DON	17.00	18.32	9.41
DON+N	16.51	18.57	8.38
FNO	123.45	18.64	13.12
FNO+N	94.89	18.65	9.65
WNO	–	18.63	115.60
NOMAD	456.58	25.88	8.01

UGR ($\delta = 1$)	Air	Burg	Dar
DON	17.07	4.98	9.71
DON+N	16.84	3.83	8.94
FNO	129.11	2.68	17.08
FNO+N	80.18	2.71	9.47
WNO	–	3.00	284.80
NOMAD	341.61	24.48	8.79

LGB ($\delta = 1$)	Air	Burg	Dar
DON	10.10	2.99	8.29
DON+N	9.96	2.21	6.64
FNO	44.60	1.25	4.20
FNO+N	41.77	1.27	3.10
WNO	–	1.36	59.41
NOMAD	13.37	24.43	6.52

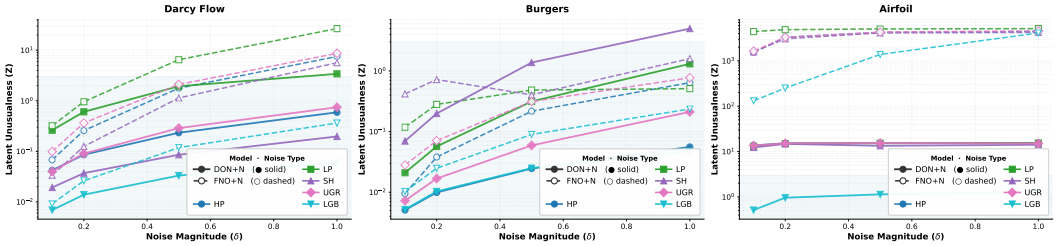


Figure 2: **Latent manifold statistics (Z-Score).** Log-scaled detection profiles showing how the internal representation drifts under corruption. The trajectories track the latent Z-score as noise magnitude δ increases. The shaded region ($Z = 3$) denote the "Stable Latent Manifold."

geometric perturbations compared to field noise. These trends are consistent across PDEs, indicating that Z-scores capture intrinsic architectural sensitivity rather than problem-specific effects.

5.4 Architecture-specific robustness signatures

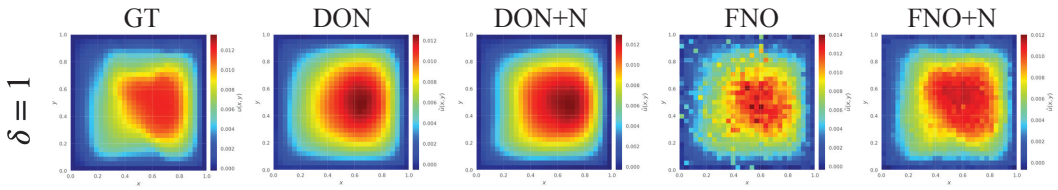


Figure 3: **Darcy Flow under unstructured gaussian noise** ($\delta = 1$)**.** The left most column displays the pressure field solutions ground truth (GT), while remaining columns represent model predictions at Unstructured Gaussian Noise (UGR). Displays distinct failures per architecture under severely corrupted input.

Figure 3 reveals architecture-specific robustness signatures under severe corruption. Vanilla DON exhibits latent oversmoothing, while FNO shows pronounced localized artifacts, reflecting distinct corruption pathways. NOIR mitigates both, sharpening DON reconstructions and restoring global coherence in FNO predictions . These qualitative patterns align with the ablation results, confirming that robustness emerges through architecture-dependent latent statistical control rather than uniform regularization effects. This behavior is consistently observed across all cases (Appendix D.3, D.4).

Table 2: Ablation study under partial random noise (UGR) on Darcy. Entries are Rel. L_2 mean $\pm$ standard deviation across different seeds.

Model	$\delta = 0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 1$
FNO	2.70 ± 0.33	2.89 ± 0.44	3.43 ± 0.70	6.99 ± 1.67	17.08 ± 2.76
FNO+N	2.44 ± 0.09	2.50 ± 0.08	2.72 ± 0.07	4.61 ± 0.34	9.91 ± 0.26
FNO+Neg	2.46 ± 0.09	2.52 ± 0.10	2.73 ± 0.11	4.46 ± 0.30	9.73 ± 0.37
FNO+W	2.47 ± 0.08	2.53 ± 0.09	2.73 ± 0.09	4.43 ± 0.22	9.39 ± 0.28
FNO+S	2.45 ± 0.07	2.50 ± 0.07	2.71 ± 0.08	4.45 ± 0.22	9.63 ± 0.38

5.5 Ablation study

To isolate the role of individual NOIR components, we perform an ablation study on FNO under UGR noise, selecting FNO as the tested due to its stronger susceptibility to perturbations in the vanilla setting. Table 2 reveals a clear noise-dependent shift in the dominant robustness mechanism. At low noise levels ($\delta \leq 0.2$), all variants perform comparably, indicating that the unmixing structure itself captures most of the robustness gain; the strong performance of the unregularized variant (+S) suggests that architectural disentanglement alone is sufficient to stabilize latent representations under mild corruption. This behavior changes sharply as noise increases. In the moderate- and high-noise regimes ($\delta \in 0.5, 1$), whitening-based regularization (+W) consistently achieves the lowest error, establishing second-order decorrelation as the dominant stabilizing mechanism under severe perturbation, while the advantage of negentropy regularization (+Neg) becomes comparatively weaker. These results expose a clear robustness hierarchy: structural disentanglement governs low-noise stability, whereas covariance conditioning becomes critical as corruption intensifies, highlighting that NOIR's effectiveness arises from noise-adaptive latent statistical control rather than a single regularization effect.

6 Conclusion and limitations

We introduced NOIR, an end-to-end regularization framework for enforcing statistically independent latent representations in neural operators. While the mechanism is broadly applicable to operator architectures and integrates naturally with diverse operator backbones, the resulting robustness gains are architecture-dependent, reflecting differences in latent inductive biases across operator families. Across Burgers' equation, NACA airfoil flow, and Darcy flow under diverse additive corruption settings, NOIR consistently improves robustness to noisy measurements while preserving clean-data performance. Our analysis further shows that latent independence induces distinctive architecture-specific statistical signatures under corruption, offering mechanistic insight into how neural operators respond to perturbations. These findings establish latent statistical independence as a principled inductive bias for robust neural operator design and highlight the broader potential of latent-space regularization for improving reliability in scientific machine learning.

Limitations. NOIR currently assumes additive corruption and provides only global robustness enhancement without localizing failure sources. Its effectiveness remains architecture-dependent, motivating future work on adaptive regularization strategies and extension to broader corruption mechanisms, including structured and non-additive perturbations.

References

[1] Bahador Bahmani, Somdatta Goswami, Ioannis G Kevrekidis, and Michael D Shields. A resolution independent neural operator. *Computer Methods in Applied Mechanics and Engineering*, 444:118113, 2025.

[2] 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.

[3] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

[4] Taylor Denouden, Rick Salay, Krzysztof Czarnecki, Vahdat Abdelzad, Buu Phan, and Sachin Vernekar. Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance. *arXiv preprint arXiv:1812.02765*, 2018.

[5] Gerald Farin. *Curves and surfaces for computer-aided geometric design: a practical guide*. Elsevier, 2014.

[6] Somdatta Goswami, Aniruddha Bora, Yue Yu, and George Em Karniadakis. Physics-informed deep neural operator networks. In *Machine learning in modeling and simulation: methods and applications*, pages 219–254. Springer, 2023.

[7] Mengdi Hou, Jianyu Yang, Mingjie Yang, Fen Hu, Rongge Zhao, Yuhang Pan, Wan Li, Mingxin Chen, Jingjun Xu, Ke Xu, et al. Fast and robust drift correction for single-molecule localization microscopy. *Nature Communications*, 16(1):9031, 2025.

[8] Aapo Hyvarinen. Fast ica for noisy data using gaussian moments. In *1999 IEEE international symposium on circuits and systems (ISCAS)*, volume 5, pages 57–61. IEEE, 1999.

[9] Aapo Hyvärinen and Patrik Hoyer. Emergence of phase-and shift-invariant features by decomposition of natural images into independent feature subspaces. *Neural computation*, 12(7):1705–1720, 2000.

[10] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.

[11] Ilyes Khemakhem, Diederik Kingma, Ricardo Monti, and Aapo Hyvarinen. Variational autoencoders and nonlinear ica: A unifying framework. In *International conference on artificial intelligence and statistics*, pages 2207–2217. PMLR, 2020.

[12] 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.

[13] Quoc V Le, Will Y Zou, Serena Y Yeung, and Andrew Y Ng. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In *CVPR 2011*, pages 3361–3368. IEEE, 2011.

[14] Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.

[15] Jeremiah Liu, John Paisley, Marianthi-Anna Kioumourtzoglou, and Brent Coull. Accurate uncertainty estimation and decomposition in ensemble learning. *Advances in neural information processing systems*, 32, 2019.

[16] Weitang Liu, Xiaoyun Wang, John Owens, and Yixuan Li. Energy-based out-of-distribution detection. *Advances in neural information processing systems*, 33:21464–21475, 2020.

[17] 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.

[18] Lu Lu, Xuhui Meng, Shengze Cai, Zhiping Mao, Somdatta Goswami, Zhongqiang Zhang, and George Em Karniadakis. A comprehensive and fair comparison of two neural operators (with practical extensions) based on fair data. *Computer Methods in Applied Mechanics and Engineering*, 393:114778, 2022.

[19] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[20] Jiquan Ngiam, Zhenghao Chen, Daniel Chia, Pang Koh, Quoc Le, and Andrew Ng. Tiled convolutional neural networks. *Advances in neural information processing systems*, 23, 2010.

[21] Jeil Oh and Matt Bartos. Scalable, adaptive and risk-informed design of hydrological sensor networks. *Nature Water*, pages 1–11, 2025.

[22] Kiho Park, Yo Joong Choe, and Victor Veitch. The linear representation hypothesis and the geometry of large language models. *arXiv preprint arXiv:2311.03658*, 2023.

[23] Eric Persoon and King-Sun Fu. Shape discrimination using fourier descriptors. *IEEE Transactions on systems, man, and cybernetics*, 7(3):170–179, 2007.

[24] Apostolos F Psaros, Xuhui Meng, Zongren Zou, Ling Guo, and George Em Karniadakis. Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons. *Journal of Computational Physics*, 477:111902, 2023.

[25] Melanie Roschewitz, Galvin Khara, Joe Yearsley, Nisha Sharma, Jonathan J James, Eva Ambrozay, Adam Heroux, Peter Kecskemethy, Tobias Rijken, and Ben Glocker. Automatic correction of performance drift under acquisition shift in medical image classification. *Nature Communications*, 14(1):6608, 2023.

[26] Subhankar Sarkar and Souvik Chakraborty. Spatio-spectral graph neural operator for solving computational mechanics problems on irregular domain and unstructured grid. *Computer Methods in Applied Mechanics and Engineering*, 435:117659, 2025.

[27] Jacob Seidman, Georgios Kissas, Paris Perdikaris, and George J Pappas. Nomad: Nonlinear manifold decoders for operator learning. *Advances in Neural Information Processing Systems*, 35:5601–5613, 2022.

[28] Tapas Tripura and Souvik Chakraborty. Wavelet neural operator: a neural operator for parametric partial differential equations. *arXiv preprint arXiv:2205.02191*, 2022.

[29] Enrico Zio. Prognostics and health management (phm): Where are we and where do we (need to) go in theory and practice. *Reliability Engineering & System Safety*, 218:108119, 2022.

A Mathematical Formulation

This appendix provides rigorous mathematical derivations and extended technical details omitted from the main text for brevity. We present complete step-by-step derivations for all major components of our framework.

A.1 Latent-channel independence, identifiability, and orthogonality

Equation (3) can be viewed as a basis–coefficient decomposition over p latent channels: the trunk network provides data-adaptive basis functions $\{t_k(\mathbf{y})\}_{k=1}^{p}$ defined over the output domain, and the branch network provides the corresponding coefficients $\{b_k(u)\}_{k=1}^{p}$ as functions of the input u. Absent structural constraints, DeepONet admits a rotational non-uniqueness [11]: for any invertible matrix $\mathbf{R} \in \mathbb{R}^{p\times p}$,

$$b'(u) \;=\; \mathbf{R}\, b(u), \qquad t'(\mathbf{y}) \;=\; \mathbf{R}^{-\top}\, t(\mathbf{y}) \quad \Rightarrow \quad \sum_{k=1}^{p} b'_k(u)\, t'_k(\mathbf{y}) \;=\; \sum_{k=1}^{p} b_k(u)\, t_k(\mathbf{y}). \tag{11}$$

Thus, the latent representation is not identifiable without further constraints. Moreover, correlated channels can lead to redundancy and hinder interpretability and data efficiency.

Enforcing orthogonality is a cornerstone of many efficient computational methods for analyzing dynamical systems, as exemplified by techniques such as SVD, DMD, and the FFT. These methods leverage orthogonal bases to achieve stable, interpretable, and compressed representations. However, directly imposing such structural constraints within a deep neural network during backpropagation is severely intrusive. Methods that enforce orthogonality on weight matrices can interfere with the gradient descent optimization that governs the learning process. Such "hard" constraints can disrupt the flow of gradients and hinder the network's ability to converge to an optimal solution for the primary operator learning task.

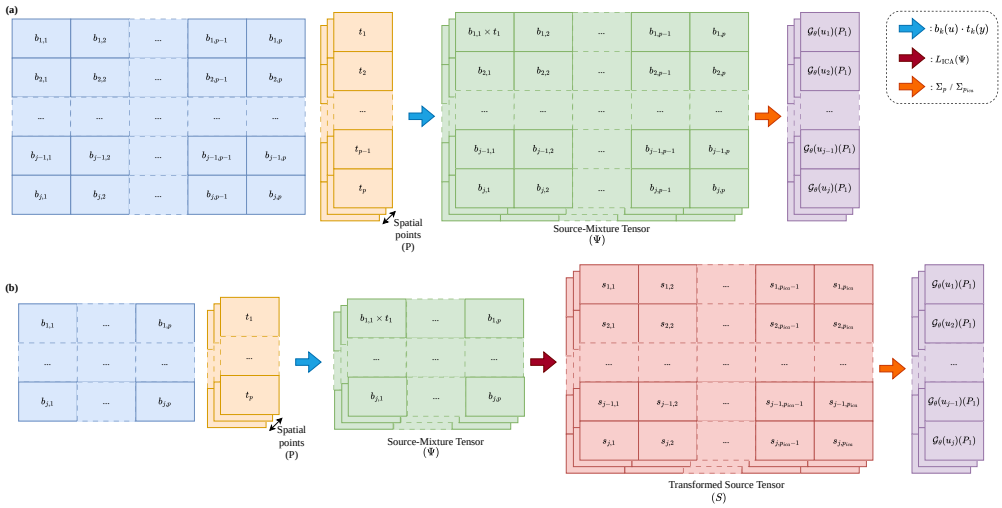


Figure 4: **Visualized transformed source tensor for DeepONet. (a)** The branch network outputs for J input functions are linearized into p latent features, and the trunk network outputs for P spatial query points are linearized into the same p latent features. Rather than collapsing these features immediately with an Einstein summation, we expose the intermediate source-mixture tensor Ψ, whose channel-wise slices quantify the contribution of each latent channel to the final prediction $\mathcal{G}_\theta(u)(\mathbf{y})$. **(b)** ICA-layer unmixing: a learnable unmixing layer transforms the source-mixture tensor into p_{ica} source features, which are then regularized by NOIR to encourage statistically independent latent channels. This interpretation naturally extends to kernel-based neural operators, including FNO, by applying the same idea to latent component tensors induced by their final projection or downlifting MLP layers. This figure consider the output channel dimension $C = 1$ for simplicity.

A more effective and compatible approach is to introduce a "soft" constraint by regularizing the network's objective function. We propose adding a regularization term to the main loss function that explicitly encourages the desired property, in this case, the orthogonality of the latent representations produced by the branch and the trunk network. This regularization term penalizes correlations among the latent channels. This method guides the optimization process toward solutions where the latent channels are decorrelated, thereby improving identifiability and reducing redundancy, without fundamentally altering the backpropagation mechanism. Therefore, the network learns to satisfy the orthogonality condition as part of its overall objective.

For arbitrary functions f_i, f_j in a Hilbert space $\mathcal{H}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$,

$$\langle f_i, f_j \rangle_{\mathcal{H}} \;=\; \int f_i(x)\, f_j(x)\, d\mu(x) \;=\; 0 \quad \text{for } i \neq j, \tag{12}$$

defines the orthogonality property, i.e., the functions are uncorrelated with respect to the measure μ.

In practice, this continuous integral can be approximated empirically using discrete samples. To formalize this, we examine the computation within the branch network's final layer. The penultimate layer produces a feature matrix $\mathbf{Z} \coloneqq \mathbf{X}_{h-1} \in \mathbb{R}^{N_b \times d_{h-1}}$ for a batch of N_b input functions. The final layer is defined by a weight matrix $\mathbf{W} \coloneqq \mathbf{W}_h \in \mathbb{R}^{p \times d_{h-1}}$. Each of the p rows of this weight matrix, denoted $\mathbf{w}_k \in \mathbb{R}^{d_{h-1}}$ for $k \in \{1, \ldots, p\}$, defines a linear functional that projects the high-dimensional feature space onto the k-th latent dimension.

The calculation of the k-th coefficient function, $b_k(u)$, for all N_b samples in the batch is achieved by applying the k-th weight vector $\mathbf{w}_k$ to the entire penultimate feature matrix $\mathbf{Z}$. This yields the k-th column of the final branch output matrix $\mathbf{F} \in \mathbb{R}^{N_b \times p}$:

$$\mathbf{f}_k = \mathbf{Z}\mathbf{w}_k \in \mathbb{R}^{N_b}. \tag{13}$$

Conversely, we can consider the calculation for a single input sample from the batch, represented by the j-th row of the feature matrix, $\mathbf{z}_j \in \mathbb{R}^{d_{h-1}}$ (as a column vector). The full p-dimensional

coefficient vector for this single sample is obtained by applying all p linear functionals (the rows of $\mathbf{W}$) to $\mathbf{z}_j$. This corresponds to the j-th row of the output matrix $\mathbf{F}$:

$$\mathbf{F}_{j,:} = \mathbf{W}\mathbf{z}_j \in \mathbb{R}^p. \tag{14}$$

The individual scalar element F_{jk}, representing the k-th coefficient for the j-th input sample, can be expressed by unifying the row- and column-wise perspectives from Eq. (13) and Eq. (14). It is computed as the inner product of the k-th channel's weight vector and the j-th sample's feature vector:

$$F_{jk} = \mathbf{w}_k^\top \mathbf{z}_j. \tag{15}$$

This value, F_{jk}, is the empirical realization of the coefficient function $b_k(u)$ for the specific input function u_j, i.e., $F_{jk} = b_k(u_j)$. Substituting this computation back into the operator definition, Equation 3, for a single sample u_j reveals the underlying structure:

$$\mathcal{G}(u_j)(\mathbf{y}) = \sum_{k=1}^{p} b_k(u_j)t_k(\mathbf{y}) = \sum_{k=1}^{p} (\mathbf{w}_k^\top \mathbf{z}_j)t_k(\mathbf{y}). \tag{16}$$

The structure reveals that each latent channel's contribution is computed independently from the other features before being combined. The term $(\mathbf{w}_k^\top \mathbf{z}_j)$ isolates the computation for the k-th coefficient of the j-th sample. The strict one-to-one coupling with the corresponding trunk function $t_k(\mathbf{y})$ implies that the weight vectors $\{\mathbf{w}_k\}_{k=1}^p$ defining the coefficient functions should form a non-redundant, and ideally orthogonal, set.

To formalize this requirement, we seek orthogonality of the weight vectors $\{\mathbf{w}_k\}_{k=1}^p$ with respect to the distribution of penultimate features $\mathbf{z}_j$ across the batch. Specifically, for any two distinct latent channels $i \neq j$, we require that the corresponding weight vectors produce uncorrelated outputs when applied to the feature distribution:

$$\mathbb{E}_{\mathbf{z}\sim\mathcal{P}_{\text{data}}}[(\mathbf{w}_i^\top \mathbf{z})(\mathbf{w}_j^\top \mathbf{z})] \approx 0, \quad \text{for } i \neq j, \tag{17}$$

where $\mathcal{P}_{\text{data}}$ denotes the empirical distribution of penultimate features. This condition is empirically tested by examining the batch-wise evaluations of the coefficient functions, which are precisely the columns of the matrix $\mathbf{F} = \mathbf{Z}\mathbf{W}^\top$. The Gram matrix provides a systematic tool to quantify these correlations.

Define the centered coefficients $\tilde{\mathbf{F}} \coloneqq \mathbf{F} - \mathbf{1}\boldsymbol{\mu}^\top$, where $\boldsymbol{\mu} \coloneqq \frac{1}{N_b}\mathbf{F}^\top\mathbf{1} \in \mathbb{R}^p$ is the vector of batch means for each channel and $\mathbf{1} \in \mathbb{R}^{N_b}$ is the vector of ones. The empirical Gram (or sample covariance) matrix is then:

$$\mathbf{G}_b \;\coloneqq\; \frac{1}{N_b}\,\tilde{\mathbf{F}}^\top\tilde{\mathbf{F}} \;\in\; \mathbb{R}^{p\times p}, \tag{18}$$

where N_b is the batch size. The element $(\mathbf{G}_b)_{ij}$, which represents the entry in the i-th row and j-th column of the Gram matrix, is the sample covariance between the coefficient functions $b_i(u)$ and $b_j(u)$. More explicitly, let $\tilde{\mathbf{f}}_i$ and $\tilde{\mathbf{f}}_j$ be the i-th and j-th columns of the centered matrix $\tilde{\mathbf{F}}$. The entry $(\mathbf{G}_b)_{ij}$ is computed as the scaled inner product of these two vectors:

$$(\mathbf{G}_b)_{ij} = \frac{1}{N_b}\tilde{\mathbf{f}}_i^\top\tilde{\mathbf{f}}_j = \frac{1}{N_b}\sum_{k=1}^{N_b}\tilde{F}_{ki}\tilde{F}_{kj} = \frac{1}{N_b}\sum_{k=1}^{N_b}(F_{ki}-\mu_i)(F_{kj}-\mu_j), \tag{19}$$

where F_{ki} is the evaluation of the i-th coefficient function for the k-th batch sample, and μ_i is the sample mean of that coefficient function over the batch. This expression is the definition of the sample covariance between the empirical evaluations of functions $b_i(u)$ and $b_j(u)$. The principle of orthogonality requires that these off-diagonal elements be zero. We could form a regularization loss, penalizing their magnitude:

$$\mathcal{L}_{\text{ortho-branch}} \;\coloneqq\; \left\|\text{OffDiag}\left(\mathbf{G}_b\right)\right\|_F^2 \;=\; \sum_{i=1}^{p}\sum_{\substack{j=1\\ j\neq i}}^{p}\left(\mathbf{G}_b\right)_{ij}^2. \tag{20}$$

To avoid trivial scaling and promote well-conditioned channels, a whitening term can be added:

$$\mathcal{L}_{\text{scale-branch}} \;\coloneqq\; \left\|\text{Diag}\left(\mathbf{G}_b\right) - \mathbf{I}_p\right\|_F^2, \tag{21}$$

yielding the total branch regularization loss:

$$\mathcal{L}_{\text{branch}} \coloneqq \mathcal{L}_{\text{ortho-branch}} + \beta\, \mathcal{L}_{\text{scale-branch}}, \qquad \beta \geq 0. \tag{22}$$

Note that $\mathcal{L}_{\text{ortho-branch}} = 0$ if and only if the columns of $\tilde{\mathbf{F}}$ are mutually orthogonal under the empirical measure. The proposed $\mathcal{L}_{\text{branch}}$ loss thus regularizes the branch network's final linear layer. It guides the optimization to find a set of coefficient functions $\{b_k(u)\}_{k=1}^{p}$ that are empirically decorrelated (orthogonal) and normalized (unit variance) over the batch of input functions $u \sim \mathcal{U}_{\text{data}}$. This promotes a more identifiable and non-redundant representation of the input function manifold.

Symmetrically, we can enforce an analogous constraint on the trunk network. Let the final trunk output be $\mathbf{T} = \mathbf{Z}^t(\mathbf{W}^{\mathbf{t}})^\top \in \mathbb{R}^{P\times p}$ (for $C = 1$), computed from its penultimate features $\mathbf{Z}^t$ and final weights $\mathbf{W}^{\mathbf{t}}$. We compute the centered trunk outputs $\tilde{\mathbf{T}}$ (centering over the P query points). The empirical Gram matrix, analogous to Equation 18, for the trunk is calculated as:

$$\mathbf{G}_t \coloneqq \frac{1}{P}\tilde{\mathbf{T}}^\top\tilde{\mathbf{T}} \in \mathbb{R}^{p\times p}. \tag{23}$$

The off-diagonal elements $(\mathbf{G}_t)_{ij}$ represent the sample covariance between the basis functions $t_i(\mathbf{y})$ and $t_j(\mathbf{y})$ over the set of query locations $\mathbf{Y}$. We can define a corresponding trunk regularization loss:

$$\mathcal{L}_{\text{trunk}} \coloneqq \left\|\text{OffDiag}\left(\mathbf{G}_t\right)\right\|_F^2 + \gamma\, \left\|\text{Diag}\left(\mathbf{G}_t\right) - \mathbf{I}_p\right\|_F^2, \qquad \gamma \geq 0. \tag{24}$$

This loss encourages the trunk network to learn a set of basis functions $\{t_k(\mathbf{y})\}_{k=1}^{p}$ that are empirically orthogonal and normalized over the coordinate query manifold $\mathcal{Y}$.

While the branch and trunk regularization losses in Equations (22) and (24) enforce orthogonality within their respective domains, a naive approach that applies these constraints *separately* on the input manifold $\mathcal{U}$ (via branch coefficients) and the coordinate manifold $\mathcal{Y}$ (via trunk bases) is insufficient. Such decoupled regularization fails to impose structural constraints on the *interaction* between these manifolds, which fundamentally defines the operator itself through the outer product structure $\mathbf{b}_k\mathbf{t}_k^\top$.

A more principled approach is to analyze the *combined operator manifold* $\mathcal{U}\times\mathcal{Y}$ by examining the source-mixture tensor $\mathbf{\Psi} \in \mathbb{R}^{N_b\times P\times p\times C}$ defined in the vector-valued case. This tensor encapsulates the unsummed contributions of all p latent channels before aggregation. Each slice $\mathbf{\Psi}_{:,:,k,c} \in \mathbb{R}^{N_b\times P}$ represents the output of the k-th latent channel for the c-th output component across all input functions and query locations.

For vector-valued outputs with $C > 1$, we compute the measure separately for each output channel dimension and then aggregate. For a fixed output channel $c \in \{1,\ldots,C\}$, we extract the corresponding slice $\mathbf{\Psi}^{(c)} \in \mathbb{R}^{N_b\times P\times p}$. Reshaping this slice to $\mathbf{F}^{(c)}_{\text{global}} \in \mathbb{R}^{N\times p}$ (where $N = N_bP$), we obtain a matrix where each of the p columns represents the k-th latent channel's contribution evaluated at all N combinations of input functions and query points for output channel c.

The empirical Gram matrix for output channel c is then:

$$\mathbf{G}^{(c)}_{\text{global}} \coloneqq \frac{1}{N}(\tilde{\mathbf{F}}^{(c)}_{\text{global}})^\top\tilde{\mathbf{F}}^{(c)}_{\text{global}} \in \mathbb{R}^{p\times p}, \tag{25}$$

where $\tilde{\mathbf{F}}^{(c)}_{\text{global}}$ denotes the centered version of $\mathbf{F}^{(c)}_{\text{global}}$. The orthogonality loss for output channel c is:

$$\mathcal{L}^{(c)}_{\text{ortho-global}} \coloneqq \left\|\text{OffDiag}\left(\mathbf{G}^{(c)}_{\text{global}}\right)\right\|_F^2. \tag{26}$$

The final combined orthogonality loss is the average over all output channels:

$$\mathcal{L}_{\text{ortho-global}} \coloneqq \frac{1}{C}\sum_{c=1}^{C}\mathcal{L}^{(c)}_{\text{ortho-global}}. \tag{27}$$

A.2 Independent Component Analysis for latent channel disentanglement

The orthogonality loss function $\mathcal{L}_{\text{ortho-global}}$ penalizes second-order statistics, driving the sample covariance $(\mathbf{G}_b)_{ij}$ toward zero for $i \neq j$. However, decorrelation (zero covariance) is a strictly weaker condition than statistical independence. Two random variables can be uncorrelated yet

remain statistically dependent through higher-order moments. For instance, if $b_i \sim \mathcal{N}(0,1)$ and $b_j = b_i^2 - 1$, these channels are perfectly dependent (one is a deterministic function of the other), yet their covariance is exactly zero: $\text{Cov}(b_i, b_j) = \mathbb{E}[b_i(b_i^2 - 1)] = \mathbb{E}[b_i^3] - \mathbb{E}[b_i] = 0$. Such higher-order dependencies are not captured by the Gram matrix but can lead to redundant and uninterpretable latent representations. Moreover, the computational cost $O(N^2 p)$ scales quadratic to the multiple of number of spatial points and the size of the batch. This presents a strong computational scalability limitation.

Therefore, we clarify that the orthogonality regularization (Eqs. 20–27) was introduced as a mathematical steppingstone toward the following formulation.

To enforce a stronger structural constraint, we employ principles from Independent Component Analysis (ICA) [9]. ICA seeks an unmixing transformation that recovers statistically independent source signals $\mathbf{s} = [s_1, \ldots, s_{p_{\text{ica}}}]^\top$ from observed mixtures $\mathbf{x}$. The theoretical foundation rests on the Central Limit Theorem: a linear mixture of independent, non-Gaussian signals tends toward a more Gaussian distribution than the individual sources. Conversely, to recover the original independent sources from their mixture, one must find a projection that maximizes the *non-Gaussianity* of the resulting components.

A rigorous measure of non-Gaussianity is the negentropy, defined as

$$J(s_k) \;:=\; H(\nu) - H(s_k), \tag{28}$$

where $H(\cdot)$ denotes differential entropy, s_k is a source component, and ν is a Gaussian random variable with the same mean and variance as s_k. Negentropy is always non-negative and equals zero if and only if s_k is Gaussian. Maximizing negentropy is equivalent to maximizing non-Gaussianity. Since differential entropy is computationally intractable, we employ FastICA method, a widely-used approximation based on contrast functions [8]:

$$J(s_k) \;\approx\; \rho \cdot \left(\mathbb{E}_{s_k}[\mathcal{G}_{ICA}(s_k)] - \mathbb{E}_\nu[\mathcal{G}_{ICA}(\nu)]\right)^2, \tag{29}$$

where s_k is assumed to be centered (zero mean) and whitened (unit variance), $\nu \sim \mathcal{N}(0,1)$ is a standard normal variable, $\rho > 0$ is a normalization constant, and $\mathcal{G}_{\mathcal{ICA}} : \mathbb{R} \to \mathbb{R}$ is a non-quadratic contrast function. Common choices for $\mathcal{G}_{ICA}$ include:

1. **Kurtosis:** $\mathcal{G}_{ICA}(x) = x^4$, which measures the "tailedness" of the distribution. For a standard Gaussian, $\mathbb{E}_\nu[\nu^4] = 3$.
2. **LogCosh:** $\mathcal{G}_{ICA}(x) = \log(\cosh(x))$, a robust general-purpose contrast function with $\mathbb{E}_\nu[\mathcal{G}(\nu)] \approx 0.3746$.

We integrate ICA principles directly into the operator network's forward pass through a learnable **unmixing ICA layer** $L_{\text{ICA}} : \mathbb{R}^p \to \mathbb{R}^{p_{\text{ica}}}$, parameterized by a weight matrix $\mathbf{A} \in \mathbb{R}^{p_{\text{ica}} \times p}$. This layer transforms the p latent channels into p_{ica} statistically independent non-gaussian source components. The architecture proceeds as follows:

Intermediate component computation. The model first computes the source-mixture tensor $\mathbf{\Psi} \in \mathbb{R}^{N_b \times P \times p \times C}$, where each element is given by

$$\mathbf{\Psi}_{jqkc} = B_{jk} \cdot T_{qkc}, \tag{30}$$

representing the contribution of the k-th latent channel for the j-th input function at the q-th query location for the c-th output component. This tensor encapsulates the complete operator output before the final summation over latent channels.

Unmixing transformation. We permute $\mathbf{\Psi}$ to shape (N_b, P, C, p) and apply the unmixing layer L_{ICA} broadcast-style to the last dimension. This yields the transformed source tensor:

$$\mathbf{S}_{\text{4D}} = L_{\text{ICA}}(\mathbf{\Psi}) \in \mathbb{R}^{N_b \times P \times C \times p_{\text{ica}}}, \tag{31}$$

where $[\mathbf{S}_{\text{4D}}]_{jqc} = \mathbf{A}[\mathbf{\Psi}]_{jqc} \in \mathbb{R}^{p_{\text{ica}}}$ for each (j, q, c) triple.

Final output aggregation. The final operator output is obtained by summing over the transformed latent dimension:

$$\mathcal{G}_\theta(u_j)(\mathbf{y}_q) = \sum_{k=1}^{p_{\text{ica}}} [\mathbf{S}_{\text{4D}}]_{jqk:} \in \mathbb{R}^C. \tag{32}$$

A.2.1 Whitening and ICA Loss formulation

As in the main text, we treat each output channel $c \in \{1, \dots, C\}$ as defining a separate statistical ensemble. For a fixed channel c, we extract the corresponding slice $\mathbf{S}^{(c)} \in \mathbb{R}^{N_b \times P \times p_{\text{ica}}}$ from $\mathbf{S}_{\text{4D}}$ and reshape it into a 2D matrix

$$\mathbf{S}_{\text{2D}}^{(c)} = \text{reshape}\big(\mathbf{S}^{(c)}, (N, p_{\text{ica}})\big) \in \mathbb{R}^{N \times p_{\text{ica}}}, \qquad N = N_b \cdot P. \tag{33}$$

Negentropy maximization loss. We define the ICA non-Gaussianity loss on these sources. Let $\hat{s}_{jk}^{(c)}$ denote the (j,k)-entry of the whitened source matrix $\hat{\mathbf{S}}^{(c)}$. To prevent numerical overflow in the contrast function, we apply clamping:

$$\bar{s}_{jk}^{(c)} = \text{clamp}(\hat{s}_{jk}^{(c)}, -10, 10). \tag{34}$$

Using the LogCosh contrast function $\mathcal{G}_{\text{ICA}}(x) = \log(\cosh(x))$ and the standard normal reference variable $\nu \sim \mathcal{N}(0,1)$, the negentropy approximation for source k is:

$$J_k^{(c)} \approx \left(\frac{1}{N} \sum_{j=1}^{N} \mathcal{G}_{\text{ICA}}\big(\bar{s}_{jk}^{(c)}\big) - \mathbb{E}_\nu[\mathcal{G}_{\text{ICA}}(\nu)] \right)^2, \tag{35}$$

where $\mathbb{E}_\nu[\log(\cosh(\nu))] \approx 0.3746$ is computed analytically for $\nu \sim \mathcal{N}(0,1)$.

The ICA regularization loss for output channel c is the negative sum of negentropies:

$$\mathcal{L}_{\text{ICA}}^{(c)} = -\sum_{k=1}^{p_{\text{ica}}} J_k^{(c)}. \tag{36}$$

Minimizing $\mathcal{L}_{\text{ICA}}^{(c)}$ (equivalently, maximizing $\sum_k J_k^{(c)}$) drives each source toward maximal non-Gaussianity, enforcing statistical independence.

Finally, we average over all output channels to obtain the total ICA regularization term:

$$\mathcal{L}_{\text{ICA}} = \frac{1}{C} \sum_{c=1}^{C} \mathcal{L}_{\text{ICA}}^{(c)}. \tag{37}$$

Simplified whitening via standardization. From Eq. (31), our objective is to promote non-Gaussianity of the p_{ica} source channels in $\mathbf{S}_{\text{4D}}$. To satisfy the assumptions of classical ICA, we apply a whitening transform to the sources before evaluating the negentropy surrogate.

Classical ICA whitening employs eigendecomposition of the empirical covariance matrix to simultaneously decorrelate and normalize sources [9]. However, for mini-batch stochastic optimization in deep networks, eigendecomposition introduces computational overhead and numerical instability due to small batch sizes and shifting statistics. Following the principled simplifications introduced by [10], we adopt two key approximations:

Approximation 1: diagonal covariance assumption. We assume the empirical covariance of latent sources is approximately diagonal:

$$\mathbf{C}^{(c)} \approx \text{diag}(\sigma_1^2, \dots, \sigma_{p_{\text{ica}}}^2), \tag{38}$$

where σ_k^2 is the empirical variance of the k-th source component. This assumption is justified by the ICA regularization itself, which explicitly penalizes off-diagonal covariances through Eq. (37), driving $\mathbf{C}^{(c)}$ toward diagonality during training.

Approximation 2: per-source standardization. Under diagonal covariance, the whitening transform reduces to independent standardization of each source component. We compute the empirical mean and standard deviation for each source k over the batch dimension (actuation dimension):

$$\mu_k^{(c)} = \frac{1}{N} \sum_{j=1}^{N} S_{jk}^{(c)}, \qquad \sigma_k^{(c)} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (S_{jk}^{(c)} - \mu_k^{(c)})^2}, \tag{39}$$

where $S_{jk}^{(c)}$ denotes the (j,k)-entry of $\mathbf{S}_{\text{2D}}^{(c)}$.

The whitened sources are then obtained via element-wise standardization:

$$\hat{S}_{jk}^{(c)} = \frac{S_{jk}^{(c)} - \mu_k^{(c)}}{\sigma_k^{(c)} + \epsilon}, \tag{40}$$

where $\epsilon = 10^{-8}$ is a small constant for numerical stability. This yields the whitened source matrix

$$\hat{\mathbf{S}}^{(c)} = [\hat{S}_{jk}^{(c)}] \in \mathbb{R}^{N \times p_{\text{ica}}}, \tag{41}$$

satisfying the approximate whitening condition:

$$\frac{1}{N}\sum_{j=1}^{N} \hat{S}_{jk}^{(c)} \approx 0, \qquad \frac{1}{N}\sum_{j=1}^{N} (\hat{S}_{jk}^{(c)})^2 \approx 1, \quad \forall k. \tag{42}$$

Remark A.1 (Computational efficiency). This simplified whitening requires $\mathcal{O}(Np_{\text{ica}})$ operations versus $\mathcal{O}(Np_{\text{ica}}^2 + p_{\text{ica}}^3)$ for eigendecomposition-based whitening, enabling efficient mini-batch training.

A.3 Latent Gaussianization under corruption

The central hypothesis of NOIR is that corrupted inputs produce latent representations that are statistically closer to Gaussian than those induced by clean inputs. This motivates the use of latent negentropy as an intrinsic indicator of corruption.

Let u denote a clean input function and let

$$\tilde{u} = u + \epsilon$$

denote its corrupted counterpart, where ϵ represents an additive sensor perturbation independent of the clean input-generating process. Let the encoder $\mathcal{E}$ map the input to a latent representation

$$z = \mathcal{E}(u).$$

Under a first-order local linearization of the encoder around u,

$$\mathcal{E}(\tilde{u}) = \mathcal{E}(u + \epsilon) \approx \mathcal{E}(u) + J_{\mathcal{E}}(u)\epsilon + r(u, \epsilon), \tag{43}$$

where $J_{\mathcal{E}}(u)$ is the encoder Jacobian and $r(u,\epsilon)$ is a higher-order remainder satisfying

$$\|r(u,\epsilon)\| = o(\|\epsilon\|) \qquad \text{as } \|\epsilon\| \to 0.$$

This yields the approximate decomposition

$$\tilde{z} \approx z_{\text{phys}} + z_{\text{noise}}, \qquad z_{\text{phys}} := \mathcal{E}(u), \qquad z_{\text{noise}} := J_{\mathcal{E}}(u)\epsilon. \tag{44}$$

Conditioned on the clean input u, the perturbation term z_{noise} acts as an additive random component superimposed on the physical latent representation z_{phys}. Since NOIR applies per-source standardization following Eq. (40), the resulting latent channels are approximately zero-mean and unit-variance with weak residual cross-channel dependence. The relevant object is therefore the empirical distribution of each standardized latent channel across the input-query ensemble rather than the raw latent vector itself.

Under the additive perturbation model in Eq. (44), corruption effectively smooths the distribution of each latent channel. From an information-theoretic perspective, additive perturbations are known to reduce higher-order non-Gaussian structure and move distributions toward Gaussianity. Consequently, after standardization, corrupted latent channels are expected to exhibit lower excess non-Gaussianity relative to the standard Gaussian reference used in the FastICA objective. Equivalently, corruption is expected to reduce channelwise negentropy:

$$J(\hat{S}_{\ell,\text{corr}}^{(c)}) \;\lesssim\; J(\hat{S}_{\ell,\text{clean}}^{(c)}).$$

This interpretation provides a mechanistic explanation for the NOIR detection criterion. During training, the ICA objective explicitly promotes statistically independent and non-Gaussian latent

channels for clean data. Corruption perturbs this learned latent structure through additive smoothing and partial disruption of the independence pattern, thereby producing measurable deviations from the clean latent baseline. NOIR therefore monitors latent Z-scores as intrinsic indicators of corruption.

The Gaussianization interpretation is most reliable in the mild-to-moderate corruption regime, where the local linearization in Eq. (43) remains accurate and the latent covariance structure is approximately preserved after standardization. At higher corruption levels, nonlinear encoder effects, structured perturbations, and covariance distortion may weaken the additive perturbation approximation.

A.4 Generalization to kernel-based neural operators

While DeepONet employs a branch-trunk factorization to learn operator mappings, FNO leverages spectral convolutions in the Fourier domain to capture global dependencies. Despite these architectural differences, both frameworks share a common structure: a final linear projection that maps learned representations to output predictions. This structure enables the application of our ICA constraints to kernel-based operators with minimal modification.

The FNO architecture processes input functions through a sequence of Fourier layers, each consisting of a spectral convolution operator $\mathcal{K}$ and a local linear transformation $\mathcal{W}$. For a 2D spatial domain, the ℓ-th layer computes:

$$\mathbf{v}_{\ell+1}(\mathbf{x}) = \sigma\left(\mathcal{K}_\ell(\mathbf{v}_\ell)(\mathbf{x}) + \mathcal{W}_\ell(\mathbf{v}_\ell)(\mathbf{x})\right), \tag{45}$$

where $\mathbf{v}_\ell : \Omega \to \mathbb{R}^{d_\ell}$ is the feature function at layer ℓ, σ is a nonlinear activation, and $\Omega \subset \mathbb{R}^2$ is the spatial domain. The spectral convolution $\mathcal{K}_\ell$ operates in the Fourier domain:

$$\mathcal{K}_\ell(\mathbf{v})(\mathbf{x}) = \mathcal{F}^{-1}\left(\mathbf{R}_\ell \cdot (\mathcal{F}\mathbf{v})\right)(\mathbf{x}), \tag{46}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier transform and its inverse, and $\mathbf{R}_\ell \in \mathbb{C}^{d_{\ell+1} \times d_\ell \times M_1 \times M_2}$ is a learnable tensor of Fourier coefficients truncated to the first $M_1 \times M_2$ modes. The local transformation $\mathcal{W}_\ell$ is implemented as a pointwise 1×1 convolution.

After L Fourier layers, the network applies a sequence of fully connected layers to project the learned representation to the output space. Let $\mathbf{v}_L(\mathbf{x}) \in \mathbb{R}^{d_L}$ denote the final feature representation at spatial location $\mathbf{x}$. The output is computed as:

$$\mathbf{u}(\mathbf{x}) = \mathbf{W}_2\sigma(\mathbf{W}_1\mathbf{v}_L(\mathbf{x}) + \mathbf{b}_1) + \mathbf{b}_2, \tag{47}$$

where $\mathbf{W}_1 \in \mathbb{R}^{d_{\text{hidden}} \times d_L}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times d_{\text{hidden}}}$ are weight matrices, $\mathbf{b}_1, \mathbf{b}_2$ are optional bias vectors, and C is the number of output channels.

To apply our disentanglement framework to FNO, we modify the final projection layers to expose the intermediate component structure. Let $\mathbf{z}(\mathbf{x}) \coloneqq \sigma(\mathbf{W}_1\mathbf{v}_L(\mathbf{x}) + \mathbf{b}_1) \in \mathbb{R}^{d_{\text{hidden}}}$ denote the penultimate feature representation. The standard output computation in Eq. (47) can be rewritten as:

$$u_c(\mathbf{x}) = \sum_{i=1}^{d_{\text{hidden}}} [\mathbf{W}_2]_{ci} \cdot z_i(\mathbf{x}) = \mathbf{w}_c^\top \mathbf{z}(\mathbf{x}), \tag{48}$$

where $\mathbf{w}_c \in \mathbb{R}^{d_{\text{hidden}}}$ is the c-th row of $\mathbf{W}_2$, representing the linear functional that produces the c-th output channel.

This decomposition reveals a structure that each output channel c is computed as a weighted sum over d_{hidden} latent components. To expose these components explicitly, we compute the element-wise product:

$$\Phi_{c,i}(\mathbf{x}) = [\mathbf{W}_2]_{ci} \cdot z_i(\mathbf{x}), \tag{49}$$

representing the contribution of the i-th latent dimension to the c-th output channel at location $\mathbf{x}$.

For a discretized spatial domain with $N_x \times N_y$ grid points and a batch of N_b input functions, we construct the component tensor $\mathbf{\Phi} \in \mathbb{R}^{N_b \times N_x \times N_y \times C \times d_{\text{hidden}}}$, where:

$$\Phi_{bjxyci} = [\mathbf{W}_2]_{ci} \cdot z_i(\mathbf{x}_j, \mathbf{y}_x), \tag{50}$$

with $\mathbf{x}_j = (x_j, y_x) \in \Omega$ denoting the spatial grid point at indices (j, x) for the b-th batch sample.

To enforce statistical independence among the latent components, we introduce an unmixing layer $L_{\text{ICA}} : \mathbb{R}^{d_{\text{hidden}}} \to \mathbb{R}^{d_{\text{ica}}}$, parameterized by $\mathbf{A} \in \mathbb{R}^{d_{\text{ica}} \times d_{\text{hidden}}}$. This layer is applied to the last dimension of $\mathbf{\Phi}$:

$$\mathbf{\Phi}_{\text{ICA}} = L_{\text{ICA}}(\mathbf{\Phi}) \in \mathbb{R}^{N_b \times N_x \times N_y \times C \times d_{\text{ica}}}, \tag{51}$$

and the final output is obtained by summing over the transformed latent dimension:

$$u_c(\mathbf{x}_j) = \sum_{k=1}^{d_{\text{ica}}} [\mathbf{\Phi}_{\text{ICA}}]_{bjxck}. \tag{52}$$

A.4.1 ICA Loss for kernel-based operators

The ICA regularization for FNO follows the same per-channel formulation as in Eq. (36). For each output channel $c \in \{1, \ldots, C\}$, we extract the slice $\mathbf{\Phi}_{\text{ICA}}^{(c)} \in \mathbb{R}^{B \times N_x \times N_y \times d_{\text{ica}}}$ and reshape it to:

$$\mathbf{S}_{\text{FNO}}^{(c)} = \text{reshape}(\mathbf{\Phi}_{\text{ICA}}^{(c)}, (N_b \cdot N_x \cdot N_y, d_{\text{ica}})) \in \mathbb{R}^{N_{\text{spatial}}^{(c)} \times d_{\text{ica}}}, \tag{53}$$

where $N_{\text{spatial}}^{(c)} = N_b \cdot N_x \cdot N_y$ represents the total number of spatial samples across the batch for output channel c. Each row of $\mathbf{S}_{\text{FNO}}^{(c)}$ corresponds to a spatial location $\mathbf{x}$ in the discretized domain, and each column represents one of the d_{ica} source components.

Following the same centering, whitening, and negentropy approximation procedure, we compute:

$$\mathcal{L}_{\text{ICA-FNO}}^{(c)} \coloneqq \sum_{k=1}^{d_{\text{ica}}} \left(\frac{1}{N_{\text{spatial}}^{(c)}} \sum_{j=1}^{N_{\text{spatial}}^{(c)}} \mathcal{G}\left([\tilde{\mathbf{S}}_{\text{FNO,norm}}^{(c)}]_{jk} \right) - \mathbb{E}_\nu[\mathcal{G}(\nu)] \right)^2, \tag{54}$$

and the total FNO-ICA loss is:

$$\mathcal{L}_{\text{ICA-FNO}} \coloneqq \frac{1}{C} \sum_{c=1}^{C} \mathcal{L}_{\text{ICA-FNO}}^{(c)}. \tag{55}$$

A.5 Robustness via Latent Decorrelation

A.5.1 Proof of Lemma 4.3

Proof. From the definition of $\mathcal{G}_\theta$,

$$\mathcal{G}_\theta(u)(\mathbf{y}) = \sum_{k=1}^{p} s_k(u_{X_s}) \psi_k(\mathbf{y}),$$

and therefore

$$\mathcal{G}_\theta(u + \epsilon)(\mathbf{y}) = \sum_{k=1}^{p} s_k(u_{X_s} + \epsilon) \psi_k(\mathbf{y}).$$

Hence,

$$\Delta = \mathcal{G}_\theta(u + \epsilon)(\mathbf{y}) - \mathcal{G}_\theta(u)(\mathbf{y}) = \sum_{k=1}^{p} \Big(s_k(u_{X_s} + \epsilon) - s_k(u_{X_s}) \Big) \psi_k(\mathbf{y}).$$

Using the definition

$$\delta_k := s_k(u_{X_s} + \epsilon) - s_k(u_{X_s}),$$

we obtain

$$\Delta = \sum_{k=1}^{p} \delta_k \psi_k(\mathbf{y}).$$

Expanding the squared Euclidean norm,

$$\|\Delta\|_2^2 = \left\| \sum_{k=1}^{p} \delta_k \psi_k(\mathbf{y}) \right\|_2^2.$$

Using the identity

$$\left\| \sum_{k=1}^{p} a_k \right\|_2^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} \langle a_i, a_j \rangle,$$

with

$$a_k := \delta_k \psi_k(\mathbf{y}),$$

gives

$$\|\Delta\|_2^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} \langle \delta_i \psi_i(\mathbf{y}), \delta_j \psi_j(\mathbf{y}) \rangle .$$

Since scalar multiplication factors out of the inner product,

$$\|\Delta\|_2^2 = \sum_{i=1}^{p} \sum_{j=1}^{p} \delta_i^\top \delta_j \ \langle \psi_i(\mathbf{y}), \psi_j(\mathbf{y}) \rangle .$$

Separating diagonal and off-diagonal terms yields

$$\|\Delta\|_2^2 = \sum_{k=1}^{p} \|\psi_k(\mathbf{y})\|_2^2 \|\delta_k\|_2^2 + \sum_{i \neq j} \langle \psi_i(\mathbf{y}), \psi_j(\mathbf{y}) \rangle \, \delta_i^\top \delta_j.$$

Finally, taking expectation with respect to the perturbation ϵ and using linearity of expectation,

$$\mathbb{E}\|\Delta\|_2^2 = \sum_{k=1}^{p} \|\psi_k(\mathbf{y})\|_2^2 \mathbb{E}\|\delta_k\|_2^2 + \sum_{i \neq j} \langle \psi_i(\mathbf{y}), \psi_j(\mathbf{y}) \rangle \, \mathbb{E}[\delta_i^\top \delta_j].$$

This completes the proof. □

A.5.2 Proof of Theorem 4.4

Proof. From Lemma 4.3,

$$\mathbb{E}\|\Delta\|_2^2 = \sum_{k=1}^{p} \|\psi_k(y)\|_2^2 \mathbb{E}\|\delta_k\|_2^2 + \sum_{i \neq j} \langle \psi_i(y), \psi_j(y) \rangle \mathbb{E}[\delta_i^\top \delta_j].$$

Under Assumption 4.1, the latent perturbations are decorrelated, i.e.,

$$\mathbb{E}[\delta_i^\top \delta_j] = 0, \qquad i \neq j.$$

Hence all cross-covariance terms vanish, yielding

$$\mathbb{E}\|\Delta\|_2^2 = \sum_{k=1}^{p} \|\psi_k(y)\|_2^2 \mathbb{E}\|\delta_k\|_2^2.$$

Now apply Proposition 4.2, which states that each latent coordinate satisfies

$$\|\delta_k\|_2 = \|s_k(u_{X_s} + \epsilon) - s_k(u_{X_s})\|_2 \leq L_k \|\epsilon\|_2.$$

Squaring both sides gives

$$\|\delta_k\|_2^2 \leq L_k^2 \|\epsilon\|_2^2.$$

Taking expectation preserves the inequality:

$$\mathbb{E}\|\delta_k\|_2^2 \leq L_k^2 \|\epsilon\|_2^2.$$

Substituting into the previous expression yields

$$\mathbb{E}\|\Delta\|_2^2 \leq \sum_{k=1}^{p} \|\psi_k(y)\|_2^2 L_k^2 \|\epsilon\|_2^2.$$

Factoring out $\|\epsilon\|_2^2$,

$$\mathbb{E}\|\Delta\|_2^2 \leq \left(\sum_{k=1}^{p} \|\psi_k(y)\|_2^2 L_k^2\right) \|\epsilon\|_2^2.$$

Define

$$L^2 := \sum_{k=1}^{p} \|\psi_k(y)\|_2^2 L_k^2.$$

Then

$$\mathbb{E}\|\Delta\|_2^2 \leq L^2 \|\epsilon\|_2^2.$$

Finally, if $\|\epsilon\|_2 \leq \eta$, then

$$\mathbb{E}\|\Delta\|_2^2 \leq \eta^2 L^2.$$

Taking square roots gives

$$\|\mathcal{G}_\theta(u+\epsilon) - \mathcal{G}_\theta(u)\|_2 \leq \eta L.$$

This completes the proof. □

B Detailed experimental setup

2D Darcy Flow. We evaluate NOIR on the steady-state 2D Darcy Flow, an elliptic PDE modeling pressure $u(\mathbf{x})$ distribution in a porous medium with spatially-varying permeability $a(\mathbf{x})$:

$$\begin{aligned} -\nabla \cdot (a(\mathbf{x})\nabla u(\mathbf{x})) &= f(\mathbf{x}), && \mathbf{x} \in (0,1)^2 \\ u(\mathbf{x}) &= 0, && \mathbf{x} \in \partial(0,1)^2 \end{aligned} \tag{56}$$

with uniform source $f(\mathbf{x}) = 1$ and homogeneous Dirichlet boundary conditions. The task is to learn the solution operator $G^\dagger : a(\cdot) \mapsto u(\cdot)$ mapping permeability fields to pressure distributions. Input permeability fields are generated as log-Gaussian Random Fields, $a = \exp(z)$ where $z \sim \mathcal{N}(0, C)$ with Matérn covariance ($s = 2.5, \tau = 7$). Solutions are computed via finite difference discretization on a 29×29 grid. The dataset contains 1,000 training and 200 testing samples.

Flow around Airfoils. To evaluate operator learning on geometry-parameterized systems, we consider transonic flow around NACA-0012 airfoil variants governed by the 2D compressible Euler equations:

$$\begin{aligned} \partial_t \rho + \nabla \cdot (\rho \boldsymbol{v}) &= 0, \\ \partial_t (\rho \boldsymbol{v}) + \nabla \cdot (\rho \boldsymbol{v} \otimes \boldsymbol{v} + p\boldsymbol{I}) &= 0, \\ \partial_t E + \nabla \cdot ((E+p)\boldsymbol{v}) &= 0, \\ \boldsymbol{x} \in \Omega_g, \quad t \in (0, \infty) & \end{aligned} \tag{57}$$

where ρ, $\boldsymbol{v}$, p, E denote density, velocity, pressure, and total energy. Far-field conditions specify $\rho_\infty = 1$, $p_\infty = 1$, Mach number $M_\infty = 0.8$, and zero angle of attack. Airfoil geometries are parameterized via Bézier control points [5], with each shape g defining a domain Ω_g discretized on a body-fitted curvilinear mesh of $\sim 11{,}000$ points. Steady-state solutions are computed using high-fidelity CFD solvers. The objective is to learn the operator $\mathcal{G}^\dagger : g \mapsto M_g(\boldsymbol{x})$ mapping airfoil geometry to local Mach number field at mesh coordinates $\boldsymbol{x} = (x, y)$. Since DeepONet requires fixed functional input for the branch network, we encode geometry via Fourier descriptors [23] with 64 harmonics (128 features), while FNO operates directly on varying mesh coordinates. The dataset contains 2,490 samples with 1,000 used for training and 200 for testing.

1D Burgers' Equation The 1-dimensional Burgers' equation is a canonical non-linear PDE modeling phenomena such as one-dimensional fluid flow and shock formation. We consider the viscous Burgers' equation with periodic boundary conditions:

$$\begin{aligned} \partial_t u + u\partial_x u &= \nu \partial_{xx} u, && x \in (0,1), t \in (0,1] \\ u(x,0) &= u_0(x), && x \in (0,1) \end{aligned} \tag{58}$$

where $u(x,t)$ is the velocity field, u_0 is the initial condition, and $\nu \in \mathbb{R}_+$ is the viscosity coefficient. The initial conditions u_0 are drawn from a Gaussian Random Field (GRF) with zero mean and

a covariance structure defined by a Matérn kernel, following the setup in [14, 18]. Specifically, $u_0 \sim \mathcal{N}(0, C)$ where $C = (-\Delta + \tau^2)^{-s}$ with $s = 2.5$ and $\tau = 7$. We use the viscosity $\nu = 0.01$. The objective is to learn the solution operator $G^\dagger : u_0(\cdot) \mapsto u(\cdot, 1)$, which maps the initial state $u_0 \in L^2_{\text{per}}((0,1))$ to the final solution state $u(\cdot, 1) \in L^2_{\text{per}}((0,1))$. The training dataset consists of 1,000 input-output pairs, and we test on a separate set of 200 pairs.

All contamination strategies are governed by a noise level parameter $\delta \in \{0.0, 0.1, 0.2, 0.5, 1.0\}$, which controls the magnitude of contamination relative to the intrinsic variability of the training data. We first compute the standard deviation of input features across the entire training dataset, denoted σ_{data}. The target standard deviation for any generated noise field is:

$$\sigma_{\text{noise}} = \delta \cdot \sigma_{\text{data}}. \tag{59}$$

For a clean input function $\mathbf{u}_{\text{clean}}$, the contaminated input is:

$$\mathbf{u}_{\text{noisy}} = \mathbf{u}_{\text{clean}} + \boldsymbol{\epsilon}_{\text{noise}}, \tag{60}$$

where $\boldsymbol{\epsilon}_{\text{noise}}$ is the noise vector generated according to one of five strategies designed to emulate distinct failure modes:

Spatial Harmonic noise. Persistent, structured interference patterns (e.g., fixed electromagnetic noise). Defined as:

$$\boldsymbol{\epsilon}_{\text{noise}}(\mathbf{x}, t) = \sigma_{\text{noise}} \cdot \sin(k_x x + k_y y + \varphi) \tag{61}$$

with wave numbers $k_x, k_y \sim \text{Uniform}[1, 16]$ and phase $\varphi \sim \text{Uniform}[0, 2\pi]$ randomly sampled per test instance.

Band-limited noise. Spectral contamination via Fourier domain filtering with cutoff fraction $f_c = 0.1$. Two variants:

- **Low-pass**: Smooth, large-scale perturbations. Apply Gaussian noise in spatial domain, then Fourier transform, zero out modes with $|\mathbf{k}| > f_c \cdot k_{\max}$, and inverse transform.
- **High-pass**: Rapid, fine-scale fluctuations. Apply Gaussian noise, then zero out modes with $|\mathbf{k}| < f_c \cdot k_{\max}$.

Unstructured Gaussian noise. Random, uncorrelated sensor failures. Governed by noise level δ and noise fraction $\phi \in \{0.25, 0.5, 1.0\}$. At each spatial location, noise is applied with probability ϕ:

$$[\boldsymbol{\epsilon}_{\text{noise}}]_i = \begin{cases} \mathcal{N}(0, \sigma^2_{\text{noise}}) & \text{with probability } \phi \\ 0 & \text{otherwise} \end{cases} \tag{62}$$

Localized Gaussian Block noise. Localized contamination over a contiguous $25\% \times 25\%$ spatial region with randomized location:

$$[\boldsymbol{\epsilon}_{\text{noise}}]_{i,j} = \begin{cases} \mathcal{N}(0, \sigma^2_{\text{noise}}) & \text{if } (i,j) \in \text{Block} \\ 0 & \text{otherwise} \end{cases} \tag{63}$$

where Block location is uniformly sampled from valid positions ensuring the block remains within domain boundaries.

Critically, target outputs remain the clean, ground-truth solutions corresponding to $\mathbf{u}_{\text{clean}}$, ensuring evaluation measures the model's ability to recover correct operator outputs despite input corruption.

C Training specifications

C.1 Model architectures

Table 3 reports architectural hyperparameters for each benchmark and model configuration. Following the architecture specifications, all computations are done with Nvidia A100 GPU. All models employ the ReLU activation function. Training is conducted using the Adam optimizer with a learning rate of 1×10^{-3}, a weight decay factor of 1×10^{-4}, and a cosine annealing learning rate scheduler with $T_{\max} = 1500$, for a total of 2000 epochs in each training run.

Table 3: Consolidated architecture specifications across 1D Burgers', 2D Darcy Flow, and 2D NACA Airfoil benchmarks. All configurations utilize ReLU activations with a hidden dimension (HD) of 128 for projection layers.

Model Category	Architecture Details	Params
1D Burgers' Equation		
DeepONet	Branch: $[1024, 128^3, 128]$; Trunk: $[1, 128^3, 128]$	230,784
NOMAD	Branch: $[1024, 128^3, 128]$; Combined: $[256, 128^2, 1]$	280,065
FNO	4 Spectral Layers (Width 32), Modes: 12	221,889
WNO	4 Wavelet Layers (Width 32), Level 5, Omega 4	320,193
DeepONet + NOIR	Baseline + ICA Layer: $[128, 256]$	263,808 (+14.3%)
FNO + NOIR	Baseline + ICA Layer: $[128, 256]$	254,913 (+14.9%)
2D Darcy Flow		
DeepONet	Branch: $[841, 128^3, 128]$; Trunk: $[2, 128^3, 128]$	207,488
NOMAD	Combined: $[256, 128^2, 1]$; Trunk: $[2, 128^3]$	256,769
FNO	3 Fourier Layers (Width 32), Lifting: $[3, 32]$	892,385
WNO	3 Wavelet Layers (Width 32), Level 5	87,521
DeepONet + NOIR	Baseline + ICA Layer: $[128, 32]$	211,616 (+1.9%)
FNO + NOIR	Baseline + ICA Layer: $[128, 32]$	896,513 (+0.4%)
2D NACA Airfoil		
DeepONet	Branch: $[128, 128^3, 128]$; Trunk: $[2, 128^3, 128]$	116,224
NOMAD	Branch: $[128, 128^3, 128]$; Combined: $[256, 128^2, 1]$	165,505
FNO	2 Spectral Layers (Width 32), Modes: 12×8	597,441
DeepONet + NOIR	Baseline + ICA Layer: $[128, 24]$	119,320 (+2.7%)
FNO + NOIR	Baseline + ICA Layer: $[128, 24]$	600,537 (+0.5%)

C.2 Hyperparameter sensitivity

Table 4 reports performance for key hyperparameter variations on the Darcy Flow benchmark.

Table 4: Hyperparameter sensitivity (DON+N, 2D Darcy Flow, high-pass noise). Unscaled Relative L_2 Errors are reported per hyperparameter configuration

ICA Dimensionality ($\lambda_{\text{ICA}} = 0.01$)					
p_{ICA}	$\delta = 0.0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 1.0$
16	0.0675	0.0682	0.0697	0.0774	0.0981
24	0.0671	0.0678	0.0692	0.0767	0.0958
32	0.0668	0.0672	0.0682	0.0748	0.0938
64	0.0672	0.0675	0.0683	0.0728	0.0857
128	0.0642	0.0646	0.0656	0.0721	0.0901
Regularization Strength ($p_{\text{ICA}} = 24$)					
λ_{ICA}	$\delta = 0.0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 1.0$
0.001	0.0642	0.0646	0.0656	0.0721	0.0901
0.01	0.0671	0.0678	0.0692	0.0767	0.0958
0.1	0.0679	0.0684	0.0697	0.0773	0.0979
1.0	0.0671	0.0678	0.0692	0.0767	0.0958

Performance is stable across reasonable parameter ranges, with degradation only at extreme regularization ($\lambda \geq 1.0$).

D Additional experimental results

D.1 Noise sensitivity

Table 5-7 we present detailed results on noise sensitivity with and without the proposed structure.

Table 5: Performance on the NACA airfoil dataset. Relative L_2 error ($\times 10^{-2}$) is reported across all corruption magnitudes $\delta \in \{0, 0.1, 0.2, 0.5, 1.0\}$. Each entry shows mean $\pm$ standard deviation over seeds. Sens. denotes the sensitivity at $\delta = 1.0$. DON+N denotes DeepONet+NOIR; FNO+N denotes FNO+NOIR.

Noise	Model	Rel. L_2 Error ($\times 10^{-2}$)					Sens.
		$\delta = 0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 1.0$	
HP	DON	7.25±0.07	16.29±0.74	17.04±0.93	17.06±0.86	17.15±0.87	2.37
	DON+N	7.15±0.09	16.68±1.04	17.06±0.86	16.87±0.72	16.90±0.77	2.36
	FNO	1.06±0.03	23.14±7.61	48.10±17.78	93.32±36.09	121.89±47.15	115.18
	FNO+N	1.04±0.05	20.10±3.02	35.93±5.96	58.03±12.57	69.42±15.33	66.44
	NOMAD	7.47±0.10	25.72±4.22	50.46±15.15	147.63±77.32	344.83±211.19	46.17
LP	DON	7.25±0.07	16.52±0.76	17.07±0.89	17.19±0.96	17.18±0.92	2.37
	DON+N	7.15±0.09	16.56±0.88	17.09±0.85	16.95±0.64	16.92±0.67	2.37
	FNO	1.06±0.03	73.45±14.99	128.33±31.22	269.85±71.96	491.77±131.97	464.71
	FNO+N	1.04±0.05	64.89±8.30	108.20±18.92	225.31±49.70	417.09±102.35	399.14
	NOMAD	7.47±0.10	29.91±7.48	55.40±18.76	157.44±75.42	355.75±202.02	47.63
SH	DON	7.25±0.07	16.61±0.60	17.05±0.94	16.84±0.66	17.00±0.72	2.35
	DON+N	7.15±0.09	16.50±0.80	16.92±0.68	16.29±0.70	16.51±0.76	2.31
	FNO	1.06±0.03	18.24±5.17	36.74±11.75	82.54±28.57	123.45±39.81	116.65
	FNO+N	1.04±0.05	20.40±5.93	35.52±6.27	66.95±11.52	94.89±15.94	90.81
	NOMAD	7.47±0.10	28.19±7.78	56.91±24.05	192.84±81.48	456.58±244.68	61.13
UGR	DON	7.25±0.07	16.27±0.70	16.98±0.89	17.10±1.04	17.07±0.87	2.35
	DON+N	7.15±0.09	16.67±1.10	17.06±0.80	16.89±0.70	16.84±0.73	2.36
	FNO	1.06±0.03	25.04±6.82	49.73±16.96	96.12±34.99	129.11±45.50	122.00
	FNO+N	1.04±0.05	21.84±2.76	37.82±6.45	62.85±13.31	80.18±16.27	76.73
	NOMAD	7.47±0.10	25.56±4.08	51.27±15.58	148.94±75.17	341.61±212.18	45.74
LGB	DON	7.25±0.07	9.77±0.25	10.06±0.30	10.07±0.34	10.10±0.34	1.39
	DON+N	7.15±0.09	9.72±0.43	9.94±0.41	10.00±0.31	9.96±0.33	1.39
	FNO	1.06±0.03	7.19±0.46	11.69±1.09	24.31±4.43	44.60±9.97	42.15
	FNO+N	1.04±0.05	7.20±0.29	11.18±0.30	22.47±2.10	41.77±7.33	39.97
	NOMAD	7.47±0.10	11.29±1.37	12.28±2.12	13.22±2.85	13.37±2.98	1.79

Table 6: Performance on the Burgers' equation dataset. Relative L_2 error ($\times 10^{-2}$) is reported across all corruption magnitudes $\delta \in \{0, 0.1, 0.2, 0.5, 1.0\}$. Each entry shows mean $\pm$ standard deviation over seeds. Sens. denotes the sensitivity at $\delta = 1.0$. DON+N denotes DeepONet+NOIR; FNO+N denotes FNO+NOIR.

Noise	Model	Rel. L_2 Error ($\times 10^{-2}$)					Sens.
		$\delta = 0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 1.0$	
HP	DON	1.86±0.55	1.91±0.53	2.05±0.49	2.74±0.39	4.27±0.50	2.29
	DON+N	1.28±0.12	1.31±0.12	1.38±0.14	1.80±0.21	2.77±0.39	2.16
	FNO	0.08±0.01	0.08±0.01	0.09±0.01	0.15±0.01	0.26±0.03	3.27
	FNO+N	0.12±0.01	0.13±0.01	0.15±0.02	0.25±0.03	0.45±0.05	3.63
	WNO	0.31±0.02	0.33±0.03	0.39±0.03	0.65±0.07	1.22±0.16	3.90
	NOMAD	24.42±5.31	24.42±5.31	24.42±5.31	24.42±5.31	24.42±5.31	1.00
LP	DON	1.86±0.55	2.10±0.50	2.63±0.41	4.84±0.25	8.99±0.22	4.82
	DON+N	1.28±0.12	1.57±0.10	2.19±0.08	4.53±0.08	8.74±0.15	6.83
	FNO	0.08±0.01	0.85±0.01	1.69±0.03	4.23±0.07	8.45±0.14	104.85
	FNO+N	0.12±0.01	0.86±0.01	1.70±0.03	4.24±0.07	8.47±0.13	67.80
	WNO	0.31±0.02	0.93±0.02	1.77±0.03	4.37±0.08	8.72±0.15	27.98
	NOMAD	24.42±5.31	24.42±5.30	24.43±5.28	24.53±5.16	24.87±4.74	1.02
SH	DON	1.86±0.55	2.72±0.41	4.24±0.29	9.44±0.22	18.32±0.51	9.83
	DON+N	1.28±0.12	2.32±0.07	3.99±0.08	9.42±0.21	18.57±0.47	14.50
	FNO	0.08±0.01	1.87±0.04	3.73±0.09	9.33±0.22	18.64±0.44	231.21
	FNO+N	0.12±0.01	1.87±0.04	3.73±0.09	9.33±0.22	18.65±0.44	149.25
	WNO	0.31±0.02	1.89±0.05	3.74±0.09	9.32±0.22	18.63±0.43	59.79
	NOMAD	24.42±5.31	24.44±5.28	24.51±5.19	24.90±4.70	25.88±3.47	1.06
UGR	DON	1.86±0.55	1.93±0.53	2.12±0.49	3.03±0.39	4.98±0.44	2.67
	DON+N	1.28±0.12	1.34±0.12	1.49±0.12	2.24±0.16	3.83±0.25	2.99
	FNO	0.08±0.01	0.28±0.01	0.55±0.01	1.35±0.03	2.68±0.05	33.30
	FNO+N	0.12±0.01	0.30±0.01	0.56±0.01	1.36±0.03	2.71±0.06	21.71
	WNO	0.31±0.02	0.43±0.02	0.67±0.01	1.52±0.03	3.00±0.08	9.63
	NOMAD	24.42±5.31	24.42±5.31	24.42±5.30	24.44±5.29	24.48±5.24	1.00
LGB	DON	1.86±0.55	1.88±0.54	1.93±0.53	2.23±0.46	2.99±0.38	1.60
	DON+N	1.28±0.12	1.30±0.12	1.34±0.12	1.58±0.12	2.21±0.15	1.73
	FNO	0.08±0.01	0.15±0.01	0.27±0.01	0.63±0.03	1.25±0.06	15.49
	FNO+N	0.12±0.01	0.18±0.01	0.29±0.01	0.65±0.02	1.27±0.05	10.16
	WNO	0.31±0.02	0.34±0.02	0.41±0.02	0.73±0.03	1.36±0.05	4.36
	NOMAD	24.42±5.31	24.42±5.31	24.42±5.31	24.42±5.31	24.43±5.29	1.00

Table 7: Performance on the Darcy flow dataset. Relative L_2 error ($\times 10^{-2}$) is reported across all corruption magnitudes $\delta \in \{0, 0.1, 0.2, 0.5, 1.0\}$. Each entry shows mean $\pm$ standard deviation over seeds. Sens. denotes the sensitivity at $\delta = 1.0$. DON+N denotes DeepONet+NOIR; FNO+N denotes FNO+NOIR.

Noise	Model	Rel. L_2 Error ($\times 10^{-2}$)					Sens.
		$\delta = 0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.5$	$\delta = 1.0$	
HP	DON	8.19±0.83	8.21±0.82	8.26±0.79	8.57±0.63	9.50±0.37	1.16
	DON+N	6.42±0.19	6.46±0.21	6.55±0.23	7.12±0.32	8.73±0.55	1.36
	FNO	2.70±0.33	2.89±0.43	3.42±0.69	6.97±1.66	17.02±2.79	6.30
	FNO+N	2.47±0.12	2.52±0.12	2.70±0.12	4.29±0.22	9.20±0.50	3.72
	WNO	4.22±0.11	6.71±2.04	14.34±6.63	81.79±90.08	284.69±387.38	67.49
	NOMAD	6.31±0.27	6.35±0.27	6.44±0.26	7.00±0.24	8.59±0.35	1.36
LP	DON	8.19±0.83	8.59±0.74	9.72±0.54	15.58±0.70	28.87±2.11	3.53
	DON+N	6.42±0.19	7.02±0.12	8.48±0.16	14.87±0.35	25.98±0.87	4.05
	FNO	2.70±0.33	3.39±0.23	4.90±0.33	12.35±0.85	28.64±3.46	10.61
	FNO+N	2.47±0.12	3.40±0.15	5.27±0.27	13.03±0.80	29.13±2.00	11.79
	WNO	4.22±0.11	6.91±1.40	14.21±4.92	68.15±61.88	251.77±302.91	59.68
	NOMAD	6.31±0.27	6.82±0.22	8.07±0.21	13.48±0.47	21.52±0.72	3.41
SH	DON	8.19±0.83	8.20±0.84	8.24±0.83	8.52±0.78	9.41±0.57	1.15
	DON+N	6.42±0.19	6.44±0.19	6.52±0.19	6.99±0.22	8.38±0.34	1.31
	FNO	2.70±0.33	2.83±0.37	3.19±0.49	5.50±0.99	13.12±1.95	4.86
	FNO+N	2.47±0.12	2.56±0.13	2.81±0.15	4.48±0.21	9.65±0.38	3.91
	WNO	4.22±0.11	5.47±1.22	8.63±3.03	29.57±15.08	115.60±121.57	27.40
	NOMAD	6.31±0.27	6.32±0.27	6.38±0.28	6.78±0.32	8.01±0.40	1.27
UGR	DON	8.19±0.83	8.20±0.83	8.25±0.81	8.60±0.70	9.71±0.41	1.19
	DON+N	6.42±0.19	6.45±0.23	6.55±0.27	7.17±0.40	8.94±0.61	1.39
	FNO	2.70±0.33	2.89±0.44	3.43±0.70	6.99±1.67	17.08±2.76	6.32
	FNO+N	2.44 ± 0.09	2.50 ± 0.08	2.72 ± 0.07	4.61 ± 0.34	9.91 ± 0.26	4.06
	WNO	4.22±0.11	6.70±2.03	14.29±6.49	81.64±89.57	284.80±386.51	67.52
	NOMAD	6.31±0.27	6.35±0.27	6.44±0.27	7.06±0.27	8.79±0.41	1.39
LGB	DON	8.19±0.83	8.19±0.83	8.19±0.83	8.21±0.81	8.29±0.77	1.01
	DON+N	6.42±0.19	6.42±0.19	6.43±0.19	6.48±0.19	6.64±0.20	1.03
	FNO	2.70±0.33	2.71±0.34	2.76±0.36	3.08±0.40	4.20±0.81	1.55
	FNO+N	2.47±0.12	2.47±0.12	2.48±0.12	2.59±0.13	3.10±0.19	1.25
	WNO	4.22±0.11	4.36±0.16	5.16±0.89	17.27±17.28	59.41±78.45	14.08
	NOMAD	6.31±0.27	6.31±0.27	6.32±0.27	6.37±0.25	6.52±0.23	1.03

D.2 Performance degradation heatmap

Figures 5–7 present comprehensive error landscapes across all benchmarks, visualizing baseline-to-maximum error transitions (Baseline / Max) for each model-noise combination. Color gradients (red to green) encode degradation severity.

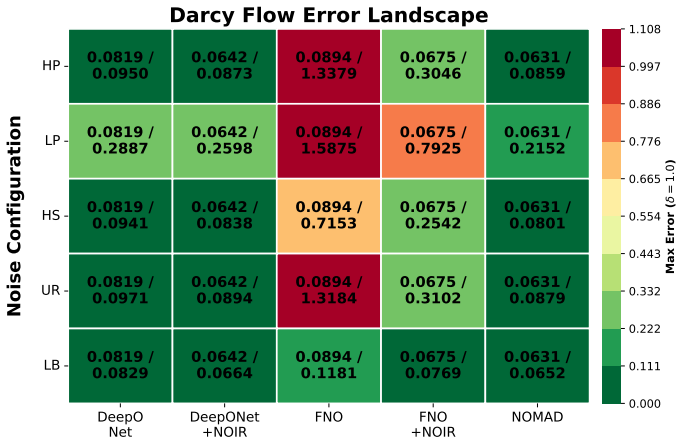


Figure 5: **Error landscape for Darcy Flow.** A quantitative breakdown of predictive degradation. Cell values report the transition from baseline error to maximum noise error.

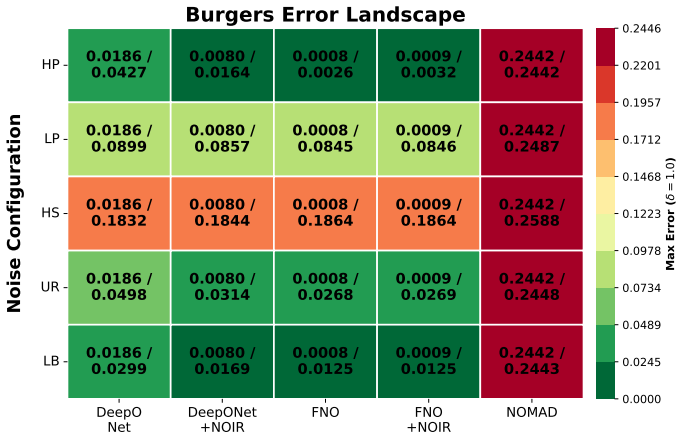


Figure 6: **Error landscape for Burgers' Equation.** Heatmap showing the transition from baseline to maximum error for each model and noise type on the Burgers' equation benchmark.

D.3 Qualitative results: 2D Darcy Flow

Figures 8–12 present complete visualization results for the Darcy Flow benchmark, demonstrating consistent architectural vulnerability patterns across various contamination strategies.

D.4 Qualitative results: Airfoil

Figures 13–17 present complete visualization results for the Airfoil benchmark, demonstrating severe FNO vulnerability to spectral contamination and NOIR's partial mitigation effects.

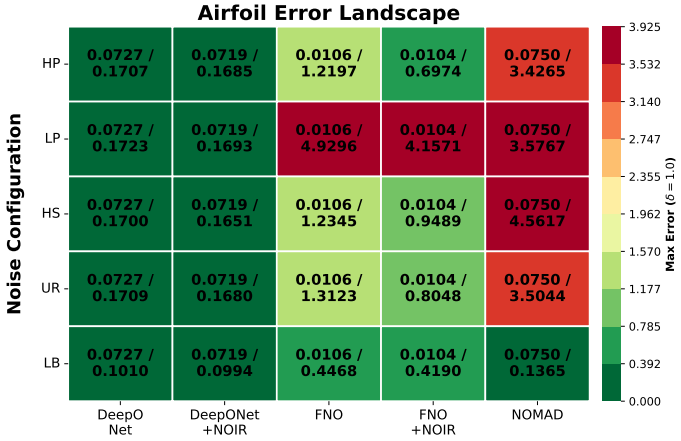


Figure 7: **Error landscape for NACA Airfoil.** Heatmap showing the transition from baseline to maximum error for each model and noise type on the NACA Airfoil benchmark.

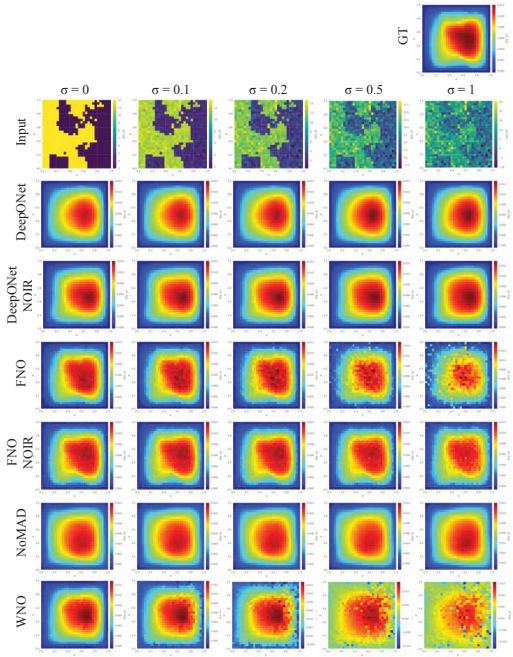


Figure 8: **Darcy Flow under high-frequency contamination.** Pressure field predictions when permeability inputs are corrupted with bandlimited high-pass (HP) noise ($\delta = 1.0$, $f_c = 0.1$). The color bars share the same minimum and maximum values within each model. Ground truth (top right inset) shows the target solution. At lower noise regime, minimal visual effects are observed while at higher noise regime, structure-based disturbance characters are revealed.

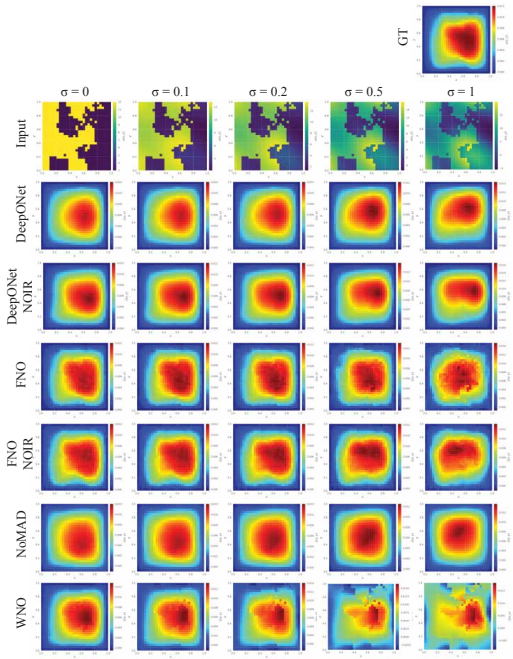


Figure 9: **Darcy Flow under low-frequency contamination.** Pressure field predictions when permeability inputs are corrupted with bandlimited low-pass noise ($\delta = 1.0$, $f_c = 0.1$). The color bars share the same minimum and maximum values within each model. This contamination type introduces smooth, large-scale perturbations that corrupt the learned representational manifold.

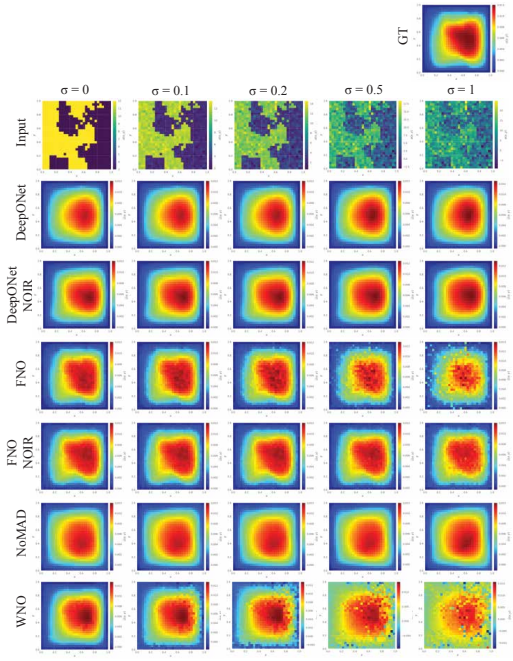


Figure 10: **Darcy Flow under unstructured gaussian noise.** Each row displays the pressure field solutions for a specific model, while columns represent increasing noise magnitudes from $\delta = 0$ (clean) to $\delta = 1$ (maximum). The color bars share the same minimum and maximum values within each model. The ground truth pressure field is shown in the top-right inset. The color bars share the same minimum and maximum values within each model.

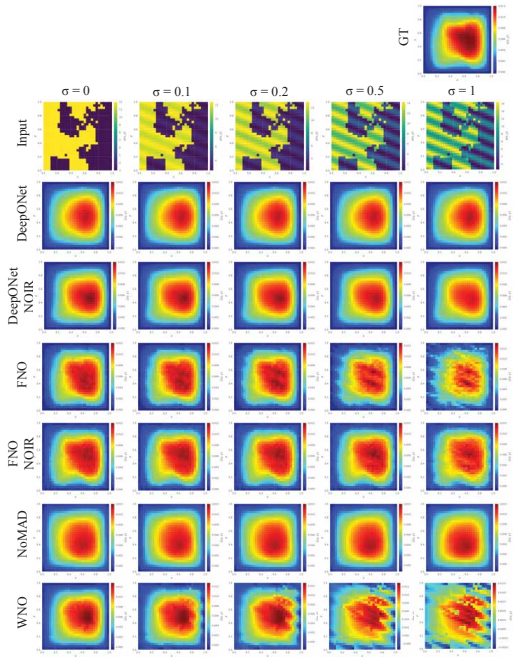


Figure 11: **Darcy Flow under harmonic spatial interference.** Each row displays the pressure field solutions for a specific model, while columns represent increasing noise magnitudes from $\delta = 0$ (clean) to $\delta = 1$ (maximum). The color bars share the same minimum and maximum values within each model. The ground truth pressure field is shown in the top-right inset.

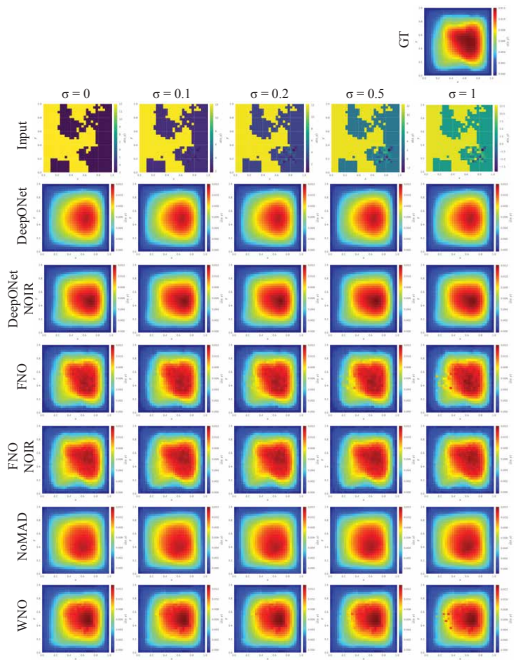


Figure 12: **Darcy Flow under localized spatial block failure.** Pressure field predictions when a contiguous $25\% \times 25\%$ region of permeability measurements is corrupted ($\delta = 1.0$, random location). The color bars share the same minimum and maximum values within each model. This scenario simulates complete sensor array failure in a local region, requiring spatial inpainting capability.

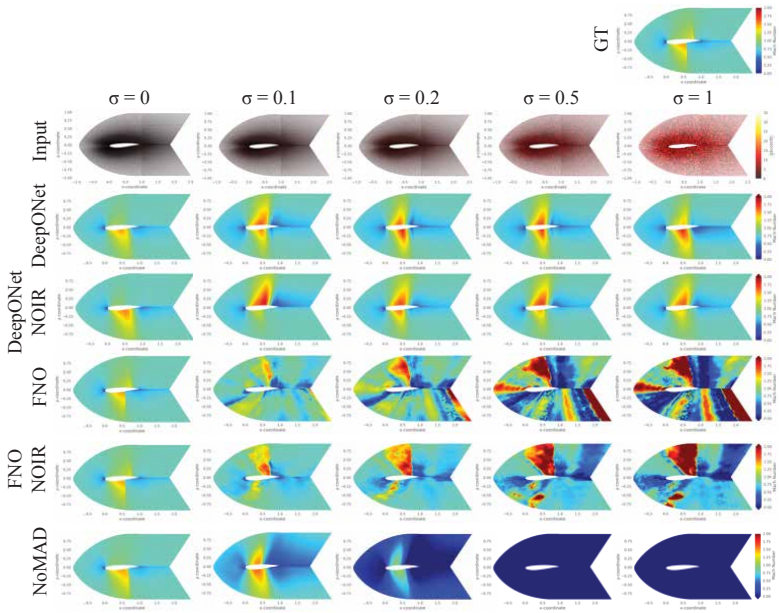


Figure 13: **Airfoil flow field under high-frequency contamination.** Velocity magnitude predictions when geometry/flow boundary conditions are corrupted with bandlimited high-pass noise ($\delta = 1.0$, $f_c = 0.1$). The color bars share the same minimum and maximum values within each model. Ground truth (top right inset) shows the target flow field around NACA airfoil. FNO exhibits catastrophic failure. DeepONet-based models maintain structural stability (sensitivity: $2.35\times$) through coordinate-based representations, yet physically implausible solution for contaminated geometric inputs.

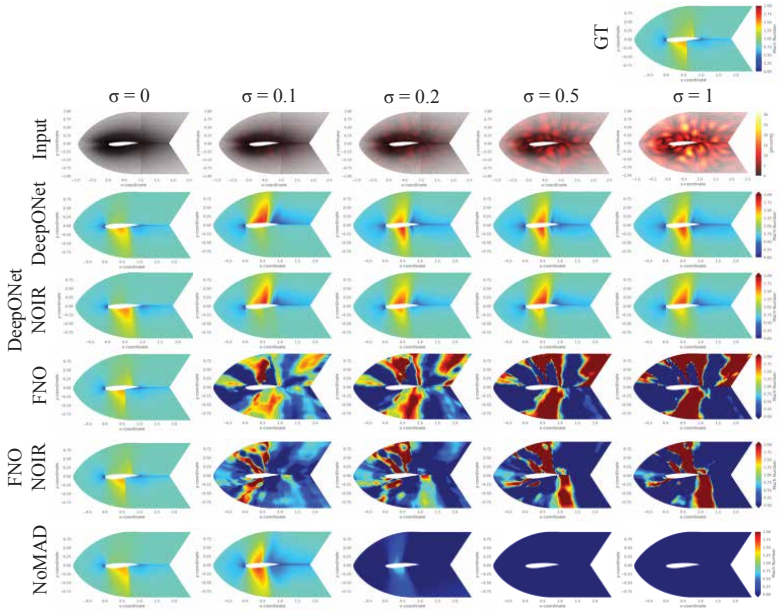


Figure 14: **Airfoil flow field under low-frequency contamination.** Velocity magnitude predictions when inputs are corrupted with bandlimited low-pass noise ($\delta = 1.0$, $f_c = 0.1$). The color bars share the same minimum and maximum values within each model. This represents the smooth, large-scale perturbations fundamentally corrupt the learned solution manifold. FNO demonstrates chaotic collapse, producing entirely unphysical predictions. DeepONet maintains relative stability (sensitivity: $2.37\times$) but shows persistent geometric artifacts. NOMAD exhibits complete collapse, highlighting universal vulnerability.

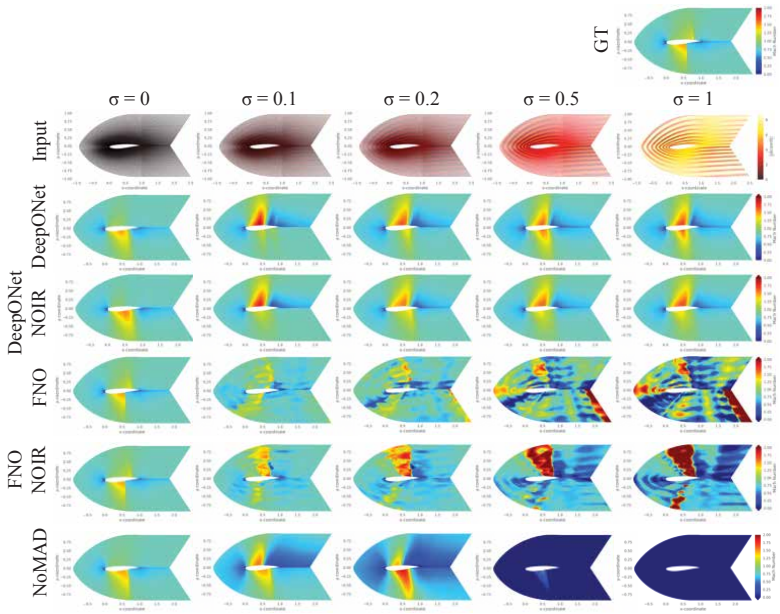


Figure 15: **Airfoil flow field under stationary harmonic interference.** Velocity magnitude predictions with stationary sinusoidal pattern ($\delta = 1.0$, wavelength $\lambda = L/4$) superimposed on geometric inputs. The color bars share the same minimum and maximum values within each model. This structured contamination tests resilience to periodic measurement artifacts or systematic sensor interference. FNO shows severe degradation following the stationalry harmonic noise pattern.

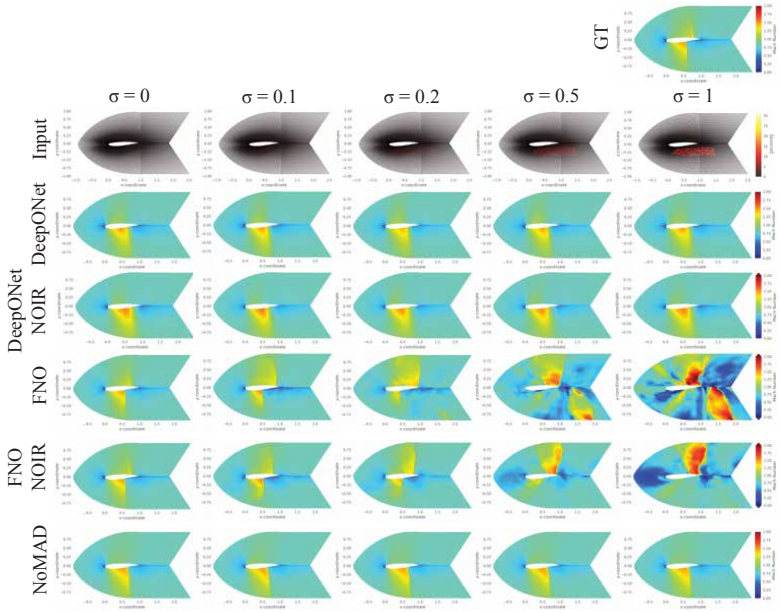


Figure 16: **Airfoil: Localized spatial block failure on flow field predictions.** Ground truth (top right inset) shows the target mach number field around the NACA airfoil. A contiguous $25\% \times 25\%$ region of geometry measurements is corrupted ($\delta = 1.0$, random location). The color bars share the same minimum and maximum values within each model. FNO shows the highest level of degradation as localized corruption disrupts global spectral modes.

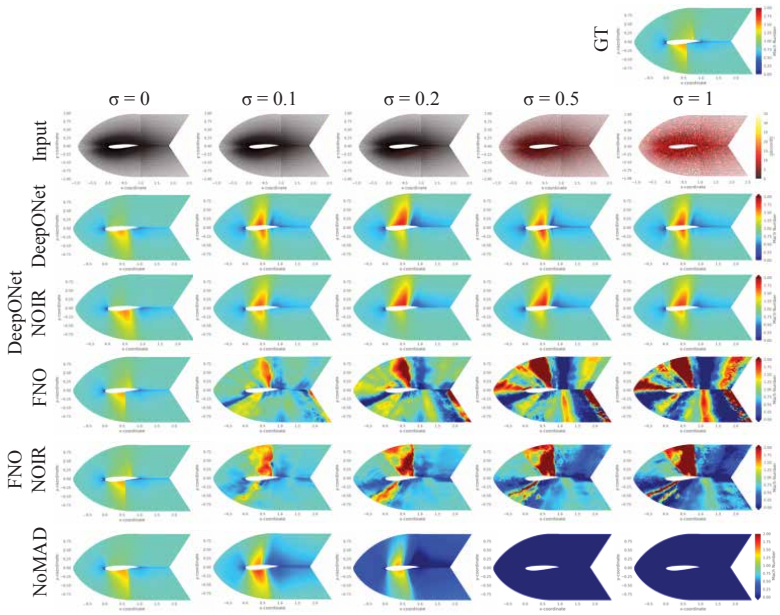


Figure 17: **Airfoil flow field under unstructured gaussian noise.** Velocity magnitude predictions with Unstructured Gaussian contamination ($\delta = 1.0$). The color bars share the same minimum and maximum values within each model. This emulates uncorrelated sensor measurement errors or data acquisition noise in geometric/boundary condition specifications. FNO demonstrates severe vulnerability with random high-frequency components contaminating spectral representations. DeepONet maintains stability (sensitivity: $2.35\times$).

E Extended discussion on related work

This section provides additional detail on how NOIR relates to and differs from existing work in Out-of-Distribution Methods, which was summarized in the main text for brevity.

Recent work has explored robust training frameworks for operator learning through Bayesian approaches [24], ensemble methods [15], and calibration techniques [12]. Unlike these methods, which require multiple forward passes (ensembles) or expensive inference (Bayesian), NOIR explores single-pass latent statistics by exploiting architectural structure.

Relationship to Out-of-Distribution methods. A key clarification precedes this discussion: NOIR is not intended as a standalone anomaly detector. Neural operators are regression surrogates, and the natural evaluation protocol in this literature is based on predictive error, typically relative L_2 error. The Z-score produced by NOIR is a *latent structural statistic* that quantifies independence violations in the learned source-mixture representation, rather than a binary corruption classifier.

With this framing established, NOIR shares the high-level goal of identifying OOD features. However, structural distinctions separate our approach from model-agnostic alternatives such as Mahalanobis distance [4] and energy-based detection [16].

Directly applying Mahalanobis distance to unregularized DeepONet or FNO latent spaces faces a practical obstacle that requires a well-conditioned covariance structure, but vanilla neural operator latent spaces often contain dead feature channels with near-zero variance that dynamically shift during training, inducing numerical instabilities in covariance estimation. This limitation prevents a principled direct comparison on raw latent representations.

However, the source-mixture tensor Ψ (Eqs. 30) for DeepONet and the component tensor Φ (Eqs. 50) for FNO provide a principled, fixed-dimensional representation over which Mahalanobis distance and energy scores can be computed by masking dead feature channels at inference time. We include quantitative comparisons of Mahalanobis distance and energy score applied to the NOIR source-mixture tensor across all three benchmarks (NACA airfoil, Burgers, Darcy flow) and all five noise types (HP, LP, SH, UGR, LGB); see Tables 8–22.

The results reveal a consistent pattern. The Mahalanobis distance does respond to noise degradation, e.g., on the airfoil benchmark under low-pass noise ($\delta = 1$), FNO's Mahalanobis distance increases from 4.43 to 391.34 (Table 9). The Energy scores exhibit less informative; for DON+N, the energy score remains essentially constant across all noise types and noise levels with very small variance. The ICA loss and corresponding Z-scores respond monotonically and differentially to noise severity and architecture. Moreover, across all model structures for all noise types, we observe significant decrease in Mahalanobis distance in latent space, with the increase of Energy score when regulated with the proposed method. This suggests the effectiveness of the regularization, i.e., the latent features are more independent, compared to the non-regulated vanilla models.

Correspondingly, we provide Pearson correlation between noise levels and our ICA-based statistic alongside Mahalanobis distance and energy scores, demonstrating the signal's consistency relative to the baselines (Tables 23–25).

Table 8: Baseline comparison on the NACA airfoil dataset under bandlimited highpass (HP) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$7.25_{\pm 0.07}$	1.00	-0.7280	2.97	-4.8595
	0.1	$16.29_{\pm 0.74}$	2.25	-0.6349	4.54	-4.8598
	0.2	$17.04_{\pm 0.93}$	2.35	-0.6058	4.83	-4.8597
	0.5	$17.06_{\pm 0.86}$	2.35	-0.5801	4.91	-4.8596
	1	$17.15_{\pm 0.87}$	2.37	-0.5691	4.97	-4.8596
DON+N	0	$7.19_{\pm 0.08}$	1.00	-1.9357	0.79	-3.2150
	0.1	$16.80_{\pm 1.00}$	2.34	-1.9068	1.02	-3.2152
	0.2	$17.00_{\pm 0.85}$	2.36	-1.9027	1.05	-3.2152
	0.5	$16.91_{\pm 0.74}$	2.35	-1.9024	1.06	-3.2153
	1	$16.85_{\pm 0.74}$	2.34	-1.9022	1.06	-3.2153
FNO	0	$1.06_{\pm 0.03}$	1.00	-1.4900	4.43	-4.8228
	0.1	$23.14_{\pm 7.61}$	21.87	-1.3610	17.49	-4.8317
	0.2	$48.10_{\pm 17.78}$	45.45	-1.3502	36.74	-4.9737
	0.5	$93.32_{\pm 36.09}$	88.18	-1.3495	71.43	-5.4683
	1	$121.89_{\pm 47.15}$	115.18	-1.3475	93.59	-5.9279
FNO+N	0	$1.04_{\pm 0.05}$	1.00	-3.3440	0.43	-3.2167
	0.1	$20.16_{\pm 3.05}$	19.30	-2.6254	0.94	-3.2487
	0.2	$36.09_{\pm 6.11}$	34.54	-1.9896	2.06	-3.3145
	0.5	$58.27_{\pm 12.85}$	55.76	-1.6801	4.21	-3.5026
	1	$69.74_{\pm 15.79}$	66.74	-1.6031	5.34	-3.6314
NOMAD	0	$7.47_{\pm 0.10}$	1.00	-10.6218	2.67	-4.9369
	0.1	$25.72_{\pm 4.22}$	3.44	-10.4507	7.26	-4.9546
	0.2	$50.46_{\pm 15.15}$	6.76	-10.5104	16.73	-5.0078
	0.5	$147.63_{\pm 77.32}$	19.77	-10.5213	52.01	-5.4094
	1	$344.83_{\pm 211.19}$	46.17	-10.4989	115.21	-6.8683

Table 9: Baseline comparison on the NACA airfoil dataset under bandlimited lowpass (LP) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$7.25_{\pm 0.07}$	1.00	-0.7280	2.97	-4.8595
	0.1	$16.52_{\pm 0.76}$	2.28	-0.6663	4.54	-4.8599
	0.2	$17.07_{\pm 0.89}$	2.35	-0.6186	4.80	-4.8597
	0.5	$17.19_{\pm 0.96}$	2.37	-0.5930	4.94	-4.8597
	1	$17.18_{\pm 0.92}$	2.37	-0.5807	4.97	-4.8596
DON+N	0	$7.19_{\pm 0.08}$	1.00	-1.9357	0.79	-3.2150
	0.1	$16.59_{\pm 0.92}$	2.31	-1.9080	1.01	-3.2152
	0.2	$17.00_{\pm 0.76}$	2.36	-1.9031	1.05	-3.2152
	0.5	$16.98_{\pm 0.74}$	2.36	-1.9022	1.06	-3.2153
	1	$16.93_{\pm 0.64}$	2.35	-1.9019	1.06	-3.2153
FNO	0	$1.06_{\pm 0.03}$	1.00	-1.4900	4.43	-4.8228
	0.1	$73.45_{\pm 14.99}$	69.41	-1.3841	53.29	-5.1658
	0.2	$128.33_{\pm 31.22}$	121.27	-1.3718	98.30	-6.0164
	0.5	$269.85_{\pm 71.96}$	255.01	-1.3693	212.83	-9.1830
	1	$491.77_{\pm 131.97}$	464.71	-1.3709	391.34	-15.0614
FNO+N	0	$1.04_{\pm 0.05}$	1.00	-3.3440	0.43	-3.2167
	0.1	$64.26_{\pm 8.69}$	61.49	-1.6242	3.75	-3.4106
	0.2	$107.71_{\pm 20.83}$	103.07	-1.4507	7.24	-3.6929
	0.5	$224.37_{\pm 53.58}$	214.71	-1.3998	16.57	-4.4831
	1	$415.71_{\pm 107.56}$	397.82	-1.3753	31.68	-5.8478
NOMAD	0	$7.47_{\pm 0.10}$	1.00	-10.6218	2.67	-4.9369
	0.1	$29.91_{\pm 7.48}$	4.00	-10.4077	8.71	-4.9589
	0.2	$55.40_{\pm 18.76}$	7.42	-10.4447	17.90	-5.0126
	0.5	$157.44_{\pm 75.42}$	21.08	-10.4886	53.38	-5.4294
	1	$355.75_{\pm 202.02}$	47.63	-10.4672	117.49	-6.9672

Table 10: Baseline comparison on the NACA airfoil dataset under harmonic spatial (SH) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$7.25_{\pm 0.07}$	1.00	-0.7280	2.97	-4.8595
	0.1	$16.61_{\pm 0.60}$	2.29	-0.6805	4.52	-4.8599
	0.2	$17.05_{\pm 0.94}$	2.35	-0.6476	4.82	-4.8598
	0.5	$16.84_{\pm 0.66}$	2.32	-0.6105	4.77	-4.8596
	1	$17.00_{\pm 0.72}$	2.35	-0.6487	4.90	-4.8596
DON+N	0	$7.19_{\pm 0.08}$	1.00	-1.9357	0.79	-3.2150
	0.1	$16.49_{\pm 0.81}$	2.29	-1.9096	1.00	-3.2151
	0.2	$16.93_{\pm 0.68}$	2.35	-1.9035	1.04	-3.2152
	0.5	$16.29_{\pm 0.70}$	2.26	-1.9061	1.04	-3.2153
	1	$16.51_{\pm 0.77}$	2.30	-1.9047	1.05	-3.2153
FNO	0	$1.06_{\pm 0.03}$	1.00	-1.4900	4.43	-4.8228
	0.1	$18.24_{\pm 5.17}$	17.23	-1.3794	12.30	-4.8156
	0.2	$36.74_{\pm 11.75}$	34.72	-1.3482	27.16	-4.8866
	0.5	$82.54_{\pm 28.57}$	78.00	-1.3392	63.54	-5.3399
	1	$123.45_{\pm 39.81}$	116.65	-1.3459	96.20	-5.9847
FNO+N	0	$1.04_{\pm 0.05}$	1.00	-3.3440	0.43	-3.2167
	0.1	$20.40_{\pm 5.93}$	19.52	-2.6172	0.94	-3.2519
	0.2	$35.52_{\pm 6.27}$	33.99	-2.0345	1.94	-3.3227
	0.5	$66.95_{\pm 11.52}$	64.07	-1.6761	4.71	-3.5847
	1	$94.89_{\pm 15.94}$	90.81	-1.6222	7.20	-3.8980
NOMAD	0	$7.47_{\pm 0.10}$	1.00	-10.6218	2.67	-4.9369
	0.1	$28.19_{\pm 7.78}$	3.77	-10.4158	8.02	-4.9546
	0.2	$56.91_{\pm 24.05}$	7.62	-10.4373	18.27	-5.0094
	0.5	$192.84_{\pm 81.48}$	25.82	-10.4778	58.97	-5.6485
	1	$456.58_{\pm 244.68}$	61.13	-10.4407	139.48	-7.6569

Table 11: Baseline comparison on the NACA airfoil dataset under unstructured Gaussian (UGR) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$7.25_{\pm 0.07}$	1.00	-0.7280	2.97	-4.8595
	0.1	$16.27_{\pm 0.70}$	2.24	-0.6399	4.50	-4.8598
	0.2	$16.98_{\pm 0.89}$	2.34	-0.6049	4.78	-4.8597
	0.5	$17.10_{\pm 1.04}$	2.36	-0.5831	4.92	-4.8596
	1	$17.07_{\pm 0.87}$	2.35	-0.5689	4.96	-4.8596
DON+N	0	$7.19_{\pm 0.08}$	1.00	-1.9357	0.79	-3.2150
	0.1	$16.64_{\pm 1.03}$	2.31	-1.9069	1.02	-3.2152
	0.2	$16.95_{\pm 0.81}$	2.36	-1.9028	1.05	-3.2152
	0.5	$16.97_{\pm 0.79}$	2.36	-1.9022	1.06	-3.2153
	1	$16.80_{\pm 0.75}$	2.34	-1.9025	1.06	-3.2153
FNO	0	$1.06_{\pm 0.03}$	1.00	-1.4900	4.43	-4.8228
	0.1	$25.04_{\pm 6.82}$	23.66	-1.3608	18.32	-4.8355
	0.2	$49.73_{\pm 16.96}$	47.00	-1.3430	37.91	-4.9855
	0.5	$96.12_{\pm 34.99}$	90.83	-1.3383	74.10	-5.5216
	1	$129.11_{\pm 45.50}$	122.00	-1.3353	100.36	-6.0924
FNO+N	0	$1.04_{\pm 0.05}$	1.00	-3.3440	0.43	-3.2167
	0.1	$22.05_{\pm 2.77}$	21.10	-2.5817	1.02	-3.2515
	0.2	$38.09_{\pm 6.56}$	36.45	-1.9762	2.20	-3.3209
	0.5	$63.06_{\pm 13.58}$	60.35	-1.6552	4.49	-3.5166
	1	$80.48_{\pm 16.99}$	77.01	-1.5590	6.01	-3.6710
NOMAD	0	$7.47_{\pm 0.10}$	1.00	-10.6218	2.67	-4.9369
	0.1	$25.56_{\pm 4.08}$	3.42	-10.4436	7.25	-4.9540
	0.2	$51.27_{\pm 15.58}$	6.86	-10.5012	16.75	-5.0063
	0.5	$148.94_{\pm 75.17}$	19.94	-10.5246	51.64	-5.3856
	1	$341.61_{\pm 212.18}$	45.74	-10.5081	114.90	-6.8263

Table 12: Baseline comparison on the NACA airfoil dataset under local Gaussian block (LGB) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$7.25_{\pm 0.07}$	1.00	-0.7280	2.97	-4.8595
	0.1	$9.77_{\pm 0.25}$	1.35	-0.7492	3.33	-4.8596
	0.2	$10.06_{\pm 0.30}$	1.39	-0.7454	3.38	-4.8596
	0.5	$10.07_{\pm 0.34}$	1.39	-0.7504	3.42	-4.8596
	1	$10.10_{\pm 0.34}$	1.39	-0.7446	3.46	-4.8596
DON+N	0	$7.19_{\pm 0.08}$	1.00	-1.9357	0.79	-3.2150
	0.1	$9.73_{\pm 0.45}$	1.35	-1.9341	0.81	-3.2150
	0.2	$9.99_{\pm 0.24}$	1.39	-1.9330	0.81	-3.2150
	0.5	$10.02_{\pm 0.38}$	1.39	-1.9327	0.82	-3.2150
	1	$9.94_{\pm 0.34}$	1.38	-1.9324	0.82	-3.2150
FNO	0	$1.06_{\pm 0.03}$	1.00	-1.4900	4.43	-4.8228
	0.1	$7.19_{\pm 0.46}$	6.79	-1.4904	4.95	-4.8203
	0.2	$11.69_{\pm 1.09}$	11.05	-1.4983	6.35	-4.8171
	0.5	$24.31_{\pm 4.43}$	22.97	-1.4503	13.57	-4.8211
	1	$44.60_{\pm 9.97}$	42.15	-1.3907	29.19	-4.9041
FNO+N	0	$1.04_{\pm 0.05}$	1.00	-3.3440	0.43	-3.2167
	0.1	$7.15_{\pm 0.22}$	6.85	-3.2782	0.46	-3.2195
	0.2	$11.17_{\pm 0.15}$	10.69	-3.2152	0.50	-3.2217
	0.5	$22.59_{\pm 2.22}$	21.62	-2.7395	0.89	-3.2364
	1	$41.90_{\pm 7.39}$	40.10	-1.7765	2.44	-3.3346
NOMAD	0	$7.47_{\pm 0.10}$	1.00	-10.6218	2.67	-4.9369
	0.1	$11.29_{\pm 1.37}$	1.51	-10.5483	3.43	-4.9382
	0.2	$12.28_{\pm 2.12}$	1.64	-10.5275	3.72	-4.9391
	0.5	$13.22_{\pm 2.85}$	1.77	-10.5123	4.01	-4.9402
	1	$13.37_{\pm 2.98}$	1.79	-10.5016	4.11	-4.9404

Table 13: Baseline comparison on the Burgers dataset under bandlimited highpass (HP) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$1.86_{\pm 0.55}$	1.00	-0.0932	7.05	-4.8703
	0.1	$1.91_{\pm 0.53}$	1.03	-0.0931	7.05	-4.8703
	0.2	$2.05_{\pm 0.49}$	1.10	-0.0931	7.06	-4.8703
	0.5	$2.74_{\pm 0.39}$	1.47	-0.0930	7.14	-4.8703
	1	$4.27_{\pm 0.50}$	2.29	-0.0928	7.40	-4.8703
DON+N	0	$0.80_{\pm 0.18}$	1.00	-0.0365	4.51	-5.5523
	0.1	$0.81_{\pm 0.17}$	1.02	-0.0365	4.52	-5.5523
	0.2	$0.86_{\pm 0.16}$	1.07	-0.0364	4.52	-5.5523
	0.5	$1.09_{\pm 0.10}$	1.36	-0.0364	4.54	-5.5523
	1	$1.64_{\pm 0.11}$	2.05	-0.0364	4.61	-5.5523
FNO	0	$0.08_{\pm 0.01}$	1.00	-6.6251	3.50	-4.9077
	0.1	$0.08_{\pm 0.01}$	1.05	-6.6251	3.50	-4.9077
	0.2	$0.09_{\pm 0.01}$	1.17	-6.6251	3.51	-4.9077
	0.5	$0.15_{\pm 0.01}$	1.81	-6.6250	3.53	-4.9077
	1	$0.26_{\pm 0.03}$	3.27	-6.6250	3.63	-4.9077
FNO+N	0	$0.09_{\pm 0.00}$	1.00	-0.3454	0.76	-5.5485
	0.1	$0.10_{\pm 0.01}$	1.06	-0.3450	0.76	-5.5485
	0.2	$0.11_{\pm 0.01}$	1.21	-0.3441	0.76	-5.5485
	0.5	$0.18_{\pm 0.02}$	1.97	-0.3395	0.77	-5.5485
	1	$0.32_{\pm 0.05}$	3.53	-0.3283	0.77	-5.5486
NOMAD	0	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.1	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.2	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.5	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	1	$24.42_{\pm 5.31}$	1.00	-9.3316	1.68	-4.9173
WNO	0	$0.31_{\pm 0.02}$	1.00	-0.1499	6.33	-4.8520
	0.1	$0.33_{\pm 0.03}$	1.07	-0.1499	6.34	-4.8520
	0.2	$0.39_{\pm 0.03}$	1.24	-0.1496	6.36	-4.8520
	0.5	$0.65_{\pm 0.07}$	2.08	-0.1481	6.49	-4.8521
	1	$1.22_{\pm 0.16}$	3.90	-0.1435	6.93	-4.8525

Table 14: Baseline comparison on the Burgers dataset under bandlimited lowpass (LP) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$1.86_{\pm 0.55}$	1.00	-0.0932	7.05	-4.8703
	0.1	$2.10_{\pm 0.50}$	1.13	-0.0932	7.06	-4.8703
	0.2	$2.63_{\pm 0.41}$	1.41	-0.0932	7.08	-4.8703
	0.5	$4.84_{\pm 0.25}$	2.60	-0.0935	7.25	-4.8704
	1	$8.99_{\pm 0.22}$	4.82	-0.0945	7.82	-4.8705
DON+N	0	$0.80_{\pm 0.18}$	1.00	-0.0365	4.51	-5.5523
	0.1	$1.20_{\pm 0.13}$	1.50	-0.0364	4.52	-5.5523
	0.2	$1.91_{\pm 0.08}$	2.40	-0.0364	4.52	-5.5523
	0.5	$4.35_{\pm 0.07}$	5.45	-0.0362	4.58	-5.5523
	1	$8.57_{\pm 0.16}$	10.73	-0.0355	4.79	-5.5523
FNO	0	$0.08_{\pm 0.01}$	1.00	-6.6251	3.50	-4.9077
	0.1	$0.85_{\pm 0.01}$	10.57	-6.6251	3.50	-4.9077
	0.2	$1.69_{\pm 0.03}$	21.03	-6.6250	3.51	-4.9077
	0.5	$4.23_{\pm 0.07}$	52.47	-6.6250	3.54	-4.9077
	1	$8.45_{\pm 0.14}$	104.85	-6.6248	3.64	-4.9078
FNO+N	0	$0.09_{\pm 0.00}$	1.00	-0.3454	0.76	-5.5485
	0.1	$0.85_{\pm 0.01}$	9.40	-0.3420	0.76	-5.5485
	0.2	$1.70_{\pm 0.03}$	18.67	-0.3380	0.77	-5.5485
	0.5	$4.23_{\pm 0.07}$	46.60	-0.3337	0.77	-5.5485
	1	$8.46_{\pm 0.14}$	93.15	-0.3463	0.80	-5.5486
NOMAD	0	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.1	$24.42_{\pm 5.30}$	1.00	-9.3317	1.67	-4.9173
	0.2	$24.43_{\pm 5.28}$	1.00	-9.3317	1.67	-4.9173
	0.5	$24.53_{\pm 5.16}$	1.00	-9.3312	1.70	-4.9173
	1	$24.87_{\pm 4.74}$	1.02	-9.3307	1.77	-4.9173
WNO	0	$0.31_{\pm 0.02}$	1.00	-0.1499	6.33	-4.8520
	0.1	$0.93_{\pm 0.02}$	2.99	-0.1498	6.34	-4.8520
	0.2	$1.77_{\pm 0.03}$	5.70	-0.1495	6.36	-4.8521
	0.5	$4.37_{\pm 0.08}$	14.03	-0.1473	6.54	-4.8523
	1	$8.72_{\pm 0.15}$	27.98	-0.1406	7.12	-4.8530

Table 15: Baseline comparison on the Burgers dataset under harmonic spatial (SH) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$1.86_{\pm 0.55}$	1.00	-0.0932	7.05	-4.8703
	0.1	$2.72_{\pm 0.41}$	1.46	-0.0931	7.07	-4.8703
	0.2	$4.24_{\pm 0.29}$	2.28	-0.0932	7.14	-4.8703
	0.5	$9.44_{\pm 0.22}$	5.07	-0.0945	7.60	-4.8705
	1	$18.32_{\pm 0.51}$	9.83	-0.0987	9.03	-4.8708
DON+N	0	$0.80_{\pm 0.18}$	1.00	-0.0365	4.51	-5.5523
	0.1	$2.07_{\pm 0.10}$	2.59	-0.0364	4.52	-5.5523
	0.2	$3.84_{\pm 0.10}$	4.81	-0.0362	4.55	-5.5523
	0.5	$9.34_{\pm 0.23}$	11.70	-0.0353	4.71	-5.5523
	1	$18.44_{\pm 0.47}$	23.10	-0.0324	5.25	-5.5523
FNO	0	$0.08_{\pm 0.01}$	1.00	-6.6251	3.50	-4.9077
	0.1	$1.87_{\pm 0.04}$	23.19	-6.6251	3.50	-4.9077
	0.2	$3.73_{\pm 0.09}$	46.31	-6.6251	3.51	-4.9077
	0.5	$9.33_{\pm 0.22}$	115.70	-6.6251	3.54	-4.9078
	1	$18.64_{\pm 0.44}$	231.21	-6.6254	3.65	-4.9083
FNO+N	0	$0.09_{\pm 0.00}$	1.00	-0.3454	0.76	-5.5485
	0.1	$1.87_{\pm 0.04}$	20.58	-0.3608	0.76	-5.5485
	0.2	$3.73_{\pm 0.09}$	41.10	-0.3664	0.77	-5.5485
	0.5	$9.32_{\pm 0.22}$	102.69	-0.3663	0.80	-5.5485
	1	$18.64_{\pm 0.44}$	205.28	-0.3121	0.90	-5.5486
NOMAD	0	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.1	$24.44_{\pm 5.28}$	1.00	-9.3318	1.67	-4.9173
	0.2	$24.51_{\pm 5.19}$	1.00	-9.3317	1.67	-4.9173
	0.5	$24.90_{\pm 4.70}$	1.02	-9.3317	1.68	-4.9173
	1	$25.88_{\pm 3.47}$	1.06	-9.3311	1.73	-4.9174
WNO	0	$0.31_{\pm 0.02}$	1.00	-0.1499	6.33	-4.8520
	0.1	$1.89_{\pm 0.05}$	6.08	-0.1498	6.34	-4.8520
	0.2	$3.74_{\pm 0.09}$	12.01	-0.1493	6.36	-4.8521
	0.5	$9.32_{\pm 0.22}$	29.92	-0.1462	6.50	-4.8527
	1	$18.63_{\pm 0.43}$	59.79	-0.1364	7.01	-4.8547

Table 16: Baseline comparison on the Burgers dataset under unstructured Gaussian (UGR) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$1.86_{\pm 0.55}$	1.00	-0.0932	7.05	-4.8703
	0.1	$1.93_{\pm 0.53}$	1.04	-0.0931	7.05	-4.8703
	0.2	$2.12_{\pm 0.49}$	1.14	-0.0930	7.07	-4.8703
	0.5	$3.03_{\pm 0.39}$	1.62	-0.0928	7.16	-4.8703
	1	$4.98_{\pm 0.44}$	2.67	-0.0927	7.45	-4.8703
DON+N	0	$0.80_{\pm 0.18}$	1.00	-0.0365	4.51	-5.5523
	0.1	$0.86_{\pm 0.17}$	1.08	-0.0365	4.51	-5.5523
	0.2	$1.02_{\pm 0.14}$	1.28	-0.0364	4.52	-5.5523
	0.5	$1.74_{\pm 0.09}$	2.18	-0.0364	4.54	-5.5523
	1	$3.14_{\pm 0.09}$	3.93	-0.0363	4.62	-5.5523
FNO	0	$0.08_{\pm 0.01}$	1.00	-6.6251	3.50	-4.9077
	0.1	$0.28_{\pm 0.01}$	3.52	-6.6250	3.50	-4.9077
	0.2	$0.55_{\pm 0.01}$	6.77	-6.6250	3.51	-4.9077
	0.5	$1.35_{\pm 0.03}$	16.69	-6.6250	3.53	-4.9077
	1	$2.68_{\pm 0.05}$	33.30	-6.6249	3.63	-4.9077
FNO+N	0	$0.09_{\pm 0.00}$	1.00	-0.3454	0.76	-5.5485
	0.1	$0.29_{\pm 0.01}$	3.17	-0.3453	0.76	-5.5485
	0.2	$0.55_{\pm 0.01}$	6.05	-0.3447	0.76	-5.5485
	0.5	$1.35_{\pm 0.03}$	14.87	-0.3415	0.77	-5.5485
	1	$2.69_{\pm 0.05}$	29.66	-0.3324	0.77	-5.5486
NOMAD	0	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.1	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.2	$24.42_{\pm 5.30}$	1.00	-9.3317	1.67	-4.9173
	0.5	$24.44_{\pm 5.29}$	1.00	-9.3316	1.67	-4.9173
	1	$24.48_{\pm 5.24}$	1.00	-9.3313	1.69	-4.9173
WNO	0	$0.31_{\pm 0.02}$	1.00	-0.1499	6.33	-4.8520
	0.1	$0.43_{\pm 0.02}$	1.39	-0.1499	6.34	-4.8520
	0.2	$0.67_{\pm 0.01}$	2.16	-0.1497	6.36	-4.8520
	0.5	$1.52_{\pm 0.03}$	4.87	-0.1481	6.50	-4.8521
	1	$3.00_{\pm 0.08}$	9.63	-0.1434	6.95	-4.8525

Table 17: Baseline comparison on the Burgers dataset under local Gaussian block (LGB) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$1.86_{\pm 0.55}$	1.00	-0.0932	7.05	-4.8703
	0.1	$1.88_{\pm 0.54}$	1.01	-0.0932	7.05	-4.8703
	0.2	$1.93_{\pm 0.53}$	1.04	-0.0932	7.05	-4.8703
	0.5	$2.23_{\pm 0.46}$	1.19	-0.0932	7.07	-4.8703
	1	$2.99_{\pm 0.38}$	1.60	-0.0932	7.15	-4.8703
DON+N	0	$0.80_{\pm 0.18}$	1.00	-0.0365	4.51	-5.5523
	0.1	$0.82_{\pm 0.18}$	1.02	-0.0365	4.51	-5.5523
	0.2	$0.86_{\pm 0.17}$	1.08	-0.0364	4.52	-5.5523
	0.5	$1.11_{\pm 0.12}$	1.39	-0.0364	4.52	-5.5523
	1	$1.69_{\pm 0.08}$	2.12	-0.0364	4.54	-5.5523
FNO	0	$0.08_{\pm 0.01}$	1.00	-6.6251	3.50	-4.9077
	0.1	$0.15_{\pm 0.01}$	1.90	-6.6251	3.50	-4.9077
	0.2	$0.27_{\pm 0.01}$	3.31	-6.6251	3.50	-4.9077
	0.5	$0.63_{\pm 0.03}$	7.83	-6.6251	3.51	-4.9077
	1	$1.25_{\pm 0.06}$	15.49	-6.6251	3.53	-4.9077
FNO+N	0	$0.09_{\pm 0.00}$	1.00	-0.3454	0.76	-5.5485
	0.1	$0.16_{\pm 0.00}$	1.76	-0.3451	0.76	-5.5485
	0.2	$0.27_{\pm 0.01}$	2.99	-0.3447	0.76	-5.5485
	0.5	$0.63_{\pm 0.03}$	6.99	-0.3427	0.76	-5.5485
	1	$1.25_{\pm 0.05}$	13.81	-0.3380	0.77	-5.5485
NOMAD	0	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.1	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.2	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	0.5	$24.42_{\pm 5.31}$	1.00	-9.3317	1.67	-4.9173
	1	$24.43_{\pm 5.29}$	1.00	-9.3316	1.67	-4.9173
WNO	0	$0.31_{\pm 0.02}$	1.00	-0.1499	6.33	-4.8520
	0.1	$0.34_{\pm 0.02}$	1.09	-0.1499	6.33	-4.8520
	0.2	$0.41_{\pm 0.02}$	1.33	-0.1498	6.34	-4.8520
	0.5	$0.73_{\pm 0.03}$	2.35	-0.1492	6.37	-4.8521
	1	$1.36_{\pm 0.05}$	4.36	-0.1473	6.48	-4.8522

Table 18: Baseline comparison on the Darcy dataset under bandlimited highpass (HP) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$8.19_{\pm 0.83}$	1.00	-0.0216	0.64	-4.8533
	0.1	$8.21_{\pm 0.82}$	1.00	-0.0216	0.64	-4.8533
	0.2	$8.26_{\pm 0.79}$	1.01	-0.0215	0.64	-4.8533
	0.5	$8.57_{\pm 0.63}$	1.05	-0.0214	0.64	-4.8533
	1	$9.50_{\pm 0.37}$	1.16	-0.0213	0.64	-4.8533
DON+N	0	$6.42_{\pm 0.19}$	1.00	-0.0066	0.23	-3.4673
	0.1	$6.46_{\pm 0.21}$	1.01	-0.0066	0.23	-3.4673
	0.2	$6.55_{\pm 0.23}$	1.02	-0.0066	0.23	-3.4673
	0.5	$7.12_{\pm 0.32}$	1.11	-0.0065	0.24	-3.4673
	1	$8.73_{\pm 0.55}$	1.36	-0.0064	0.24	-3.4673
FNO	0	$2.70_{\pm 0.33}$	1.00	-3.5200	1.85	-4.8637
	0.1	$2.89_{\pm 0.43}$	1.07	-3.5189	1.85	-4.8637
	0.2	$3.42_{\pm 0.69}$	1.27	-3.5156	1.86	-4.8637
	0.5	$6.97_{\pm 1.66}$	2.58	-3.4999	1.91	-4.8636
	1	$17.02_{\pm 2.79}$	6.30	-3.4931	2.28	-4.8632
FNO+N	0	$2.47_{\pm 0.12}$	1.00	-2.3059	0.03	-3.4672
	0.1	$2.52_{\pm 0.12}$	1.02	-2.3053	0.03	-3.4672
	0.2	$2.70_{\pm 0.12}$	1.09	-2.3042	0.03	-3.4672
	0.5	$4.29_{\pm 0.22}$	1.74	-2.2951	0.03	-3.4672
	1	$9.20_{\pm 0.50}$	3.72	-2.2522	0.03	-3.4672
NOMAD	0	$6.31_{\pm 0.27}$	1.00	-8.8778	1.58	-4.8634
	0.1	$6.35_{\pm 0.27}$	1.01	-8.8784	1.58	-4.8634
	0.2	$6.44_{\pm 0.26}$	1.02	-8.8790	1.58	-4.8634
	0.5	$7.00_{\pm 0.24}$	1.11	-8.8808	1.58	-4.8634
	1	$8.59_{\pm 0.35}$	1.36	-8.8836	1.59	-4.8634
WNO	0	$4.22_{\pm 0.11}$	1.00	-0.0222	4.38	-4.8577
	0.1	$6.71_{\pm 2.04}$	1.59	-0.0232	4.44	-4.8578
	0.2	$14.34_{\pm 6.63}$	3.40	-0.0601	4.71	-4.8579
	0.5	$81.79_{\pm 90.08}$	19.39	-0.2745	6.70	-4.8595
	1	$284.69_{\pm 387.38}$	67.49	-0.4612	11.68	-4.8662

Table 19: Baseline comparison on the Darcy dataset under bandlimited lowpass (LP) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$8.19_{\pm 0.83}$	1.00	-0.0216	0.64	-4.8533
	0.1	$8.59_{\pm 0.74}$	1.05	-0.0214	0.64	-4.8533
	0.2	$9.72_{\pm 0.54}$	1.19	-0.0211	0.65	-4.8533
	0.5	$15.58_{\pm 0.70}$	1.90	-0.0201	0.67	-4.8533
	1	$28.87_{\pm 2.11}$	3.53	-0.0290	0.74	-4.8533
DON+N	0	$6.42_{\pm 0.19}$	1.00	-0.0066	0.23	-3.4673
	0.1	$7.02_{\pm 0.12}$	1.09	-0.0064	0.24	-3.4673
	0.2	$8.48_{\pm 0.16}$	1.32	-0.0063	0.24	-3.4673
	0.5	$14.87_{\pm 0.35}$	2.32	-0.0057	0.25	-3.4673
	1	$25.98_{\pm 0.87}$	4.05	-0.0055	0.28	-3.4673
FNO	0	$2.70_{\pm 0.33}$	1.00	-3.5200	1.85	-4.8637
	0.1	$3.39_{\pm 0.23}$	1.25	-3.5185	1.86	-4.8637
	0.2	$4.90_{\pm 0.33}$	1.81	-3.5135	1.87	-4.8637
	0.5	$12.35_{\pm 0.85}$	4.57	-3.4949	1.95	-4.8636
	1	$28.64_{\pm 3.46}$	10.61	-3.5078	2.38	-4.8631
FNO+N	0	$2.47_{\pm 0.12}$	1.00	-2.3059	0.03	-3.4672
	0.1	$3.40_{\pm 0.15}$	1.38	-2.3049	0.03	-3.4672
	0.2	$5.27_{\pm 0.27}$	2.13	-2.3016	0.03	-3.4672
	0.5	$13.03_{\pm 0.80}$	5.27	-2.2773	0.03	-3.4672
	1	$29.13_{\pm 2.00}$	11.79	-2.1617	0.03	-3.4672
NOMAD	0	$6.31_{\pm 0.27}$	1.00	-8.8778	1.58	-4.8634
	0.1	$6.82_{\pm 0.22}$	1.08	-8.8782	1.58	-4.8634
	0.2	$8.07_{\pm 0.21}$	1.28	-8.8789	1.59	-4.8634
	0.5	$13.48_{\pm 0.47}$	2.14	-8.8789	1.63	-4.8634
	1	$21.52_{\pm 0.72}$	3.41	-8.8831	1.78	-4.8635
WNO	0	$4.22_{\pm 0.11}$	1.00	-0.0222	4.38	-4.8577
	0.1	$6.91_{\pm 1.40}$	1.64	-0.0229	4.48	-4.8578
	0.2	$14.21_{\pm 4.92}$	3.37	-0.0466	4.89	-4.8580
	0.5	$68.15_{\pm 61.88}$	16.16	-0.1864	7.69	-4.8598
	1	$251.77_{\pm 302.91}$	59.68	-0.3233	14.14	-4.8678

Table 20: Baseline comparison on the Darcy dataset under harmonic spatial (SH) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$8.19_{\pm 0.83}$	1.00	-0.0216	0.64	-4.8533
	0.1	$8.20_{\pm 0.84}$	1.00	-0.0216	0.64	-4.8533
	0.2	$8.24_{\pm 0.83}$	1.01	-0.0215	0.64	-4.8533
	0.5	$8.52_{\pm 0.78}$	1.04	-0.0214	0.64	-4.8533
	1	$9.41_{\pm 0.57}$	1.15	-0.0216	0.64	-4.8533
DON+N	0	$6.42_{\pm 0.19}$	1.00	-0.0066	0.23	-3.4673
	0.1	$6.44_{\pm 0.19}$	1.00	-0.0066	0.23	-3.4673
	0.2	$6.52_{\pm 0.19}$	1.02	-0.0066	0.23	-3.4673
	0.5	$6.99_{\pm 0.22}$	1.09	-0.0066	0.24	-3.4673
	1	$8.38_{\pm 0.34}$	1.31	-0.0065	0.24	-3.4673
FNO	0	$2.70_{\pm 0.33}$	1.00	-3.5200	1.85	-4.8637
	0.1	$2.83_{\pm 0.37}$	1.05	-3.5195	1.85	-4.8637
	0.2	$3.19_{\pm 0.49}$	1.18	-3.5178	1.86	-4.8637
	0.5	$5.50_{\pm 0.99}$	2.03	-3.5082	1.88	-4.8637
	1	$13.12_{\pm 1.95}$	4.86	-3.4956	2.03	-4.8635
FNO+N	0	$2.47_{\pm 0.12}$	1.00	-2.3059	0.03	-3.4672
	0.1	$2.56_{\pm 0.13}$	1.04	-2.3059	0.03	-3.4672
	0.2	$2.81_{\pm 0.15}$	1.14	-2.3053	0.03	-3.4672
	0.5	$4.48_{\pm 0.21}$	1.81	-2.2994	0.03	-3.4672
	1	$9.65_{\pm 0.38}$	3.91	-2.2681	0.03	-3.4672
NOMAD	0	$6.31_{\pm 0.27}$	1.00	-8.8778	1.58	-4.8634
	0.1	$6.32_{\pm 0.27}$	1.00	-8.8781	1.58	-4.8634
	0.2	$6.38_{\pm 0.28}$	1.01	-8.8784	1.58	-4.8634
	0.5	$6.78_{\pm 0.32}$	1.07	-8.8796	1.58	-4.8634
	1	$8.01_{\pm 0.40}$	1.27	-8.8827	1.59	-4.8634
WNO	0	$4.22_{\pm 0.11}$	1.00	-0.0222	4.38	-4.8577
	0.1	$5.47_{\pm 1.22}$	1.30	-0.0221	4.40	-4.8578
	0.2	$8.63_{\pm 3.03}$	2.05	-0.0225	4.51	-4.8578
	0.5	$29.57_{\pm 15.08}$	7.01	-0.0478	5.67	-4.8584
	1	$115.60_{\pm 121.57}$	27.40	-0.0901	9.36	-4.8614

Table 21: Baseline comparison on the Darcy dataset under unstructured Gaussian (UGR) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$8.19_{\pm 0.83}$	1.00	-0.0216	0.64	-4.8533
	0.1	$8.20_{\pm 0.83}$	1.00	-0.0216	0.64	-4.8533
	0.2	$8.25_{\pm 0.81}$	1.01	-0.0216	0.64	-4.8533
	0.5	$8.60_{\pm 0.70}$	1.05	-0.0215	0.64	-4.8533
	1	$9.71_{\pm 0.41}$	1.19	-0.0214	0.64	-4.8533
DON+N	0	$6.42_{\pm 0.19}$	1.00	-0.0066	0.23	-3.4673
	0.1	$6.45_{\pm 0.23}$	1.01	-0.0066	0.23	-3.4673
	0.2	$6.55_{\pm 0.27}$	1.02	-0.0066	0.23	-3.4673
	0.5	$7.17_{\pm 0.40}$	1.12	-0.0065	0.24	-3.4673
	1	$8.94_{\pm 0.61}$	1.39	-0.0063	0.24	-3.4673
FNO	0	$2.70_{\pm 0.33}$	1.00	-3.5200	1.85	-4.8637
	0.1	$2.89_{\pm 0.44}$	1.07	-3.5187	1.85	-4.8637
	0.2	$3.43_{\pm 0.70}$	1.27	-3.5152	1.86	-4.8637
	0.5	$6.99_{\pm 1.67}$	2.59	-3.4997	1.91	-4.8636
	1	$17.08_{\pm 2.76}$	6.32	-3.4928	2.28	-4.8632
FNO+N	0	$2.44_{\pm 0.09}$	1.00	-2.3059	0.03	-3.4672
	0.1	$2.50_{\pm 0.08}$	1.02	-2.3053	0.03	-3.4672
	0.2	$2.72_{\pm 0.07}$	1.11	-2.3037	0.03	-3.4672
	0.5	$4.61_{\pm 0.34}$	1.79	-2.2920	0.03	-3.4672
	1	$9.91_{\pm 0.26}$	3.83	-2.2411	0.03	-3.4672
NOMAD	0	$6.31_{\pm 0.27}$	1.00	-8.8778	1.58	-4.8634
	0.1	$6.35_{\pm 0.27}$	1.01	-8.8779	1.58	-4.8634
	0.2	$6.44_{\pm 0.27}$	1.02	-8.8783	1.58	-4.8634
	0.5	$7.06_{\pm 0.27}$	1.12	-8.8804	1.58	-4.8634
	1	$8.79_{\pm 0.41}$	1.39	-8.8840	1.59	-4.8634
WNO	0	$4.22_{\pm 0.11}$	1.00	-0.0222	4.38	-4.8577
	0.1	$6.70_{\pm 2.03}$	1.59	-0.0231	4.44	-4.8578
	0.2	$14.29_{\pm 6.49}$	3.39	-0.0595	4.71	-4.8579
	0.5	$81.64_{\pm 89.57}$	19.35	-0.2726	6.71	-4.8595
	1	$284.80_{\pm 386.51}$	67.52	-0.4583	11.71	-4.8662

Table 22: Baseline comparison on the Darcy dataset under local Gaussian block (LGB) noise. Relative L_2 Error ($\times 10^{-2}$), Sensitivity (Sens.), mean ICA Loss, mean Mahalanobis Distance (M. Dist), and mean Energy Score (Energy) are reported. $\pm$ denotes standard deviation.

Model	δ	Rel. L_2 ($\times 10^{-2}$)	Sens.	ICA Loss	M. Dist	Energy
DON	0	$8.19_{\pm 0.83}$	1.00	-0.0216	0.64	-4.8533
	0.1	$8.19_{\pm 0.83}$	1.00	-0.0216	0.64	-4.8533
	0.2	$8.19_{\pm 0.83}$	1.00	-0.0216	0.64	-4.8533
	0.5	$8.21_{\pm 0.81}$	1.00	-0.0216	0.64	-4.8533
	1	$8.29_{\pm 0.77}$	1.01	-0.0216	0.64	-4.8533
DON+N	0	$6.42_{\pm 0.19}$	1.00	-0.0066	0.23	-3.4673
	0.1	$6.42_{\pm 0.19}$	1.00	-0.0066	0.23	-3.4673
	0.2	$6.43_{\pm 0.19}$	1.00	-0.0066	0.23	-3.4673
	0.5	$6.48_{\pm 0.19}$	1.01	-0.0066	0.24	-3.4673
	1	$6.64_{\pm 0.20}$	1.03	-0.0066	0.24	-3.4673
FNO	0	$2.70_{\pm 0.33}$	1.00	-3.5200	1.85	-4.8637
	0.1	$2.71_{\pm 0.34}$	1.01	-3.5199	1.85	-4.8637
	0.2	$2.76_{\pm 0.36}$	1.02	-3.5197	1.85	-4.8637
	0.5	$3.08_{\pm 0.40}$	1.14	-3.5180	1.85	-4.8637
	1	$4.20_{\pm 0.81}$	1.55	-3.5118	1.87	-4.8637
FNO+N	0	$2.47_{\pm 0.12}$	1.00	-2.3059	0.03	-3.4672
	0.1	$2.47_{\pm 0.12}$	1.00	-2.3058	0.03	-3.4672
	0.2	$2.48_{\pm 0.12}$	1.01	-2.3057	0.03	-3.4672
	0.5	$2.59_{\pm 0.13}$	1.05	-2.3051	0.03	-3.4672
	1	$3.10_{\pm 0.19}$	1.25	-2.3035	0.03	-3.4672
NOMAD	0	$6.31_{\pm 0.27}$	1.00	-8.8778	1.58	-4.8634
	0.1	$6.31_{\pm 0.27}$	1.00	-8.8779	1.58	-4.8634
	0.2	$6.32_{\pm 0.27}$	1.00	-8.8779	1.58	-4.8634
	0.5	$6.37_{\pm 0.25}$	1.01	-8.8782	1.58	-4.8634
	1	$6.52_{\pm 0.23}$	1.03	-8.8788	1.58	-4.8634
WNO	0	$4.22_{\pm 0.11}$	1.00	-0.0222	4.38	-4.8577
	0.1	$4.36_{\pm 0.16}$	1.03	-0.0222	4.38	-4.8578
	0.2	$5.16_{\pm 0.89}$	1.22	-0.0266	4.39	-4.8578
	0.5	$17.27_{\pm 17.28}$	4.09	-0.1946	4.49	-4.8578
	1	$59.41_{\pm 78.45}$	14.08	-0.7695	4.75	-4.8582

Table 23: Pearson correlation of metric Z-scores with respect to noise level $\delta = 1$ on the Burgers dataset. ICA Loss, Mahalanobis Distance (M. Dist), and Energy Score (Energy) are reported. Models include DeepONet (DON), DeepONet+NOIR (DON+N), FNO, FNO+NOIR (FNO+N), NOMAD, and WNO. Noise types are abbreviated as HP (Bandlimited Highpass), LP (Bandlimited Lowpass), SH (Spatial Harmonic), UGR (Unstructured Gaussian), and LGB (Local Gaussian Block).

	DON			DON+N			FNO			FNO+N			NOMAD			WNO		
Noise	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy
HP	0.9992	0.9691	0.9585	0.9715	0.9697	0.8574	0.9797	0.9709	0.9617	0.9863	0.9672	0.9742	0.9758	0.9690	0.9998	0.9756	0.9723	0.9677
LP	0.9752	0.9686	0.9703	0.9682	0.9682	0.9801	0.9986	0.9705	0.9611	0.9388	0.9682	0.9890	0.9921	0.9823	0.9998	0.9744	0.9716	0.9687
SH	0.9995	0.9747	0.9705	0.9685	0.9719	0.9876	0.9926	0.9676	0.9661	0.9976	0.9656	0.9412	0.9655	0.9649	0.9998	0.9712	0.9683	0.9663
UGR	0.9992	0.9696	0.9746	0.9631	0.9653	0.9856	0.9981	0.9707	0.9588	0.9891	0.9607	0.9675	0.9830	0.9775	0.9994	0.9739	0.9721	0.9670
LGB	0.9940	0.9659	0.9866	0.9680	0.9674	0.3365	0.9970	0.9711	0.9989	0.9834	0.9641	0.9884	0.9651	0.9725	0.9996	0.9742	0.9724	0.9703

Table 24: Pearson correlation of metric Z-scores with respect to noise level $\delta = 1$ on the Darcy dataset. ICA Loss, Mahalanobis Distance (M. Dist), and Energy Score (Energy) are reported. Models include DeepONet (DON), DeepONet+NOIR (DON+N), FNO, FNO+NOIR (FNO+N), NOMAD, and WNO. Noise types are abbreviated as HP (Bandlimited Highpass), LP (Bandlimited Lowpass), SH (Spatial Harmonic), UGR (Unstructured Gaussian), and LGB (Local Gaussian Block)

	DON			DON+N			FNO			FNO+N			NOMAD			WNO		
Noise	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy
HP	0.9962	0.9859	0.9480	0.9941	0.9667	0.9886	0.8831	0.9341	0.9392	0.9643	0.9922	0.9747	0.9990	0.9665	0.9813	0.9873	0.9804	0.9645
LP	0.9853	0.9773	0.9770	0.9963	0.9840	0.9858	0.9764	0.9522	0.9512	0.9659	0.9771	0.9629	0.9936	0.9736	0.9832	0.9887	0.9837	0.9649
SH	0.9900	0.9885	0.8700	0.9976	0.9723	0.9596	0.9952	0.9391	0.9555	0.9552	0.9929	0.9572	0.9887	0.9794	0.9974	0.9788	0.9689	0.9596
UGR	0.9839	0.9751	0.5255	0.9929	0.9755	0.9751	0.8904	0.9347	0.9385	0.9663	0.9950	0.9757	0.9778	0.9615	0.9815	0.9873	0.9804	0.9644
LGB	0.9996	0.9983	-0.0161	0.9988	0.9964	0.9592	0.9630	0.9477	0.9461	0.9833	0.9741	0.9558	0.9998	0.9998	0.9672	0.9562	0.9800	0.9631

Table 25: Pearson correlation of metric Z-scores with respect to noise level $\delta = 1$ on the Airfoil dataset. ICA Loss, Mahalanobis Distance (M. Dist), and Energy Score (Energy) are reported. Models include DeepONet (DON), DeepONet+NOIR (DON+N), FNO, FNO+NOIR (FNO+N), NOMAD, and WNO. Noise types are abbreviated as HP (Bandlimited Highpass), LP (Bandlimited Lowpass), SH (Spatial Harmonic), UGR (Unstructured Gaussian), and LGB (Local Gaussian Block)

	DON			DON+N			FNO			FNO+N			NOMAD		
Noise	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy	ICA	M. Dist	Energy
HP	0.7618	0.6206	0.2083	0.5493	0.5875	0.5779	0.5061	0.9639	0.9908	0.8226	0.9596	0.9652	0.5442	0.9966	0.9757
LP	0.6680	0.6275	0.2010	0.5686	0.6123	0.6674	0.5565	0.9988	0.9968	0.5892	0.9998	0.9986	0.5889	0.9973	0.9747
SH	0.7247	0.6002	0.0959	0.5348	0.6147	0.7170	0.5787	0.9884	0.9945	0.8199	0.9897	0.9962	0.5998	0.9946	0.9696
UGR	0.7418	0.6349	0.1820	0.5382	0.5796	0.5750	0.5669	0.9704	0.9933	0.8305	0.9708	0.9773	0.5428	0.9966	0.9748
LGB	0.5950	0.6850	0.7245	0.8483	0.8066	0.4010	0.9965	0.9890	0.9430	0.9841	0.9615	0.9375	0.7372	0.8040	0.8262

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately describe the paper's core contributions: NOIR as an independence-promoting regularization framework for neural operators, improved robustness under corrupted sensing, and latent-statistic-based monitoring. These claims are supported by the methodology, theoretical analysis, experiments, and the stated limitations in the main paper and appendix.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Answer: [Yes]

   Justification: The paper includes a dedicated limitations paragraph in the conclusion and further discusses important assumptions and scope restrictions in the appendix, including additive corruption assumptions, architecture dependence, and the local nature of the latent Gaussianization interpretation under higher noise.

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: The paper states the main assumptions clearly and provides proof sketches for the principal theoretical results in the main text, with additional derivations and technical discussion in Appendix A.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper discloses the datasets, corruption processes, model families, evaluation metrics, architecture configurations, and training protocol needed to reproduce the main experimental results. Additional reproducibility details are provided in the experimental setup, training specifications, hyperparameter sensitivity analysis, and appendix tables.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The paper provides public access to code, datasets, checkpoints, and trained models via the repository linked in the abstract, and the accompanying materials include the information needed to reproduce the main experiments.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies the benchmark problems, data splits, corruption types and magnitudes, model architectures, optimizer, learning-rate schedule, training epochs, and hyperparameter choices. Further details are provided in the appendix sections on detailed experimental setup, training specifications, and hyperparameter sensitivity.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

Justification: The experimental tables report mean performance together with standard deviations, and the captions state that the entries are aggregated over seeds. These statistics are provided for the main performance tables as well as the appendix results, which support the main empirical claims.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: The paper specifies the primary hardware used (NVIDIA A100 GPU). The cloud provider is hidden for the initial submission in respect to the double-blinded review process. The camera ready version acknowledge Delta and Delta AI clusters as the cloud providers.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

Answer: [Yes]

Justification: To the best of our knowledge, the research complies with the NeurIPS Code of Ethics. The work is a methodological study on operator learning using established scientific-machine-learning benchmarks and does not involve deceptive practices, human subjects, or high-risk deployment settings.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [No]

    Justification: The work is primarily foundational methodological research in scientific machine learning, and the current manuscript does not include a dedicated discussion of broader societal impacts. Potential positive impacts include improved reliability of scientific surrogate models, while possible negative impacts could arise from over-trust in imperfect surrogate predictions in safety-critical settings.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release high-risk generative models, scraped web-scale datasets, or other assets with an obvious dual-use or misuse profile that would require special release safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.

- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper properly credits the benchmark datasets and prior methods through citations.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.

- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs are not an important, original, or non-standard component of the core methodology, experiments, or scientific contributions of this research.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.