MULTISCALE SURFACE EQUILIBRIUM RESPONSES FOR UNIFIED 3D ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Two geometric deviations of the same magnitude can carry different evidence of a defect: a narrow bump is often anomalous where a broad variation is not. Reliable 3D anomaly detection therefore requires interpreting not only the size of a deviation but its spatial pattern, relative to the variation observed in normal shapes. We propose a unified framework that uses a nominal surface as the analysis domain and represents each query through four multiscale equilibrium responses. Signed query-to-surface discrepancies are lifted into a scalar field, and screened-Poisson equilibria reveal how this field responds to increasing spatial coupling. This four-channel decomposition is exactly additive: it preserves the input discrepancy while exposing its structure across spatial scales. A single geometry-conditioned Gaussian potential, shared across categories and trained only on normal scans, evaluates the joint responses against the variation expected at each surface location. Anomaly scores measure the energy gap from the expected normal response, assigning high cost to departures along directions with little normal variation. Despite using one shared model per dataset, our method surpasses not only unified baselines but also the strongest category-specific methods: it achieves point-level AUROC of 92.7% and 97.2% on Real3D-AD and Anomaly-ShapeNet, exceeding the previous best by 4.9 and 3.6 points, respectively, while also attaining the best mean object-level AUROC among unified methods on both benchmarks.

1 INTRODUCTION

3D anomaly detection aims to identify defective objects and localize defects using only normal examples. Recent methods learn normal representations through point-offset prediction (Ye et al., 2025) and geometry-aware reconstruction (Cheng et al., 2025; Kim et al., 2026), with unified models extending across object categories. These advances highlight the value of geometric structure, yet they share a common indirection: the measured query is compared to a learned reconstruction or embedding of it, so where the query actually departs from a normal shape, and over what spatial extent, is not necessarily represented explicitly. Unified models further trade per-category accuracy for generality. We ask whether making the spatial structure of measured discrepancies explicit, and evaluating it against the variation observed in normal scans, can close both gaps.

As illustrated in Fig. 1, a narrow bump and a broad, smooth variation can have similar deviation magnitudes yet carry different evidence of a defect. In a setting where broad variations recur in normal scans, the localized bump is an unusual departure despite its comparable magnitude. The interpretation also depends on the surrounding geometry and the patterns of variation permitted there. Scoring rules driven by per-point deviation magnitude cannot make this distinction, and learned embeddings entangle it with appearance factors that are difficult to control. This motivates a central question: *how can geometric discrepancies be represented together with their spatial extent, and assessed relative to the joint variation expected from normal shapes?*

We propose a unified framework that uses a nominal surface as the computational domain for this analysis. For each category, a surface reconstructed from a normal reference scan provides a fixed geometric basis. We align each query point cloud and lift its signed normal discrepancies into a scalar field on that surface. Crucially, the query itself is never reconstructed or embedded: its measured points enter the analysis directly, and only the reference geometry is fixed. We construct multiscale responses from a common variational formulation that balances fidelity to the observed

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

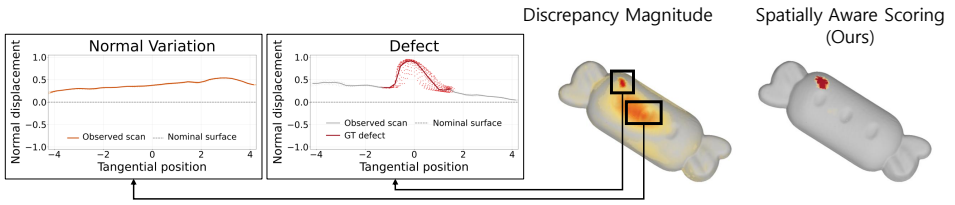


Figure 1: **Beyond discrepancy magnitude**. Two geometric deviations of similar magnitude can carry different evidence of a defect: the left-hand profiles from the boxed regions contrast a narrow bump (right) with broad, smooth normal variation (left). Our method interprets these spatial patterns relative to the variation observed in normal shapes, highlighting the anomalous region.

discrepancy with Dirichlet regularization on the nominal surface. Varying the interaction length yields screened-Poisson equilibria that reveal spatial structure at different scales. Differences between these equilibria, together with the finest residual and the coarsest field, form an exactly additive representation of the discrepancy.

To interpret these responses, we estimate their joint distribution over normal scans, conditioned on nominal surface geometry. A conditional Gaussian defines a quadratic potential whose minimum is the expected normal response and whose precision determines how strongly departures are penalized . Its full covariance captures which scale responses vary together in normal scans, penalizing departures along tightly constrained directions. The potential is trained with no anomaly examples or synthesis; anomalies are scored by the energy gap from its normal reference state, which is zero at the expected response for every conditioning geometry. A single shared model serves all categories within each dataset; only the nominal surface is category-specific.

Experiments on Real3D-AD and Anomaly-ShapeNet show that a single shared model surpasses not only unified baselines but also the strongest category-specific methods: point-level AUROC reaches 92.7% and 97.2%, exceeding the previous best by 4.9 and 3.6 points, and mean object-level AUROC is the highest among unified methods on both benchmarks. Controlled ablations isolate the contributions of the multiscale responses and geometry conditioning, with the largest gains on the noisy real-scan benchmark where spatial structure matters most. Our contributions are:

- A nominal-surface formulation that analyzes measured query discrepancies directly on a fixed reference domain, avoiding query reconstruction and making the spatial interaction of deviations explicit and analyzable.
- An exactly additive multiscale equilibrium representation, derived from a single screened-Poisson operator, that exposes the spatial organization of a discrepancy across scales without discarding any of its amplitude.
- A geometry-conditioned potential, shared across categories and trained on normal scans, whose energy-gap scores achieve state-of-the-art point-level localization on Real3D-AD and Anomaly-ShapeNet while surpassing category-specific methods with one model.

2 Related Work

2.1 2D Anomaly Detection

Feature embedding methods model normal feature distributions or construct representative memory banks for anomaly comparison (Defard et al., 2021; Roth et al., 2022; Gudovskiy et al., 2022), while reconstruction-based approaches identify anomalies from discrepancies between inputs and reconstructed normal counterparts (Zavrtanik et al., 2021; Yao et al., 2024; Zhang et al., 2024). Recent studies further extend anomaly detection to unified multi-category settings (You et al., 2022; Guo et al., 2025; He et al., 2024; Lee et al., 2026). Compared with regular image grids, however, point clouds are sparse and unordered and require explicit modeling of 3D geometric structure.

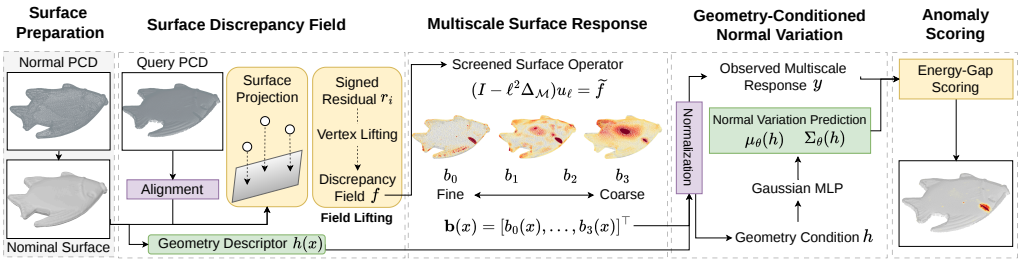


Figure 2: **Overview of the proposed framework.** Query residuals are lifted into a nominal-surface discrepancy field and analyzed through multiscale equilibrium responses. A shared geometry-conditioned Gaussian model localizes anomalies via energy gaps from expected normal responses.

2.2 3D Anomaly Detection

Recent 3D anomaly detection methods exploit handcrafted or pretrained geometric representations for point-cloud anomaly comparison (Bergmann & Sattlegger, 2023; Horwitz & Hoshen, 2023; Wang et al., 2023; Cao et al., 2024), combine geometric alignment and learned point representations into memory-based models (Liu et al., 2023; Chu et al., 2023; Zhu et al., 2024; Liang et al., 2025), or detect anomalies from reconstruction errors or predicted geometric corrections (Li et al., 2024; Zhou et al., 2024b; Ye et al., 2025), alongside zero-shot, unified, and registration-aware formulations (Zhou et al., 2024a; Cheng et al., 2025; Kim et al., 2026; Yu et al., 2025). Recent methods incorporate geometric priors more explicitly: CASL exploits multi-scale curvature for representation learning (Zha et al., 2026), AARD aligns diffusion reconstruction with surface normals (Wu et al., 2026), and PASDF employs continuous geometric representations for localization and repair (Zheng et al., 2025); all use geometry to shape a learned representation of the query. In contrast, we treat the nominal surface as a computational domain for measured query-to-reference discrepancies, without learning a representation of the query itself: the multiscale structure we compute is that of the measured discrepancy, not of the shape. A common surface equilibrium operator characterizes its spatial organization across scales, while a shared geometry-conditioned potential evaluates its joint departure from normal variation.

2.3 Multiscale signal processing

Multiscale signal analysis spans Laplacian pyramids (Burt & Adelson, 1987), surface filtering (Taubin, 1995; Desbrun et al., 1999), and diffusion and spectral graph wavelets (Coifman & Maggioni, 2006; Hammond et al., 2011). Building on these ideas, we construct an additive multiscale representation of query-to-surface discrepancy fields using screened-Poisson equilibria, and combine it with a shared geometry-conditioned model of normal variation for 3D anomaly detection.

3 Methodology

Our method (Fig. 2) evaluates both the spatial pattern of a geometric discrepancy and its consistency with normal shape variation. We first express the differences between a scan and a nominal surface as a scalar field on that surface. We then examine how this field changes under smoothing at several spatial scales, obtaining a multiscale set of response channels. Normal scans establish which combinations of these responses are expected at different surface geometries. At inference, we score query responses by their departure from this learned normal reference. Normal training scans and test queries follow the same field and response construction.

3.1 Surface discrepancy representation

A common domain for geometric differences. Analyzing the spatial pattern of a discrepancy requires a domain that defines how nearby measurements are connected along the object surface. For each category, we reconstruct a nominal triangular surface $\mathcal{M}$ from one predetermined normal scan using pretrained NKSR (Huang et al., 2023). We rigidly align each input scan to this reference using

geometric initialization and robust ICP refinement (Besl & McKay, 1992). Coordinates are normalized by the nominal bounding-box diagonal, and each aligned scan retains its measured points; no resampling or reconstruction is applied to the scan itself. For a scan $Q = \{q_i\}_{i=1}^{N_Q}$ with N_Q points, we measure its signed displacement along the nominal surface normal:

$$p_i = \Pi_{\mathcal{M}}(q_i), \qquad r_i = \boldsymbol{n}(p_i)^{\mathsf{T}}(q_i - p_i). \tag{1}$$

Here p_i is the closest surface point returned by the projection $\Pi_{\mathcal{M}}$, and $\boldsymbol{n}(p_i)$ is the oriented unit normal of its triangle. The scalar r_i records the magnitude and sign of the local normal displacement: positive where the scan lies outside the nominal surface and negative where it lies inside.

From point measurements to a surface field. We lift these measurements to surface vertices by robust aggregation of nearby projected residuals with compatible normals, weighted by distance; see Appendix A for details. The resulting vertex values are linearly interpolated within each triangle, defining a discrepancy field $f(x)$ for $x \in \mathcal{M}$, with spatial relationships supplied by the nominal mesh. Where observations are incomplete, we keep observed values fixed and fill unobserved regions by minimizing squared surface gradients, yielding a harmonic extension $\widetilde{f}$, restricted to connected components containing observations. We retain the observation mask for learning and scoring, so extended values support the surface solves but are not treated as observations.

3.2 Multiscale surface responses

A narrow bump and a broad deformation can have the same maximum discrepancy but different spatial extents. Smoothing over increasing distances separates them. The resulting changes describe the spatial structure of $\widetilde{f}$; normal training data will determine which response patterns are expected.

Balancing fidelity and surface smoothness. At a spatial scale $\ell > 0$, we seek a field u that stays close to $\widetilde{f}$ while varying smoothly along the surface. These two requirements give the objective

$$\mathcal{J}_\ell(u; \widetilde{f}) = \frac{1}{2}\int_{\mathcal{M}} (u - \widetilde{f})^2 \, dA + \frac{\ell^2}{2}\int_{\mathcal{M}} \|\nabla_{\mathcal{M}} u\|_2^2 \, dA, \qquad u_\ell = \arg\min_u \mathcal{J}_\ell(u; \widetilde{f}). \tag{2}$$

Here dA is surface area and $\nabla_{\mathcal{M}} u$ is the tangential gradient, which measures how u changes along the surface. The first term preserves the input values. The second, the Dirichlet energy (Pinkall & Polthier, 1993), penalizes rapid spatial variation. Increasing ℓ strengthens this penalty and smooths the field over a broader range. The minimizing field u_ℓ balances these two terms and is referred to as the equilibrium response at scale ℓ. Here ℓ specifies an analysis length on the nominal surface rather than a physical material constant. The optimality condition for this minimization gives the screened-Poisson equation (Yukawa, 1935; Chuang & Kazhdan, 2011)

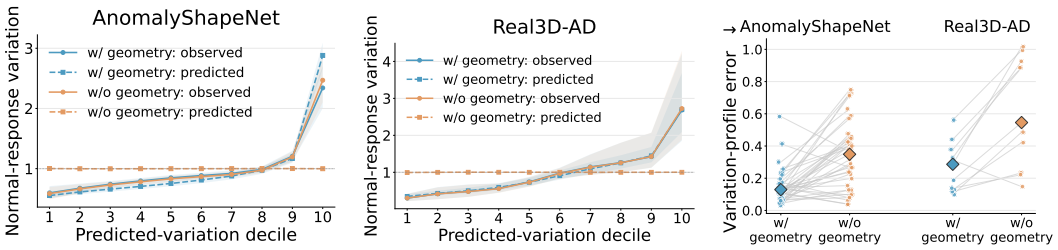
$$(I - \ell^2 \Delta_{\mathcal{M}}) u_\ell = \widetilde{f}, \tag{3}$$

where I is the identity and $\Delta_{\mathcal{M}}$ is the surface Laplacian, defined as the divergence of the tangential gradient. We solve this equation on the nominal mesh, following implicit surface filtering (Desbrun et al., 1999). Each scale produces a separate response to the same input field.

An additive multiscale decomposition. With J dyadic lengths $\ell_j = 2^j h_0$, $j = 0, \ldots, J-1$, we write $u_j = u_{\ell_j}$. The base length h_0 is determined from the nominal mesh spacing and source sampling resolution and remains fixed for all scans of that category. To record the changes introduced by smoothing, we construct

$$\begin{aligned} b_0 &= \widetilde{f} - u_0, \qquad b_J = u_{J-1}, \\ b_j &= u_{j-1} - u_j, \qquad j = 1, \ldots, J-1, \\ \boldsymbol{b}(x) &= [b_0(x), \ldots, b_J(x)]^{\mathsf{T}}. \end{aligned} \tag{4}$$

The first channel records the difference introduced by the finest smoothing. The intermediate channels record additional changes as the smoothing length increases, and the final channel retains the coarsest field. Adding these channels cancels the intermediate fields, giving $\sum_{j=0}^{J} b_j(x) = \widetilde{f}(x)$. Thus, the $(J+1)$-dimensional response retains the input field value while describing how it separates across spatial scales. The decomposition is exactly additive for any J: no component of the input discrepancy is discarded. This allows discrepancies of similar amplitude to have distinct response patterns. We use $J = 3$, giving four response channels, and examine this choice in Table 5b.



(a) Normal-variation tracking on Anomaly-ShapeNet

(b) Normal-variation tracking on Real3D-AD

(c) Category-wise profile error

Figure 3: **Geometry conditioning captures relative variation in normal responses.** (a,b) Held-out normal locations, grouped into deciles of the conditioned model's predicted total variance; the same groups are used with and without conditioning. Solid: observed squared deviations from each model's predicted mean; dashed: predicted variances, each normalized by its overall mean. The conditioned model rises with the observed variation; the unconditioned one cannot. (c) Per-category error between predicted and observed variation profiles, lower with conditioning on both datasets.

3.3 Learning geometry-conditioned normal variation

Normal scans also yield nonzero responses because of ordinary shape variation and measurement differences. The same response can carry different evidence in different geometric contexts. We therefore learn a normal reference for the joint responses, conditioned on the nominal geometry.

Geometry as a shared conditioning variable. At each location, a fixed descriptor $\boldsymbol{h}(x) \in \mathbb{R}^2$ summarizes normal variation and relative departure from the local tangent plane. Both statistics are computed over a single neighborhood size K; see Appendix A for details. Normal variation measures changes in surface orientation, while the normal-height ratio measures out-of-plane displacement relative to the neighborhood's spatial extent. For example, nearly planar and curved regions provide different contexts for a measured discrepancy. Conditioning on nominal geometry allows the model to share information across locations and categories with similar geometric contexts.

A conditional model of normal responses. A shared network with parameters θ predicts the mean response $\boldsymbol{\mu}_\theta(\boldsymbol{h}) \in \mathbb{R}^{J+1}$ and a full positive-definite covariance $\Sigma_\theta(\boldsymbol{h}) \in \mathbb{R}^{(J+1)\times(J+1)}$. We model the normal response with the conditional Gaussian density

$$p_\theta(\boldsymbol{b} \mid \boldsymbol{h}) = \mathcal{N}(\boldsymbol{b}; \boldsymbol{\mu}_\theta(\boldsymbol{h}), \Sigma_\theta(\boldsymbol{h}))\,. \tag{5}$$

The mean specifies the expected response, while the covariance describes normal variability and cross-channel relationships, such as two channels typically increasing together. Conditioning on the fixed nominal descriptor gives all queries the same normal reference at a given surface location.

Fitting the normal distribution. We train one network across categories within each dataset. Response and geometry-descriptor normalization use statistics estimated from normal training scans pooled across all categories and fixed at inference; see Appendix A for details. Below, the predicted mean and covariance are expressed in the band coordinates $\boldsymbol{b}$. Let $\mathcal{D}_{\mathrm{N}}$ denote the normal training response–geometry pairs and let $\boldsymbol{\delta}_\theta = \boldsymbol{b} - \boldsymbol{\mu}_\theta(\boldsymbol{h})$ be the departure from the predicted mean. Minimizing the Gaussian negative log likelihood gives, up to constants independent of θ,

$$\mathcal{L}(\theta) = \mathbb{E}_{(\boldsymbol{b},\boldsymbol{h})\sim\mathcal{D}_{\mathrm{N}}}\left[\frac{1}{2}\boldsymbol{\delta}_\theta^{\mathsf{T}}\Sigma_\theta(\boldsymbol{h})^{-1}\boldsymbol{\delta}_\theta + \frac{1}{2}\log\det\Sigma_\theta(\boldsymbol{h})\right]. \tag{6}$$

The two terms act in opposition on the covariance: the quadratic term alone could be reduced by inflating Σ_θ, flattening the model, and the log-determinant term penalizes exactly this inflation. At their balance, the model estimates the descriptor-conditional mean and covariance of normal responses, capturing their expected values and variability at each geometric context. Training gives categories and scans equal weight and uses surface-area weighting within each scan. Surface operators and response construction remain fixed while the network learns the normal distribution. Model selection uses held-out normal scans. Figure 3 supports this on held-out normal data: locations are

grouped into deciles of the conditioned model's predicted variance, both models are evaluated on the same groups, and category-wise profiles are compared by bin-mass-weighted absolute log error. Conditioning captures variation across these groups and reduces profile error on both datasets.

3.4 NORMAL-REFERENCED ANOMALY SCORING

A potential relative to the normal reference. The equilibrium operators describe how discrepancies are spatially organized, and the learned distribution determines which joint responses are expected at a given nominal geometry. To evaluate a query, we use the Gaussian negative log density as a response-space potential:

$$\Phi_\theta(\boldsymbol{y}; \boldsymbol{h}) = -\log p_\theta(\boldsymbol{y} \mid \boldsymbol{h}). \tag{7}$$

For fixed $\boldsymbol{h}$, this potential is minimized at $\boldsymbol{\mu}_\theta(\boldsymbol{h})$. Its minimum depends on the covariance through the Gaussian's peak density. To measure departure relative to each geometry's normal reference, we subtract this minimum and retain the potential increase caused by the observed response. Next, at each valid projected query point p_i, we interpolate the raw band responses and nominal descriptors linearly from the vertices of the projected triangle, then apply the fixed training normalizations and evaluate the network. Let $\boldsymbol{y}_i$ and $\boldsymbol{h}_i$ be the resulting standardized quantities, and let $\boldsymbol{\delta}_i = \boldsymbol{y}_i - \boldsymbol{\mu}_\theta(\boldsymbol{h}_i)$. The point score is

$$\begin{aligned} s_i &= \Phi_\theta(\boldsymbol{y}_i; \boldsymbol{h}_i) - \Phi_\theta(\boldsymbol{\mu}_\theta(\boldsymbol{h}_i); \boldsymbol{h}_i) \\ &= \log \frac{p_\theta(\boldsymbol{\mu}_\theta(\boldsymbol{h}_i) \mid \boldsymbol{h}_i)}{p_\theta(\boldsymbol{y}_i \mid \boldsymbol{h}_i)} = \frac{1}{2} \boldsymbol{\delta}_i^\mathsf{T} \Sigma_\theta(\boldsymbol{h}_i)^{-1} \boldsymbol{\delta}_i. \end{aligned} \tag{8}$$

The Gaussian normalization terms cancel because both potentials use the same conditioning geometry. We call this score the energy gap, following the convention $p \propto \exp(-E)$ (LeCun et al., 2006). The dimensionless score is zero at the predicted normal mean and increases most along directions with little normal variation. Cross-coordinate correlations determine whether the observed combination follows the learned normal relationships. The fixed invertible affine transformation between $\mathbf{y}$ and $\mathbf{b}$ (Appendix A), together with the corresponding transformations of the mean and covariance, leaves this quadratic energy gap unchanged. The scores s_i serve directly as point-level predictions; the object score is the mean of the highest-scoring 1%, retaining localized evidence without depending on a single extreme point.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

We evaluate on two benchmarks: Real3D-AD (Liu et al., 2023) , with high-resolution real-world scans from 12 categories and 100 test samples per category, and the synthetic Anomaly-ShapeNet (Li et al., 2024), with 1,600 samples across 40 categories; all baseline methods compared below, published since 2024, evaluate on these two benchmarks. Both provide four complete ($360°$) normal point clouds per category for training, while Real3D-AD provides single-view test scans, introducing partial observations at inference. Here, we use only these normal training samples.

Evaluation Metrics. We report the area under the receiver operating characteristic curve at the object level, O-AUROC, for detecting anomalous objects, and at the point level, P-AUROC, for localizing anomalous regions. Both are reported per category and as the mean (%; higher is better).

Baseline Methods. We compare against prior approaches : PatchCore(FPFH) (Roth et al., 2022), M3DM (Wang et al., 2023), Reg3D-AD (Liu et al., 2023), IMRNet (Li et al., 2024), R3D-AD (Zhou et al., 2024b), ISMP (Liang et al., 2025), PO3AD (Ye et al., 2025), PASDF (Zheng et al., 2025), Reg2Inv (Yu et al., 2025), PA3AD (Ning et al., 2026), AF3AD (Balapour & Hach, 2026), and AARD (Wu et al., 2026), which train per-category models, and the unified MC3D-AD (Cheng et al., 2025) and SeDiR (Kim et al., 2026), which, like ours, learn one model across categories within each dataset, with category labels used for supervised losses in SeDiR but for reference selection in MC3D-AD and ours. The results are obtained from publicly available code or their papers.

Table 1: Point-level AUROC (%) on Real3D-AD across 12 categories. Best in bold, second-best underlined. The upper block trains and evaluates one model per category; the lower block uses one unified model for all categories.

Method	Airplane	Car	Candybar	Chicken	Diamond	Duck	Fish	Gemstone	Seahorse	Shell	Starfish	Toffees	Mean
Category-specific Learning													
PatchCore(FPFH) (CVPR22')	56.2	75.4	78.0	42.9	82.8	26.4	82.9	91.0	73.9	73.9	60.6	74.7	68.2
M3DM (CVPR23')	54.7	60.2	67.9	67.8	60.8	66.7	60.6	67.4	56.0	73.8	53.2	68.2	62.0
Reg3D-AD (NeurIPS23')	63.1	71.8	72.4	67.6	83.5	50.3	82.6	54.5	81.7	81.1	61.7	75.9	70.5
R3D-AD (ECCV24')	59.4	55.7	59.3	62.0	55.5	63.5	57.3	66.8	56.2	57.8	60.8	56.8	59.2
ISMP (AAAI25')	75.3	83.6	89.5	79.8	92.6	87.6	88.6	85.7	81.3	83.9	64.1	89.5	83.6
PO3AD (CVPR25')	71.5	57.4	76.3	56.3	61.4	64.4	87.7	53.0	58.5	54.2	65.9	72.9	65.0
PASDF (ICCV25')	77.7	80.2	54.6	76.8	69.9	78.2	83.7	65.4	88.7	64.8	70.3	83.8	74.5
Reg2Inv (NeurIPS25')	**92.3**	94.4	96.9	91.0	97.9	**93.7**	84.6	90.7	64.5	**90.6**	84.0	73.7	87.8
PA3AD (PR26')	74.2	88.0	74.9	89.0	86.1	87.8	91.0	78.7	88.5	88.0	76.8	83.2	83.9
AF3AD (ECCV26')	78.3	75.9	93.5	75.4	**98.4**	82.7	95.7	91.6	81.9	84.8	80.0	95.0	86.1
AARD (CVPR26')	83.5	80.1	85.7	90.6	88.6	89.4	91.7	**92.6**	85.2	84.6	78.5	80.6	86.0
Unified Learning													
MC3D-AD (IJCAI25')	62.8	81.9	91.0	64.0	94.2	82.2	93.2	45.8	65.9	77.8	69.0	93.4	76.8
SeDiR (CVPR26')	66.6	90.8	95.1	63.2	97.7	83.4	96.7	54.7	73.9	79.8	70.9	94.0	80.6
Ours	88.6	**98.2**	**98.2**	**92.3**	97.2	89.9	**99.5**	86.9	**98.9**	66.5	**97.7**	**98.6**	**92.7**

Table 2: Comparison of point-level AUROC(%) on the Anomaly-ShapeNet.

Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5	bucket0	bucket1	cap0	cap3	cap4	cap5	cup0	cup1	eraser0	headset0
Category-specific Learning																					
PatchCore(FPFH)	59.7	57.4	65.4	68.7	51.2	52.4	53.1	62.5	32.7	72.0	35.8	45.9	57.1	47.2	65.3	59.5	79.5	65.5	59.6	81.0	58.3
M3DM	57.7	63.7	66.3	63.7	53.2	65.8	66.3	69.4	65.7	62.4	48.9	69.8	69.9	53.1	60.5	71.8	65.5	71.5	55.6	71.0	58.1
Reg3D-AD	59.7	71.5	88.6	69.6	52.5	77.5	61.5	59.3	65.4	80.0	69.1	61.9	75.2	63.2	71.8	81.5	46.7	68.5	69.8	75.5	58.0
IMRNet	67.1	66.8	55.6	70.2	64.1	78.1	70.5	68.4	59.9	57.6	71.5	58.5	77.4	71.5	70.6	75.3	74.2	64.3	68.8	54.8	70.5
ISMP	86.5	73.4	72.2	86.9	74.0	76.2	70.2	70.6	85.1	75.3	73.3	54.5	68.3	67.2	77.5	66.1	77.0	55.2	85.1	52.4	47.2
PO3AD	96.2	94.9	91.2	84.4	88.0	97.8	91.4	91.8	93.5	96.7	94.1	75.5	89.9	95.7	94.8	94.0	86.4	90.9	93.2	97.4	82.3
PASDF	91.9	95.8	95.1	92.6	94.8	96.3	90.0	81.6	93.9	86.5	90.9	87.5	82.4	94.8	86.1	89.4	92.0	94.8	88.4	94.5	86.3
Reg2Inv	78.5	**99.1**	**99.5**	84.9	81.7	98.3	82.8	82.2	76.1	78.8	82.4	61.0	85.5	86.1	94.5	86.4	97.0	79.8	88.1	98.0	94.6
PA3AD	96.3	96.5	95.8	91.5	94.0	97.1	88.3	93.4	96.4	98.4	94.8	90.0	92.0	95.9	95.5	94.5	96.3	97.5	95.8	98.0	86.1
AF3AD	83.7	93.0	98.6	94.0	88.7	94.1	89.2	**97.3**	97.0	95.9	94.0	75.9	95.4	95.7	96.4	**98.0**	95.9	93.8	88.2	94.6	96.6
AARD	97.1	93.1	95.5	92.7	93.4	97.3	98.2	96.6	95.8	98.1	**97.7**	88.2	95.1	96.0	96.2	96.8	94.9	94.5	95.7	97.0	86.6
Unified Learning																					
MC3D-AD	80.1	81.5	89.5	88.3	90.1	82.7	53.4	60.7	78.3	65.4	55.2	79.0	89.5	86.7	92.2	88.0	88.5	82.9	71.5	85.6	66.6
SeDiR	76.4	89.8	91.1	89.8	94.5	86.6	70.0	79.5	86.9	80.6	64.6	71.7	88.4	90.0	97.2	94.0	95.0	87.1	84.2	74.8	70.8
Ours	**98.9**	**97.9**	**99.5**	**98.4**	**97.7**	**99.3**	**98.8**	93.0	**99.7**	**98.5**	96.0	**92.5**	**95.9**	**97.5**	**97.6**	95.7	**98.4**	**99.4**	**99.2**	**99.4**	**98.8**

Method	headset1	helmet0	helmet1	helmet2	helmet3	jar0	phone	shelf0	tap0	tap1	vase0	vase1	vase2	vase3	vase4	vase5	vase7	vase8	vase9	Mean
Category-specific Learning																				
PatchCore(FPFH)	46.4	54.8	48.9	45.5	73.7	47.8	48.8	61.3	73.3	76.8	65.5	45.3	72.1	43.0	50.5	44.7	69.3	57.5	66.3	58.0
M3DM	58.5	59.9	42.7	62.3	65.5	54.1	35.8	55.4	65.4	71.2	60.8	60.2	73.7	65.8	65.5	64.2	51.7	55.1	66.3	61.6
Reg3D-AD	62.6	60.0	62.4	82.5	62.0	59.9	59.9	68.8	58.9	74.1	54.8	60.2	40.5	51.1	75.5	62.4	88.1	81.1	69.4	66.8
IMRNet	47.6	59.8	60.4	64.4	66.3	76.5	74.2	60.5	68.1	69.9	53.5	68.5	61.4	40.1	52.4	68.2	59.3	63.5	69.1	65.0
ISMP	84.3	61.5	60.3	56.8	52.2	66.1	60.0	70.1	84.4	67.8	68.7	53.4	77.3	62.2	54.6	58.0	74.7	73.6	82.3	69.1
PO3AD	90.7	87.8	94.8	93.2	84.6	87.1	81.0	66.3	78.3	69.2	95.5	88.2	97.8	88.4	90.2	93.7	98.2	95.0	95.2	89.8
PASDF	89.1	81.6	64.6	80.9	95.8	95.9	95.1	86.5	88.4	90.2	94.4	79.7	95.6	86.8	89.9	91.5	95.9	90.9	86.3	89.7
Reg2Inv	97.0	92.5	90.6	89.1	95.6	98.2	**99.2**	63.2	91.8	86.9	98.0	70.5	99.7	84.4	92.7	87.9	86.3	93.4	97.1	88.2
PA3AD	92.2	84.6	96.1	95.2	89.9	95.3	87.9	78.0	89.4	81.7	97.4	**93.9**	98.3	88.9	89.1	96.2	99.2	95.0	92.0	93.1
AF3AD	90.0	**94.1**	78.0	**97.4**	84.5	92.9	94.3	80.3	86.1	90.7	99.0	89.3	97.9	96.1	96.5	86.1	98.4	**97.2**	97.5	92.5
AARD	96.7	86.2	90.7	86.1	89.2	90.5	86.4	89.1	85.7	82.6	98.3	92.4	98.8	93.6	93.6	96.2	98.5	97.1	**98.7**	93.6
Unified Learning																				
MC3D-AD	70.9	74.4	58.2	81.7	60.2	84.8	85.4	65.8	52.4	53.7	87.0	72.6	82.2	79.7	78.6	64.7	61.6	88.5	77.6	75.9
SeDiR	68.7	81.0	59.6	88.5	70.1	91.1	89.1	67.9	59.7	63.0	84.6	72.2	85.0	83.0	88.4	76.6	77.7	90.0	83.5	81.0
Ours	**99.0**	92.4	**99.5**	89.6	**98.2**	**99.6**	98.3	**94.0**	**96.9**	**97.4**	**99.5**	86.6	**99.8**	**96.3**	**97.1**	**99.7**	**99.7**	95.7	96.7	**97.2**

4.2 Experimental Results

Point-level results on Real3D-AD. In Table 1, our method achieves the highest mean P-AUROC of 92.7% among all compared methods. It exceeds Reg2Inv , the strongest category-specific baseline, by 4.9 percentage points and SeDiR , the strongest unified baseline, by 12.1 percentage points. Importantly, these gains are obtained using a single geometry-conditioned response potential shared across categories, rather than separately trained category-specific detectors.

Point-level results on Anomaly-ShapeNet. As shown in Table 2, our method achieves a mean P-AUROC of 97.2%, surpassing the strongest category-specific baseline, AARD, by 3.6 percentage points and the strongest unified baseline, SeDiR, by 16.2 percentage points. Together with the real-

Table 3: Comparison of object-level AUROC (%) on Real3D-AD.

Method	Airplane	Car	Candybar	Chicken	Diamond	Duck	Fish	Gemstone	Seahorse	Shell	Starfish	Toffees	Mean
Category-specific Learning													
PatchCore(FPFH) (CVPR22')	**88.2**	59.0	54.1	83.7	57.4	54.6	67.5	37.0	50.5	58.9	44.1	56.5	59.3
Reg3D-AD (NeurIPS23')	71.6	69.7	68.5	85.2	90.0	58.4	91.5	41.7	76.2	58.3	50.6	82.7	70.4
IMRNet (CVPR24')	76.2	71.1	75.5	78.0	90.5	51.7	88.0	67.4	60.4	66.5	67.4	77.4	72.5
PO3AD (CVPR25')	80.4	65.4	78.5	68.6	80.1	82.0	85.9	69.3	75.6	80.0	75.8	77.1	76.5
PASDF (ICCV25')	62.8	**95.9**	78.8	73.9	89.4	65.8	98.9	63.4	**100.0**	**85.0**	61.7	86.6	80.2
Reg2Inv (NeurIPS25')	81.8	75.8	**100.0**	**94.4**	**100.0**	75.0	67.2	73.5	53.2	69.2	84.1	62.6	78.0
PA3AD (PR26')	76.0	76.9	75.2	67.7	77.6	78.2	95.4	72.8	87.5	80.8	79.6	78.9	78.9
AF3AD (ECCV26')	83.0	79.1	91.4	76.1	99.8	76.4	96.0	**80.9**	82.5	80.8	83.7	92.3	**85.2**
AARD (CVPR26')	**88.2**	72.5	86.4	87.1	91.7	84.6	89.1	76.5	80.7	81.1	75.9	85.5	82.0
Unified Learning													
MC3D-AD (IJCAI25')	85.0	74.9	83.0	71.5	95.5	83.1	86.5	56.0	71.6	80.3	76.6	73.8	78.2
SeDiR (CVPR26')	86.0	78.3	81.9	72.9	94.8	**86.2**	93.8	62.7	67.4	77.9	85.4	84.5	81.0
Ours	69.1	91.9	99.0	85.6	98.5	66.0	**100.0**	72.3	98.6	37.1	**96.8**	**94.6**	84.1

Table 4: Comparison of object-level AUROC (%) on Anomaly-ShapeNet.

Method	ashtray0	bag0	bottle0	bottle1	bottle3	bowl0	bowl1	bowl2	bowl3	bowl4	bowl5	bucket0	bucket1	cap0	cap3	cap4	cap5	cup0	cup1	eraser0	headset0
Category-specific Learning																					
Reg3D-AD	59.7	70.6	48.6	69.5	52.5	67.1	52.5	49.0	34.8	66.3	59.3	61.0	75.2	69.3	72.5	64.3	46.7	51.0	53.8	34.3	53.7
IMRNet	67.1	66.0	55.2	70.0	64.0	68.1	70.2	68.5	59.9	67.6	71.0	58.0	77.1	73.7	77.5	65.2	65.2	64.3	75.7	54.8	72.0
PO3AD	**100.0**	83.3	90.0	93.3	92.6	92.2	82.9	83.3	88.1	98.1	84.9	85.3	78.7	87.7	85.9	79.2	67.0	87.1	83.3	99.5	80.8
PASDF	**100.0**	99.5	**100.0**	**100.0**	**100.0**	**100.0**	94.8	**100.0**	**100.0**	93.3	91.2	**96.8**	77.5	85.2	64.9	64.6	85.3	97.1	85.7	95.2	**100.0**
Reg2Inv	90.0	**100.0**	**100.0**	**100.0**	**100.0**	**100.0**	80.7	65.6	58.5	85.2	81.8	81.3	90.2	65.9	86.3	68.1	90.2	73.3	93.3	**100.0**	**100.0**
PA3AD	**100.0**	93.8	98.1	93.3	97.1	98.9	89.6	95.6	90.4	98.9	92.6	**96.8**	89.5	94.4	93.3	90.5	**92.6**	**100.0**	95.7	**100.0**	**100.0**
AF3AD	99.7	89.7	97.9	99.5	97.9	99.0	83.6	96.6	96.4	93.8	90.0	92.6	88.6	93.1	84.6	92.3	91.9	98.6	87.4	97.6	98.9
Unified Learning																					
MC3D-AD	96.2	80.5	79.5	70.9	75.6	93.0	97.8	71.9	88.5	91.1	75.4	89.8	78.4	79.3	70.1	83.5	76.1	74.3	95.2	77.6	86.2
SeDiR	97.6	88.6	92.9	91.6	98.4	95.9	95.6	92.6	96.7	98.5	**97.9**	91.1	**93.3**	**98.5**	**98.9**	**99.6**	87.4	99.5	**100.0**	66.7	84.9
Ours	**100.0**	98.1	**100.0**	**100.0**	**100.0**	**100.0**	**98.9**	94.1	**100.0**	**100.0**	94.0	96.5	88.9	88.5	90.2	94.0	91.2	**100.0**	**100.0**	**100.0**	**100.0**

Method	headset1	helmet0	helmet1	helmet2	helmet3	jar0	phone	shelf0	tap0	tap1	vase0	vase1	vase2	vase3	vase4	vase5	vase7	vase8	vase9	Mean
Category-specific Learning																				
Reg3D-AD	61.0	60.0	38.1	61.4	36.7	59.2	41.4	68.8	67.6	64.1	48.6	70.2	60.5	65.0	50.0	65.0	46.2	62.0	59.4	57.2
IMRNet	67.6	59.7	60.0	64.1	57.3	78.0	75.5	60.3	67.6	69.6	53.3	75.7	61.4	70.0	52.4	67.6	63.5	63.0	59.4	66.1
PO3AD	92.3	76.2	96.1	86.9	75.4	86.6	77.6	57.3	74.5	68.1	85.8	74.2	95.2	82.1	67.5	85.2	96.6	73.9	83.0	83.9
PASDF	79.5	81.2	93.8	76.5	84.6	**100.0**	**100.0**	71.3	88.2	79.3	**100.0**	92.9	**100.0**	80.6	91.2	**100.0**	**100.0**	92.4	83.6	90.0
Reg2Inv	84.3	81.7	98.6	87.5	87.6	**100.0**	**100.0**	57.7	**94.8**	80.4	99.6	60.5	**100.0**	84.5	81.8	**100.0**	64.3	81.8	87.3	86.1
PA3AD	98.1	80.3	**100.0**	90.7	87.6	95.2	86.2	70.4	92.4	78.2	97.5	90.5	98.1	91.8	78.2	98.1	**100.0**	**93.3**	**98.4**	93.3
AF3AD	96.0	84.8	83.0	**91.1**	82.7	94.3	97.4	72.2	82.8	85.4	93.3	86.1	98.3	**92.5**	88.4	89.9	98.6	86.9	85.9	91.5
Unified Learning																				
MC3D-AD	88.6	67.2	**100.0**	60.9	97.9	97.1	91.9	84.1	94.5	**97.0**	79.5	85.7	92.9	76.1	87.6	76.1	93.8	67.0	73.6	84.2
SeDiR	89.0	**96.5**	**100.0**	87.0	**100.0**	**100.0**	90.5	**88.7**	92.7	89.3	88.8	97.1	97.6	78.8	**92.4**	94.8	**100.0**	87.3	94.2	93.3
Ours	**100.0**	87.8	**100.0**	84.6	90.6	**100.0**	91.4	84.9	82.4	85.9	**100.0**	**100.0**	**100.0**	89.1	82.1	**100.0**	**100.0**	88.8	87.9	**94.8**

scan results, these gains over both baseline groups demonstrate the localization advantage of the multiscale surface equilibrium framework across real-world scans and synthetic shapes.

Object-level results. Our method achieves mean O-AUROCs of 84.1% on Real3D-AD and 94.8% on Anomaly-ShapeNet (Tables 3 and 4), leading unified methods on both benchmarks. It ranks second overall on Real3D-AD, 1.1 percentage points behind AF3AD, and first on Anomaly-ShapeNet, 1.5 points above PA3AD and SeDiR. Compared with category-specific methods published in 2026, including AF3AD and PA3AD, our shared model leads in three of the four benchmark–metric combinations. For Shell, high-scoring normal regions concentrate near repeated ridge endings and the outer boundary. Appendix examines this pattern and its potential impact on top-1% aggregation.

4.3 Ablation and Diagnostic Analysis

Ablation studies. In Table 5, removing equilibrium responses reduces P-AUROC by 12.5 points on Real3D-AD and 3.6 on Anomaly-ShapeNet, highlighting the importance of multiscale spatial information for localization. O-AUROC decreases by 11.8 and 1.1 points, respectively. Removing geometry conditioning most affects object-level detection, lowering O-AUROC by 7.5 and 1.9 points and P-AUROC by 1.3 and 0.3 points. This supports adapting normal-response distributions to local nominal geometry. Restricting covariance to a diagonal matrix also reduces both metrics on both benchmarks, supporting correlation modeling beyond per-coordinate variances. Table 5c shows that joint multiscale modeling outperforms every single-band variant in both detection and localization on both benchmarks. The larger Real3D-AD gains are consistent with the importance of spatially

Table 5: Ablation studies on model components, equilibrium scales, and single-band responses. O/P denotes object-level and point-level AUROC (%), respectively. (a) evaluates components. (b) varies the number of equilibrium scales J. (c) compares single-band and multiscale configuration (Ours).

(a) Components ablation.

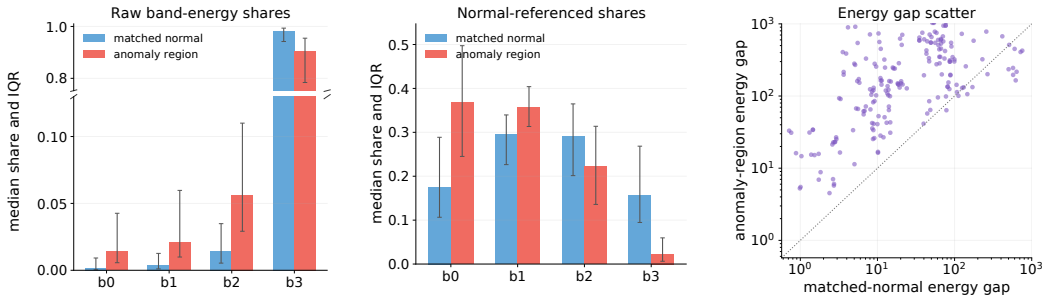
Setting	Real3D-AD	AnomalyShapeNet
w/o Equilibrium responses	72.3 / 80.2	93.7 / 93.6
w/o Geometry conditioning	76.6 / 91.4	92.9 / 96.9
w/o Full covariance	82.5 / 92.0	93.3 / 97.0
Full model	**84.1 / 92.7**	**94.8 / 97.2**

(b) Number of scales.

Setting	Real3D-AD	AnomalyShapeNet
$J = 2$	84.2 / 92.3	94.6 / 97.0
$J = 3$ (Ours)	84.1 / 92.7	94.8 / 97.2
$J = 4$	83.8 / 92.5	94.8 / 97.1
$J = 5$	83.6 / 92.0	94.5 / 96.8

(c) Single-band responses.

Setting	Real3D-AD	AnomalyShapeNet
b_0	83.3 / 86.6	93.9 / 92.2
b_1	83.6 / 89.5	93.4 / 93.7
b_2	81.4 / 88.5	94.0 / 94.6
b_3	79.1 / 86.6	92.7 / 92.8
Ours	84.1 / 92.7	94.8 / 97.2



(a) Raw band-energy shares. (b) Normal-referenced shares. (c) Energy gap scatter.

Figure 4: **Similar deviation magnitudes, different spatial patterns.** Normal–anomaly region pairs on Real3D-AD, matched by category, deviation magnitude and sign, local geometry, and support. Bands range from fine, b_0, to coarse, b_3. (a) Median band-energy shares with interquartile ranges: normal regions concentrate in b_3, anomalies shift toward finer bands. (b) The same shares after centering and scaling by the geometry-conditioned normal statistics, sharpening the contrast. (c) Energy gaps of paired regions; points above the identity line score the anomaly higher.

informed, geometry-conditioned scoring for partial real-world scans. Beyond improving AUROC, geometry conditioning enables closer tracking of held-out normal variation (Figure 3).

Analysis of equilibrium responses. We examine whether multiscale responses distinguish discrepancies of comparable magnitude, using matched normal–anomaly region pairs on Real3D-AD: pairs share category, residual sign, and local geometry, with support-fraction differences below 0.05 and top-5% residual magnitudes within a factor of 1.25, using neither localization metrics nor energy-gap scores. Figure 4a shows normal regions are dominated by the coarse band b_3 and anomalous regions shift toward finer bands, a contrast sharpened by geometry-conditioned normalization (Figs. 4b, 4c). Across 221 pairs, the median residual-magnitude ratio is 1.047, yet the median energy-gap ratio is 8.23, and the anomaly scores higher in 92.8% of cases: pairs indistinguishable by magnitude are separated by their multiscale organization. This holds even when magnitude points the wrong way: among 87 pairs where the normal region's residual is larger, the anomaly still attains the higher energy gap in 79 cases , 90.8%, with median ratio 6.83. These matched-pair results support Fig. 1: spatial organization provides discriminative evidence beyond residual magnitude.

5 Conclusion

We presented a unified framework that directly analyzes measured query-to-surface discrepancies on a nominal surface without query reconstruction. Screened-Poisson equilibria yield an exactly additive multiscale decomposition, preserving the discrepancy and exposing its spatial organization. A geometry-conditioned Gaussian potential trained on normal scans and shared across categories scores joint departures through energy gaps. On both benchmarks, one model per dataset achieves the highest mean P-AUROC overall and O-AUROC among unified methods, while matched normal–anomaly pairs of near-identical magnitude exhibit an 8.23-fold median energy-gap ratio. These results support interpreting geometric deviations not by magnitude alone, but by their spatial organization relative to expected normal variation.

AI USE STATEMENT

In this work, we used generative AI tools for polishing the English text, grammar correction, and improving readability. We have not used generative AI tools for generating core research ideas, deriving mathematical proofs, synthesizing primary literature, or formulating novel scientific claims, and generating experimental data with AI is not applicable to this work. Additionally, we used generative AI tools for formatting LaTeX tables and generating boilerplate code for data visualization. We have reviewed all AI-assisted work. Specifically, all AI-assisted text revisions were meticulously cross-checked by the authors to ensure scientific accuracy and preserve the original meaning. Furthermore, any AI-generated code snippets were thoroughly verified, executed, and tested for correctness by the authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work studies 3D anomaly detection for object inspection using the publicly available Real3D-AD and Anomaly-ShapeNet benchmarks. The study does not involve human participants or the collection of personal data.

REPRODUCIBILITY STATEMENT

Section 3 describes the proposed framework, including surface discrepancy construction, multi-scale equilibrium responses, geometry-conditioned normal-response modeling, and anomaly scoring. Section 4.1 specifies the benchmarks, training data, and evaluation metrics. Appendix A provides implementation details for the model architecture, surface preparation and field lifting, discrete operators and spatial scales, geometry descriptors, and response normalization. It also specifies the optimization settings, random seed, model selection using held-out normal scans, and inference procedure. Appendix B describes the component-ablation settings and the diagnostic protocols for matched-pair response analysis and normal-variation tracking.

REFERENCES

Ali Balapour and Faraz Hach. Anomaly factory 3d: A modular framework for diverse pseudo-anomaly synthesis in unsupervised 3d anomaly detection. *arXiv preprint arXiv:2606.29181*, 2026.

Paul Bergmann and David Sattlegger. Anomaly detection in 3d point clouds using deep geometric descriptors. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2612–2622. IEEE, 2023.

Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pp. 586–606. Spie, 1992.

Peter J Burt and Edward H Adelson. The laplacian pyramid as a compact image code. In *Readings in computer vision*, pp. 671–679. Elsevier, 1987.

Yunkang Cao, Xiaohao Xu, and Weiming Shen. Complementary pseudo multimodal feature for point cloud anomaly detection. *Pattern Recognition*, 156:110761, 2024.

Jiayi Cheng, Can Gao, Jie Zhou, Jiajun Wen, Tao Dai, and Jinbao Wang. Mc3d-ad: A unified geometry-aware reconstruction model for multi-category 3d anomaly detection. *arXiv preprint arXiv:2505.01969*, 2025.

Yu-Min Chu, Chieh Liu, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3D anomaly detection. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 6185–6194. PMLR, 23–29 Jul 2023.

Ming Chuang and Michael Kazhdan. Interactive and anisotropic geometry processing using the screened poisson equation. In *ACM SIGGRAPH 2011 papers*, pp. 1–10. 2011.

Ronald R Coifman and Mauro Maggioni. Diffusion wavelets. *Applied and computational harmonic analysis*, 21(1):53–94, 2006.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International conference on pattern recognition*, pp. 475–489. Springer, 2021.

Mathieu Desbrun, Mark Meyer, Peter Schröder, and Alan H Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pp. 317–324, 1999.

Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *2022 IEEE/CVF winter conference on applications of computer vision (WACV)*, pp. 1819–1828. IEEE, 2022.

Jia Guo, Shuai Lu, Weihang Zhang, Fang Chen, Huiqi Li, and Hongen Liao. Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20405–20415. IEEE, 2025.

David K Hammond, Pierre Vandergheynst, and Rémi Gribonval. Wavelets on graphs via spectral graph theory. *Applied and computational harmonic analysis*, 30(2):129–150, 2011.

Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. *Advances in Neural Information Processing Systems*, 37:71162–71187, 2024.

Eliahu Horwitz and Yedid Hoshen. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2968–2977. IEEE, 2023.

Jiahui Huang, Zan Gojcic, Matan Atzmon, Or Litany, Sanja Fidler, and Francis Williams. Neural kernel surface reconstruction. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4369–4379. IEEE, 2023.

SuYeon Kim, Wongyu Lee, and MyeongAh Cho. A semantically disentangled unified model for multi-category 3d anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 33036–33045, 2026.

Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, Fujie Huang, et al. A tutorial on energy-based learning. *Predicting structured data*, 1(0), 2006.

Geonu Lee, Yujeong Oh, Geonhui Jang, Soyoung Lee, Jeonghyo Song, Sungmin Cha, and YoungJoon Yoo. Continual-mega: A large-scale benchmark for generalizable continual anomaly detection. *Neurocomputing*, pp. 134460, 2026.

Wenqiao Li, Xiaohao Xu, Yao Gu, Bozhong Zheng, Shenghua Gao, and Yingna Wu. Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 22207–22216, 2024.

Hanzhe Liang, Guoyang Xie, Chengbin Hou, Bingshu Wang, Can Gao, and Jinbao Wang. Look inside for more: Internal spatial modality perception for 3d anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 5146–5154, 2025.

Jiaqi Liu, Guoyang Xie, Ruitao Chen, Xinpeng Li, Jinbao Wang, Yong Liu, Chengjie Wang, and Feng Zheng. Real3d-ad: A dataset of point cloud anomaly detection. *Advances in Neural Information Processing Systems*, 36:30402–30415, 2023.

Jian Ning, Qin Zou, Linchun Wu, Yuanhao Yue, Kunmo Li, Shoubin Chen, and Zhongyuan Wang. Physics-inspired pseudo anomaly generation and prototype feature guidance for 3d anomaly detection. *Pattern Recognition*, pp. 114391, 2026.

Ulrich Pinkall and Konrad Polthier. Computing discrete minimal surfaces and their conjugates. *Experimental mathematics*, 2(1):15–36, 1993.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14298–14308. IEEE, 2022.

Gabriel Taubin. A signal processing approach to fair surface design. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 351–358, 1995.

Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8032–8041. IEEE, 2023.

Linchun Wu, Qin Zou, Yuanhao Yue, and Zhongyuan Wang. Geometry-aligned and anomaly-aware reconstruction for 3d anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14648–14657, 2026.

Hang Yao, Ming Liu, Zhicun Yin, Zifei Yan, Xiaopeng Hong, and Wangmeng Zuo. Glad: Towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection. In *European Conference on Computer Vision*, pp. 1–17. Springer, 2024.

Jianan Ye, Weiguang Zhao, Xi Yang, Guangliang Cheng, and Kaizhu Huang. Po3ad: Predicting point offsets toward better 3d point cloud anomaly detection. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1353–1362. IEEE, 2025.

Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35:4571–4584, 2022.

Yuyang Yu, Zhengwei Chen, Xuemiao Xu, Lei Zhang, Haoxin Yang, Yongwei Nie, and Shengfeng He. Registration is a powerful rotation-invariance learner for 3d anomaly detection. *Advances in Neural Information Processing Systems*, 38:5806–5832, 2025.

Hideki Yukawa. On the interaction of elementary particles. i. *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series*, 17:48–57, 1935.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dræm–a discriminatively trained reconstruction embedding for surface anomaly detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8310–8319. IEEE, 2021.

Yaohua Zha, Xue Yuerong, Chunlin Fan, Yuansong Wang, Tao Dai, Ke Chen, and Shu-Tao Xia. Casl: Curvature-augmented self-supervised learning for 3d anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 12340–12348, 2026.

Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16699–16708. IEEE, 2024.

Bozhong Zheng, Jinye Gan, Xiaohao Xu, Xintao Chen, Wenqiao Li, Xiaonan Huang, Na Ni, and Yingna Wu. Bridging 3d anomaly localization and repair via high-quality continuous geometric representation. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27063–27072. IEEE, 2025.

Qihang Zhou, Jiangtao Yan, Shibo He, Wenchao Meng, and Jiming Chen. Pointad: Comprehending 3d anomalies from points and pixels for zero-shot 3d anomaly detection. *Advances in Neural Information Processing Systems*, 37:84866–84896, 2024a.

Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3d-ad: Reconstruction via diffusion for 3d anomaly detection. In *European conference on computer vision*, pp. 91–107. Springer, 2024b.

Hongze Zhu, Guoyang Xie, Chengbin Hou, Tao Dai, Can Gao, Jinbao Wang, and Linlin Shen. Towards high-resolution 3d anomaly detection via group-level feature contrastive learning. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 4680–4689, 2024.

A Methodological Details

A.1 Implementation Details

Model setup. The geometry-conditioned response predictor is an MLP with a 2-dimensional geometry input, two hidden layers of width 64 with SiLU activations, and a 14-dimensional output. The outputs parameterize a four-dimensional mean and the 10 entries of a lower-triangular matrix C. Its diagonal entries use $\mathrm{softplus}(\cdot) + 10^{-4}$. The covariance in standardized coordinates is

$$V = CC^{\mathsf{T}}, \qquad \Sigma_y = 0.9V + 0.1\,\mathrm{diag}(\mathrm{diag}\,V) + 0.05^2 I_4. \tag{1}$$

Here the diagonal term shrinks cross-channel covariance and the identity term provides a variance floor. The output layer is initialized to predict zero mean and identity covariance. Network evaluation uses float32, while likelihoods and energy gaps use float64 Cholesky factorization and triangular solves.

Training and inference setup. We train a separate model for each dataset, sharing its parameters across all categories. Four normal scans per category are used, including the scan supplying the nominal surface. We use AdamW with learning rate 10^{-3}, weight decay 10^{-4}, batch size 4096, 100 updates per epoch, and train for 100 epochs. The gradient norm is clipped at 5, and we use base seed 42. Sampling selects categories and scans uniformly, followed by vertices in proportion to their surface areas. Model selection uses held-out normal scans and fold-specific normalization. For Real3D-AD and Anomaly-ShapeNet, three folds each hold out one non-reference scan per category while retaining the reference scan for fitting. We select the epoch minimizing mean validation negative log likelihood and refit a new model using all training scans. At inference, raw responses and geometry descriptors are barycentrically interpolated to projected query points before normalization and network evaluation. Point scores are the energy gaps; the object score averages their largest 1%.

Surface preparation and field lifting. We reconstruct one nominal mesh per category using pretrained NKSR and normalize lengths by its bounding-box diagonal. Each scan undergoes rigid geometric initialization and robust ICP refinement. Signed point-to-surface residuals are lifted using the 32 nearest projected samples at each nominal vertex. We retain samples whose nominal face normal differs from the vertex normal by less than 60°. Weights combine a Gaussian distance kernel and squared normal agreement. The Gaussian bandwidth is the larger of twice the local vertex spacing and the median neighbor distance. Starting from the weighted mean, we apply three Huber reweighting updates, with the threshold set to 1.5×1.4826 times the weighted median of absolute deviations from the current estimate. Harmonic extension keeps observed values fixed and minimizes Dirichlet energy elsewhere on connected components containing observations.

Surface operators and spatial scales. We use a lumped vertex-area mass matrix M and a symmetric positive-semidefinite cotangent stiffness matrix L. Each equilibrium is computed by solving the sparse system

$$(M + \ell_j^2 L)\mathbf{u}_j = M\widetilde{\mathbf{f}}, \qquad \ell_j = 2^j h_0, \quad j = 0, 1, 2. \tag{2}$$

The base length is $h_0 = \max(2h_{\mathrm{mesh}}, h_{\mathrm{sampling}})$, where h_{mesh} is the median nominal vertex spacing. For h_{sampling}, we project the reference scan onto the mesh, compute each vertex's median distance to its 32 nearest projected samples, and take the area-weighted median across vertices. Operators and lengths remain fixed across scans; open boundaries use the natural zero-flux condition.

Nominal geometry descriptors. We compute two deterministic statistics on each category's nominal mesh. Let $\mathcal{N}_K(v)$ contain vertex v and its $K-1$ nearest vertices under shortest-path distance along mesh edges, with edge lengths as costs. We use $K = 512$ by default; Appendix C.2 reports sensitivity to this choice. Let p_v, n_v, and a_v denote the normalized vertex position, unit vertex normal, and lumped vertex area, respectively. The neighborhood weights are

$$w_{vu} = \frac{a_u}{\sum_{z \in \mathcal{N}_K(v)} a_z}. \tag{3}$$

We define normal variation and the normal-height ratio as

$$h_{\mathrm{nv}}(v) = \sum_{u \in \mathcal{N}_K(v)} w_{vu}\big(1 - |n_u^\top n_v|\big),$$
$$h_{\mathrm{nh}}(v) = \left(\frac{\sum_{u \in \mathcal{N}_K(v)} w_{vu} \big[(p_u - p_v)^\top n_v\big]^2}{\sum_{u \in \mathcal{N}_K(v)} w_{vu} \|p_u - p_v\|_2^2} \right)^{1/2}. \tag{4}$$

The raw descriptor is $h(v) = [h_{\mathrm{nv}}(v), h_{\mathrm{nh}}(v)]^\top \in \mathbb{R}^2$. The first statistic measures normal disagreement without depending on normal orientation signs. The second is the RMS displacement along the central normal divided by the RMS spatial displacement within the neighborhood.

Descriptors are computed once on the nominal mesh and shared by all scans of that category. Each component is standardized using normal-training statistics pooled across all categories within each dataset, with equal scan weights and vertex-area weighting within scans; standard deviations are floored at 10^{-6}. At projected query points, raw descriptors are barycentrically interpolated before applying the fixed training normalization.

Response coordinates and normalization. The implementation models $\mathbf{a} = [f, b_0, b_1, b_2]^\top$. Since $f = \sum_{j=0}^{3} b_j$ at the modeled locations, this is an invertible change of coordinates from the four bands, with $b_3 = f - \sum_{j=0}^{2} b_j$. We standardize responses as $\mathbf{y} = \mathbf{D}^{-1}(\mathbf{a} - \mathbf{m})$, where $\mathbf{m}$ and the diagonal scale matrix $\mathbf{D}$ are estimated from pooled normal training scans across all categories within each dataset. We use equal category and scan weights and vertex-area weighting within each scan. Standard deviations are floored at 10^{-8}. The same normalization is applied to every category and remains fixed at inference. Transforming the predicted mean and covariance back to band coordinates preserves the quadratic energy gap exactly.

A.2 Properties of the surface response operator

We summarize the variational and spectral properties underlying the multiscale responses on a fixed nominal surface $\mathcal{M}$.

Optimality and boundary condition. For every test function $v \in H^1(\mathcal{M})$, stationarity of the fidelity–Dirichlet objective and surface integration by parts give

$$\begin{aligned} 0 &= \int_{\mathcal{M}} (u_\ell - \widetilde{f}) v \, dA + \ell^2 \int_{\mathcal{M}} \nabla_{\mathcal{M}} u_\ell \cdot \nabla_{\mathcal{M}} v \, dA \\ &= \int_{\mathcal{M}} (u_\ell - \widetilde{f} - \ell^2 \Delta_{\mathcal{M}} u_\ell) v \, dA + \ell^2 \int_{\partial \mathcal{M}} v \, \partial_\nu u_\ell \, ds. \end{aligned} \tag{5}$$

Thus $(I - \ell^2 \Delta_{\mathcal{M}}) u_\ell = \widetilde{f}$ in the weak sense, with the natural condition $\partial_\nu u_\ell = 0$ on open boundaries. Here ν is the outward unit co-normal, tangent to the surface and normal to its boundary. The boundary term is absent on a closed surface.

Unique discrete equilibrium. With positive lumped vertex areas, the mass matrix satisfies $M \succ 0$, while the symmetric cotangent stiffness matrix satisfies $L \succeq 0$. For any nonzero vertex vector z,

$$z^{\mathsf{T}}(M + \ell^2 L) z = z^{\mathsf{T}} M z + \ell^2 z^{\mathsf{T}} L z > 0. \tag{6}$$

Consequently, $(M + \ell^2 L) u_\ell = M \widetilde{f}$ has a unique solution. The fidelity term removes the ambiguity associated with the Laplacian's constant modes, including on disconnected components.

Spectral response and stability. Let $L\phi_k = \lambda_k M \phi_k$ with $\phi_k^{\mathsf{T}} M \phi_r = \delta_{kr}$ and $\lambda_k \geq 0$. Writing $\widetilde{f} = \sum_k a_k \phi_k$, where $a_k = \phi_k^{\mathsf{T}} M \widetilde{f}$, gives

$$u_\ell = \sum_k H_\ell(\lambda_k) a_k \phi_k, \qquad H_\ell(\lambda) = \frac{1}{1 + \ell^2 \lambda}. \tag{7}$$

Larger λ represents more rapid spatial variation. Increasing ℓ attenuates these modes more strongly, while $H_\ell(0) = 1$ preserves constants. Since $0 < H_\ell(\lambda) \leq 1$, the response does not amplify input perturbations in the area-weighted norm: $\|\delta u_\ell\|_M \leq \|\delta \widetilde{f}\|_M$, where $\|z\|_M^2 = z^\mathsf{T} M z$. This stability statement holds for fixed surface operators and scales.

Band reconstruction and energy. The bands defined in the main text have spectral responses

$$\begin{aligned} G_0(\lambda) &= 1 - H_{\ell_0}(\lambda), & G_J(\lambda) &= H_{\ell_{J-1}}(\lambda), \\ G_j(\lambda) &= H_{\ell_{j-1}}(\lambda) - H_{\ell_j}(\lambda), & j &= 1, \ldots, J-1. \end{aligned} \tag{8}$$

This spectral-multiplier interpretation connects our multiscale responses to spectral graph wavelet constructions (Hammond et al., 2011). Here, the band filters arise from differences between screened-Poisson equilibrium responses. The intermediate bands capture changes between successive smoothing scales. Telescoping yields $\sum_{j=0}^{J} G_j(\lambda) = 1$ and hence $\sum_{j=0}^{J} b_j = \widetilde{f}$. The same identity holds after linear barycentric interpolation; at scoring locations, $\widetilde{f} = f$. However, the filters overlap, so the bands are generally not orthogonal in the M-inner product:

$$\|\widetilde{f}\|_M^2 = \sum_{j=0}^{J} \|b_j\|_M^2 + 2 \sum_{0 \leq j < k \leq J} b_j^\mathsf{T} M b_k. \tag{9}$$

Exact reconstruction therefore preserves the input field but does not imply an orthogonal energy decomposition or an additive decomposition of the learned joint anomaly score.

B Ablation and Diagnostic Analyses

B.1 Component ablations

Each component-ablation variant in Table 5 is trained independently from scratch. The modifications are summarized below.

Setting	Modification
w/o Equilibrium responses	Replace the response target with $\boldsymbol{t} = [f]^\mathsf{T}$, and remove equilibrium response channels during both training and evaluation.
w/o Geometry conditioning	Fix the geometry-conditioning input to zero during both training and evaluation.
w/o Full covariance	Restrict the covariance to a diagonal matrix during training and evaluation in the standardized response coordinates $\boldsymbol{y}$, retaining geometry-conditioned means and per-coordinate variances.

B.2 Normal-variation tracking

We assess whether geometry conditioning captures the relative variation of normal responses across surface locations (Fig. 3). Evaluation uses only the normal scans held out from model fitting in the three-fold protocol; no anomaly or test scans enter this analysis. The geometry-conditioned model uses the nominal descriptor $\boldsymbol{h}(x)$. The comparison model is trained separately on the same response targets with its geometry-conditioning input fixed to zero. Response and geometry normalization statistics are estimated globally across categories using only the training normals of the corresponding fold, and both models are evaluated in the same standardized target coordinates and at identical held-out locations.

Predicted and observed variation. Let $m \in \{\mathrm{geom}, \mathrm{zero}\}$ index the two models, with the appropriate fold model used at each evaluated location i. In standardized response coordinates, we compute

$$V_i^{\mathrm{pred},m} = \operatorname{tr} \Sigma_i^{(m)}, \qquad V_i^{\mathrm{obs},m} = \|\boldsymbol{y}_i - \boldsymbol{\mu}_i^{(m)}\|_2^2. \tag{10}$$

Thus, each model's observed squared deviation is measured about its own predicted mean.

Common deciles and relative profiles. We partition the pooled held-out locations into ten deciles using $V_i^{\mathrm{pred,geom}}$. Both models use these same location sets $B_1, \ldots, B_{10}$; the zero-input model does not define its own bins. Let w_i denote the evaluation weights, which apply surface-area weighting within scans and balance scans and categories so that their contributions are not determined by vertex counts. For $z \in \{\mathrm{pred}, \mathrm{obs}\}$, define

$$\overline{V}_k^{z,m} = \frac{\sum_{i \in B_k} w_i V_i^{z,m}}{\sum_{i \in B_k} w_i}, \qquad \overline{V}_{\mathrm{all}}^{z,m} = \frac{\sum_i w_i V_i^{z,m}}{\sum_i w_i}, \qquad \widetilde{V}_k^{z,m} = \frac{\overline{V}_k^{z,m}}{\overline{V}_{\mathrm{all}}^{z,m}}. \tag{11}$$

Each curve in panels (a) and (b) is divided by its own benchmark-wide mean, yielding a relative variation profile. This analysis evaluates relative variation tracking rather than absolute covariance calibration.

Category-wise profile error and uncertainty. For panel (c), we retain the same bin membership and form weighted profiles within each category c. Let $\overline{V}_{ck}^{z,m}$ be the weighted mean over category-c locations in bin k, and $\overline{V}_c^{z,m}$ the corresponding category-wide mean. We define

$$\widetilde{V}_{ck}^{z,m} = \frac{\overline{V}_{ck}^{z,m}}{\overline{V}_c^{z,m}}, \qquad \epsilon_c^{(m)} = \frac{\sum_{k: w_{ck} > 0} w_{ck} \left| \log \widetilde{V}_{ck}^{\mathrm{obs},m} - \log \widetilde{V}_{ck}^{\mathrm{pred},m} \right|}{\sum_k w_{ck}}, \tag{12}$$

where w_{ck} is the total evaluation weight of category-c locations in bin k. Lower error indicates closer agreement between the predicted and observed relative profiles. Each point represents one category, connecting lines link the same category across models, and the large diamonds indicate the median across categories. A one-sided paired Wilcoxon signed-rank test supports lower category-wise profile errors with geometry conditioning on both Anomaly-ShapeNet ($p = 8.6153 \times 10^{-7}$) and Real3D-AD ($p = 0.0024414$). The 95% confidence intervals are estimated with 2,000 hierarchical bootstrap replicates, resampling categories and held-out scans within categories rather than individual vertices.

Held-out normal likelihood. We additionally evaluate mean negative log likelihood on the held-out normal validation folds, without decile binning. Compared with the zero-input model, geometry conditioning reduces mean NLL by 0.551 nats on Anomaly-ShapeNet and 0.777 nats on Real3D-AD.

B.3 Strict matched-pair response analysis

We evaluate whether normal and anomalous regions with comparable discrepancy magnitudes and nominal geometry exhibit different multiscale response profiles and normal-referenced energy gaps (Fig. 4). Here, a region denotes a selected set of response points.

Pair construction. For each anomalous scan, we form an anomalous region R_{A} from at most 512 ground-truth-positive response points. For each candidate normal scan in the same category, we select a normal region R_{N} containing the same number of points using nearest-neighbor selection in geometry-descriptor space. We define the regional discrepancy magnitude as

$$r(R) = \mathrm{TopMean}_{5\%}\big(\{|f(x)| : x \in R\}\big), \tag{13}$$

where $\mathrm{TopMean}_{5\%}$ averages the largest 5% of the supplied values. We retain a candidate pair only when the regional signed residuals have the same sign and all three conditions hold:

$$\rho(R_{\mathrm{N}}, R_{\mathrm{A}}) := \frac{\max\{r(R_{\mathrm{N}}), r(R_{\mathrm{A}})\}}{\min\{r(R_{\mathrm{N}}), r(R_{\mathrm{A}})\}} \le 1.25, \qquad d_h \le 0.50, \qquad |\eta_{\mathrm{N}} - \eta_{\mathrm{A}}| \le 0.05. \tag{14}$$

Here d_h is the root-mean-square difference between the standardized geometry descriptors used for matching, and η_{N} and η_{A} are the valid-support fractions of the two regions. Among eligible candidates, we choose the normal region with the smallest matching cost, obtained by summing the residual-ratio, geometry-distance, and support-difference terms after normalization by their respective tolerances. Pair selection uses neither localization performance nor energy-gap scores.

Table 6: Sensitivity to nominal-reference selection, neighborhood size K, and comparison of computational cost. (a) evaluates nominal-reference selection (mean $\pm$ standard deviation of AUROC (%) across five seeds). (b) varies the geometry descriptor's neighborhood size K. (c) compares computational cost in parameters and FLOPs.

(a) Nominal-reference selection.

Real3D-AD		AnomalyShapeNet	
O-AUROC	P-AUROC	O-AUROC	P-AUROC
84.1 ± 0.0367	92.7 ± 0.0388	94.8 ± 0.0270	97.2 ± 0.0042

(b) Neighborhood size K.

Setting	Real3D-AD	AnomalyShapeNet
$K = 128$	81.9 / 92.3	94.1 / 97.0
$K = 256$	82.8 / 92.4	94.0 / 97.0
$K = 512$	84.1 / 92.7	94.8 / 97.2
$K = 1024$	83.6 / 92.0	94.5 / 96.8

(c) Computational cost.

Method	Parameters (M)	FLOPs (G) $\downarrow$
MC3D-AD	34.16	439.21
SeDiR	33.25	508.25
Ours	0.005262	259.78

Response profiles and regional scores. For a scalar quantity z, write $\langle z \rangle_R = |R|^{-1} \sum_{x \in R} z(x)$. With the four response bands $b_0, \ldots, b_3$, the raw band-energy shares are

$$q_j^{\mathrm{raw}}(R) = \frac{\langle b_j^2 \rangle_R}{\sum_{k=0}^{3} \langle b_k^2 \rangle_R}, \qquad j = 0, \ldots, 3. \tag{15}$$

Let $\boldsymbol{\mu}_b(x)$ and $\Sigma_b(x)$ denote the learned normal mean and covariance expressed in band coordinates. We standardize each band by its marginal normal distribution and compute

$$e_j(x) = \frac{b_j(x) - \mu_{b,j}(x)}{\sqrt{[\Sigma_b(x)]_{jj}}}, \qquad q_j^{\mathrm{ref}}(R) = \frac{\langle e_j^2 \rangle_R}{\sum_{k=0}^{3} \langle e_k^2 \rangle_R}. \tag{16}$$

Both profiles are ratios of regional mean squared responses. The normal-referenced shares use marginal standardization to describe response composition; they are not an exact additive decomposition of the joint energy gap when the band-coordinate covariance has off-diagonal entries.

Writing $E(x) = s(x)$ for the pointwise joint energy gap defined in the main text, we score a region by

$$E_R = \mathrm{TopMean}_{5\%}\big(\{E(x) : x \in R\}\big). \tag{17}$$

The top-5% selection in this regional score is performed on energy gaps, separately from the discrepancy-based selection in Eq. 13.

Panels (a) and (b) report the benchmark-wide median and interquartile range of the regional shares across retained pairs, separately for normal and anomalous regions. Panel (c) shows one point per pair, with normal energy on the horizontal axis and anomalous energy on the vertical axis; the gray line marks $y = x$. We additionally summarize the fraction of pairs with $E_{R_\mathrm{A}} > E_{R_\mathrm{N}}$, the median paired ratio $E_{R_\mathrm{A}}/E_{R_\mathrm{N}}$. A one-sided paired Wilcoxon signed-rank test supports higher energy gaps in anomalous regions than in their matched normal regions ($p = 1.41 \times 10^{-23}$).

C Additional Experiments and Analyses

C.1 Reference selection and random-seed sensitivity

We assess sensitivity to nominal-reference selection using five random seeds: 42, 43, 44, 45, and 46. For each seed, one normal training scan is randomly selected per category to construct its nominal surface, and the method is evaluated with the resulting references. Table 6a reports the mean and standard deviation of category-averaged O-AUROC and P-AUROC across the five seeds. This experiment measures how the choice of nominal reference affects detection and localization performance.

C.2 Sensitivity to Descriptor Neighborhood Size

Table 6b evaluates sensitivity to the geometry descriptor's neighborhood size K. P-AUROC varies by at most 0.7 percentage points across the tested settings, with $K = 512$ achieving the highest O-AUROC and P-AUROC on both benchmarks.

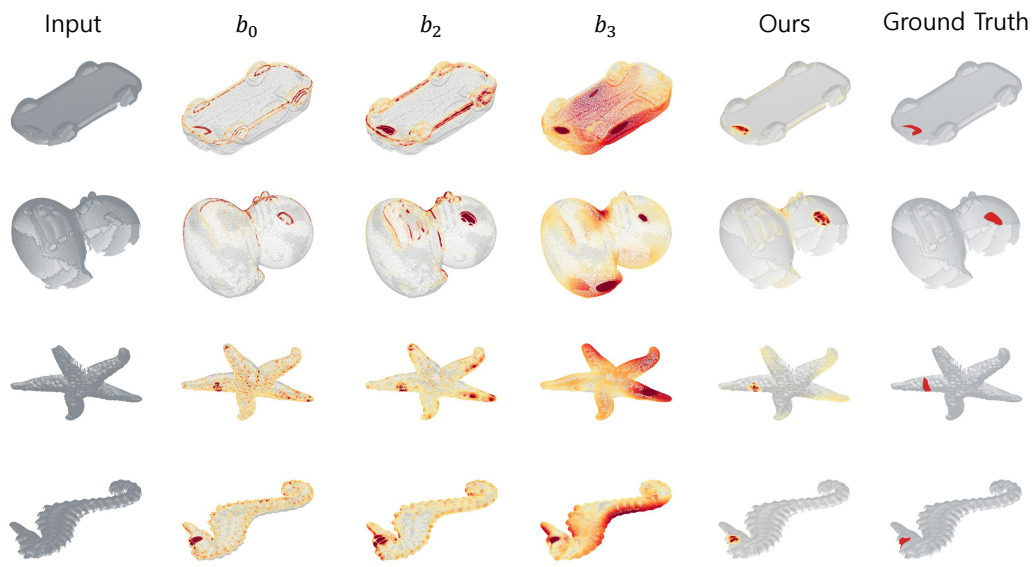


Figure 5: Qualitative results on Real3D-AD. Columns show input scans, selected band responses (b_0, b_2, b_3), final energy-gap scores, and ground-truth annotations.

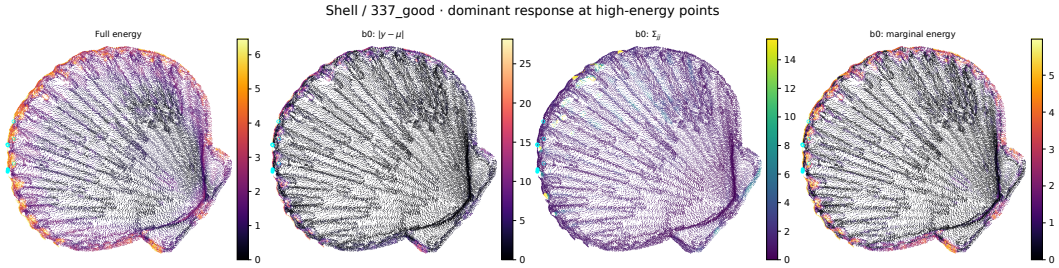


Figure 6: **Analysis on the Shell category.** The normal scan `337_good` nevertheless produces high anomaly scores near the outer rim and the terminations of its repeated ridges. The absolute deviation of the b_0 response from its predicted normal mean, the geometry-conditioned predicted variance, and the corresponding marginal energy are shown at the same surface locations. High marginal energies co-localize with large b_0 residuals without a corresponding collapse of the predicted variance, consistent with a fine-response mismatch with the learned normal reference. Since object-level scoring aggregates the top 1% of point scores, these sparse ridge-associated false positives can disproportionately increase the scan score.

C.3 Computational Cost

Table 6c compares the parameter counts and FLOPs of our method with MC3D-AD and SeDiR. Our method uses only 0.005262M trainable parameters, compared with 34.16M and 33.25M, respectively. This compact parameterization follows from computing multiscale responses with fixed surface operators and learning their geometry-conditioned normal statistics using a small MLP. The reported computational cost is 259.78 GFLOPs, corresponding to reductions of 40.9% and 48.9% relative to MC3D-AD and SeDiR, respectively. The reported FLOPs include registration, field lifting, and multiscale equilibrium solves, in addition to network evaluation.

C.4 Qualitative Results

Figure 5 shows selected Real3D-AD examples with individual band responses, final energy-gap maps, and ground-truth annotations. Individual bands respond to both defects and normal surface structures, while the final geometry-conditioned scores concentrate more closely on the defects.

C.5 Analysis of the Shell Category

Shell is an informative failure case with densely repeated ridges and numerous ridge terminations. Figure 6 shows that the normal scan `337_good` produces localized high energy gaps near ridge endings and the outer rim. These regions exhibit large deviations of the fine response b_0 from its predicted normal mean and high marginal energies, without a corresponding collapse in predicted variance, suggesting a mismatch with the learned normal reference.

Repeated narrow ridges can be sensitive to small spatial shifts arising from sampling, alignment, or surface reconstruction. Such shifts may alter responses near ridge endings while the object remains normal, with repeated structures providing multiple opportunities for localized mismatches. Under object-level aggregation, these false positives can dominate the scan score despite low energy over most of the surface. This highlights the difficulty of distinguishing localized defects from normal variation associated with repeated ridge structures.