PHYSDEFECT: A PHYSICS-BASED BENCHMARK FOR 3D ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

3D anomaly detection is central to industrial inspection, yet progress is constrained by scarce real defects and confidentiality restrictions. Existing synthesis pipelines often apply analytic displacements directly to point clouds. While they can create plausible local anomalies, they struggle to represent global shape deviations, material-dependent responses, and the scanning process used in practice. We introduce PhysDefect, a physics-based benchmark for 3D anomaly detection. Its synthesizer organizes seventeen mechanisms into material deformation, removal, and addition, modeling elastoplastic response, transport-driven interface recession, and continuous deposition and growth. Geometry and material state pass between successive events. A triangulation-based simulated scanner explicitly models occlusion, artifacts, and measurement noise, producing incomplete observations. Across 100 industrial geometries, PhysDefect contains 3,213 anomalous and 2,000 normal instances and evaluates eight established methods plus a geometric baseline. Their localization performance is substantially below published results on other benchmarks, while changes in scanning conditions affect methods in both directions. The benchmark and its components are open source to support research on physical defect synthesis, scan simulation, and transfer to real settings.

1 INTRODUCTION

Additive manufacturing, flexible small-batch production, and distributed contracting disperse manufacturing, reworking, and assembly across processes and sites. Handling, clamping, welding, and storage can cause impact deformation, wear, spatter, and corrosion, affecting assembly accuracy, contact, and service life. Inspection must therefore cover local damage and global deviations: excessive clamping permanently bends thin walls, corrosion alters edges and hole surfaces, and deposits obstruct gaps or change mating surfaces.

Dimensional metrology checks tolerances and assembly dimensions at predefined locations; visual inspection detects color and texture anomalies. Point clouds directly observe dents, bulges, and bending with little color contrast. Point-cloud anomaly detection compares scanned surfaces with normal shapes to localize defects, jointly addressing local damage and global deviations on complex parts with diverse defect locations.

Representative normal and anomalous samples underpin 3D anomaly detection, but real defects are rare and confidentiality restricts collection and release. MVTec 3D-AD, Real3D-AD, and MulSen-AD cover 10, 12, and 15 categories (Bergmann et al., 2022; Liu et al., 2023; Li et al., 2025); Anomaly-ShapeNet expands to 40 through synthesis (Li et al., 2024b). Synthesis thus extends defect coverage. Beyond the diverse local anomalies already supported, representative benchmarks require material-dependent global deformation, sequential damage, and scanning errors.

We introduce PhysDefect, connecting physical defect formation with scanning for 3D anomaly detection. We compare established methods on industrial geometries across scanning conditions and examine simulated scanning on native ShapeNet defects to measure sensitivity to changed observations.

Our main contributions are:

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.

1. **Physical synthesis.** Material coupling supports global deformation, complex surfaces, and damage chains, extending conventional local geometric synthesis.
2. **Simulated scanning.** Occlusion, artifacts, and noise yield point clouds covering incomplete observations and scanning errors in real inspection.
3. **A 3D defect detection benchmark.** Unified data and evaluation compare established methods across scanning levels and ablations, revealing scanning-error sensitivity.

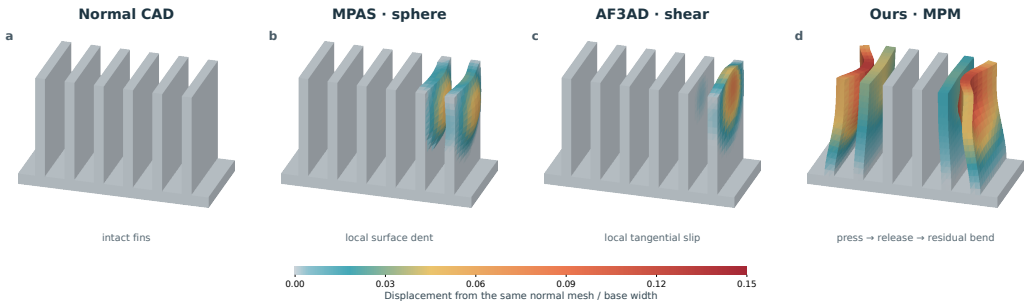


Figure 1: Comparison of defect synthesis approaches: selected presets on a shared geometry.

Figure 1 compares a normal surface, MPAS spherical defect, AF3AD local shear, and physically simulated residual bending, sharing reference geometry, viewpoint, and displacement colors. The local changes and global deviations of the constrained structure illustrate complementary challenges from different preset extents.

2 Related Work

2.1 The Development of 3D Defect Synthesis

Geometric representations and modeling approaches define four stages, progressing from local point perturbations to structural defects and physics-inspired modeling.

Stage 1: direct point displacement. Early methods displace local points of normal clouds outward or inward along prescribed directions to create bulges and dents. R3D-AD and Group3AD train detectors with such local pseudo-anomalies (Zhou et al., 2024; Zhu et al., 2024). These simple operations provide point-level supervision, but region selection and displacement direction determine morphology, with limited surface continuity and structural constraints.

Stage 2: continuous surface deformation. Neighborhoods, normals, and continuous fields organize smooth surface-attached displacements. 3D-PNAS uses principal component analysis for local projection, samples planar Perlin noise, and interpolates it back as normal displacements (Cheng & Du, 2025). DAS3D applies smooth depth increments within convex/concave depth-map masks, aligning geometry and appearance anomalies (Li et al., 2024a). Continuity improves boundaries, while projections and support regions determine spatial extent.

Stage 3: multidimensional support and unified parameterization. MPAS in Synthesis4AD uses 1D/2D/3D support primitives: geodesic skeletons for point-like and line-like defects, planes for bending and cracks, and convex hulls for free-form deformation (Sun et al., 2026). These cover point, line, surface, and volume morphologies; cracks alter point-cloud connectivity. MC4AD's DA-Gen applies displacement fields to local patches (Liang et al., 2025a). AF3AD unifies center conditioning, local coordinates, anisotropic distances, smooth kernels, and directional control in eleven configurable presets (Balapour & Hach, 2026). This broadens morphology coverage and supports comparisons through a common interface.

Stage 4: physics-inspired deformation. PA3AD connects geometry to mechanical concepts through analytical functions: radial basis functions for bulges/dents, perforation and crack functions for structural damage, and Euler–Bernoulli beams for bending (Ning et al., 2026). ScratchSim renders surface-defect images using procedural scratches, material appearance, and viewpoint

visibility (Kühn et al., 2026). These physics-inspired displacement and appearance models enrich synthesis priors.

These stages improve local continuity, morphological diversity, and structural representation, but three limitations remain.

Deformation scale and material response. Local projections and neighborhoods primarily generate local shapes; global deviations additionally depend on constraints, load transfer, and yielding. Although MPAS and PA3AD support bending, global material-dependent elastoplastic response needs further modeling. Bending can preserve near-normal local surfaces, challenging local features and global comparisons differently.

Feedback during continuous physical processes. Defects evolve during formation: corrosion changes transport fields as interfaces retreat, and wear redistributes contact pressure as surfaces flatten. Concentrated reactant or current flux preferentially retreats and rounds convex edges, forming shape through spatially varying reaction rates. Sequential damage requires state retention so subsequent corrosion responds to existing plastic zones.

Scanning after defect generation. Scans cover jointly illuminated and camera-visible surfaces. Deep cavities, holes, and narrow gaps cause missing observations; multiview fusion and noise further alter local geometry. Clean, complete sampling poorly captures these effects. Models using sampling discontinuities as anomaly cues may respond to normal occlusion gaps, motivating explicit scanning after synthesis.

Table 1: Capabilities of defect synthesis methods.

Capability	3D-PNAS	MPAS	MC4AD	AF3AD	PA3AD	ScratchSim	PhysDefect
Material coupling	✗	✗	✗	✗	✗	▲	✓
Global deformation	✗	✓	✗	✗	✓	✗	✓
Topological change	✗	✓	✗	✗	✓	✗	✓
Scanning simulation	✗	✗	✗	✗	✗	▲	✓
Numerical solver	✗	✗	✗	✗	✗	✗	✓
Event chaining	✗	✗	✗	✗	✗	✗	✓

✓ Supported; ▲ Partially supported; ✗ Unsupported. MPAS/PA3AD bending and cracks count toward deformation/topology; numerical solving is assessed separately. ScratchSim's appearance materials and camera visibility give partial material/scanning support. Criteria: Appendix D.

Table 1 summarizes capabilities. Topological change covers cracks, holes, and other connectivity changes associated with multidimensional defects. Event chaining transfers geometry and material state, for example concentrating corrosion in previously stressed regions.

2.2 3D Detection, Physical Simulation, and Scanning

Our evaluated point-cloud methods span feature memory banks BTF and PatchCore (Horwitz & Hoshen, 2023; Roth et al., 2022), multimodal M3DM (Wang et al., 2023), spatial-modality ISMP (Liang et al., 2025b), continuous geometry PASDF (Zheng et al., 2025), and Reg3D, RIF, and Simple3D (Liu et al., 2023; Liang et al., 2026; Cheng et al., 2026). The ShapeNet ablation additionally evaluates reconstruction-based R3D-AD and registration-based Reg2Inv (Zhou et al., 2024; Yu et al., 2025). These configurations use local descriptors, learned representations, or geometric correspondences to compare responses across defects and scanning conditions.

MPM and MLS-MPM solve large deformation and history-dependent material response (Sulsky et al., 1994; Hu et al., 2018); level sets model interfaces (Osher & Sethian, 1988). BlenSor simulates range sensors, while physics-based differentiable depth simulation combines light transport and stereo matching (Gschwandtner et al., 2011; Planche & Singh, 2021). These provide composable foundations for defect formation and observation.

2.3 Our Approach

Addressing shape, process, and observation requirements, PhysDefect connects physical synthesis and simulated scanning. Geometry, materials, and action conditions determine damaged surfaces,

which are then scanned for detection models, integrating object-state changes and observation in one framework.

Physical synthesis comprises seventeen mechanisms in three families. Deformation uses the moving least squares material point method (MLS-MPM) for elastoplastic impact, dropping, scratching, compression, permanent bending, and resonance overload. Removal uses transport fields and level sets for four corrosion mechanisms and contact conditions for erosion/wear. Addition models oxide scale, spatter, deposits, and excess adhesive through growth, deposition, and wetting. Common surfaces and annotations, with geometry and material states passed between events, cover local damage, global deviations, and damage chains.

Simulated scanning uses projector–stereo-camera triangulation. Illumination and visibility determine returns; common instrument parameters control density and depth noise. Paths, multiview fusion, and reconstruction determine coverage and thickness. Easy and medium are calibrated to selected real-cloud statistics; hard is an out-of-range stress test. Separate artifact labels distinguish acquisition errors from defects.

Geometry and materials extend each mechanism across industrial structures: 34,372 source CAD meshes support all seventeen mechanisms across 93 categories; 24 canonical materials supply mechanical, hardness, or oxidation parameters. We benchmark 100 industrial geometries and examine sensitivity to changed observations through a simulated-scanning ablation on native ShapeNet defects.

3 METHOD

PhysDefect combines physical synthesis and simulated scanning. A normal CAD mesh, material parameters, and action conditions produce an annotated damaged surface, which is reobserved as a point-labeled cloud.

3.1 PHYSICAL DEFECT SYNTHESIS

3.1.1 MATERIAL DEFORMATION

Material response. Loading, yielding, and unloading produce residual deformation. MLS-MPM solves elastoplastic response using Young's modulus, Poisson's ratio, density, yield strength, and hardening modulus. Hencky strain energy, von Mises yielding, and linear isotropic hardening give

$$f = \| \operatorname{dev} \tau \| - \sqrt{\tfrac{2}{3}}(\sigma_y + H\alpha) \leq 0, \tag{1}$$

where τ, σ_y, H, and α are Kirchhoff stress, yield strength, hardening modulus, and accumulated equivalent plastic strain. Plastic deformation survives unloading; contact follows Coulomb friction. Material and structure determine residual shapes under identical actions.

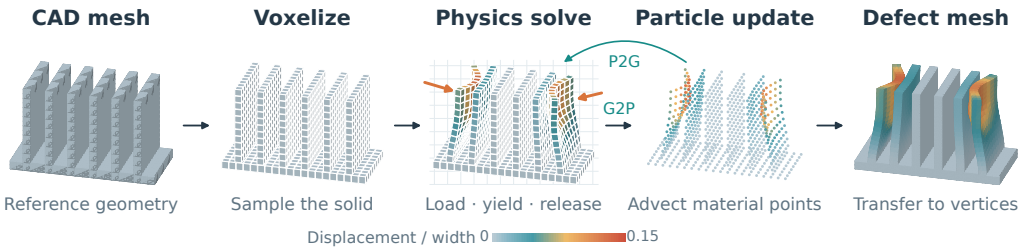


Figure 2: Physical synthesis of stress-induced deformation.

Solution and morphology. Figure 2 discretizes the normal mesh into material-state particles. Momentum is integrated on a background grid that exchanges velocity and stress with particles. After loading and unloading, residual displacements transfer to the surface. The heat-sink fins bend under compression, with root constraints and contact determining the extent; colors show displacement normalized by base width. End panels show simulation meshes; the middle illustrates particle–grid exchange.

Seven mechanisms share this material model, with loading and contact determining defect type. **Single-point impact** produces a local indentation; **ballistic impact** exchanges finite-mass projectile momentum with the part, making depth energy- and material-dependent; in **drop-and-tumble** simulations, the part strikes rigid ground from random poses, damaging colliding corners. **Plowing scratches** indent, drag, and withdraw to form grooves and side ridges. **Clamping damage** uses a broad indenter and lower support for flattening and lateral bulging. **Permanent bending** loads a constrained part beyond yield before release, causing global deviations. **Resonance overload** applies swept-frequency base excitation for mode-dependent damage.

3.1.2 Material Removal

Corrosion and interface evolution. Faraday's law converts local anodic current density i_a to retreat rate $v_n = i_a M/(nF\rho)$, where M, n, F, and ρ are molar mass, electron-transfer number, Faraday's constant, and density. Electrolyte transport determines accessibility; a level set represents the solid interface:

$$\nabla^2 c - \sigma c = 0, \qquad \frac{\partial \phi}{\partial t} = v_n |\nabla \phi|. \tag{2}$$

Here c is normalized transport, σ the near-wall reaction coefficient, and ϕ signed distance. Far-field supply and wall reactions determine local rates. Recomputing transport after interface changes couples geometry and reaction. The Damköhler number balances transport and reaction; charge-coupling scale controls local versus global electrochemical action.

Four mechanisms share this process. **Uniform corrosion** thins surfaces and rounds edges; **crevice corrosion** preferentially erodes concavities through differential aeration; **pitting corrosion** starts at passive-film breakdown sites, with deepening pits increasing shielding and changing subsequent rates. **Stress corrosion** uses preceding plastic strain to enhance erosion through mechanochemical coupling. Stress damage and subsequent localized corrosion thus share geometry and material state.

Erosion and wear. Two contact-driven mechanisms complete removal. **Particle erosion** uses particle direction, visibility, and incidence angle, combining Finnie cutting and Bitter deformation (Finnie, 1960; Bitter, 1963) for material-dependent exposed-side erosion and lee-side shielding. **Sliding wear** follows Archard's law (Archard, 1953), with pressure, sliding speed, and hardness determining rates. Contact and pressure redistribute as surfaces flatten, forming counterface-conforming scars.

3.1.3 Material Addition

Four addition mechanisms use growth, deposition, and wetting. **Oxide scale** follows diffusion-limited growth $x^2 = kt$: thickness x increases with time t but grows more slowly where thicker. The oxide-to-metal volume ratio determines expansion and blistering (Wagner, 1933; Pilling & Bedworth, 1923).

Weld spatter uses ballistic visibility for deposition sites, and volume conservation and contact angle for solidified spherical caps. Flight cooling changes wetting and height-to-width ratios. **Foreign deposits** form irregular particle aggregates flattened along surface normals. **Excess adhesive and sagging** combine surface tension, wetting, and gravity: adhesive collects in accessible concavities/corners; larger droplets sag along tangential gravity.

The families use material state, transport, and deposition conditions to output common damaged surfaces and deviation labels. Implicit addition/removal surfaces support complex connectivity; sequential events inherit surfaces and physical states. Materials enter mechanism-specific response parameters, with dimensionless transport/reaction conditions controlling corrosion.

Annotations and validation. After rigid-motion removal, deviation at damaged point x is $q(x) = |d_{\text{signed}}(x, S_{\text{ref}})|/L$, with normal surface S_{ref}, longest side L, and binary label $y(x) = \mathbb{1}[q(x) > \tau]$. Plastic strain is an auxiliary channel. Severity constrains peak deviation and anomalous fraction for observable, comparable samples. Tests yield 0.9% maximum pitting-interface error against a threefold-finer reference and a 6.48 median corrosion-depth ratio between high-/low-strain regions, supporting solver and event-coupling consistency (Appendix A.7).

Figure 3 shows seventeen mechanisms and a normal reference on one spur gear: deformation produces dents, ridges, and global deviations; removal selectively retreats edges, crevices, and contact

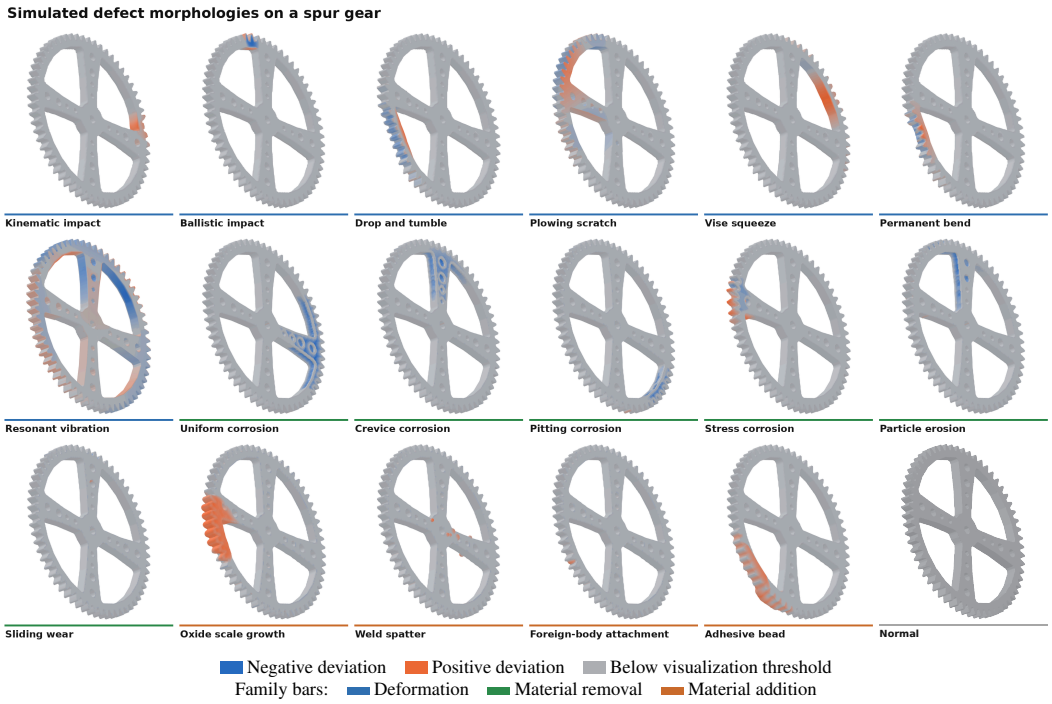


Figure 3: Seventeen defect mechanisms in three families and a normal reference. Panel-wise normalization precludes cross-panel intensity comparisons.

regions; addition produces oxide, particles, and adhesive. Blue/orange denote negative/positive deviations, normalized by each panel's peak.

3.2 Simulated Scanning

Projector–camera triangulation scans damaged surfaces. Multiview acquisition and fusion produce point clouds with occlusion, noise, and reflection artifacts. Figure 4 illustrates three components.

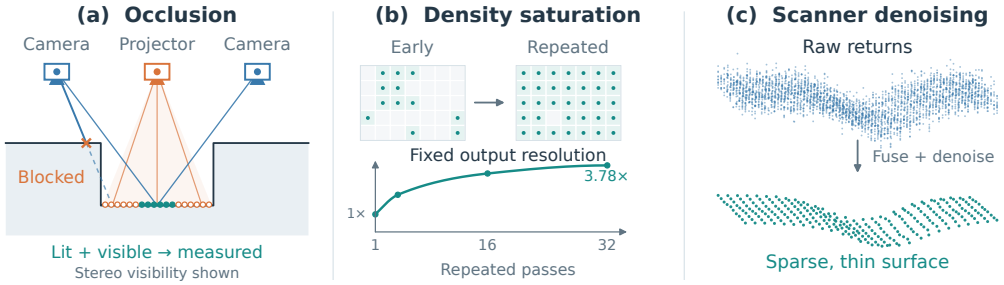


Figure 4: Occlusion, density saturation, and denoising in simulated scanning.

Occlusion and missing observations. In Figure 4(a), returns require illumination and camera visibility. Groove rims can occlude illuminated floors; stereo requires both cameras, producing missing regions in deep cavities, gaps, and holes. Green filled points are jointly visible; orange open points are occluded. Distance increases spacing and depth uncertainty; a larger effective baseline reduces uncertainty but increases cavity occlusion. Common instrument parameters thus couple density, noise, and visibility.

Paths and density saturation. In Figure 4(b), nearby stations prioritize under-covered regions. Repeated returns accumulate in fixed-resolution surface cells until saturation redirects scanning

elsewhere. Longer dwell mainly adds repetition, saturating output counts: 32 passes produce approximately 3.78 times as many fused points.

Fusion and denoising. In Figure 4(c), weighted multiview fusion merges returns; local planar reconstruction reduces thickness. Registration drift, mixed pixels, radiometric dropout, and outliers are included. Point spacing, density uniformity, and local thickness calibrate outputs. Common-view, common-scale clouds show noisy returns becoming thinner surfaces. Thickness ratio divides neighborhood normal-direction standard deviation by spacing, enabling cross-size calibration. Table 2 reports one calibration configuration; Appendix B diagnoses denoising bias.

Table 2: Statistical calibration against released real point clouds.

Statistic	Real3D-AD	easy	medium
Normalized point spacing	1.55×10^{-3}	7.8×10^{-4}	2.09×10^{-3}
Density uniformity	1.34	1.53	1.31
Local thickness ratio	0.014–0.045	0.021	0.028

To simulate different observational difficulty, easy, medium, and hard control coverage, visibility, noise, and artifacts (Table 3). They share damaged surfaces, separating observation changes from defect changes.

Table 3: Scanning difficulty levels and their main challenges.

Level	Occlusion	Multipath ghosts	Noise	Surface processing
easy	Broad multiview coverage	Off	Low	Fusion + reconstruction
medium	Fewer views	Off	Medium	Fusion + reconstruction
hard	Few views, stereo visibility	On	High	Fusion

Coverage decreases and degradation increases across levels. Easy and medium are calibrated to selected real-cloud statistics; hard stress-tests stronger errors. All include drift, mixed pixels, and outliers; hard adds multipath ghosts. Scanning conditions define difficulty, and repeated observations of the same defect isolate model responses to observation changes.

Annotations follow surface hits, while artifact types distinguish defects from scanning errors, preserving visible defect regions and valid evaluation points. Scanning implementation and parameters appear in Appendix B.

4 Experiments

We compare anomaly localization performance on PhysDefect and examine the effect of simulated scanning on fixed native ShapeNet defects.

4.1 Dataset and Evaluation Protocol

Dataset. The source asset index contains 122,858 geometries. Its CAD models eligible for synthesis screening come from Mechanical Components Benchmark, SimJEB, and the segmentation and reconstruction sets of Fusion 360 Gallery (Kim et al., 2020; Whalen et al., 2021; Lambourne et al., 2021; Willis et al., 2021). Screening leaves 34,372 geometries that support all seventeen mechanisms; we select 100 industrial geometries for evaluation. The dataset spans seventeen mechanisms and easy/medium/hard scanning, with 3,213 anomalous and 2,000 normal instances. Mesh validity and target severity govern acceptance, so mechanism counts depend on geometry. Repeated scans capture observation variability.

Normal and anomalous samples. Normal and anomalous samples scan surfaces before and after defect generation. Of 20 normal instances per geometry, 16 form the reference bank representing normal shape; four join anomalous instances for testing. Counts $K = 1/4/16$ use nested prefixes. Three levels yield 4,800 reference and 10,839 test scans. Only normal references support training/adaptation; references and tests have independent unknown poses.

Metric. Point-level AUROC (P-AUROC) is computed per scan containing both classes, averaged within geometry–mechanism cells, then equally over geometries within each mechanism, and finally over mechanisms. Type labels exclude artifacts; single-class scores are undefined. The main table reports conditional means over scorable samples, with failures/undefined scores missing and effective counts in Appendix C. This measures visible-surface localization; normal-object false alarms and artifact recognition require separate evaluation.

Compared methods. BTF-FPFH, PatchCore-FPFH, PASDF, ISMP, Reg3D, RIF (Liang et al., 2026), M3DM-PointMAE, and Simple3D (Cheng et al., 2026) cover local features, continuous geometry, and learned representations in point-cloud configurations. Training-free PTP combines estimated registration and point-to-plane residuals. In the main experiment, Simple3D uses MSND/LFSA and shared geometric registration, only at $K = 1$. Rows represent the adaptation and registration configurations in Appendix C.

4.2 Main Results

Table 4: P-AUROC of evaluated point-cloud configurations across scanning levels and reference counts.

Method	$K = 1$			$K = 4$			$K = 16$		
	easy	med.	hard	easy	med.	hard	easy	med.	hard
PTP (our baseline)	**0.801**	**0.792**	**0.727**	**0.827**	**0.811**	**0.743**	**0.839**	**0.825**	**0.748**
BTF-FPFH (CVPRW 2023)	0.646	0.633	0.634	0.593	0.584	0.588	0.568	0.563	0.571
PatchCore-FPFH (CVPR 2022)	0.562	0.583	0.585	0.556	0.576	0.578	0.550	0.567	0.569
PASDF (ICCV 2025)	0.742	0.724	0.692	0.748	0.726	0.705	0.743	0.717	0.694
ISMP (AAAI 2025)	0.679	0.642	0.639	0.709	0.621	0.597	0.730	0.617	0.595
Reg3D (NeurIPS 2023)	0.689	0.677	0.640	0.703	0.647	0.614	0.722	0.605	0.586
RIF (PR 2026)	0.476	0.497	0.492	0.471	0.488	0.487	0.472	0.480	0.485
M3DM-PointMAE (CVPR 2023)	0.576	0.578	0.570	0.621	0.617	0.602	0.658	0.649	0.625
Simple3D (AAAI 2026)	0.721	0.720	0.660	–	–	–	–	–	–

Bold: column best. Venues identify original methods (PR: Pattern Recognition). Simple3D uses only $K = 1$; – denotes unevaluated configurations. Adaptations and registration: Appendix C.

Overall performance. Table 4 compares three scanning levels and reference sizes. The evaluated PTP configuration leads all nine settings in conditional mean P-AUROC: $0.839/0.825/0.748$ at $K = 16$, versus PASDF's $0.743/0.717/0.694$. Simple3D scores $0.721/0.720/0.660$ at $K = 1$, losing about 6 percentage points from easy to hard. Common conditions for geometric comparisons, local features, and learned representations reveal room to improve localization under complex defects and degradation.

Scanning conditions. At $K = 16$, easy-to-hard decreases are 9.1, 13.5, and 13.6 percentage points for PTP, ISMP, and Reg3D. M3DM-PointMAE changes $0.658 \rightarrow 0.625$ and PatchCore-FPFH $0.550 \rightarrow 0.569$, showing different sensitivity directions and magnitudes. Levels jointly vary coverage, noise, and processing; the following paired comparison isolates the change in observation for fixed native defects.

Context from published benchmark results. Table 5 contrasts representative original-paper results with ours. All main configurations remain below 0.9, reaching at most 0.839, below high public-benchmark scores. This is cross-dataset context: resolution, adaptation, reference counts, and aggregation differ and jointly contribute to these gaps.

Number of normal references. Extra references expand surface coverage and observation variability. From $K = 1$ to $K = 16$ on hard, PTP improves $0.727 \rightarrow 0.748$, M3DM-PointMAE $0.570 \rightarrow 0.625$, and BTF-FPFH changes $0.634 \rightarrow 0.571$. Benefits depend on representation, matching, and scanning, motivating separate reference-size reporting.

Table 5: Published benchmark results and PhysDefect results (P-AUROC).

Method	Original dataset	Published	PhysDefect
M3DM (point-only) (Wang et al., 2023)	MVTec 3D-AD	**0.970**	0.658
ISMP (Liang et al., 2025b)	Real3D-AD	**0.836**	0.730
PASDF (Zheng et al., 2025)	Anomaly-ShapeNet	**0.897**	0.748
Simple3D (Cheng et al., 2026)	Real3D-AD	**0.923**	0.721

PhysDefect: each method's maximum in Table 4 (Simple3D: $K = 1$ only). Bold: row maximum. Cross-protocol context only; adaptations, resolution, reference counts, and aggregation differ.

4.3 Simulated Scanning on Native ShapeNet Defects

We evaluate eleven methods on 79 paired native defects from four Anomaly-ShapeNet categories: cabinet, chair, screen, and vase. Original point clouds use official labels; medium-level simulated scans of the same defect meshes use transferred labels (Table 6).

Both normal references and test clouds follow each observation condition. Reference-bank methods use four normal samples per category; PTP and fixed R3D-AD use one. Method settings stay fixed. P-AUROC is averaged within and then equally across categories; changes pair each defect's scanned and original scores (Appendix C.3).

Table 6: Simulated scanning on 79 native defects from four Anomaly-ShapeNet categories. P-AUROC is category macro-averaged; Δ is paired scanned minus original, using unrounded scores. Bold marks the highest score in each performance column.

Method	K	Original	Scanned	Change Δ
Simple3D	4	**0.874**	**0.835**	-0.038
PTP	1	0.770	0.734	-0.036
Reg2Inv	4	0.693	0.648	-0.045
PASDF	4	0.702	0.668	-0.034
R3D-AD†	1	0.525	0.532	+0.007
Reg3D	4	0.725	0.667	-0.058
PatchCore-FPFH	4	0.563	0.622	+0.059
M3DM-PointMAE	4	0.560	0.553	-0.007
BTF-FPFH	4	0.545	0.566	+0.021
ISMP	4	0.701	0.662	-0.039
RIF	4	0.516	0.557	+0.040

Both observations use the same 79 defects. †Fixed autoencoder transfer configuration; see Appendix C.3.

Simple3D leads in both conditions but changes from 0.874 to 0.835 ($\Delta = -0.038$); PTP changes by -0.036. Seven methods decline and four improve, spanning Reg3D's -0.058 to PatchCore-FPFH's $+0.059$. The direction of the scanning effect thus depends on the method.

5 Conclusion

PhysDefect unifies physical defect formation and simulated scanning for 3D anomaly detection. Seventeen mechanisms cover material deformation, removal, and addition, passing geometry and material state across damage events. A triangulation-based scanner adds occlusion, artifacts, and noise, enabling paired observation tests.

Across eight methods and a geometric baseline, localization falls substantially below reported scores on MVTec 3D-AD, Real3D-AD, and Anomaly-ShapeNet. On native ShapeNet defects, simulated scans change relative rankings: some methods improve, others degrade. This exposes the interaction between defect morphology and acquisition as a practical challenge.

The synthesizer, scanner, and evaluation protocol are open source. Future work will scale synthesis for cross-category and zero-shot detectors, accelerate GPU simulation for embodied-learning data, extend coupled damage and material models, and test transfer to measured scans and real defects.

REFERENCES

J. F. Archard. Contact and rubbing of flat surfaces. *Journal of Applied Physics*, 24(8):981–988, 1953.

Ali Balapour and Faraz Hach. Anomaly factory 3d: A modular framework for diverse pseudo-anomaly synthesis in unsupervised 3d anomaly detection. *CoRR*, abs/2606.29181, 2026. doi: 10.48550/arXiv.2606.29181.

Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022, Volume 5: VISAPP*, pp. 202–213, 2022.

J. G. A. Bitter. A study of erosion phenomena part i. *Wear*, 6(1):5–21, 1963.

Yifeng Cheng and Juan Du. 3d-pnas: 3d industrial surface anomaly synthesis with perlin noise. *CoRR*, abs/2504.12856, 2025. doi: 10.48550/arXiv.2504.12856.

Yuqi Cheng, Yihan Sun, Hui Zhang, Weiming Shen, and Yunkang Cao. Towards high-resolution 3d anomaly detection: A scalable dataset and real-time framework for subtle industrial defects. In *Fortieth AAAI Conference on Artificial Intelligence*, 2026.

Iain Finnie. Erosion of surfaces by solid particles. *Wear*, 3(2):87–103, 1960.

Michael Gschwandtner, Roland Kwitt, Andreas Uhl, and Wolfgang Pree. BlenSor: Blender sensor simulation toolbox. In *Advances in Visual Computing*, volume 6939 of *Lecture Notes in Computer Science*, pp. 199–208. Springer, 2011. doi: 10.1007/978-3-642-24031-7_20.

E. M. Gutman. Mechanochemistry of solid surfaces. *World Scientific*, 1994.

Eliahu Horwitz and Yedid Hoshen. Back to the feature: Classical 3d features are (almost) all you need for 3d anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 2968–2977, 2023. doi: 10.1109/CVPRW59228.2023.00298.

Yuanming Hu, Yu Fang, Ziheng Ge, Ziyin Qu, Yixin Zhu, Andre Pradhana, and Chenfanfu Jiang. A moving least squares material point method with displacement discontinuity and two-way rigid body coupling. *ACM Transactions on Graphics*, 37(4):150:1–150:14, 2018.

Denny A. Jones. *Principles and Prevention of Corrosion*. Prentice Hall, 2nd edition, 1996.

Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32:922–923, 1976.

Sangpil Kim, Hyung-gun Chi, Xiao Hu, Qixing Huang, and Karthik Ramani. A large-scale annotated mechanical components benchmark for classification and retrieval tasks with deep neural networks. In *Computer Vision – ECCV 2020*, 2020. doi: 10.1007/978-3-030-58523-5_11.

Paul Julius Kühn, Saptarshi Neil Sinha, Tiago Kleist, Richard Hoffmann, Arjan Kuijper, and Michael Weinmann. Scratchsim: A procedural synthetic data pipeline for surface scratch detection. `https://arxiv.org/abs/2607.27065`, 2026.

Joseph G. Lambourne, Karl D. D. Willis, Pradeep Kumar Jayaraman, Aditya Sanghi, Peter Meltzer, and Hooman Shayani. BRepNet: A topological message passing system for solid models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12773–12782, 2021.

Kecen Li, Bingquan Dai, Jingjing Fu, and Xinwen Hou. DAS3D: dual-modality anomaly synthesis for 3d anomaly detection. In *Computer Vision - ECCV 2024 Workshops*, volume 15626 of *Lecture Notes in Computer Science*, pp. 148–165. Springer, 2024a.

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 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22207–22216, 2024b.

Wenqiao Li, Bozhong Zheng, Xiaohao Xu, Jinye Gan, Fading Lu, Xiang Li, Na Ni, Zheng Tian, Xiaonan Huang, Shenghua Gao, and Yingna Wu. Multi-sensor object anomaly detection: Unifying appearance, geometry, and internal properties. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9984–9993, 2025.

Hanzhe Liang, Aoran Wang, Jie Zhou, Xin Jin, Can Gao, and Jinbao Wang. Examining the source of defects from a mechanical perspective for 3d anomaly detection. *CoRR*, abs/2505.05901, 2025a. doi: 10.48550/arXiv.2505.05901.

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 *Thirty-Ninth AAAI Conference on Artificial Intelligence*, 2025b.

Hanzhe Liang, Jie Zhou, Can Gao, Bingyang Guo, Jinbao Wang, and Linlin Shen. A lightweight 3d anomaly detection method with rotationally invariant features. *Pattern Recognition*, 2026. doi: 10.1016/j.patcog.2025.112924.

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. In *Advances in Neural Information Processing Systems 36 (NeurIPS), Datasets and Benchmarks Track*, 2023.

William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. *ACM SIGGRAPH Computer Graphics*, 21(4):163–169, 1987.

John Newman and Karen E. Thomas-Alyea. *Electrochemical Systems*. Wiley-Interscience, 3rd edition, 2004.

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*, 2026. doi: 10.1016/j.patcog.2026.114391.

Stanley Osher and James A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

N. B. Pilling and R. E. Bedworth. The oxidation of metals at high temperatures. *Journal of the Institute of Metals*, 29:529–582, 1923.

Benjamin Planche and Rajat Vikram Singh. Physics-based differentiable depth sensor simulation. *arXiv preprint arXiv:2103.16563*, 2021. URL `https://arxiv.org/abs/2103.16563`.

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

Giovanni Russo and Peter Smereka. A remark on computing distance functions. *Journal of Computational Physics*, 163(1):51–67, 2000.

Deborah Sulsky, Zhen Chen, and Howard L. Schreyer. A particle method for history-dependent materials. *Computer Methods in Applied Mechanics and Engineering*, 118(1–2):179–196, 1994.

Yihan Sun, Yuqi Cheng, Junjie Zu, Yuxiang Tan, Guoyang Xie, Yucheng Wang, Yunkang Cao, and Weiming Shen. Synthesis4ad: Synthetic anomalies are all you need for 3d anomaly detection. *CoRR*, abs/2604.04658, 2026. doi: 10.48550/ARXIV.2604.04658.

Mark Sussman, Peter Smereka, and Stanley Osher. A level set approach for computing solutions to incompressible two-phase flow. *Journal of Computational Physics*, 114(1):146–159, 1994.

Carl Wagner. Beitrag zur theorie des anlaufvorgangs. *Zeitschrift für physikalische Chemie*, 21B: 25–41, 1933.

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

Eamon Whalen, Azariah Beyene, and Caitlin Mueller. SimJEB: Simulated jet engine bracket dataset. *Computer Graphics Forum*, 40(5):9–17, 2021. doi: 10.1111/cgf.14353.

Karl D. D. Willis, Yewen Pu, Jieliang Luo, Hang Chu, Tao Du, Joseph G. Lambourne, Armando Solar-Lezama, and Wojciech Matusik. Fusion 360 gallery: A dataset and environment for programmatic CAD construction from human design sequences. *ACM Transactions on Graphics*, 40 (4):54:1–54:24, 2021. doi: 10.1145/3450626.3459818.

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. *CoRR*, abs/2510.16865, 2025.

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 *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 27063–27072, 2025.

Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3D-AD: reconstruction via diffusion for 3d anomaly detection. In *Computer Vision - ECCV 2024 - 18th European Conference*, 2024.

Hongze Zhu, Guoyang Xie, Chengbin Hou, Tao Dai, Can Gao, Jinbao Wang, and Linlin Shen. Towards multi-class pre-trained models for 3d anomaly detection: A group-based approach. In *Proceedings of the ACM International Conference on Multimedia (MM)*, 2024.

A Implementation Details

This appendix supplements material configurations, numerical solvers, annotation transfer, scanner parameters, and evaluation settings. All mechanisms output a common damaged surface and annotations for subsequent events and scanning.

A.1 Geometry and Material Configurations

The source asset index contains 122,858 geometries: 121,419 CAD models eligible for synthesis screening and 1,439 evaluation-only geometries from Real3D-AD and MulSen-AD (Liu et al., 2023; Li et al., 2025). Screening for topology, watertightness, distance-field applicability, and other requirements identifies 34,372 CAD geometries supporting all seventeen mechanisms, spanning 93 part categories. Evaluation uses 100 independent geometries. Each is centered at its bounding-box center and normalized by its longest side. Different defects and scans of the same geometry retain a common object identity.

The material table contains 24 canonical materials spanning steels, stainless steels, cast irons, aluminum, copper, titanium, and zinc alloys, and engineering plastics, with sampling weights assigned by part category. Stress deformation uses Young's modulus, Poisson's ratio, density, yield strength, and hardening modulus, selecting materials according to plastic applicability. Erosion uses material family and hardness; oxide-scale growth uses the Pilling–Bedworth volume ratio. Corrosion controls local rates through dimensionless transport and reaction parameters, whose configuration determines the extent of material coupling.

Yield strength and hardening modulus are nondimensionalized by Young's modulus, preserving yield-strain ratio and relative hardening strength. Ballistic contact also preserves the projectile-to-part density ratio.

A.2 Material Point Method and Surface Transfer

Particles sampled at interior voxel centers carry velocity, deformation gradient, and accumulated plastic strain. MLS-MPM transfers state between particles and the background grid using quadratic B-spline weights (Hu et al., 2018), integrates momentum on the grid, and updates constitutive response on particles. Contact uses tool implicit surfaces and Coulomb friction; plastic updates use von Mises yield projection with linear hardening.

Local shallow defects combine a coarse global solve with local refinement, maintaining a common timeline and boundary-velocity history. Integration continues through unloading after tool loading, dragging, or free collision. Particle displacement and plastic strain are interpolated to surface vertices. Initial–final vertex correspondences support rigid-motion removal. A weighted Kabsch solution (Kabsch, 1976) with Tukey biweight iterations focuses rigid fitting on regions with small geometric changes.

A.3 Interface Evolution and Material Protection

Signed distance is negative inside the solid, and removal follows $\partial\phi/\partial t = v_n|\nabla\phi|$ with $v_n > 0$. Godunov upwinding discretizes interface motion, Sussman reinitialization maintains distance-function properties, and Russo–Smereka subcell correction preserves the zero level set (Sussman et al., 1994; Russo & Smereka, 2000). Marching cubes reconstructs the surface (Lorensen & Cline, 1987).

The narrow band covers the expected retreat or addition over the entire event. Rate fields with rapid normal-direction variation, such as contact pressure, are projected onto the interface before propagation. Sufficient grid cells resolve defect depth and local wall thickness. Conforming red–green refinement with shared edge midpoints maintains triangle connectivity.

Before removal, inward conical rays estimate wall thickness and construct a protected material core, which is united with the resulting solid after the event. Conventional removal retains the larger of 30% of the maximum wall thickness and three grid-cell widths; separate configurations govern penetrating defects. Retreat is also measured from the original surface to the output solid to cover locations where material has been removed.

A.4 Electrochemical Transport and Event Coupling

Electrochemical transport follows far-field supply and wall-absorption conditions (Newman & Thomas-Alyea, 2004; Jones, 1996). Robin wall reactions are represented as near-wall volumetric sinks in $\nabla^2 c - \sigma c = 0$, and accessibility A is defined by the normalized wall field. The Damköhler number controls relative reaction and transport strength, while the charge-coupling length controls electrochemical interaction range.

In the local microcell limit, rates increase with accessibility, preferentially retreating exposed edges. In the global macrocell limit, open surfaces supply cathodic current and shielded regions become dominant anodes. The two limiting rates are $v \propto A$ and $v \propto (A_c - A)^+$, normalized to preserve the corresponding total removal. Transport is updated after interface evolution, feeding increased pitting shielding back into subsequent rates.

Stress corrosion reads accumulated plastic strain from preceding events and enhances local reactions by $w \leftarrow w(1 + \gamma\alpha/\alpha_{\max})$ (Gutman, 1994). Deformation, transport, and corrosion share damaged geometry within a sample; plastic zones guide subsequent erosion.

A.5 Shape Constraints for Material Addition

Oxide scale follows $x^2 = kt$, with outward displacement $(\rho_{\mathrm{PB}} - 1)x$, where ρ_{PB} is the volume ratio of oxide to consumed metal. Spatter droplets are sampled according to visibility and distance from the weld. A droplet of diameter d solidifies into a spherical cap of radius R and contact angle θ, conserving volume:

$$\frac{\pi d^3}{6} = \frac{\pi R^3}{3}(1 - \cos\theta)^2(2 + \cos\theta). \tag{3}$$

Contact angle varies with cooling during flight. Foreign deposits are unions of offset spheres flattened along surface normals. Excess adhesive uses constant-mean-curvature surfaces for local wetting and morphological closing for capillary-accessible locations. The capillary length $l_c = \sqrt{\gamma_s/(\rho g)}$ controls the relative importance of gravity and surface tension, where γ_s is surface tension.

A.6 Annotations and Severity

The primary geometric annotation is signed distance to the normal surface, normalized by the object's longest side and converted to its absolute value. Thresholding produces a binary anomaly mask. Accumulated plastic strain is retained as an additional physical channel. Multi-event samples record mechanism identities and overlap regions.

Severity control jointly constrains maximum geometric deviation and anomalous fraction. Corrosion uses anomalous triangle-surface area; other mechanisms use vertex fractions. The controller adjusts action strength using measured deviation and lateral extent using anomalous area. Mesh validity and target severity jointly determine sample acceptance.

A.7 Numerical and Mechanism Validation

Table 7 summarizes tests of interface accuracy, capillary morphology, and event coupling. The stress-corrosion depth ratio compares corrosion depth in the highest and lowest plastic-strain deciles. These tests support numerical accuracy and mechanism consistency under the corresponding settings. Coverage of real defect distributions and detection transfer require evaluation with measured data.

Table 7: Tests of numerical accuracy and physical mechanisms.

Test	Reference or measurement	Result
Pitting interface	Threefold-finer reference	Maximum error 0.9%
Capillary corner	Analytical 90° corner	Error 0–0.02 cells
Stress-corrosion chain	High-/low-strain depth ratio	Median 6.48
Stress-corrosion chain	Samples with strain coupling	91%

B SCANNER PARAMETERS AND VALIDATION

B.1 INSTRUMENT AND PATHS

The scanner explicitly positions a projector and cameras, checking camera visibility after raster rays hit the surface. Single-camera visibility mode accepts returns visible to at least one camera; stereo mode requires visibility to both. Working distance, focal length, baseline, and disparity precision determine point spacing and depth uncertainty:

$$s = zp/f, \qquad \sigma_z = z^2\sigma_d/(b_{\text{eff}}f). \tag{4}$$

Here z is working distance, p pixel stride, f focal length, b_{eff} effective baseline, and σ_d disparity noise.

Scan stations are selected according to residual measurement demand in unsaturated surface cells; travel-distance penalties connect nearby stations. Each cell accepts at most n_{sat} measurements, and fused output resolution matches cell size. Repeated returns use inverse-variance weighting, with correspondences established by ray geometry. Easy and medium additionally use local planar projection over 16 neighbors to reduce surface thickness.

Table 8: Scanner parameters.

Level	Stations	Visible cameras	Triangulation angle	n_{sat}	Disparity noise
easy	48	At least one	25°	2	0.010
medium	28	At least one	25°	2	0.020
hard	14	Both	15°	1	0.080

All levels share normalized working distance 2.4, field of view 28°, a 300×300 raster, and fusion voxel size 3.99×10^{-3}. The pipeline includes per-frame registration drift, mixed pixels, radiometric dropout, and outliers; hard additionally includes multipath ghosts. Multipath uses statistical perturbations, reconstruction uses local planes, and scan sampling uses a 2D raster.

B.2 OCCLUSION AND SATURATION VALIDATION

For a slit of width w and depth d, the limiting depth-to-width ratios under single-camera and stereo visibility are $2z_0/(b-w)$ and $z_0/(b-w)$, respectively. Analytical predictions of 10.67 and 5.33 agree with measured cutoff intervals $(10, 11]$ and $(5, 6]$. A step of height h and thickness t creates a shadow band of width $(b/2 + t/2)h/(z_0 - h)$. Its measured stereo width is 0.0441, compared with the analytical 0.0421, differing by one raster step.

Table 9: Point counts under repeated scanning.

Scan passes	Direct concatenation	Fused output
1	$1.00\times$	$1.00\times$
4	$3.99\times$	$2.11\times$
16	$16.00\times$	$3.32\times$
32	$32.00\times$	$3.78\times$

The four fused counts in Table 9 form the curve in Figure 4(b). Its surface cells and denoising point clouds illustrate acquisition and fusion principles.

B.3 CALIBRATION AND LABEL TRANSFER

Calibration uses scale-independent point spacing, density uniformity, and local thickness ratio. Point spacing is normalized by the bounding-box diagonal; density uniformity is the $p95/p50$ ratio of nearest-neighbor distances; local thickness is the neighborhood normal-direction standard deviation relative to point spacing. Table 2 uses one calibration configuration. Easy and medium thickness ratios fall within measurements of released real clouds, while hard provides an out-of-range stress test. Statistical proximity supports similarity only in the corresponding observation-quality dimensions.

Table 10 compares local denoising under medium conditions. Removing local planar reconstruction increases the thickness ratio from 0.057 to 0.112, while the ratio of mean surface distances for anomalous versus normal points changes from 1.14 to 1.12. Reconstruction thus mainly reduces measurement scatter in this configuration, with similar regional distance ratios. This diagnostic supports local denoising on the measured clouds; Table 2 separately reports calibration statistics for released real and synthetic clouds.

Table 10: Regional bias diagnostic for local denoising.

Configuration	Thickness ratio	Anomalous/normal surface-distance ratio
medium	0.057	1.14
Without registration jitter	0.044	0.92
Without planar reconstruction	0.112	1.12

Ray hits record triangles and barycentric coordinates. Geometric deviations and defect labels are interpolated from vertices to scan points. Valid surface returns, random outliers, multipath ghosts, and mixed pixels receive distinct labels; the latter three are excluded from P-AUROC.

C Evaluation Settings and Scoring Scope

C.1 Method Configurations

Table 4 uses point-cloud adaptations. PTP estimates normals from observed points, estimates registration, and scores point-to-plane residuals. BTF-FPFH uses dense 33-dimensional FPFH representations. PatchCore-FPFH combines FPFH with an approximate greedy coreset and normalizes using the observed scale of the first normal reference.

PASDF uses a narrow-band signed-distance representation of normal clouds. ISMP uses PointMAE and FPFH features, with SIE fusion outside this configuration. Reg3D, introduced with Real3D-AD (Liu et al., 2023), uses coordinate and PointMAE banks; it is distinct from Reg2Inv in the related work. RIF uses PCM and a CTF pretrained on ModelNet40, normalizes through its intrinsic coordinate construction, and excludes S3DA. M3DM-PointMAE uses its point-cloud branch. Each row represents these adaptations; multimodal branches, pose handling, and pretraining sources are part of the method configuration.

Simple3D uses the released multi-scale neighborhood descriptors (MSND) and local feature spatial aggregation (LFSA), with 8,192 sampled points per cloud, a 5% coreset, and Euclidean nearest-neighbor scoring, only at $K = 1$. It reuses shared geometric registration estimated from observed points; normals are likewise estimated from observations. Its registration configuration is recorded separately from those of the first eight rows, so comparisons include joint effects of features and registration. True poses, CAD normals, and anomaly labels remain on the evaluation side for every method.

C.2 P-AUROC Aggregation

Each level contains 3,213 anomalous scans and 400 normal test scans. P-AUROC requires both labels among sampled valid surface points; it remains undefined for fully occluded defects or single-class samples. The main table averages within geometry–mechanism cells, then equally over defined geometry cells within each mechanism, then equally over defined mechanisms; Table 11 lists effective counts. Inference failures and single-class scans remain missing, and aggregation uses only defined entries, without zero imputation or filtering by registration quality. Effective sample sets may therefore differ between rows, so denominators must accompany interpretation. Multiple scanning levels and reference configurations of one object share its geometry; comparisons are interpreted under these paired conditions.

For Simple3D at $K = 1$, 10,839 test scans yield 10,754 valid predictions and 85 inference failures: 1 at easy, 32 at medium, and 52 at hard. P-AUROC remains undefined for successful predictions on entirely normal or single-class anomalous scans. The three levels contain 3,168, 3,132, and 3,110 scorable scans, respectively.

Table 11: Scorable samples and inference failures for P-AUROC.

Method	K	easy valid	medium valid	hard valid	Failures
PTP (geometric baseline)	1	3,150	3,140	3,099	38
PTP (geometric baseline)	4	3,150	3,140	3,099	38
PTP (geometric baseline)	16	3,150	3,140	3,099	38
BTF-FPFH	1	3,150	3,140	3,099	38
BTF-FPFH	4	3,150	3,140	3,099	38
BTF-FPFH	16	3,150	3,140	3,099	38
PatchCore-FPFH	1	3,150	3,140	3,099	38
PatchCore-FPFH	4	3,150	3,140	3,099	38
PatchCore-FPFH	16	3,150	3,140	3,099	38
PASDF	1	3,150	3,140	3,099	38
PASDF	4	3,150	3,140	3,099	38
PASDF	16	3,150	3,140	3,099	38
ISMP	1	3,150	3,140	3,099	38
ISMP	4	3,150	3,140	3,099	38
ISMP	16	3,150	3,140	3,099	38
Reg3D	1	3,150	3,106	3,099	76
Reg3D	4	3,150	3,106	3,099	76
Reg3D	16	3,150	3,106	3,099	76
RIF	1	3,169	3,161	3,122	38
RIF	4	3,169	3,161	3,122	38
RIF	16	3,169	3,161	3,122	38
M3DM-PointMAE	1	3,169	3,127	3,122	76
M3DM-PointMAE	4	3,169	3,127	3,122	76
M3DM-PointMAE	16	3,169	3,127	3,122	76
Simple3D	1	3,168	3,132	3,110	85

Each level contains 3,213 anomalous and 400 normal test scans. The final column sums inference failures across levels. Scorable scans contain both normal and anomalous labels among non-artifact points. P-AUROC is undefined when the defect is unobserved or unsampled.

C.3 Simulated Scanning of Native ShapeNet Defects

This study uses four Anomaly-ShapeNet categories: cabinet0, chair0, screen0, and vase6. Native defects retain the original point clouds and official annotations. Their meshes are scanned with the medium configuration to obtain the second observation. One vase sample has no corresponding mesh and is omitted from both observations, leaving 79 native defects: 20 cabinets, 20 chairs, 18 screens, and 21 vases. Normal test clouds are also processed with both observation settings.

Each category uses four original normal training templates. Original point clouds form the reference bank for the original observation; simulated scans of the same templates form the bank for the scanned observation. Thus the comparison covers acquisition of both normal references and test objects. Nine reference-bank methods use $K = 4$, while PTP and the fixed R3D-AD reconstruction configuration use the first reference, $K = 1$. Each method retains its own registration or coordinate-normalization procedure across the two conditions; the methods need not use the same procedure.

Native labels are transferred from the original labelled points to scan hits by nearest-neighbor correspondence. The 99th-percentile correspondence distance must be within 3% of the point-cloud bounding-box diagonal. All 79 native defects pass this check, with no additional exclusions. The medium scanner models the same acquisition effects as in the main experiment. Artificial returns remain in model inputs and are excluded from surface-localization scoring using their point-type labels.

PTP, BTF-FPFH, PASDF, ISMP, Reg3D, PatchCore-FPFH, and Reg2Inv receive 4,096 points; the other methods receive 8,192. Internal sampling and pretrained parameters follow the existing point-cloud implementations. Simple3D uses the released MSND/LFSA implementation. Reg2Inv uses released ShapeNet pretraining (Yu et al., 2025). R3D-AD uses the released autoencoder architecture, trained locally for 40,000 updates on one normal Real3D-AD airplane scan; its parameters remain fixed for transfer (Zhou et al., 2024). These model configurations are retained across both observations.

Each method evaluates 278 test clouds: 79 native-defect clouds and 60 normal clouds under each of the two observations. All eleven methods return predictions for all 278 clouds, and P-AUROC is defined for all 79 defects in both observations. Table 6 averages defect-cloud P-AUROC within each category and then equally across categories. The change is computed for each paired defect before applying the same averaging, using unrounded scores.

The normal reference files are `category_template0` through `category_template3` for each category. Scan seeds for normal references and test clouds are fixed by sample identifiers; each native defect uses the same defect mesh in its two observations.

Pose handling. PTP, Simple3D, PASDF, ISMP, Reg3D, M3DM-PointMAE, and BTF-FPFH use FPFH–RANSAC and principal-axis initializations, followed by robust point-to-plane ICP. The voxel width is $v = 0.025$ times the observed bounding-box diagonal; normal and FPFH radii are $2v$ and $5v$. RANSAC uses at most 100,000 iterations, confidence 0.999, and distance $1.8v$. The best four initializations under a 90% trimmed nearest-point distance are refined at $3v$, $1.5v$, and $0.8v$, with at most 60 iterations per scale; an update is retained when that distance does not increase.

PatchCore-FPFH scales the first reference diagonal to 100, uses voxel width 0.5, normal/feature radii $1/2.5$, and FPFH–RANSAC with distance 0.75, 100,000 iterations, and confidence 0.999. R3D-AD uses the first reference radius as the common scale, voxel width 0.05, normal/feature radii $0.1/0.25$, RANSAC distance 0.1, and 50 point-to-plane ICP iterations at distance 0.03. Its RANSAC uses 100,000 iterations and confidence 0.999. Reg2Inv uses its released learned correspondences and 50,000 three-point RANSAC iterations. RIF independently canonicalizes each point cloud using its intrinsic coordinate representation. All pose estimates use observed geometry, without true poses or anomaly labels.

D Capability Comparison and Morphology Examples

In Table 1, material coupling requires material-dependent geometry (appearance alone gives partial support); global deformation is object-scale change; topology concerns connectivity, e.g., cracks or holes; scanning requires point-cloud acquisition (image visibility alone gives partial support); numerical solving computes responses through discretized physical equations; and event chaining transfers states between processes.

MPAS uses 1D/2D/3D support primitives for point, line, surface, and volumetric defects; planar cracks can change point-cloud connectivity, hence support topological change. MPAS and PA3AD bending can produce object-scale shape changes, so global deformation is supported regardless of whether physics is solved. PA3AD deletes hole-center points and separates points across cracks, so topological change is also supported. Its analytical deformation functions do not solve discretized physical equations, so numerical solving is unsupported. ScratchSim uses material parameters for appearance and camera visibility for observation, giving partial support for material coupling and scanning simulation. Our implicit surfaces for addition and removal support connectivity changes, with the generated range governed by wall-thickness protection and event configurations.

Figures 3 and 5 show the same mechanisms on two industrial geometries. Blue and orange indicate negative and positive geometric deviations; gray indicates deviations below 5% of the panel peak. Each panel is normalized by its own peak, precluding cross-panel intensity comparisons. Structural differences influence action locations and defect extents.

E Reproducibility

Generation seeds, sample indices, event indices, and attempt numbers jointly determine sample randomness. Accepted parameters, generation configurations, code versions, and geometry digests are stored with each sample. MPM results are checkpointed so repeated evaluations use consistent damaged surfaces. Scanning levels share physical surfaces, and reference sizes use nested prefixes of normal instances. Raw point-level predictions and sample correspondences support metric recomputation.

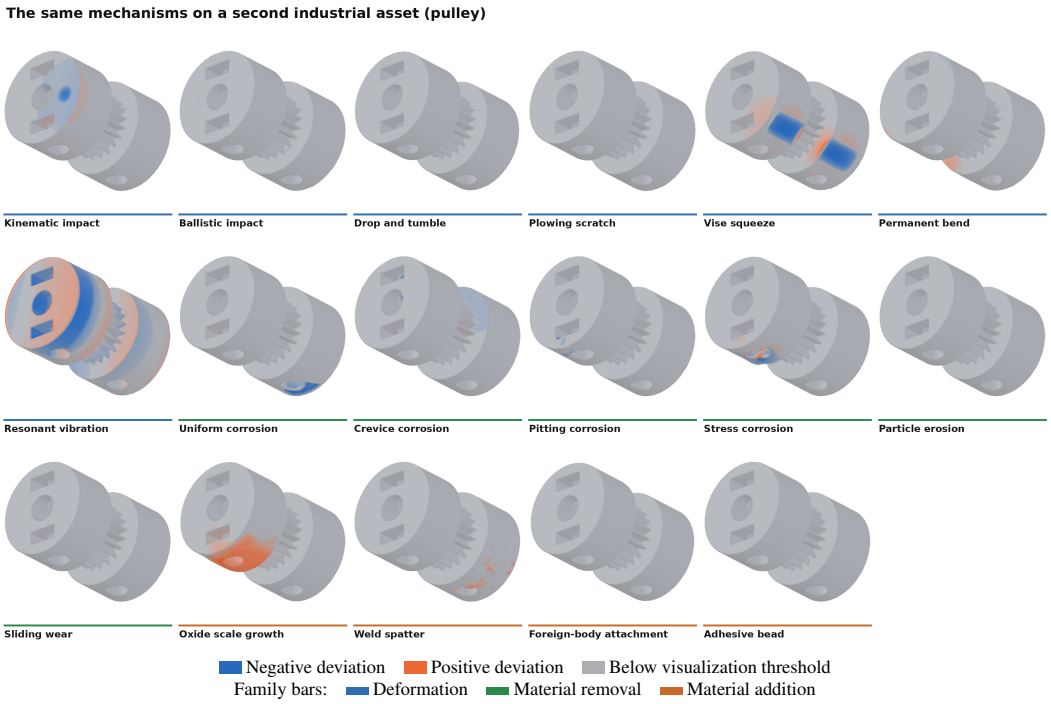


Figure 5: Seventeen defect mechanisms on a pulley.

F AI USE STATEMENT

Generative AI assisted with drafting, editing, and translating the manuscript between Chinese and English, as well as implementing, revising, and debugging method and experiment code. It also reconstructed and organized some mathematical derivations from the original papers cited here; those derivations are not novel results generated by AI. The core research ideas, experimental program, and experimental design were developed by the human authors. The authors take responsibility for AI-assisted text, code, derivations, experimental results, and the final manuscript.