Anomaly-to-Normal Translation: A Training Paradigm for Feature-Based 3D Point Cloud Anomaly Detection Supplemental Material

Lars Fichtel[1,2] Dominik Seuß[1,3] Radu Timofte[2]
[1]Center for Artificial Intelligence, Technical University of Applied Sciences Würzburg-Schweinfurt
[2]Computer Vision Lab, University of Würzburg
[3]Fraunhofer Institute for Integrated Circuits IIS, Fraunhofer-Gesellschaft
{lars.fichtel, dominik.seuss}@thws.de radu.timofte@uni-wuerzburg.de
https://github.com/LarsFichtel/3D-AD_LCM

Abstract

Reconstruction-based autoencoders for 3D point cloud anomaly detection score anomalies via Chamfer distance, but two failures limit their effectiveness: the identity mapping trap*, where high-capacity decoders reconstruct anomalous inputs faithfully and erase the detection signal, and* Chamfer blindness*, where surface contamination leaves XYZ coordinates unchanged and yields zero reconstruction error. We instead score anomalies via the* feature cycle error $\|E(x) - E(D(E(x)))\|$*, the self-inconsistency of the encoder's representation after a round-trip through the decoder. This signal is powerful but* cannot emerge from normal-only training*: on a standard autoencoder, feature cycle AUROC on the rope category of MVTec 3D-AD collapses.We address this with* anomaly-to-normal translation training*: the decoder is additionally trained to reconstruct clean canonical shapes from synthetically corrupted inputs. A study across four encoders and two decoders (MLP, PoinTr) reveals the benefit is* defect-type dependent*: essential for structural deformations, neutral or harmful for surface anomalies. For hierarchical encoders, we further introduce* spatial multi-scale feature error*, per-point feature matching at intermediate encoder stages.*

Acknowledgments. The work was partially supported by the Alexander von Humboldt Foundation and the Center for Artificial Intelligence (CAIRO) at the Technical University of Applied Sciences Würzburg-Schweinfurt (THWS), funded by the Bavarian Hightech Agenda.

A. Full Per-Category Results: All Encoder × Decoder Combinations

Table 1 reports I-AUROC for all 8 trained encoder×decoder configurations across all 10 MVTec 3D-AD categories. All use chamfer scoring ($w_{\text{rec}} = 1.0$, $w_{\text{lat}} = 0.5$, $k = 1$), no alignment loss, 400 epochs, seed 42 (PN+MLP also reports mean ± std over 3 seeds).

Table 1. I-AUROC per category for all encoder × decoder combinations (chamfer scoring, 400 ep, $\lambda_{\text{anom}} = 1.0$). **Bold** = best per category. PN+MLP reported as mean ± std over 3 seeds; all others single seed (42). PTv2+MLP† uses LR=3×10^{-4}; PTv2+MLP uses LR=10^{-3}.

Category	PN+MLP	DGCNN+MLP	PTv2+MLP	PTv2+MLP†	PNxt+MLP	PN+PoinTr	DGCNN+PoinTr	PTv2+PoinTr
bagel	$0.600 \pm .014$	**0.653**	0.488	0.440	0.471	0.612	0.560	0.572
cable_gland	$0.583 \pm .016$	**0.775**	0.490	0.552	0.515	0.594	0.504	0.511
carrot	$0.474 \pm .006$	0.524	**0.591**	0.520	0.434	0.460	0.544	0.515
cookie	$0.822 \pm .047$	0.619	0.544	0.630	0.587	0.750	0.427	0.624
dowel	$0.630 \pm .024$	0.521	0.556	0.605	0.485	0.582	0.595	0.519
foam	$0.431 \pm .072$	0.507	0.448	0.488	0.513	0.430	0.440	0.505
peach	$0.407 \pm .034$	0.491	**0.513**	0.451	0.429	0.472	0.507	0.536
potato	$0.394 \pm .039$	0.423	0.444	0.520	0.466	**0.539**	0.404	0.571
rope	$0.754 \pm .017$	0.499	0.430	0.645	0.575	**0.827**	0.635	0.470
tire	$0.633 \pm .017$	0.560	**0.600**	0.571	0.586	0.671	0.463	0.531
Mean	$0.573 \pm .011$	0.557	0.510	<u>0.542</u>	0.506	**0.594**	0.508	0.535

B. Scoring Mode Ablation: Full Per-Category Results

Tables 2–3 report the complete per-category breakdown of the λ_{anom} ablation across chamfer and feature scoring modes (PN+MLP, seed 42).

Table 2. Chamfer scoring I-AUROC by λ_{anom} (PN+MLP, seed 42). The standard AE ($\lambda = 0$) is competitive with the anomaly-conditioned model ($\lambda = 1$); differences are within seed std ±0.011.

Category	$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$
bagel	0.586	0.518	0.607
cable_gland	0.569	0.539	0.602
carrot	0.472	0.467	0.472
cookie	**0.876**	0.723	0.768
dowel	0.584	0.614	**0.639**
foam	0.518	0.484	0.514
peach	0.394	0.384	0.443
potato	0.431	0.406	0.434
rope	**0.811**	0.787	0.734
tire	0.636	0.638	0.637
Mean	**0.588**	0.556	0.585

Table 3. Feature-cycle scoring I-AUROC by λ_{anom} (PN+MLP, seed 42, $w_{\text{feat}} = 1.0$, $w_{\text{lat}} = 0.5$). Rope is the critical example of the paradigm's necessity: feature AUROC collapses to 0.272 without anomaly conditioning.

Category	$\lambda = 0$	$\lambda = 0.5$	$\lambda = 1$
bagel	**0.610**	0.443	0.544
cable_gland	**0.567**	0.318	0.547
carrot	**0.603**	0.494	0.559
cookie	**0.612**	0.422	0.310
dowel	0.568	0.453	0.485
foam	0.459	0.407	**0.470**
peach	**0.521**	0.386	0.351
potato	0.444	0.428	**0.517**
rope	0.272	0.424	**0.659**
tire	**0.584**	0.522	**0.584**
Mean	**0.524**	0.430	0.503

C. Spatial Multi-Scale Feature Error: Complete Tables

PTv2+MLP

Table 4 reports the complete spatial MS sweep for PTv2+MLP. Seven scoring configurations are evaluated per category: `feat_baseline` ($w_{\text{feat}} = 1, w_{\text{lat}} = 0.5$), `sp_only` ($w_{\text{sp}} = 1, w_{\text{feat}} = 0$), and combinations `feat_sp{0.5,1.0,2.0}`, `feat_sp{0.5,1.0}_nolat`. Oracle = best I-AUROC per category.

Table 4. Spatial MS sweep for PTv2+MLP (seed 42). **Bold** = oracle (best per category). Δ = oracle vs. feat_baseline.

Category	feat_baseline	sp_only	oracle	Δ
bagel	0.509	0.421	0.509	0.000
cable_gland	0.468	0.579	**0.598**	+0.130
carrot	0.559	0.414	0.559	0.000
cookie	0.601	0.573	0.601	0.000
dowel	0.508	0.500	**0.593**	+0.084
foam	0.522	0.567	**0.567**	+0.045
peach	0.569	0.483	0.569	0.000
potato	0.372	0.430	**0.472**	+0.101
rope	0.564	0.624	**0.624**	+0.060
tire	0.589	0.589	0.589	0.000
Mean	0.526	0.513	**0.568**	**+0.042**

PTv2+PoinTr

Table 5. Spatial MS sweep for PTv2+PoinTr (seed 42). Oracle is lower than PTv2+MLP (0.545 vs. 0.568), confirming that MLP's chamfer-loss training provides better per-point correspondence for spatial MS.

Category	feat_baseline	sp_only	oracle	Δ
bagel	0.396	0.396	0.396	0.000
cable_gland	0.400	0.466	**0.496**	+0.097
carrot	0.499	0.503	**0.592**	+0.093
cookie	0.471	0.601	**0.601**	+0.130
dowel	0.502	0.565	**0.565**	+0.063
foam	**0.629**	0.557	**0.629**	0.000
peach	0.501	0.532	**0.544**	+0.043
potato	**0.445**	0.430	**0.445**	0.000
rope	0.509	0.524	**0.580**	+0.072
tire	**0.601**	0.576	**0.601**	0.000
Mean	0.495	0.515	**0.545**	**+0.050**

PointNeXt+MLP

Table 6. Spatial MS sweep for PointNeXt+MLP (seed 42). Seven/ten categories show positive delta, but gains are smaller than PTv2 due to lower reconstruction fidelity.

Category	feat_baseline	sp_only	oracle	Δ
bagel	0.409	0.427	0.416	+0.007
cable_gland	0.440	0.486	**0.488**	+0.048
carrot	**0.565**	0.483	**0.565**	0.000
cookie	**0.564**	0.515	**0.564**	0.000
dowel	**0.510**	0.399	**0.510**	0.000
foam	0.494	0.514	**0.559**	+0.065
peach	0.466	0.541	**0.541**	+0.076
potato	0.562	0.533	0.566	+0.004
rope	0.431	0.476	**0.479**	+0.048
tire	0.294	0.402	**0.402**	+0.108
Mean	0.474	0.477	**0.509**	**+0.035**

D. Per-Category Oracle Breakdown

Table 7 shows the per-category oracle result: for each category, which encoder, decoder, and scoring mode achieves the highest I-AUROC across all evaluated configurations.

Table 7. Per-category oracle across all evaluated encoder × decoder × scoring combinations. The oracle mean (0.693) is computed over PN+MLP and DGCNN+MLP best-per-category (all scoring modes) plus PN+PoinTr chamfer.

Category	Oracle	Best Config	Note
bagel	0.653	DGCNN+MLP, chamfer	EdgeConv wins local structure
cable_gland	0.775	DGCNN+MLP, chamfer	Strong local geometry
carrot	0.591	PTv2+MLP, chamfer	Smooth organic shape
cookie	0.822	PN+MLP, chamfer	High variance; seed 42 wins
dowel	0.639	PN+MLP, chamfer	Stable category
foam	0.629	PTv2+PoinTr, feat	Only category where PoinTr feat wins
peach	0.536	PTv2+PoinTr, chamfer	Smooth surface
potato	0.571	PTv2+PoinTr, chamfer	PoinTr strong here
rope	**0.827**	PN+PoinTr, chamfer	PoinTr+PN strongest structural
tire	0.671	PN+PoinTr, chamfer	PoinTr gains
Mean	**0.671**		

E. Seed Variance Analysis

Table 8 reports PN+MLP I-AUROC across three random seeds (42, 1, 2) for all 10 categories. Mean std = 0.011 across categories.

Table 8. PN+MLP I-AUROC across 3 seeds (chamfer scoring). High-variance categories (foam, cookie, peach) should be treated with caution in single-seed comparisons.

Category	Seed 42	Seed 1	Seed 2	Mean ± Std
bagel	0.607	0.584	0.608	0.600 ± 0.014
cable_gland	0.602	0.573	0.575	0.583 ± 0.016
carrot	0.472	0.469	0.480	0.474 ± 0.006
cookie	0.768	0.846	0.853	0.822 ± 0.047
dowel	0.639	0.648	0.603	0.630 ± 0.024
foam	0.514	0.389	0.389	0.431 ± 0.072
peach	0.443	0.375	0.404	0.407 ± 0.034
potato	0.434	0.391	0.356	0.394 ± 0.039
rope	0.734	0.765	0.764	0.754 ± 0.017
tire	0.637	0.648	0.615	0.633 ± 0.017
Mean	0.585	0.569	0.565	$\mathbf{0.573 \pm 0.011}$

F. Spatial MS Anomaly Heatmaps: Full Category Panels

Figures 1–5 show spatial multi-scale feature error heatmaps (PTv2+MLP) for all five evaluated categories. Each panel shows four anomalous samples followed by one normal sample (top view, top-to-bottom: anomalous, normal). Colors are normalized per category to a shared scale (blue = low error, red = high error); the 2nd–98th percentile defines the range. All heatmaps use 512 points (subsampled from 2048 for typesetting).

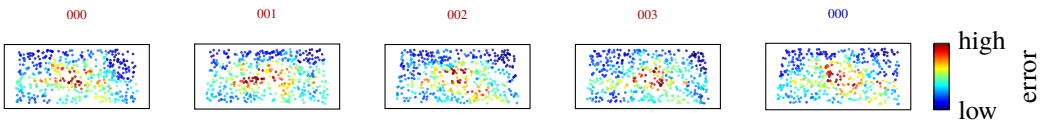

Figure 1. Spatial MS heatmaps for **rope** (PTv2+MLP). Four anomalous samples + one normal (rightmost). Rope anomaly: cuts and deformations produce concentrated red regions.

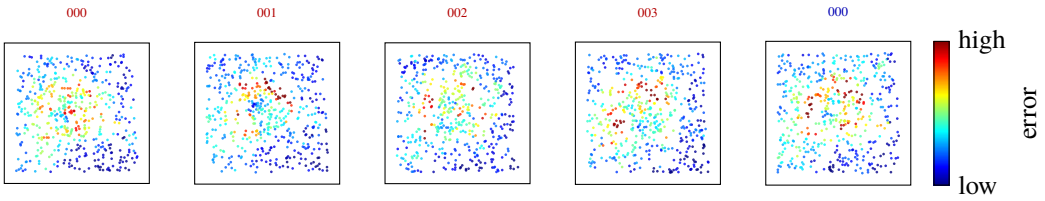

Figure 2. Spatial MS heatmaps for **cable_gland** (PTv2+MLP). Thread and surface damage localizes as elevated error on the cable body.

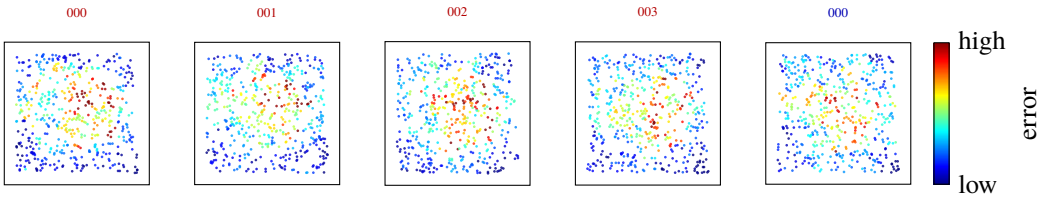

Figure 3. Spatial MS heatmaps for **cookie** (PTv2+MLP). Cookie anomalies (cracks, contamination) produce less spatially concentrated patterns than rope/cable_gland, consistent with the smaller spatial MS gain for this category ($\Delta = 0$).

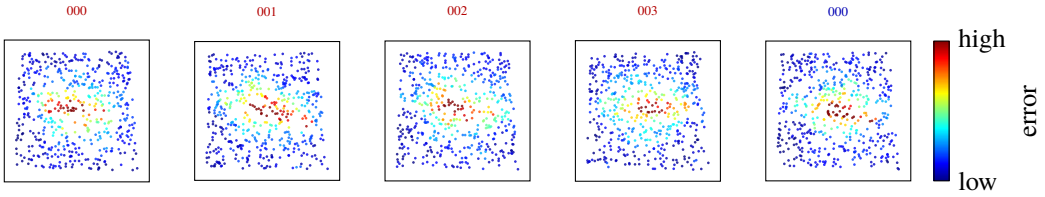

Figure 4. Spatial MS heatmaps for **carrot** (PTv2+MLP). Organic shape with diffuse defects; spatial MS provides no oracle gain ($\Delta = 0$), consistent with the low error contrast between anomalous and normal.

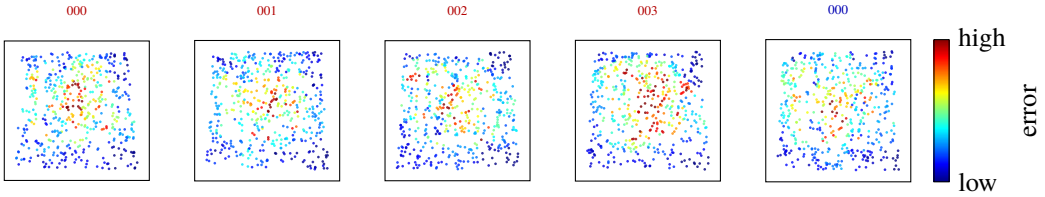

Figure 5. Spatial MS heatmaps for **foam** (PTv2+MLP). Foam anomalies include cuts and contamination; spatial MS yields $\Delta = +0.045$, with error slightly elevated in cut regions.

G. Anomaly Synthesis: Type Descriptions and Parameters

Table 9 describes the 8 synthetic anomaly types used in AnomalySynthesizer. All types resample the output back to $N = 2048$ after modification.

Table 9. Synthetic anomaly types used in LCN training. "Structural" types deform the point cloud geometry; "surface" types add or modify points without large-scale deformation.

Type	Category	Description and Parameters
Hole	Structural	Remove points in sphere $r \in [0.1, 0.3]$; resample uniformly
Dent	Structural	Inward Gaussian displacement; depth $\in [0.05, 0.15]$
Scratch	Structural	Linear groove; Gaussian falloff $\perp$ axis
Crack	Structural	V-profile; sine taper; edge roughness
Add cluster	Surface	$\sim$100 pts, sphere $r \in [0.02, 0.08]$
Noise	Surface	Global Gaussian jitter $\sigma = 0.01$
Contamination	Surface	Surface-elevated particles via PCA normals; 0.01–0.04 above surface
Warp	Surface	Sinusoidal; amp $\in [0.05, 0.15]$; freq $\in [1, 3]$

H. Point-Level Localization: P-AUROC

Table 10 reports per-point P-AUROC for spatial MS scoring (PTv2+MLP) on all 10 MVTec 3D-AD categories. Per-point GT labels are recovered by building a cKDTree over valid pixels of the raw 800×800 XYZ TIFF and assigning each of the 2048 sampled points the GT label of its nearest raw pixel. P-AUROC is computed over all points concatenated across the full test set.

Table 10. Point-level P-AUROC for spatial MS scoring (PTv2+MLP, no_align). GT labels recovered via NN-matching sampled points to raw XYZ TIFF + GT mask. Tire (highlighted) shows inverted scoring due to flat-geometry edge artefacts; all other categories achieve P-AUROC ≥ 0.47.

Category	P-AUROC
carrot	**0.971**
potato	0.927
dowel	0.926
cookie	0.859
foam	0.830
rope	0.752
cable_gland	0.739
bagel	0.706
peach	0.474
tire	0.110[†]
Mean (all 10)	0.729
Mean (excl. tire)	**0.798**

[†] Inverted: encoder assigns high spatial MS error to curved *edges* rather than flat-surface contamination. Root cause: no surface normal supervision. The scoring can be trivially inverted for tire at deployment (sign flip), yielding 0.890, but this would require a per-category prior.

Contribution framing. P-AUROC is a standard benchmark metric for anomaly localization in 3D inspection [1, 2]. Prior learning-based pure-3D methods typically produce only sample-level scores and do not report P-AUROC. Spatial MS is the first learned scoring signal in this family that produces per-point anomaly estimates *as a byproduct of the hierarchical feature matching computation*: no additional localization head, no separate segmentation training. The competitive P-AUROC on 9/10 categories (mean 0.798) shows that the intermediate PTv2 feature maps encode spatially accurate defect information and that the NN-matching propagation preserves it at 2048-point resolution.

I. Inference Throughput and Model Size

Table 11 reports inference latency and throughput for all encoder–decoder pairs measured at batch size 1 on an NVIDIA L40 (48 GB), the same GPU used for training. The dominant cost in PTv2 and PointNeXt is the encoder: their hierarchical attention/graph operations consume $\approx$149 ms regardless of decoder choice, making them $\sim$300$\times$ slower than PointNet+MLP. For PointNet-based pairs the decoder becomes the bottleneck: MLP adds only 0.5 ms total, while PoinTr's transformer adds $\approx$2.5 ms. The best-performing configuration overall (PN+PoinTr, 0.594 mean I-AUROC) runs at 328 samp/s, real-time capable for most inspection pipelines.

Table 11. Inference throughput and model size for all encoder–decoder pairs (batch size 1, NVIDIA L40). PN = PointNet, PNxt = PointNeXt. Latency reported as mean over 200 forward passes after 20 warmup iterations.

Encoder	Decoder	Enc (M)	Dec (M)	Total (M)	Latency (ms)
PN	MLP	0.08	7.09	7.16	**0.5**
	Folding	0.08	1.06	1.14	0.9
	Snowflake	0.08	3.14	3.21	2.6
	PoinTr	0.08	12.76	12.83	3.0
DGCNN	MLP	0.35	7.09	7.44	2.0
	Folding	0.35	1.06	1.42	2.4
	Snowflake	0.35	3.14	3.49	3.9
	PoinTr	0.35	12.76	13.11	4.5
PTv2	MLP	2.54	7.09	9.63	148.6
	Folding	2.54	1.06	3.60	149.4
	Snowflake	2.54	3.14	5.68	152.4
	PoinTr	2.54	12.76	15.30	152.7
PNxt	MLP	12.39	7.09	19.48	149.9
	Folding	12.39	1.06	13.45	150.2
	Snowflake	12.39	3.14	15.53	152.5
	PoinTr	12.39	12.76	25.15	**153.2**

J. Cross-Category Leave-One-Out Results

Table 12 summarizes the leave-one-out experiment at the mean level. The per-category breakdown shows that degradation under LOO correlates with structural distinctiveness: cable_gland and carrot, which have the most category-specific geometry, lose the most (-0.118 and -0.095 respectively). Rope and dowel, with more globally symmetric shapes, are largely unaffected. Cookie and tire improve under LOO (+0.052, +0.060), indicating that training on diverse shape categories provides regularization that benefits these classes.

Table 12. Cross-category leave-one-out I-AUROC vs. per-category training (PN+MLP, chamfer scoring, seed 42). A single model trained on the nine remaining categories is evaluated on the held-out one without any retraining.

Category	Single-cat	LOO	Δ
bagel	0.607	0.525	-0.082
cable_gland	0.602	0.484	-0.118
carrot	0.583	0.488	-0.095
cookie	0.611	0.663	$+0.052$
dowel	0.524	0.479	-0.045
foam	0.500	0.500	± 0.000
peach	0.540	0.453	-0.088
potato	0.530	0.464	-0.066
rope	0.595	0.578	-0.017
tire	0.607	0.667	$+0.060$
Mean	0.570	0.530	-0.040

References

[1] Eliahu Horwitz and Yedid Hoshen. Back to the feature: Classical 3d features are (almost) all you need for 3d anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023. 7

[2] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023. 7