Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.36628

A Comparative Study of Local-Global Anomaly-Map Fusion Strategies in EfficientAD

Qinyu Xiao

School of Computer Science and Engineering, Beihang University, Beijing, China
23373513@buaa.edu.cn

Abstract. Industrial anomaly detection must identify both local structural defects and global logical violations while preserving localization quality and practical inference cost. EfficientAD provides complementary local student-teacher and global autoencoder-student anomaly maps, yet its fixed equal average assumes that both maps are equally reliable at every location. This paper evaluates whether the joint-objective score-preserving spatial gate (A3) improves the fixed equal-averaging baseline (B2) under a Test-isolated protocol containing 14 frozen fusion configurations, three detector seeds, and three fusion seeds where applicable. On the frozen 3,651-image MVTec Logical Constraints Anomaly Detection (MVTec LOCO AD) archive, A3 achieved 83.35% macro image-level area under the receiver operating characteristic curve (AUROC) compared with 83.67% for B2, yielding −0.32 percentage points with a paired 95% confidence interval of [−2.56, +1.74]. Relative to B2, A3 changed structural AUROC by +0.95 points, logical AUROC by −1.58 points, and the area under the saturated per-region overlap curve up to a 0.05 false-positive rate (AU-sPRO@0.05) by −0.88 points. Although A3 added only 769 parameters, its fusion-only median latency was 1.463 ms versus 0.757 ms for B2 at batch size 1 on an RTX 4070 SUPER, a 1.93-fold ratio. A3 therefore did not demonstrate a reliable advantage or meet the pre-specified success criteria. This negative conclusion is limited to EfficientAD, MVTec LOCO AD, and the evaluated structural-proxy supervision.

Keywords: industrial anomaly detection, EfficientAD, anomaly-map fusion, spatial gating, MVTec LOCO AD

1. Introduction

Industrial visual inspection has to detect two qualitatively different failure modes. Structural anomalies include scratches, cracks, deformations, and misplaced local parts, whereas logical anomalies violate object count, composition, or arrangement without necessarily containing an unusual texture. MVTec LOCO AD formalizes this distinction and evaluates image-level detection together with spatial localization [1]. The distinction matters in production because a detector that is sensitive to local texture damage can still miss a missing component, while a context-aware detector can overlook a small surface defect. A useful system must therefore combine accurate image-level ranking, interpretable localization, and sufficiently low inference cost. EfficientAD is an attractive foundation because it was designed for millisecond-level anomaly detection and produces two

DOI: 10.54254/2755-2721/2026.36628

internal evidence maps with different receptive and training characteristics [2]. Its local student-teacher branch emphasizes fine discrepancies, while its global autoencoder-student branch responds to broader reconstruction inconsistencies. How these two maps are combined is consequently not a cosmetic post-processing choice: it determines which evidence survives into the deployed image score and anomaly localization.

Industrial anomaly detection research has improved representation quality, coverage, and logical reasoning through several complementary directions. PatchCore uses memory-bank retrieval to retain normal feature diversity [3]. CSAD decomposes objects into components for logical reasoning [4]. SALAD uses semantic awareness to identify invalid configurations [5]. LA-EAD strengthens logical-anomaly capability through targeted processing changes [6]. ObjectCore further studies object representations for efficient few-shot logical anomaly detection [7]. Together, multi-score integration [8], ranking-sensitive objectives [9], and position-aware, logic-synthesis, dual-channel, multimodal, calibrated, and unified formulations [10-15] motivate controlled fusion comparisons. EfficientAD instead uses a fixed local-global rule [2]. Fixed equal averaging is attractive because it is deterministic, parameter free, and difficult to overfit. Its limitation is equally clear: it assumes equal branch reliability across categories, anomaly types, images, and spatial locations. Statistical weights, learned global scalars, and pixel-wise gates relax progressively more of that assumption, but added flexibility can also learn unstable proxy correlations. The present study does not claim that synthetic supervision, gating, or score fusion is new. Instead, it asks a narrower empirical question under a frozen EfficientAD detector: when the branch maps, data roles, selection budget, image-score readout, and evaluation protocol are held constant, does spatial freedom add dependable value beyond simpler score-preserving fusion? This framing treats the complete comparison, including unsuccessful configurations, as evidence rather than selecting only methods with favorable point estimates.

Three factors motivate the experiment. First, bounding or clipping anomaly maps before fusion may compress extreme evidence and alter image rankings. Second, a pixel-wise gate can appear expressive while behaving almost like a global scalar if the two branches provide little identifiable complementarity. Third, pixel-level supervision from synthetic proxy masks may not align with the max-based image score used for AUROC. The experiment therefore separates score representation, spatial adaptivity, and objective alignment instead of attributing every change to the gate. Section 2 describes the MVTec LOCO AD data roles, the frozen EfficientAD foundation, four fusion-strategy families containing 14 frozen configurations, the training objectives, and the leakage-controlled evaluation protocol. Section 3 reports the complete configuration matrix, the pre-specified A3-B2 effect, structural and logical subgroup behavior, localization, ranking diagnostics, gate behavior, and efficiency. Section 4 synthesizes the controlled negative result, its practical implications, and the limits of inference from one dataset and one detector family.

2. Data and methods

2.1. Data

MVTec LOCO AD has five industrial categories and normal, structural-anomaly, and logical-anomaly test cases [1]. Against the official 3,644-image count, the frozen local archive contains 3,651 unique red-green-blue (RGB) inputs: 1,778 train, 305 validation, and 1,568 test. The seven extras are all in splicing connectors (six train and one validation). Test is unchanged. Formal runs bind the retained manifest and hashes.

DOI: 10.54254/2755-2721/2026.36628

Train-D and Fusion-B-train share the 1,778 normal training images. The latter adds one deterministic structural CutPaste proxy per source. The 305 validation images were split deterministically into normal-only Cal-A (153) for calibration and proxy Fusion-B-val (152) for checkpoint selection. The two training roles overlap and are not additive.

Test-LOCO stayed sealed until detector and fusion checkpoints. The 14-method registry, seeds, evaluator, readout, analysis, and success criteria were frozen. Only Test RGB was then opened for prediction. Anomaly types and masks were opened after prediction lock. No test labels, types, masks, or metrics informed calibration, tuning, early stopping, or recovery. Table 1 summarizes the fixed roles and leakage controls.

Table 1. Fixed data roles and leakage controls

Role	Count	Purpose	Test-feedback constraint
Train-D	1,778	EfficientAD detector fitting	Normal training images only; no Test feedback
Cal-A	153	Branch calibration and normal statistics	Normal-only; no checkpoint selection
Fusion-B-train	1,778*	Gate/scalar fitting with frozen proxies	Shares the normal train pool; no Test access
Fusion-B-val	152	Fusion checkpoint selection	Proxy criteria only; no Test labels
Test-LOCO	1,568	One-time formal prediction and evaluation	RGB after freeze; annotations after prediction lock

**Fusion-B-train reuses the 1,778 Train-D images and is not an additional sample set.*

2.2. Model and algorithm

The detector follows EfficientAD-S [2]. Pretrained teacher T and student S form the local discrepancy branch; autoencoder A and second student S' form the global reconstruction branch. After channel normalization, spatial alignment, and calibration, the frozen detector yields E_{L} and E_{G}, defined here as calibrated but unbounded maps; 'raw evidence' means no later clipping. For each category, seeds 401001-401003 were trained for 70,000 steps with batch size 1, learning rate 10^{-4}, weight decay 10^{-5}, a step-66,500 learning-rate drop, and float32 without automatic mixed precision, yielding 15 detector-category models.

Detector training preceded fusion and all detector parameters were then frozen. Normal Cal-A pixels and linear quantiles fitted the branch-specific, no-clamp transform in equation (1).

$$Q_{\mathrm{b}}(R_{\mathrm{b}}) = 0.1\frac{R_{\mathrm{b}} - q_{90,\mathrm{b}}}{q_{99.5,\mathrm{b}} - q_{90,\mathrm{b}} + \varepsilon} \tag{1}$$

With $\varepsilon = 10^{-6}$. Its outputs define E_{L} and E_{G}. Processing followed detector map, Cal-A transform, optional evidence or conditioning transform, convex fusion, then image readout. Each configuration therefore received identical calibrated inputs within a category and detector seed.

DOI: 10.54254/2755-2721/2026.36628

2.2.1. Fusion strategies

The 14 configurations are B0-B7, A1-A3, and D1-D3 (Table 2). Fixed methods have no fusion seed; trainable methods use 402001-402003. B2 averages calibrated raw evidence. B3 first applies $B(E) = \min(1, \max(0, E))$. B4 learns one category scalar shared by all pixels. For B5, an empirical cumulative distribution function (ECDF) tail statistic $t_{\mathrm{b}}(E_{\mathrm{b}}) = -\log(\max(P_{\mathrm{Cal-A,b}}[X \geq E_{\mathrm{b}}], 10^{-6}))$ is fitted from Cal-A and $\alpha_{\mathrm{tail}} = t_{\mathrm{L}}/(t_{\mathrm{L}} + t_{\mathrm{G}} + \varepsilon)$ it uses 0.5 when their sum is at most 2ε. Equations (2)-(4) define B2, B5, and B4, respectively.

$$F_{\mathrm{avg}} = \frac{E_{\mathrm{L}} + E_{\mathrm{G}}}{2} \tag{2}$$

$$F_{\mathrm{stat}} = \alpha_{\mathrm{tail}} E_{\mathrm{L}} + (1 - \alpha_{\mathrm{tail}}) E_{\mathrm{G}} \tag{3}$$

$$F_{\mathrm{global}} = \sigma(a) E_{\mathrm{L}} + [1 - \sigma(a)] E_{\mathrm{G}} \tag{4}$$

A3 computes $C_{\mathrm{ss}}(E) = 0.5[1 + (E/\tau_c)/(1 + E/\tau_c)]$, with $\tau_c = 0.1$, defines $C_{\mathrm{L}} = C_{\mathrm{ss}}(E_{\mathrm{L}})$ and $C_{\mathrm{G}} = C_{\mathrm{ss}}(E_{\mathrm{G}})$, then supplies $[C_{\mathrm{L}}, C_{\mathrm{G}}, |C_{\mathrm{L}} - C_{\mathrm{G}}|, C_{\mathrm{L}} C_{\mathrm{G}}]$ to a 769-parameter gate: 3×3 convolution (4 to 16 channels), Gaussian error linear unit (GELU), 16-group depthwise 3×3 convolution, GELU, 1×1 convolution, and sigmoid. Zero-initializing the output convolution gives $W = 0.5$ initially. B6, B7, A1-A3, and D3 share this 769-parameter topology. B6/B7 form the four-channel vector from clipped branch conditions and also fuse clipped evidence. D3 uses the same clipped vector but fuses raw evidence. D1 retains only the two softsign branch channels and changes the first convolution accordingly (481 parameters). D2 keeps four softsign channels, adaptive-average-pools the two-convolution features to 1×1, applies the final convolution/sigmoid, and broadcasts the scalar weight. It remains parameter-matched at 769. Only conditioning is compressed for A1-A3/D1/D2/D3. Deployed evidence uses the raw convex fusion in equation (5). The threshold-0.1, temperature-0.025 sigmoid link serves only L_{map}. Rank loss, validation pair accuracy, and inference all use the same raw-map maximum. Top-0.5% mean is sensitivity analysis only. Figure 1 summarizes the controlled branch-map and fusion workflow.

$$F_{\mathrm{gate}}(x, y) = W(x, y) E_{\mathrm{L}}(x, y) + [1 - W(x, y)] E_{\mathrm{G}}(x, y) \tag{5}$$

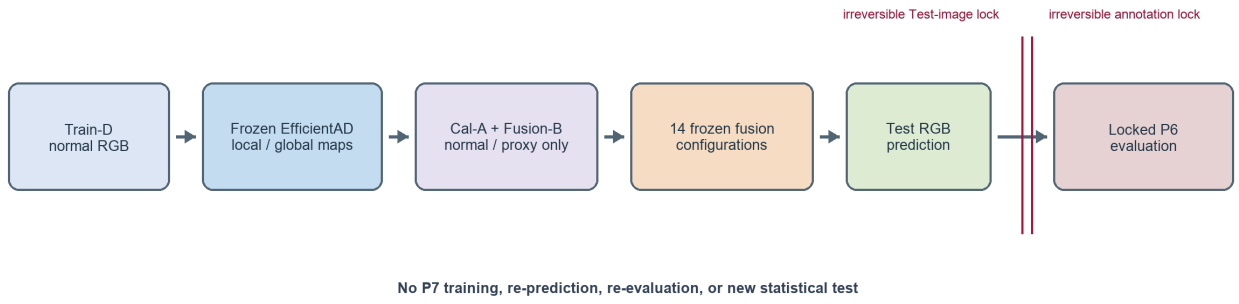

Figure 1. Controlled design. Frozen EfficientAD branch maps feed one of 14 fusion configurations; Test RGB prediction is completed and locked before annotation-bearing evaluation (picture credit: original)

DOI: 10.54254/2755-2721/2026.36628

2.2.2. Training objective and model selection

Trainable fusion models use normal images and one deterministic 256×256 structural CutPaste proxy per source; proxy area is 2-15% before mask-quality checks. The map objective is $L_{\text{map}} = 0.5(L_{\text{reg}} + L_{\text{grad}})$, combining per-sample L_2 map and Sobel-gradient errors normalized by spatial size. Paired softplus ranking encourages the raw maximum proxy score to exceed its matched normal score, scaled by 1.4826 times the frozen B2 normal-score median absolute deviation. A3 uses $L = L_{\text{map}} + 0.1L_{\text{rank}}$. A1 and A2 isolate the terms [9]. Table 2 assigns all controls. Structural proxies are training instruments, not evidence that logical relations were learned. The joint A3 objective is summarized in equation (6).

Fusion used seeds 402001-402003, batch size 8, Adam (learning rate 10^{-4}, zero weight decay), float32, a 100-epoch cap, patience 5, and earliest-minimum Fusion-B-val checkpoint selection. A no-Test pilot stopped after zero rank gradients in two strata. The subsequent engineering revision changed only the trainability diagnostic: for each unweighted objective and frozen training batch at initialization it computed $g_{\text{q,b}} = ||\nabla_{\varphi} L_{\text{q,b}}(\varphi_0)||_2$ and $G_{\text{q,RMS}} = [B^{-1}\Sigma_b {g_{\text{q,b}}}^2]^{1/2}$. both map and rank root-mean-square (RMS) values had to be finite and positive, without a post hoc magnitude cutoff. New design seeds verified this diagnostic before Test access. it did not modify the loss or optimizer and was not treated as detection performance or a paper contribution. Environment is Python 3.11.9, PyTorch 2.6.0+cu126, torchvision 0.21.0+cu126, CUDA 12.6, and RTX 4070 SUPER.

$$L = L_{\text{map}} + \lambda_{\text{rank}} L_{\text{rank}} \tag{6}$$

Table 2. Four fusion-strategy families and all 14 frozen configurations

ID	Fused evidence	Weight / conditioning	Objective / comparison role
B0	Raw local map	w=1	Fixed local-branch reference
B1	Raw global map	w=0	Fixed global-branch reference
B2	Raw local/global maps	w=0.5	Fixed primary baseline
B3	Clipped local/global maps	w=0.5	Fixed bounded-evidence control
B4	Raw local/global maps	Learned category scalar	L_{map}+0.1L_{rank}; non-spatial
B5	Raw local/global maps	Cal-A ECDF tail weight	Fixed; normal-only statistics
B6	Clipped local/global maps	4 clipped channels; spatial	L_{map}

DOI: 10.54254/2755-2721/2026.36628

Table 2. (continued)

B7	Clipped local/global maps	4 clipped channels; spatial	$L_{map}+0.1L_{rank}$
A1	Raw local/global maps	4 softsign channels; spatial	L_{map}
A2	Raw local/global maps	4 softsign channels; spatial	$0.1L_{rank}$
A3	Raw local/global maps	4 softsign channels; spatial	$L_{map}+0.1L_{rank}$; primary
D1	Raw local/global maps	2 softsign channels; spatial	Joint; no difference/product
D2	Raw local/global maps	4 softsign channels; global-pooled	Joint; non-spatial control
D3	Raw local/global maps	4 clipped channels; spatial	Joint; conditioning-only clip

2.2.3. Pre-specified evaluation protocol

The primary comparison, specified and frozen before Test access, was A3 minus B2 macro image-level AUROC. Point estimates averaged structural and logical AUROCs within each category, fusion seeds within detector seed, detector seeds within category, and categories last. For the 10,000-replicate paired hierarchical percentile bootstrap (seed 404001), each A3 fusion seed was paired sample-wise with the same seed-free B2 prediction. Each replicate sampled categories, detector seeds within category, fusion seeds within detector, and images within the good, structural, and logical groups, all with replacement. Paired A3-B2 scores moved together. Tie-safe AUROCs were recomputed in every cell and aggregated in the same order. Eight secondary A3 comparisons formed one two-sided Holm family.

Secondary outcomes were structural and logical AUROC, category and detector-seed consistency, official LOCO evaluator v2.0 AU-sPRO@0.05, top-0.5% mean readout sensitivity, ranking reversals, gate variability, added parameters, and fusion-only median latency. Success required at least +1.0 macro-AUROC point, a positive lower confidence bound, no more than a 0.5-point AU-sPRO@0.05 decrease, positive effects in at least three of five categories and two of three detector seeds, fewer than 100,000 added parameters, and a latency ratio no greater than 1.10. A3 failed the latency gate before annotation access. A documented post-gate, pre-annotation amendment authorized completion of the frozen evaluation while retaining that failure and every original success criterion.

DOI: 10.54254/2755-2721/2026.36628

3. Results

3.1. Experimental execution and data integrity

The formal matrix contained 15 category-specific detector models, 405 trainable fusion runs, 75 fixed specifications, and 480 prediction/evaluation configurations. It generated 150,528 sealed prediction records for 1,568 Test images. A code- and hash-based automated integrity audit verified all 480 evaluator results, configuration identities, checkpoint and seed mappings, prediction-to-image mappings, annotation joins, official-evaluator hashes, and frozen aggregation definitions. No completed configuration was replaced by a more favorable rerun.

Execution events were retained rather than erased, including evaluator portability, tied-threshold handling detected before any metric was produced, a host restart during configuration 214, and an exact low-memory replacement for a post-evaluation Kendall computation. These recoveries changed execution compatibility or memory use, not predictions, completed metrics, seeds, thresholds, the method registry, or success criteria. The latency-gate continuation was separately recorded as the pre-annotation protocol amendment described in Section 2.2.

The complete matrix also separates confirmatory and diagnostic evidence. The A3-B2 comparison, confidence interval, Holm family, subgroup consistency, localization tolerance, parameter gate, and latency gate are confirmatory. Rank correlations, pair reversals, label-leaking oracles, and representative cases are diagnostics used to explain the frozen result, not to redefine success after Test access.

3.2. Primary comparison

A3 achieved 83.35% macro image-level AUROC, whereas B2 fixed equal averaging achieved 83.67%. The pre-specified effect was −0.32 percentage points, with a paired 95% bootstrap confidence interval of [−2.56, +1.74] points. Because the interval included zero and the estimate missed the +1.0-point criterion in both direction and magnitude, A3 did not demonstrate a reliable advantage. This finding is not an equivalence result. Figure 2 shows the estimate, interval, zero line, and pre-specified +1-point threshold. Throughout Section 3, effects and efficiency quantities were calculated from full-precision aggregates before display rounding; direct subtraction of Table 3 percentages can differ by 0.01 percentage points, and subtraction of displayed latencies by 0.001 ms.

The full 14-configuration table prevents a favorable subset from defining the narrative. A3 ranked approximately seventh by macro point estimate. B7 had the highest descriptive estimate at 83.96%, followed by B3 at 83.89%, but these rankings do not establish superiority because the primary method and comparison family were frozen earlier. None of the eight Holm-adjusted comparisons rejected its null; for example, A3-A1 had Holm-adjusted $p_{\mathrm{Holm}} = 0.3688$.

No strategy family dominated every outcome. B3 and B5 slightly exceeded B2 on macro AUROC, B4 had the highest AU-sPRO@0.05 among the displayed non-spatial methods, and several spatial or diagnostic variants traded structural gains against logical or localization losses. These patterns are descriptive only and were not used to replace the failed A3 hypothesis after Test access.

DOI: 10.54254/2755-2721/2026.36628

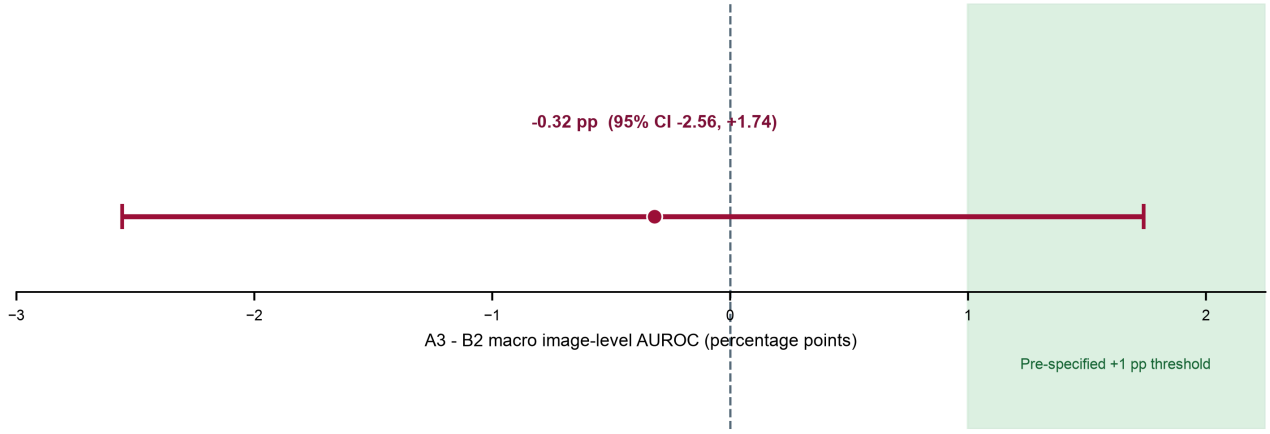

Figure 2. Primary A3-B2 effect on macro image-level AUROC. The point is the paired estimate, the horizontal line is the 95% bootstrap confidence interval, and the shaded area begins at the pre-specified +1-percentage-point threshold (picture credit: original)

Table 3. Complete frozen results for all 14 configurations (%). Values are descriptive point estimates. A3 versus B2 is the pre-specified primary comparison

Family	ID	Macro	Structural	Logical	AU-sPRO@0.05	Top-0.5%macro AUROC
Single	B0	81.54	90.69	72.38	58.80	81.39
Single	B1	72.53	71.06	74.01	47.59	73.29
Fixed/stat.	B2	83.67	89.62	77.72	67.07	83.90
Fixed/stat.	B3	83.89	89.61	78.17	66.88	84.20
Fixed/stat.	B4	83.70	89.79	77.62	67.08	83.92
Fixed/stat.	B5	83.83	88.93	78.74	65.92	84.17
Spatial	B6	79.15	83.17	75.13	65.76	81.18
Spatial	B7	83.96	90.47	77.44	66.93	84.19
Spatial	A1	80.10	84.38	75.82	64.32	80.51
Spatial	A2	83.28	90.99	75.57	64.89	83.19
Spatial	A3	83.35	90.56	76.14	66.20	83.40
Ablation	D1	83.31	90.55	76.07	66.00	83.34
Ablation	D2	83.27	90.35	76.19	65.81	83.33
Ablation	D3	83.56	90.57	76.56	66.69	83.67

3.3. Structural, logical and category-level effects

Relative to B2, A3 changed structural AUROC by +0.95 percentage points but logical AUROC by −1.58 points. The opposing directions explain why the structural gain did not produce a positive macro effect. They also show why the aggregate alone is insufficient: the gate altered the balance between anomaly regimes rather than providing a uniform improvement.

DOI: 10.54254/2755-2721/2026.36628

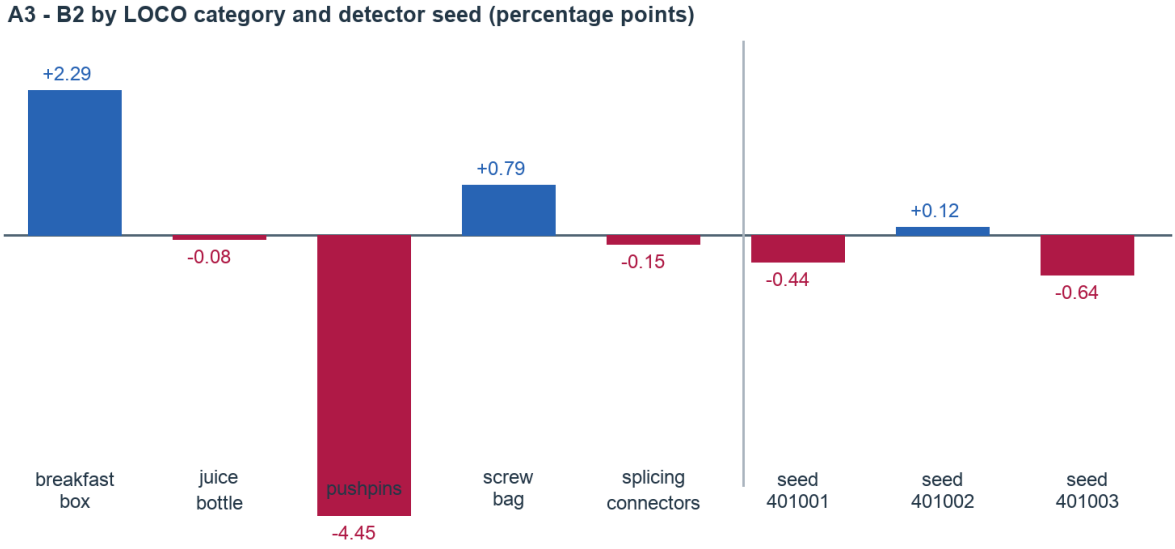

Figure 3. A3-B2 macro image-level AUROC effects by five MVTec LOCO AD categories and three detector seeds. Positive and negative bars show within-dataset heterogeneity; they are not independent-dataset estimates (picture credit: original)

Category-level A3-B2 effects were +2.29 points for breakfast box, −0.08 for juice bottle, −4.45 for pushpins, +0.79 for screw bag, and −0.15 for splicing connectors. Only two of five categories were positive, below the required three. Detector-seed effects were −0.44, +0.12, and −0.64 points, so only one of three was positive and the two-seed consistency criterion also failed.

All strata belonged to the same MVTec LOCO AD dataset and the same detector family. The mixed signs therefore indicate within-dataset heterogeneity rather than success on independent datasets. The −4.45-point pushpins effect further shows that a favorable result in one regime can coexist with substantial category-specific harm, which materially bounds the interpretation. Figure 3 displays these category- and detector-seed effects.

3.4. Localization, readout and efficiency

Localization and readout sensitivity did not reverse the primary result. A3 changed AU-sPRO@0.05 by −0.88 percentage points relative to B2, exceeding the allowed 0.5-point decrease. Under the top-0.5% mean readout, the A3-B2 effect was −0.50 points. Score bounding largely preserved global order (median Spearman 0.9857; median Kendall 0.9281), but 29,546 pairwise reversals remained across 15 detector-category strata, so high correlation did not imply identical rankings.

The learned gate was frequently close to spatially constant. Across 45 A3 configurations, the median within-map gate-weight standard deviation was 0.000943. This weakens, but does not disprove, the interpretation that pixel-wise freedom supplied useful local adaptation. Fusion-B-val pair accuracy had Spearman correlations of 0.0327 with logical Test AUROC and -0.4239 with structural Test AUROC, indicating that the proxy checkpoint signal was not a reliable guide to formal performance.

DOI: 10.54254/2755-2721/2026.36628

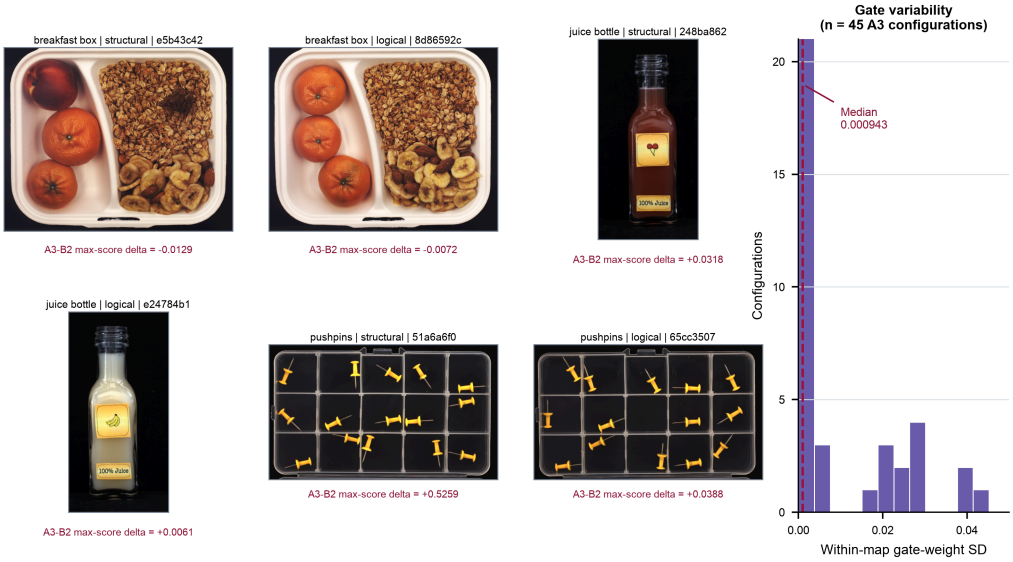

Figure 4. Representative anomalies and gate variability. Six test images were selected by the frozen median-distance rule rather than visual preference. The right panel shows the distribution of within-map gate-weight standard deviation across 45 A3 configurations; the dashed line marks the median (0.000943) (picture credit: original)

Label-leaking, non-deployable oracles were retained only as diagnostics. Selecting the better branch with Test labels reached 97.02% macro image-level AUROC, and a sampled-pixel oracle reached 95.28%. These values show that complementary branch information existed in principle, not that A3 learned a deployable selection rule. Figure 4 combines six cases selected by the frozen median-distance rule with the gate-variability distribution.

A3 added 769 parameters and therefore passed the parameter-count gate. At batch size 1 on the frozen RTX 4070 SUPER setup, however, its fusion-only median latency was 1.463 ms versus 0.757 ms for B2, an additional 0.707 ms and a 1.93-fold ratio. The ratio exceeded the 1.10 criterion. The evaluation continued only under the documented post-gate, pre-annotation amendment, and the efficiency failure remained part of the final decision.

4. Conclusion

This controlled study tested whether score-preserving spatial gating could replace fixed equal averaging of EfficientAD local and global anomaly maps under a leakage-controlled MVTec LOCO AD protocol. The primary comparison did not demonstrate a reliable advantage: A3 achieved 83.35% macro image-level AUROC, B2 achieved 83.67%, and the −0.32-point difference had a paired 95% confidence interval of [−2.56, +1.74]. Because the interval included zero and the estimate missed the pre-specified +1-point threshold, the result should not be read as equivalence or as proof that every spatial gate is ineffective.

The complete evidence explains why added flexibility did not yield stable value under the tested supervision. A3 gained +0.95 points on structural anomalies but lost −1.58 points on logical anomalies and −0.88 points on AU-sPRO@0.05; only two categories and one detector seed were positive. The gate was often nearly spatially uniform, while proxy checkpoint accuracy did not track logical Test performance. These diagnostics are consistent with a mismatch between structural proxy

DOI: 10.54254/2755-2721/2026.36628

supervision and the demands of logical anomaly detection, but they do not establish a single causal failure mechanism.

B2 is therefore the more defensible default under the evaluated conditions: it achieved a slightly higher primary point estimate without trainable fusion parameters, whereas A3 added 769 parameters and required 1.93 times the fusion-only median latency. This recommendation is conditional on one frozen MVTec LOCO AD archive, five categories, three detector seeds, one detector family, the max-based primary readout, and the evaluated proxy objective. It does not imply that fixed averaging is universally optimal or that spatial adaptation cannot help another detector, dataset, supervision source, or hardware platform.

Future studies should use prospective protocols and independent evaluation data rather than reusing the present Test outcomes for model selection. Useful tests include logic-aware proxies, pre-Test measures of branch complementarity, capacity-matched non-spatial controls, calibration uncertainty, and hardware-aware gating. The main contribution of the present experiment is therefore a complete, uncertainty-aware negative result: within the frozen EfficientAD-LOCO setting, the evaluated spatial gate did not justify replacing equal averaging.

References

[1] Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., & Steger, C. (2022). Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *International Journal of Computer Vision, 130*(4), 947-969.

[2] Batzner, K., Heckler, L., & König, R. (2024). EfficientAD: Accurate visual anomaly detection at millisecond-level latencies. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 127-137.

[3] Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., & Gehler, P. (2022). Towards total recall in industrial anomaly detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14298-14308.

[4] Hsieh, Y.-H., & Lai, S.-H. (2024). CSAD: Unsupervised component segmentation for logical anomaly detection. *35th British Machine Vision Conference (BMVC 2024)*.

[5] Fučka, M., Zavrtanik, V., & Skočaj, D. (2025). SALAD - Semantics-aware logical anomaly detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 21843-21852.

[6] Li, Z., Yang, Z., Zhang, L., Yang, L., & Liu, J. (2025). LA-EAD: Simple and effective methods for improving logical anomaly detection capability. *Sensors, 25*(16), 5016.

[7] Fučka, M., Zavrtanik, V., & Skočaj, D. (2026). ObjectCore - Efficient few-shot logical anomaly detection using object representations. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 3857-3867.

[8] Kim, M., Lee, S., Ko, B., Yoon, J.-W., Park, T., & Park, H. (2026). Heterogeneous multi-score integration for structural and logical anomaly detection. *IEEE Access, 14*, 65181-65194.

[9] Agarwal, S. (2013). Surrogate regret bounds for the area under the ROC curve via strongly proper losses. *Proceedings of the 26th Annual Conference on Learning Theory*, 338-353.

[10] Bae, J., Lee, J.-H., & Kim, S. (2023). PNI: Industrial anomaly detection using position and neighborhood information. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 6350-6360.

[11] Zhao, Y. (2024). Logical: Towards logical anomaly synthesis for unsupervised anomaly localization. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 4022-4031.

[12] Wang, C., Chen, J., & Zhang, H. (2026). DCRDF-Net: A dual-channel reverse-distillation fusion network for 3D industrial anomaly detection. *Sensors, 26*(2), 412.

[13] Tao, C., Cao, X., & Du, J. (2025). G2SF: Geometry-guided score fusion for multimodal industrial anomaly detection. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 20551-20560.

[14] Deng, M., Sun, S., Hu, X., & Wu, X. (2026). TC-MAF: Train-calibrated bounded multi-evidence fusion for multimodal industrial anomaly detection. *arXiv*, 2607.11170v1. https://arxiv.org/abs/2607.11170

[15] Li, C., Yang, C., Xiao, Y., & Tammi, K. (2026). UniSLAD: A unified framework for structural and logical industrial visual anomaly detection. *arXiv*, 2606.20768v1. https://arxiv.org/abs/2606.20768