GOMC-AD: CAPACITY-CONSTRAINED GLOBAL NORMALITY ASSOCIATION WITH TRANSPORT-CONDITIONED RESIDUAL CALIBRATION FOR FEW-SHOT INDUSTRIAL ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-shot industrial anomaly detection (FS-IAD) identifies defects from a few normal reference images. Despite advances in pretrained representations, locally plausible matches need not form a jointly consistent normal explanation, while residuals conflate defects with association ambiguity and normal appearance drift. We propose Global Optimal-Transport Memory Calibration for Anomaly Detection (GOMC-AD), a training-free framework for *capacity-constrained normality explanation*. Optimal-Transport Guided Global Memory Association (OT-GMA) couples query patches through prescribed normal capacities and a fixed rejection budget, making reference allocation part of the anomaly criterion. Transport-Residual-Aware Test-Time Gaussian Calibration (TR-TGC) incorporates transport-derived association variance into heteroscedastic Gaussian scoring and removes its modeled contribution from historical second-moment updates. Asymmetric feedback reallocates rejection mass while retaining the first-pass normal moments used to define residuals. Across MVTec AD, VisA, and Real-IAD with 1-, 2-, and 4-shot support, GOMC-AD outperforms the compared methods in mean image- and pixel-level area under the receiver operating characteristic curve (AUROC). Image-level AUROC improves by up to 5.3 percentage points and pixel-level AUROC by up to 1.3 points over the strongest compared methods. An anonymous project page is available at `https://anonymous.4open.science/r/GOMC-AD-C25C/`.

1 INTRODUCTION

Transferring pretrained visual representations to tasks with scarce supervision is a central problem in representation learning. Few-shot industrial anomaly detection (FS-IAD) addresses this problem by identifying defective images and localizing defects from a few normal references per category. Recent work exploits instance-induced prompts (Lv et al., 2025) and foundation-feature projection (Zhai et al., 2026), while feature memories enable few-shot inspection (Roth et al., 2022; Damm et al., 2025). This setting probes how transferable features become reliable normality judgments under limited reference coverage and changing test conditions.

Two modeling challenges remain. First, *local plausibility does not ensure a jointly consistent normal explanation*. Independent retrieval allows multiple query regions to reuse the same support patches, potentially concealing missing components or abnormal arrangements. Reconstruction improves reference representativeness (Fang et al., 2023). We therefore assess the extra matching cost incurred when query regions must share prescribed normal capacities (Fig. 1).

Second, *matching residuals conflate defects with distinct sources of normal variation*. Sparse references make correspondence ambiguous, while illumination and viewpoint changes alter otherwise normal appearances (Cao et al., 2023). Probabilistic adaptation models category-conditional shifts (Xu et al., 2026); patch-level calibration must additionally separate association ambiguity from historical normal variation. A shared residual Gaussian does not represent the patch-specific uncertainty of each normal correspondence.

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.

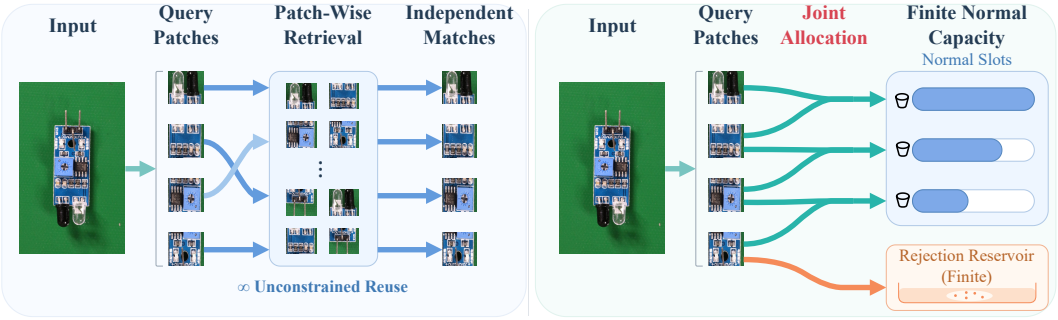


Fig. 1: From independent similarity to joint normality. Left: query patches retrieve normal neighbors independently, allowing unrestricted reuse of support patterns. Right: joint association couples assignments through shared normal capacities and a finite rejection reservoir. Competition makes the consistency of the complete query part of the anomaly criterion.

These challenges inspire us to view normal explanations through both global consistency and matching reliability. Joint correspondence estimation (Sarlin et al., 2020) and probabilistic matching (Truong et al., 2023) motivate these complementary requirements. Yet consistent assignments alone do not separate matching ambiguity from historical normal variation. We therefore link normal-reference allocation with residual calibration, so that anomaly judgments account for both shared matching constraints and reference uncertainty.

We propose Global Optimal-Transport Memory Calibration for Anomaly Detection (GOMC-AD), a training-free framework for *capacity-constrained normality explanation*. To address unrestricted reference reuse in independent retrieval, Optimal-Transport Guided Global Memory Association (OT-GMA) uses optimal transport (OT) to impose shared normal capacities and a fixed rejection budget. Competition for finite support makes the difficulty of jointly explaining all query regions part of the anomaly criterion, revealing inconsistencies that independent local similarity can overlook.

To address residual ambiguity, Transport-Residual-Aware Test-Time Gaussian Calibration (TR-TGC) models transport-derived association variance separately from historical residual drift. Incorporating association variance into scoring accounts for the uncertainty of each normal reference, while subtracting its modeled contribution from historical second moments reduces the conflation of matching variability with normal drift. Reference-preserving feedback refines rejection allocation while retaining the first-pass normal moments that define residuals, ensuring that calibrated evidence adjusts rejection against a consistent reference.

Experiments on MVTec AD, VisA, and Real-IAD under 1-, 2-, and 4-shot support show consistent gains across all 18 dataset–shot–metric combinations. Compared with the strongest baseline in each setting, image-level area under the receiver operating characteristic curve (AUROC) improves by up to 5.3 percentage points and pixel-level AUROC by up to 1.3 points. Ablations support the contributions of shared-capacity association, uncertainty-aware calibration, and asymmetric feedback.

Our contributions are summarized as follows:

- We formulate FS-IAD as capacity-constrained normality explanation, extending anomaly inference from independent local similarity to jointly constrained reference allocation and uncertainty-aware residual assessment.
- We design OT-GMA to couple query-patch assignments through prescribed normal capacities and a fixed rejection budget, deriving matching discrepancy and uncertainty-bearing normal references from the same transport plan.
- We develop TR-TGC to add association variance in residual scoring and remove its modeled contribution from historical second-moment updates, with reference-preserving feedback that refines rejection allocation while retaining first-pass normal moments.

2 Related Work

We review unsupervised anomaly detection, foundation models, and few-shot methods, comparing their reference modeling, association, and calibration.

2.1 Unsupervised Anomaly Detection

Normal-only detectors use patch-wise Gaussians in PaDiM (Defard et al., 2021) or feature memories in PatchCore (Roth et al., 2022). Limited coverage motivates stronger discrimination: dual-model distillation (Liu et al., 2024) separates normal and abnormal representations, while RealNet (Zhang et al., 2024) combines synthetic defects with selective reconstruction. DiffusionAD (Zhang et al., 2025) uses guided denoising, and AnomalyAny (Sun et al., 2025) synthesizes defects from normal images and text. MaskAD (Lu et al., 2026) compares reconstruction-error distributions across masks using OT; for composition errors, SALAD (Fučka et al., 2025) and UniVAD (Gu et al., 2025) model semantics and component relations. These methods enrich normal references; OT-GMA tests whether locally credible matches can coexist under shared normal capacities.

2.2 Foundation Models for Visual Understanding

Contrastive Language–Image Pre-training (CLIP) (Radford et al., 2021) supplies semantic priors. WinCLIP (Jeong et al., 2023) combines prompts with window features; PromptAD (Li et al., 2024) learns prompts from normal samples, and FiLo (Gu et al., 2024) refines descriptions and localization. MALM-CLIP (Chen et al., 2026) uses generative agents, InCTRL (Zhu & Pang, 2024) transfers in-context residuals, and instance-induced (Lv et al., 2025) and kernel-aware graph prompts (Tao et al., 2025) incorporate instance and cross-layer context. Visual references include dense DINOv3 features (Siméoni et al., 2026), patch retrieval in AnomalyDINO (Damm et al., 2025), normal prototypes in INP-Former (Luo et al., 2025), and manifold projection in FoundAD (Zhai et al., 2026). Strong representations still leave joint support allocation to the detector; GOMC-AD imposes shared normal capacities on query-region association.

2.3 Few-Shot Anomaly Detection

Few-shot methods address sparse references through one-prompt meta-learning (Gao, 2024), patch retrieval (Wang et al., 2025), support-based reconstruction in FastRecon (Fang et al., 2023), or normal-subspace residual scoring in SubspaceAD (Lendering et al., 2026). FOCT (Tian et al., 2024) introduces foreground-aware online conditional transport; ANoCo (Seo et al., 2026) measures conformity to anchored references through graph energy. OT-GMA prescribes normal capacities and a fixed rejection budget to jointly constrain query-region association.

For historical adaptation, RareCLIP (He et al., 2025) estimates patch rarity in zero-shot streams; probabilistic Gaussian calibration (Xu et al., 2026) models category-conditional shifts. Association ambiguity and normal drift require separate treatment. TR-TGC adds transport variance in scoring and removes its modeled contribution from historical second-moment updates. Calibrated feedback changes only rejection costs, preserving first-pass normal moments as residual references.

3 Method

3.1 Joint Normality Framework

In Fig. 2, frozen DINOv3 (Siméoni et al., 2026) encodes augmented views of K normal images into fixed, view-structured memory (top) and original/flipped queries into patches (middle). Nearest-neighbor (NN) retrieval searches all support views for local distances. OT-GMA (Sec. 3.2) selects similar views and solves per-view transport plans: normal capacities limit support reuse, while a fixed rejection reservoir provides an alternative to normal matching. Weighted view outputs supply the cost gap over NN retrieval, normal reference features, and association variance.

TR-TGC (Sec. 3.3) evaluates query-minus-reference residuals using association variance and support-initialized historical statistics. Gaussian evidence drives the orange feedback path: rejection costs change and transport runs again. Only the new rejection allocation a^* enters scoring; first-pass distances, references, and variances remain fixed. Bounded fusion combines local, gap, Gaussian, and rejection evidence while limiting score modulation. Query-view maps are aligned and averaged; pooled evidence yields image scores. After prediction is fixed, the bottom-row guard skips suspicious images; low-risk patches from accepted images supply weighted residuals to update history for subsequent queries.

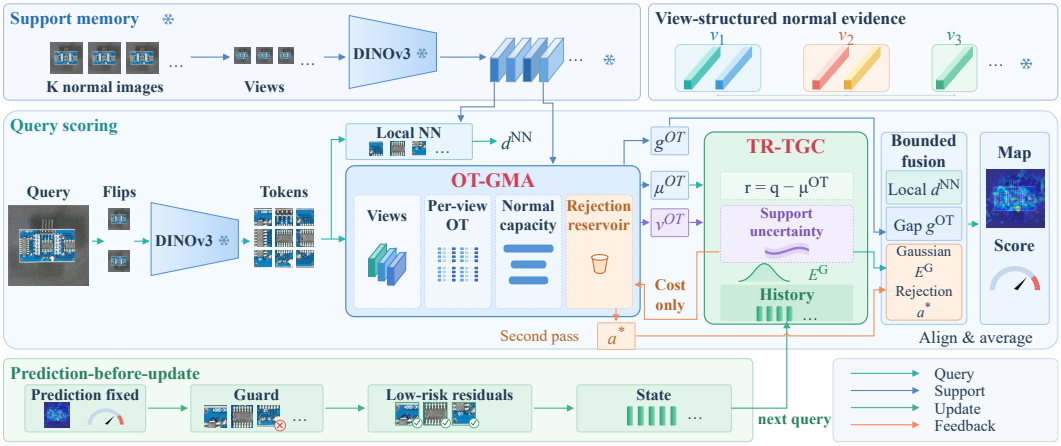


Fig. 2: GOMC-AD couples shared-capacity association with residual calibration. Nearest-neighbor (NN) retrieval supplies the local baseline. OT-GMA jointly derives normal moments and anomaly evidence. TR-TGC separates association uncertainty from historical drift and feeds calibrated evidence back to rejection costs while retaining first-pass normal moments. Predictions precede evidence-weighted history updates.

3.2 Optimal-Transport Guided Global Memory Association

Independent retrieval permits unrestricted reuse of normal patterns, potentially concealing missing components or abnormal arrangements. OT-GMA couples regional explanations through shared normal capacities: assigning support mass to one region limits its availability to others. The resulting plan supplies matching discrepancy and an uncertainty-bearing normal reference, linking joint association to residual calibration (Fig. 3).

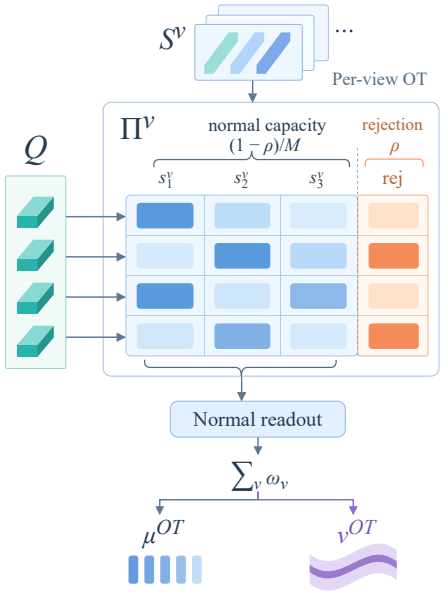


Fig. 3: OT-GMA turns shared-capacity competition into a normal prototype $\boldsymbol{\mu}^{\mathrm{OT}}$, association variance $\boldsymbol{\nu}^{\mathrm{OT}}$, and rejection evidence.

View-structured transport association Geometric augmentation retains support-image and view identities. For query t, let $\mathbf{Q}_t = \{\mathbf{q}_{t,i}\}_{i=1}^{N}$ and $\mathbf{S}^v = \{\mathbf{s}_j^v\}_{j=1}^{M}$ contain normalized query and support-view patches in $\mathbb{R}^D$, with patch counts N and M. Ranking normalized global descriptors selects support views $\mathcal{V}_t$. Within each selected view, feature similarity and spatial structure define the association cost

$$C_{t,i,j}^{v} = 1 - \mathbf{q}_{t,i}^{\top}\mathbf{s}_j^{v} + \beta\|\mathbf{p}_i - \mathbf{p}_j\|_2^2, \qquad (1)$$

where $\mathbf{p}_i, \mathbf{p}_j \in [0,1]^2$ are patch coordinates and $\beta \geq 0$ controls spatial guidance. Keeping plans view-specific prevents capacity allocation from pooling patches across different support views.

Shared-capacity reference allocation Per view, query masses are $a_i = 1/N$, normal capacities $b_j = (1-\rho)/M$ for $j \leq M$, and rejection capacity $b_\perp = \rho$, where $\rho \in [0,1)$. A pattern occupying r support patches receives total capacity $r(1-\rho)/M$; support/query composition mismatch may overconstrain matching. With costs $\widetilde{\mathbf{C}}_t^v = [\mathbf{C}_t^v, \mathbf{0}]$, the balanced entropic plan is

$$\boldsymbol{\Pi}_t^{v,*} = \underset{\substack{\boldsymbol{\Pi}\geq 0,\ \boldsymbol{\Pi}\mathbf{1}=\mathbf{a}\\ \boldsymbol{\Pi}^{\top}\mathbf{1}=\mathbf{b}}}{\arg\min} \left\{ \langle \boldsymbol{\Pi}, \widetilde{\mathbf{C}}_t^v \rangle + \varepsilon \sum_{i,j} \Pi_{i,j}(\log \Pi_{i,j} - 1) \right\}. \qquad (2)$$

Here $j \in \{1, \ldots, M, \perp\}$ and $\varepsilon > 0$ controls assignment softness. Per-view Sinkhorn (Cuturi, 2013) enforces shared capacities. Fixing ρ reserves $1-\rho$ for normal matching, so queries cannot relax this constraint by increasing rejection. A constant rejection cost c_0 adds $c_0\rho$ to the objective, leaving the

optimal plan unchanged. We prioritize a shared competition budget over adaptive mass; rejection allocation provides evidence rather than an anomaly-proportion estimate.

Transport-derived normality and uncertainty The same allocation yields a normal reference and its uncertainty. Conditioning the plan on normal columns gives

$$P^{v}_{t,i,j} = \frac{\Pi^{v,*}_{t,i,j}}{\sum_{\ell=1}^{M} \Pi^{v,*}_{t,i,\ell}}, \qquad j \le M. \tag{3}$$

Rejection is aggregated with normalized view weights ω^v_t:

$$a^{v}_{t,i} = \frac{\Pi^{v,*}_{t,i,\perp}}{\sum_{j=1}^{M} \Pi^{v,*}_{t,i,j} + \Pi^{v,*}_{t,i,\perp}}, \qquad a_{t,i} = \sum_{v \in \mathcal{V}_t} \omega^v_t a^v_{t,i}. \tag{4}$$

For each view, $d^v_{t,i} = \sum_j P^v_{t,i,j}(1 - \mathbf{q}^\top_{t,i}\mathbf{s}^v_j)$ and $\boldsymbol{\mu}^v_{t,i} = \sum_j P^v_{t,i,j}\mathbf{s}^v_j$. Setting $\omega^v_t \propto \exp(-\bar{d}^v_t/\tau)$, with normal-mass-weighted feature cost $\bar{d}^v_t$ and temperature $\tau > 0$, yields

$$d^{\mathrm{OT}}_{t,i} = \sum_v \omega^v_t d^v_{t,i}, \quad \boldsymbol{\mu}^{\mathrm{OT}}_{t,i} = \sum_v \omega^v_t \boldsymbol{\mu}^v_{t,i}, \quad \boldsymbol{\nu}^{\mathrm{OT}}_{t,i} = \mathrm{Var}_{\omega_t P_{t,i}}[\mathbf{s}]. \tag{5}$$

View sums run over $\mathcal{V}_t$; Var is elementwise variance under weights $\omega^v_t P^v_{t,i,j}$. These moments describe a normal prototype and association ambiguity within and across views. For comparison, the unconstrained nearest-neighbor distance is $d^{\mathrm{NN}}_{t,i} = \min_{v,j}(1 - \mathbf{q}^\top_{t,i}\mathbf{s}^v_j)$. The global-consistency gap $g^{\mathrm{OT}}_{t,i} = d^{\mathrm{OT}}_{t,i} - d^{\mathrm{NN}}_{t,i}$ measures extra feature cost under joint matching. It is nonnegative when both readouts use the same distance and selected supports belong to the nearest-neighbor search set. The same plan thus supplies matching discrepancy and the uncertainty needed to interpret residuals.

3.3 Transport-Residual-Aware Test-Time Gaussian Calibration

Residuals mix defects with ambiguous associations and normal appearance drift. A shared residual scale ignores patch-specific association confidence. TR-TGC applies *transport-conditioned uncertainty decomposition* in both scoring and historical updates. Its feedback adjusts rejection costs while retaining first-pass normal moments as the reference for residual calibration (Fig. 4).

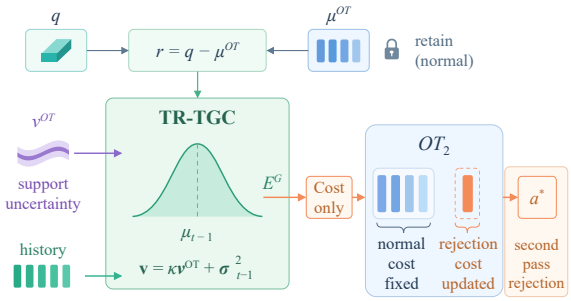


Fig. 4: TR-TGC separates association uncertainty from historical drift. Asymmetric feedback refines rejection costs while retaining first-pass normal moments as the residual reference.

Transport-conditioned uncertainty decomposition For patch i of query t, let $\mathbf{r}_{t,i} = \mathbf{q}_{t,i} - \boldsymbol{\mu}^{\mathrm{OT}}_{t,i}$. With first-pass moments fixed, a diagonal Gaussian working model uses historical mean $\boldsymbol{\mu}_{t-1}$ and patch-specific variance $\mathbf{v}_{t,i}$. Its dimension-normalized negative log-likelihood (NLL) is

$$E^G_{t,i} = \frac{1}{2D}\sum_{d=1}^{D}\left[\log(2\pi v_{t,i,d}) + \frac{(r_{t,i,d} - \mu_{t-1,d})^2}{v_{t,i,d}}\right], \quad \mathbf{v}_{t,i} = \kappa\boldsymbol{\nu}^{\mathrm{OT}}_{t,i} + \boldsymbol{\sigma}^2_{t-1}. \tag{6}$$

Here $\kappa \ge 0$ scales association variance. The D-dimensional vectors $(\boldsymbol{\mu}_{t-1}, \boldsymbol{\sigma}^2_{t-1})$ track channel-wise residual drift from support-initialized, accepted history $\mathcal{D}_{<t}$. Evaluation requires $v_{t,i,d} > 0$. The quadratic term measures drift-corrected surprise; the log-variance term penalizes diffuse explanations. Equal residual magnitudes can therefore imply different anomaly evidence when normal references have different uncertainty, rather than being judged by a shared distance threshold.

Reference-preserving rejection feedback Leave-one-view-out support queries initialize Gaussian statistics and robust scales. The normalization $\mathcal{Z}(y; m, s) = \mathrm{clip}((y - m)/\max(s, \epsilon_s), -c, c)$ produces $z^{\mathrm{OT}}, z^G, z^R$ for gap, NLL, and rejection. Here m is the median, s the scaled median

absolute deviation (MAD), and $\epsilon_s, c > 0$. Support-derived gap and rejection scales stay fixed; Gaussian median/MAD estimates use exponential moving average (EMA) updates from accepted residual NLLs after prediction. Preliminary evidence $z^{G,\mathrm{pre}}$ adjusts the second-pass rejection cost:

$$c^{v,(2)}_{t,i,\perp} = c^{v,(1)}_{t,i,\perp} - \eta \tanh\Big([z^{G,\mathrm{pre}}_{t,i}]_+\Big), \tag{7}$$

where $[x]_+ = \max(x, 0)$ and $\eta \geq 0$ controls feedback. Normal costs and marginals remain unchanged. The second plan supplies a^*, while first-pass $(\boldsymbol{\mu}^{\mathrm{OT}}, \boldsymbol{\nu}^{\mathrm{OT}}, d^{\mathrm{OT}})$ remain the scoring reference. Feedback reallocates rejection using the updated plan while residuals are evaluated against the first-pass normal moments that produced the feedback evidence.

Bounded joint-evidence fusion Prediction combines transport discrepancy, calibrated residual evidence, and refined rejection through bounded modulation. Set $B_{t,i} = d^{\mathrm{NN}}_{t,i} + \alpha[g^{\mathrm{OT}}_{t,i}]_+$, where $\alpha \in [0, 1]$. Let $e^{\mathrm{OT}} = [\tanh(z^{\mathrm{OT}})]_+$, $e^G = [\tanh(z^G)]_+$, and $e^R = a^*$. The patch score is

$$S_{t,i} = B_{t,i} \operatorname{clip}_{[\ell,u]}\big((1 + \lambda_{\mathrm{OT}} e^{\mathrm{OT}}_{t,i})(1 + \lambda_G e^G_{t,i})(1 + \lambda_R e^R_{t,i})\big). \tag{8}$$

Nonnegative $\lambda_{\mathrm{OT}}, \lambda_G, \lambda_R$ weight the evidence terms; $0 < \ell \leq u$ limits their combined multiplicative effect on the reference score. Aligned query views yield the upsampled anomaly map; image scoring applies the same bounded fusion to pooled evidence.

Association-corrected drift adaptation Predictions are fixed before statistics evolve. For queries passing an image-level guard on joint anomaly evidence, evidence ranks, association uncertainty, and view consistency select a low-risk set $\mathcal{L}_t$. Evidence-derived weights are

$$w_{t,i} = \exp\big(-[z^{\mathrm{OT}}_{t,i}]_+ - [z^G_{t,i}]_+ - [z^R_{t,i}]_+\big), \qquad i \in \mathcal{L}_t. \tag{9}$$

Discounted statistics accumulate weighted mass, residuals, and corrected second moments $\mathbf{r}^{\odot 2}_{t,i} - \kappa \boldsymbol{\nu}^{\mathrm{OT}}_{t,i}$, where $\odot 2$ denotes elementwise squaring. Subtracting the same scaled association variance keeps scoring and drift estimation consistent, avoiding its accumulation as historical variation. Selection controls which residuals enter history; variance correction controls what their second moments contribute. Support anchoring and variance shrinkage stabilize residual updates without changing the normal memory.

4 Results

We evaluate detection and localization, then analyze representation, component complementarity, uncertainty decomposition, and reference-preserving feedback.

4.1 Datasets and Evaluation Criteria

MVTec AD (Bergmann et al., 2019), VisA (Zou et al., 2022), and Real-IAD (Wang et al., 2024) contain 15/12/30 categories. We sample $K \in \{1, 2, 4\}$ normal training images per category and evaluate official test splits; Real-IAD uses INP-Former's individual-image sampling (Luo et al., 2025). We report category-macro image/pixel AUROC (I-/P-AUROC, %) with 256×256 anomaly maps, as mean $\pm$ standard deviation over five predefined support samplings. Each sampling uses one recorded-seed random test permutation, shared by matched variants. Queries are scored before statistical updates. Hyperparameters are selected on independent validation data, shared across categories within each dataset, and fixed during testing.

4.2 Implementation Details

Batch-one inference runs on Linux with three RTX 4090 graphics processing units (GPUs). Frozen DINOv3 Vision Transformer (ViT)-L/16 LVD-1689M (Siméoni et al., 2026) maps 448×448 inputs to 28×28 patches with layer normalization and ℓ_2 normalization. Support views use quarter-turn rotations and flips; query scores average original and vertically flipped views. OT uses 30 float32 log-domain Sinkhorn iterations, two retained views, entropy/spatial weights 0.05/0.05, rejection mass 0.02, and feedback strength 0.03. TR-TGC uses momentum 0.99 and transport-variance scale 1.0. Inference costs 1168.57 billion floating-point operations (GFLOPs) and approximately 1.4 GB GPU memory at 8.2 frames per second (FPS); 2-shot loading, decoding, and preprocessing average 23 ms/image.

Table 1: Few-shot detection and localization (%). Bold/underline indicate best/second-best means, including ties. All rows report reproduced five-run mean ± standard deviation.

Shot	Method	Venue	MVTec AD		VisA		Real-IAD	
			I-AUROC	P-AUROC	I-AUROC	P-AUROC	I-AUROC	P-AUROC
1	PaDiM	ICPR Workshops'21	73.3 ± 3.8	87.2 ± 1.1	58.0 ± 2.7	86.3 ± 1.0	63.5 ± 1.1	93.5 ± 0.9
	PatchCore	CVPR'22	83.7 ± 2.5	92.6 ± 0.9	70.2 ± 4.3	95.9 ± 0.7	71.3 ± 3.0	93.3 ± 1.0
	FastRecon	ICCV'23	85.9 ± 2.1	93.3 ± 0.7	75.4 ± 3.0	96.9 ± 0.5	71.0 ± 2.4	95.5 ± 0.7
	WinCLIP	CVPR'23	92.7 ± 2.1	93.8 ± 0.8	83.3 ± 4.5	95.1 ± 1.2	74.5 ± 3.1	94.7 ± 0.7
	PromptAD	CVPR'24	93.5 ± 2.2	95.5 ± 0.7	86.5 ± 3.1	95.9 ± 0.9	71.2 ± 2.7	91.8 ± 0.8
	AnomalyDINO	WACV'25	96.3 ± 1.1	95.7 ± 0.4	82.1 ± 1.6	96.5 ± 0.4	73.1 ± 1.3	95.3 ± 0.5
	INP-Former	CVPR'25	96.3 ± 0.4	96.8 ± 0.5	91.8 ± 1.4	97.6 ± 0.4	68.2 ± 1.1	95.0 ± 0.5
	FoundAD	ICLR'26	96.1 ± 0.7	96.8 ± 0.2	92.6 ± 0.6	98.0 ± 0.5	70.6 ± 1.3	95.5 ± 0.3
	SubspaceAD	CVPR'26	97.0 ± 0.6	96.9 ± 0.5	91.9 ± 1.0	98.0 ± 0.3	75.2 ± 1.4	95.5 ± 0.6
	GOMC-AD	Ours	**97.8 ± 0.4**	**97.6 ± 0.3**	**95.2 ± 0.9**	**98.4 ± 0.2**	**80.0 ± 0.7**	**96.8 ± 0.4**
2	PaDiM	ICPR Workshops'21	78.1 ± 4.2	90.3 ± 1.5	62.6 ± 4.0	90.2 ± 1.9	66.0 ± 2.8	94.2 ± 0.8
	PatchCore	CVPR'22	87.7 ± 3.7	93.6 ± 0.8	80.6 ± 4.2	96.7 ± 0.7	72.1 ± 2.5	93.6 ± 0.8
	FastRecon	ICCV'23	88.7 ± 1.7	94.2 ± 0.8	86.3 ± 2.2	97.5 ± 0.5	71.8 ± 1.9	96.1 ± 1.2
	WinCLIP	CVPR'23	94.0 ± 1.7	93.6 ± 0.5	83.9 ± 2.8	94.9 ± 0.5	75.8 ± 2.3	94.9 ± 0.6
	PromptAD	CVPR'24	94.7 ± 1.9	96.1 ± 0.3	87.8 ± 2.5	96.5 ± 0.6	75.0 ± 2.0	93.2 ± 0.4
	AnomalyDINO	WACV'25	96.5 ± 0.9	95.8 ± 0.4	82.8 ± 1.3	96.8 ± 0.3	77.0 ± 1.1	95.9 ± 0.4
	INP-Former	CVPR'25	96.8 ± 0.7	97.3 ± 0.3	94.4 ± 0.9	98.0 ± 0.4	70.8 ± 0.9	95.9 ± 0.4
	FoundAD	ICLR'26	96.8 ± 0.3	97.0 ± 0.2	93.5 ± 0.6	98.2 ± 0.5	72.1 ± 0.9	96.0 ± 0.2
	SubspaceAD	CVPR'26	97.4 ± 0.4	97.2 ± 0.4	92.6 ± 0.8	98.1 ± 0.3	77.6 ± 1.0	96.2 ± 0.5
	GOMC-AD	Ours	**98.3 ± 0.5**	**97.6 ± 0.2**	**95.9 ± 0.7**	**98.6 ± 0.3**	**82.1 ± 0.6**	**97.4 ± 0.3**
4	PaDiM	ICPR Workshops'21	79.3 ± 2.5	91.8 ± 1.5	66.0 ± 3.7	92.4 ± 0.9	70.4 ± 2.2	93.8 ± 0.8
	PatchCore	CVPR'22	89.8 ± 3.0	94.9 ± 0.7	84.4 ± 2.8	97.7 ± 0.6	75.6 ± 2.0	94.0 ± 0.7
	FastRecon	ICCV'23	91.1 ± 1.2	95.9 ± 0.5	88.6 ± 1.6	98.1 ± 0.4	73.3 ± 1.5	95.9 ± 0.7
	WinCLIP	CVPR'23	95.1 ± 1.5	94.3 ± 0.4	83.8 ± 2.1	95.6 ± 0.4	76.2 ± 1.9	94.4 ± 0.5
	PromptAD	CVPR'24	95.7 ± 1.2	96.3 ± 0.5	88.9 ± 1.8	96.9 ± 0.4	78.3 ± 1.5	94.7 ± 0.5
	AnomalyDINO	WACV'25	97.2 ± 0.5	96.5 ± 0.3	86.2 ± 0.8	97.0 ± 0.2	79.6 ± 0.6	96.2 ± 0.3
	INP-Former	CVPR'25	97.5 ± 0.3	97.0 ± 0.5	96.3 ± 0.4	98.2 ± 0.6	77.0 ± 0.9	97.3 ± 0.2
	FoundAD	ICLR'26	97.1 ± 0.2	97.2 ± 0.2	94.4 ± 0.4	98.2 ± 0.1	75.3 ± 0.6	96.3 ± 0.3
	SubspaceAD	CVPR'26	97.6 ± 0.3	97.3 ± 0.3	92.8 ± 0.7	98.1 ± 0.2	79.8 ± 0.8	96.8 ± 0.4
	GOMC-AD	Ours	**98.5 ± 0.3**	**97.9 ± 0.2**	**96.6 ± 0.5**	**98.7 ± 0.1**	**85.1 ± 0.6**	**98.2 ± 0.2**

4.3 Comparison with State-of-the-Art Methods

Table 1 reports reproduced results at 448×448. WinCLIP and PromptAD use CLIP; other baselines, including SubspaceAD (Lendering et al., 2026), use DINOv3-L.

GOMC-AD leads all 18 dataset–shot–metric combinations. Its 1-shot I-AUROC gains on MVTec AD/VisA/Real-IAD are 0.8/2.6/4.8 points over the strongest competitors and 0.8/3.3/4.8 over SubspaceAD. Both MVTec AD 1-shot metrics exceed all listed 4-shot baselines; P-AUROC improves by 0.3–1.3 points. Despite modest pixel-level gains, GOMC-AD avoids SubspaceAD's single global linear-subspace assumption through view-structured matching. Fig. 5 shows concentrated defect responses and reduced normal-region activation; matched ablations isolate association and calibration.

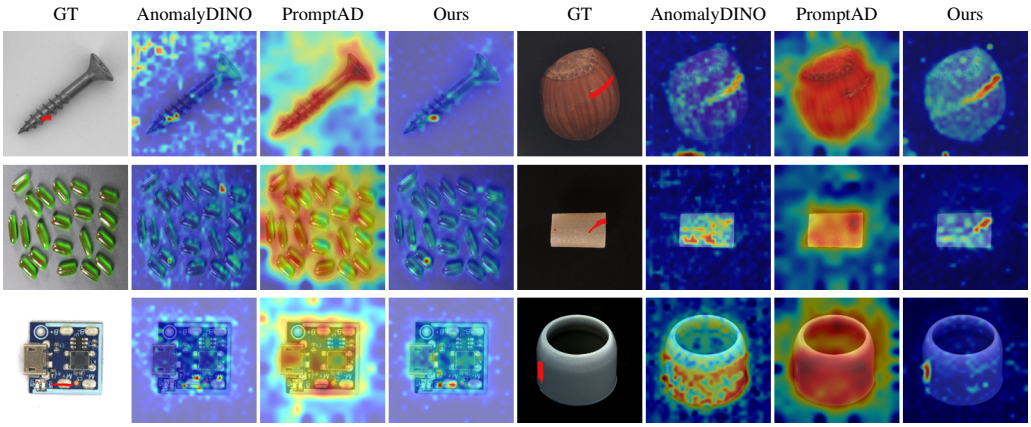


Fig. 5: Representative 1-shot localization across anomaly appearances and structures. Columns: ground truth (GT, red), AnomalyDINO (Damm et al., 2025), PromptAD (Li et al., 2024), and GOMC-AD.

4.4 Ablation Study

Backbone Ablation Table 2 compares five frozen encoders under matched 1-shot settings, including DINOv2 (Oquab et al., 2024) and WideResNet (Zagoruyko & Komodakis, 2016). CLIP uses visual features; L/B denote Large/Base. DINOv3-L leads all six metrics. The smaller DINOv3-B also exceeds DINOv2-L, suggesting that parameter count alone does not explain the encoder ranking and motivating the use of DINOv3 features.

Table 2: Backbone comparison under 1-shot support (%, higher is better).

	MVTec AD		VisA		Real-IAD	
Backbone	I-AUROC	P-AUROC	I-AUROC	P-AUROC	I-AUROC	P-AUROC
DINOv3 ViT-L/16	**97.8 ± 0.4**	**97.6 ± 0.3**	**95.2 ± 0.9**	**98.4 ± 0.2**	**80.0 ± 0.7**	**96.8 ± 0.4**
DINOv3 ViT-B/16	97.0 ± 0.3	96.9 ± 0.3	93.6 ± 0.6	97.8 ± 0.3	78.5 ± 0.7	95.8 ± 0.5
DINOv2 ViT-L/14	96.3 ± 0.4	96.2 ± 0.4	92.4 ± 0.6	97.1 ± 0.4	76.9 ± 0.8	94.8 ± 0.5
CLIP ViT-L/14	82.0 ± 1.5	91.7 ± 0.7	79.8 ± 1.8	93.6 ± 0.7	68.5 ± 1.5	91.3 ± 0.8
WideResNet-50-2	76.5 ± 2.0	89.8 ± 0.9	74.0 ± 2.0	91.8 ± 0.8	64.9 ± 1.8	89.5 ± 0.9

Component Ablation Under matched 1-shot settings (Table 3), OT-GMA yields the larger baseline gain on VisA (11.3 vs. 6.2 I-AUROC points), whereas TR-TGC yields the larger gain on Real-IAD (6.2 vs. 3.8). Combining them adds 2.9/3.2 points over the stronger component, respectively. The full model exceeds either component across all nine settings and the baseline by 4.5–14.2 points, supporting complementary roles for association and calibration. TR-TGC calibrates nearest-neighbor residuals when used alone, and transport residuals in the full model.

Table 3: Complementary association and calibration: component ablation of I-AUROC (%) on three datasets.

Method	OT-GMA	TR-TGC	MVTec AD			VisA			Real-IAD		
			1-shot	2-shot	4-shot	1-shot	2-shot	4-shot	1-shot	2-shot	4-shot
Baseline	×	×	91.4 ± 0.9	92.8 ± 0.7	94.0 ± 0.6	81.0 ± 1.4	83.4 ± 1.2	85.7 ± 1.0	70.6 ± 1.3	72.7 ± 1.1	75.5 ± 1.0
+ OT-GMA	✓	×	96.4 ± 0.5	97.0 ± 0.4	97.6 ± 0.3	92.3 ± 0.8	93.6 ± 0.7	94.9 ± 0.6	74.4 ± 0.9	76.7 ± 0.8	79.9 ± 0.8
+ TR-TGC	×	✓	94.1 ± 0.7	95.1 ± 0.6	95.9 ± 0.5	87.2 ± 0.9	89.0 ± 0.8	91.1 ± 0.7	76.8 ± 0.8	79.2 ± 0.7	82.7 ± 0.7
Full model	✓	✓	**97.8 ± 0.4**	**98.3 ± 0.5**	**98.5 ± 0.3**	**95.2 ± 0.9**	**95.9 ± 0.7**	**96.6 ± 0.5**	**80.0 ± 0.7**	**82.1 ± 0.6**	**85.1 ± 0.6**

Key Design Analysis In Table 4, removing transport variance or disabling online adaptation lowers VisA I-AUROC by 3.4 or 4.9 points, respectively, supporting patch-specific uncertainty modeling and residual adaptation. Reference-preserving feedback improves I-AUROC over no feedback (95.2 vs. 93.4), whereas full-moment feedback reduces it to 88.6. Recomputing normal moments changes the residual being assessed. This contrast supports refining rejection while retaining first-pass normal moments across passes.

Table 4: Uncertainty and feedback on VisA (1-shot, %).

Variant	I-AUROC	P-AUROC
Full (GOMC-AD)	**95.2 ± 0.9**	**98.4 ± 0.2**
w/o transport variance	91.8 ± 0.7	96.9 ± 0.2
w/o feedback	93.4 ± 0.8	97.7 ± 0.3
Full-moment feedback	88.6 ± 0.8	95.4 ± 0.3
w/o online adaptation	90.3 ± 0.7	96.2 ± 0.2

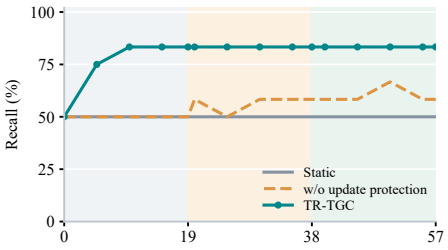


Fig. 6: Fixed-threshold calibration.

TR-TGC Calibration Fig. 6 tests whether guarded adaptation preserves fixed-threshold decisions on a representative VisA pcb4 stream. Across warm-up, anomaly exposure, and recovery (19 steps each; 10 anomalous images during exposure), TR-TGC reaches 83.3% recall, 33.3/25.0 points above static calibration/unprotected updates, with zero observed probe false positive rate (FPR). Forty additional known-normal images serve only to fix the shared threshold. This calibration set is outside the few-shot benchmark protocol. All variants share the stream and 12 normal/12 anomalous probes, evaluated without updates using separate top-50% pooling. All retain 100% probe AUROC, so recall differences reflect operating-point calibration rather than ranking improvement. The unchanged TR-TGC recall during anomaly exposure and recovery further indicates that guarded updates preserve the operating point in this stream, without increasing observed probe false positives.

OT-GMA Localization With identical support and augmentation, OT-GMA raises pixel-level average precision (AP) from 25.5% to 40.5% on the VisA pcb3 missing-component case (Fig. 7), revealing a structural discrepancy amid similar patterns. Heatmaps are independently normalized.

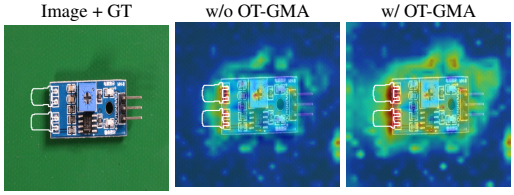


Fig. 7: OT-GMA: missing-component localization. White contours mark the missing-component region.

5 Discussion

Reference coverage and reference utilization play complementary roles. Detection and localization improve on MVTec AD, VisA, and Real-IAD under all three support budgets (Table 1). Compared with the strongest method in each setting, GOMC-AD improves I-AUROC by up to 5.3 percentage points and P-AUROC by up to 1.3 percentage points. Additional examples expand normal support, while joint association coordinates its use and calibration determines how the resulting residuals are interpreted.

OT-GMA makes reference allocation part of the normality criterion: query regions share prescribed normal capacities, so plausible matches are judged jointly. Allocating support to one region limits what remains available to others, making its normal explanation depend on the complete query. This links anomaly evidence to support competition as well as local appearance. Repeated patterns can provide locally credible matches even when the query composition differs from the support. Joint allocation exposes this structural discrepancy, as illustrated by the missing-component example (Fig. 7).

TR-TGC separates association uncertainty from historical drift in both scoring and updates. Correcting historical second moments removes modeled association variance from drift estimation, aligning adaptation with this decomposition. Residual magnitudes are thus interpreted relative to the uncertainty of their normal correspondences. Reference-preserving feedback refines rejection while retaining first-pass normal moments, allowing calibrated evidence to adjust rejection allocation against the same normal reference used to evaluate residuals.

Boundary diffusion limits localization precision. Fig. 8 shows a detected hole whose response extends onto normal surroundings, producing a broader activation than the annotated defect region. Patch-grid resolution and upsampling may contribute to this spatial mismatch: patches near a defect boundary can encode both defective and normal content, while interpolation can spread their responses across the boundary. This suggests a limitation in spatial delineation even when the defect elicits a clear anomaly response. The example motivates boundary-aware refinement that integrates edge cues with multi-scale features, using local image structure to constrain response propagation and finer spatial detail to improve boundary localization while preserving the global association criterion.

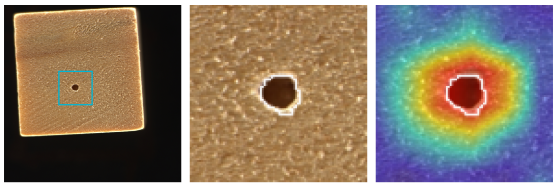


Fig. 8: Boundary diffusion on Real-IAD toy_brick (camera view C1, 1-shot). Left to right: input, enlarged defect, and our anomaly response. The cyan box marks the enlarged region; white contours mark ground truth.

6 Conclusion

Built on frozen pretrained visual representations, GOMC-AD is a training-free framework that couples capacity-constrained global association with transport-conditioned residual calibration for few-shot industrial anomaly detection. A shared variance decomposition aligns residual scoring with drift estimation by subtracting modeled association variance from historical second moments, while reference-preserving feedback refines rejection allocation using first-pass normal moments as a consistent residual reference. Experiments on MVTec AD, VisA, and Real-IAD show detection and localization gains across three support budgets, supporting the complementary roles of shared-capacity association and uncertainty-aware calibration. Extending the framework with boundary-aware refinement offers a natural direction for translating global normality evidence into more precise defect delineation.

AI use statement Generative AI tools were used only for language editing and manuscript polishing. All technical content and experimental results were produced and verified by the authors. The authors take full responsibility for the final manuscript.

Reproducibility Statement. We provide the formulation and implementation details of GOMC-AD in the main paper and appendix, including the association and calibration procedures, hyperparameter settings, dataset configurations, and few-shot evaluation protocols. Additional ablations and sensitivity analyses are provided in the appendix. The source code will be made publicly available upon publication.

REFERENCES

Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. MVTec AD – a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9592–9600, 2019. URL https://openaccess.thecvf.com/content_CVPR_2019/html/Bergmann_MVTec_AD_--_A_Comprehensive_Real-World_Dataset_for_Unsupervised_Anomaly_CVPR_2019_paper.html.

Tri Cao, Jiawen Zhu, and Guansong Pang. Anomaly detection under distribution shift. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6511–6523, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Cao_Anomaly_Detection_Under_Distribution_Shift_ICCV_2023_paper.html.

Hanzhi Chen, Jingbin Que, Kexin Zhu, Zhide Chen, Fei Zhu, Wencheng Yang, Xu Yang, and Xuechao Yang. MALM-CLIP: A generative multi-agent framework for multimodal fusion in few-shot industrial anomaly detection. *Information Fusion*, 127:103765, 2026. URL https://www.sciencedirect.com/science/article/pii/S1566253525008279.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems*, volume 26, pp. 2292–2300, 2013. URL https://proceedings.neurips.cc/paper_files/paper/2013/hash/af21d0c97db2e27e13572cbf59eb343d-Abstract.html.

Simon Damm, Mike Laszkiewicz, Johannes Lederer, and Asja Fischer. AnomalyDINO: Boosting patch-based few-shot anomaly detection with DINOv2. In *Proceedings of the Winter Conference on Applications of Computer Vision*, pp. 1319–1329, 2025. URL https://openaccess.thecvf.com/content/WACV2025/html/Damm_AnomalyDINO_Boosting_Patch-Based_Few-Shot_Anomaly_Detection_with_DINOv2_WACV_2025_paper.html.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: A patch distribution modeling framework for anomaly detection and localization. In *Pattern Recognition. ICPR International Workshops and Challenges*, volume 12664 of *Lecture Notes in Computer Science*, pp. 475–489, 2021.

Zheng Fang, Xiaoyang Wang, Haocheng Li, Jiejie Liu, Qiugui Hu, and Jimin Xiao. FastRecon: Few-shot industrial anomaly detection via fast feature reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17481–17490, 2023. URL https://openaccess.thecvf.com/content/ICCV2023/html/Fang_FastRecon_Few-shot_Industrial_Anomaly_Detection_via_Fast_Feature_Reconstruction_ICCV_2023_paper.html.

Matic Fučka, Vitjan Zavrtanik, and Danijel Skočaj. SALAD – semantics-aware logical anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 21843–21852, October 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/Fucka_SALAD_--_Semantics-Aware_Logical_Anomaly_Detection_ICCV_2025_paper.html.

Bin-Bin Gao. MetaUAS: Universal anomaly segmentation with one-prompt meta-learning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 39812–39836,

2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/463a91da3c832bd28912cd0d1b8d9974-Paper-Conference.pdf.

Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Hao Li, Ming Tang, and Jinqiao Wang. FiLo: Zero-shot anomaly detection by fine-grained description and high-quality localization. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 2041–2049, 2024.

Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, and Jinqiao Wang. UniVAD: A training-free unified model for few-shot visual anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15194–15203, June 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Gu_UniVAD_A_Training-free_Unified_Model_for_Few-shot_Visual_Anomaly_Detection_CVPR_2025_paper.html.

Jianfang He, Min Cao, Silong Peng, and Qiong Xie. RareCLIP: Rarity-aware online zero-shot industrial anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 24478–24487, October 2025. URL https://openaccess.thecvf.com/content/ICCV2025/html/He_RareCLIP_Rarity-aware_Online_Zero-shot_Industrial_Anomaly_Detection_ICCV_2025_paper.html.

Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. WinCLIP: Zero-/few-shot anomaly classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19606–19616, 2023. URL https://openaccess.thecvf.com/content/CVPR2023/html/Jeong_WinCLIP_Zero-Few-Shot_Anomaly_Classification_and_Segmentation_CVPR_2023_paper.html.

Camile Lendering, Erkut Akdag, and Egor Bondarev. SubspaceAD: Training-free few-shot anomaly detection via subspace modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28557–28566, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Lendering_SubspaceAD_Training-Free_Few-Shot_Anomaly_Detection_via_Subspace_Modeling_CVPR_2026_paper.html.

Xiaofan Li, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yuan Xie, and Lizhuang Ma. PromptAD: Learning prompts with only normal samples for few-shot anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16838–16848, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Li_PromptAD_Learning_Prompts_with_only_Normal_Samples_for_Few-Shot_Anomaly_CVPR_2024_paper.html.

Xinyue Liu, Jianyuan Wang, Biao Leng, and Shuo Zhang. Dual-modeling decouple distillation for unsupervised anomaly detection. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 5035–5044, 2024.

Ruiying Lu, Gang Liu, Kang Li, Long Tian, and Junwei Zhang. MaskAD: Parallel masked autoencoder for multi-class unsupervised anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 15457–15465, 2026. URL https://ojs.aaai.org/index.php/AAAI/article/view/38573. Issue 18.

Wei Luo, Yunkang Cao, Haiming Yao, Xiaotian Zhang, Jianan Lou, Yuqi Cheng, Weiming Shen, and Wenyong Yu. Exploring intrinsic normal prototypes within a single image for universal anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9974–9983, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Luo_Exploring_Intrinsic_Normal_Prototypes_within_a_Single_Image_for_Universal_CVPR_2025_paper.html.

Wenxi Lv, Qinliang Su, and Wenchao Xu. One-for-all few-shot anomaly detection via instance-induced prompt learning. In *International Conference on Learning Representations*, volume 2025, pp. 63657–63677, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/9f7f2f57d8eaf44b2f09020f64ff6d96-Paper-Conference.pdf.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mido Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=a68SUt6zFt.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 8748–8763, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14318–14328, 2022. URL https://openaccess.thecvf.com/content/CVPR2022/html/Roth_Towards_Total_Recall_in_Industrial_Anomaly_Detection_CVPR_2022_paper.html.

Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning feature matching with graph neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4938–4947, 2020. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Sarlin_SuperGlue_Learning_Feature_Matching_With_Graph_Neural_Networks_CVPR_2020_paper.html.

Jungwook Seo, Minjeong Kim, Younkwan Lee, Seungho Shin, and Sungyong Baik. Anomaly as non-conformity via training-free graph Laplacian energy minimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21336–21345, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Seo_Anomaly_as_Non-Conformity_via_Training-Free_Graph_Laplacian_Energy_Minimization_CVPR_2026_paper.html.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seung Eun Yi, Michael Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=2NlGyqNjns.

Han Sun, Yunkang Cao, Hao Dong, and Olga Fink. Unseen visual anomaly generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25508–25517, 2025. URL https://ieeexplore.ieee.org/document/11094985/.

Fenfang Tao, Guo-Sen Xie, Fang Zhao, and Xiangbo Shu. Kernel-aware graph prompt learning for few-shot anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 7347–7355, 2025. URL https://ojs.aaai.org/index.php/AAAI/article/view/32790.

Long Tian, Hongyi Zhao, Ruiying Lu, Rongrong Wang, Yujie Wu, Liming Wang, Xiongpeng He, and Xiyang Liu. FOCT: Few-shot industrial anomaly detection with foreground-aware online conditional transport. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 6241–6249, 2024.

Prune Truong, Martin Danelljan, Radu Timofte, and Luc Van Gool. PDC-Net+: Enhanced probabilistic dense correspondence network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(8):10247–10266, 2023.

Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jiangning Zhang, Zhihao Gu, Shuguang Qian, Mingang Chen, and Lizhuang Ma. Real-IAD: A real-world multi-view dataset for benchmarking versatile industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 22883–22892, 2024. URL https://openaccess.thecvf.com/content/CVPR2024/html/Wang_Real-IAD_A_Real-World_Multi-View_Dataset_for_Benchmarking_Versatile_Industrial_Anomaly_CVPR_2024_paper.html.

Qishan Wang, Jia Guo, Shuyong Gao, Haofen Wang, Li Xiong, Junjie Hu, Hanqi Guo, and Wenqiang Zhang. Search is all you need for few-shot anomaly detection, 2025. URL https://arxiv.org/abs/2504.11895. arXiv preprint arXiv:2504.11895.

Jinglin Xu, Yi Li, Chuxiong Sun, Xiao Xu, Jiangmeng Li, and Fanjiang Xu. Multi-modal test-time adaptation via adaptive probabilistic Gaussian calibration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30268–30277, 2026. URL https://openaccess.thecvf.com/content/CVPR2026/html/Xu_Multi-modal_Test-time_Adaptation_via_Adaptive_Probabilistic_Gaussian_Calibration_CVPR_2026_paper.html.

Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In Richard C. Wilson, Edwin R. Hancock, and William A. P. Smith (eds.), *Proceedings of the British Machine Vision Conference*, pp. 87.1–87.12. BMVA Press, 2016. URL https://bmva-archive.org.uk/bmvc/2016/papers/paper087/index.html.

Guangyao Zhai, Yue Zhou, Xinyan Deng, Lars Heckler-Kram, Nassir Navab, and Benjamin Busam. Foundation visual encoders are secretly few-shot anomaly detectors. In *International Conference on Learning Representations*, volume 2026, pp. 97353–97376, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/9e47a0bc530cc88b09b7670d2c130a29-Abstract-Conference.html.

Hui Zhang, Zheng Wang, Dan Zeng, Zuxuan Wu, and Yu-Gang Jiang. DiffusionAD: Norm-guided one-step denoising diffusion for anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(8):7140–7152, 2025. URL https://ieeexplore.ieee.org/document/11005495/.

Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. RealNet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16699–16708, 2024. URL https://ieeexplore.ieee.org/document/10658311/.

Jiawen Zhu and Guansong Pang. Toward generalist anomaly detection via in-context residual learning with few-shot sample prompts. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17826–17836, 2024. URL https://ieeexplore.ieee.org/document/10657566/.

Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. SPot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *Computer Vision – ECCV 2022*, volume 13690 of *Lecture Notes in Computer Science*, pp. 392–408, 2022.

GOMC-AD: Supplementary Material

A Complete Key Design Ablation

Table S1 reports 1-shot design ablations on MVTec AD and VisA. Values are five-run mean ± standard deviation; AUROC denotes area under the receiver operating characteristic curve; I-AUROC and P-AUROC denote its image-level and pixel-level forms.

Table S1: Complete key design ablation under 1-shot support (%).

	MVTec AD		VisA	
Variant	I-AUROC ↑	P-AUROC ↑	I-AUROC ↑	P-AUROC ↑
Full (GOMC-AD)	$\mathbf{97.8 \pm 0.4}$	$\mathbf{97.6 \pm 0.3}$	$\mathbf{95.2 \pm 0.9}$	$\mathbf{98.4 \pm 0.2}$
w/o transport variance	95.8 ± 0.4	96.5 ± 0.2	91.8 ± 0.7	96.9 ± 0.2
w/o feedback	96.8 ± 0.3	97.0 ± 0.2	93.4 ± 0.8	97.7 ± 0.3
Full-moment feedback	93.3 ± 0.5	95.4 ± 0.2	88.6 ± 0.8	95.4 ± 0.3
w/o online adaptation	94.6 ± 0.4	96.0 ± 0.2	90.3 ± 0.7	96.2 ± 0.2

B Hyperparameter Sensitivity

We vary one parameter at a time under 1-shot support on MVTec AD and VisA, keeping the others at the defaults in Table S2. The nine configurations share five support sets and identical test orders; calibration is reset before each run and updated only after prediction.

Table S2: Sensitivity parameter ranges; bold indicates the default.

Parameter	Default	Evaluated values
Rejection mass ρ	0.02	0.01, $\mathbf{0.02}$, 0.05
Optimal transport (OT) entropy ε	0.05	0.03, $\mathbf{0.05}$, 0.08
Transport variance scale κ	1.0	0.5, $\mathbf{1.0}$, 2.0
Feedback strength η	0.03	0.01, $\mathbf{0.03}$, 0.05
Other settings: DINOv3-L/16, 448×448 input, 256×256 evaluation maps; 30 Sinkhorn iterations, 2 retained views, position weight 0.05, momentum 0.99.		

Table S3: 1-shot hyperparameter sensitivity (%, five-run mean ± standard deviation). Bold indicates the default configuration.

	MVTec AD		VisA	
Configuration	I-AUROC ↑	P-AUROC ↑	I-AUROC ↑	P-AUROC ↑
Default	$\mathbf{97.76 \pm 0.42}$	$\mathbf{97.58 \pm 0.28}$	$\mathbf{95.17 \pm 0.86}$	$\mathbf{98.37 \pm 0.22}$
$\rho = 0.01$	97.45 ± 0.37	97.40 ± 0.21	94.60 ± 0.83	98.20 ± 0.24
$\rho = 0.05$	97.57 ± 0.55	97.51 ± 0.29	94.88 ± 0.79	98.29 ± 0.23
$\varepsilon = 0.03$	97.67 ± 0.31	97.62 ± 0.23	95.23 ± 0.77	98.33 ± 0.26
$\varepsilon = 0.08$	97.52 ± 0.28	97.41 ± 0.28	94.84 ± 0.96	98.23 ± 0.25
$\kappa = 0.5$	97.61 ± 0.30	97.46 ± 0.21	94.91 ± 0.97	98.27 ± 0.27
$\kappa = 2.0$	97.66 ± 0.45	97.61 ± 0.27	95.05 ± 0.97	98.41 ± 0.27
$\eta = 0.01$	97.64 ± 0.50	97.52 ± 0.29	94.95 ± 0.83	98.32 ± 0.25
$\eta = 0.05$	97.79 ± 0.43	97.54 ± 0.22	95.21 ± 0.95	98.34 ± 0.25

Across the evaluated ranges, the largest drops from default are 0.31/0.57 I-AUROC points and 0.18/0.17 P-AUROC points on MVTec AD/VisA, respectively.

C Controlled Calibration Protocol

The VisA pcb4 experiment is a controlled fixed-threshold calibration study, separate from the few-shot benchmark. With 2-shot support, 40 additional known-normal images establish a shared threshold. All variants use the same stream and 12 normal/12 anomalous probes; neither calibration images nor probes update residual statistics. This study uses top-50% image pooling. All variants retain 100% probe AUROC; recall differences measure fixed-threshold calibration.

D Selected Category Results

Table S4: Detection and localization on selected categories under 1-shot support (%, mean $\pm$ standard deviation over five random seeds).

Dataset	Category	I-AUROC ↑	P-AUROC ↑
MVTec AD	cable	94.50 ± 1.12	96.04 ± 0.29
	screw	90.00 ± 1.26	98.11 ± 0.14
	pill	98.64 ± 0.21	93.61 ± 0.26
	transistor	95.00 ± 4.19	92.34 ± 3.95
	wood	99.23 ± 0.13	96.35 ± 0.07
VisA	capsules	99.27 ± 0.39	98.19 ± 0.37
	pcb2	90.53 ± 1.96	96.87 ± 0.30
	pcb3	90.56 ± 3.94	98.13 ± 0.44
	pcb1	93.25 ± 2.57	99.22 ± 0.07
	pipe_fryum	99.38 ± 0.82	99.36 ± 0.23
Real-IAD	button_battery	86.96 ± 1.96	96.95 ± 0.15
	phone_battery	84.79 ± 4.23	98.23 ± 0.23
	porcelain_doll	84.11 ± 3.37	98.35 ± 0.14
	u_block	76.56 ± 3.52	98.51 ± 0.30
	wooden_beads	86.38 ± 1.90	98.86 ± 0.19

E Support View Count Ablation

Table S5: Retained support views on VisA under 1-shot support.

Retained views	I-AUROC (%) ↑	P-AUROC (%) ↑	Latency (ms/image) ↓
1	94.6	98.2	104
2 (default)	95.2	98.4	122
4	95.3	98.4	158

F Support View and Patch Counts

For $K = 4$ normal support images, candidate views are counted after support augmentation and before selection. N is the query patch count, and M is the normal support patch count per selected view, excluding the rejection column.

Table S6: Support view and patch counts for $K = 4$.

Support images K	Candidate views	Retained views	Query patches N	Patches per view M
4	32	2	784	784

G TRANSPORT MECHANISM COMPARISON

Table S7 compares FOCT and GOMC-AD. GOMC-AD combines shared-capacity association, explicit rejection, and transport-conditioned residual calibration.

Table S7: Transport mechanism comparison between FOCT and GOMC-AD.

Aspect	FOCT	GOMC-AD
Transport role	Conditional transport (CT) between query features and a foreground-aware memory.	Per-view query-to-support OT for joint normal association and anomaly evidence.
Mass constraints	Directional conditional normalization; aggregate memory usage is adaptive.	Fixed query mass $1/N$, normal capacity $(1-\rho)/M$, and rejection mass ρ.
Anomaly handling	Conditional-transport weighted matching cost; no explicit rejection column.	Explicit rejection allocation, OT cost gap over nearest-neighbor (NN) matching, and calibrated residual evidence.
Test-time updates	Optimize the CT metric; augment and compress episodic query memory.	Frozen support memory; update guarded residual statistics after prediction.
Residual calibration and feedback	CT matching and memory calibration; no transport-variance residual model.	Combine transport variance with historical residual variance; refine rejection costs while retaining first-pass normal moments.

Sources: FOCT (Tian et al., 2024), Sec. 4, Eqs. (5)–(8) and Sec. 4.4.1.

H STATIC AND STREAMING EVALUATION

Table S8 evaluates cross-query adaptation by comparing static and streaming GOMC-AD under matched settings across three datasets and support budgets.

Table S8: Static versus streaming GOMC-AD (%, five-run mean $\pm$ standard deviation). I/P denote image-/pixel-level AUROC.

Shot	Setting	MVTec AD		VisA		Real-IAD	
		I-AUROC	P-AUROC	I-AUROC	P-AUROC	I-AUROC	P-AUROC
1	Static	94.6 ± 0.4	96.0 ± 0.2	90.3 ± 0.7	96.2 ± 0.2	76.13 ± 1.16	95.63 ± 0.15
	Streaming	97.8 ± 0.4	97.6 ± 0.3	95.2 ± 0.9	98.4 ± 0.2	80.0 ± 0.7	96.8 ± 0.4
2	Static	95.81 ± 0.16	96.34 ± 0.13	91.97 ± 0.39	96.87 ± 0.27	79.22 ± 1.01	96.61 ± 0.14
	Streaming	98.3 ± 0.5	97.6 ± 0.2	95.9 ± 0.7	98.6 ± 0.3	82.1 ± 0.6	97.4 ± 0.3
4	Static	96.46 ± 0.37	97.20 ± 0.35	93.70 ± 0.64	97.26 ± 0.17	81.96 ± 0.15	97.38 ± 0.25
	Streaming	98.5 ± 0.3	97.9 ± 0.2	96.6 ± 0.5	98.7 ± 0.1	85.1 ± 0.6	98.2 ± 0.2

The comparison uses five paired support samplings, identical test orders, and common hyperparameters. Both settings retain within-query association, calibration, and feedback. Static freezes all support-initialized cross-query statistics, including normalization scales. Streaming updates these statistics from unlabeled queries only after prediction. State is reset for each category and run.

Streaming exceeds static GOMC-AD in all 18 dataset–shot–metric combinations, with I-AUROC gains of 2.04–4.90 percentage points and P-AUROC gains of 0.70–2.20 points, computed from the reported means.

I KEY TR-TGC PARAMETERS

Table S9 summarizes the key calibration and update settings of the unified configuration. We use negative log-likelihood (NLL), median absolute deviation (MAD), and exponential moving average (EMA).

Table S9: Key TR-TGC parameters and their roles.

Parameter	Value	Role
Residual calibration and stability		
Transport variance scale κ	1.0	Scales patch-specific association variance.
Feedback strength η	0.03	Refines rejection costs; retains first-pass normal moments.
Support anchoring weight	0.05	Mixes updated mean and second moment with support-initialized values.
Variance shrinkage weight	0.10	Shrinks diagonal drift variance toward its mean across dimensions.
Variance floor	10^{-6}	Keeps drift and predictive variances positive.
Relative variance bounds	$[0.25, 4.0]$	Bounds each drift variance relative to its support-initialized value.
Robust scale floor / z-clip	10^{-3} / 4	Floors the scaled MAD; clips normalized evidence to $[-4, 4]$.
Guarded online adaptation		
History momentum	0.99	Discounts past weighted sufficient statistics; also controls the NLL median/MAD EMA.
Guard evidence threshold	2.0	Flags jointly suspicious patches using the rule below.
Guard fraction threshold	0.15	Skips the image update if over 15% of patches are flagged.
Low-risk patch fraction	0.10	Updates from the lowest 10% of composite ranks, with at least two patches.

A patch is flagged when $z^G > 2$ and ($z^{\mathrm{OT}} > 2$ or $z^R > 2$). The composite rank sums normalized ranks of NN distance, OT, Gaussian, and rejection evidence, adding entropy/dispersion/view-disagreement ranks with weights $0.25/0.25/0.50$. Selected residuals retain the evidence weights $w_{t,i}$ defined in the main paper. Updates follow prediction; the static variant freezes the statistics.

J ESSENTIAL TR-TGC UPDATES

The first-pass normal moments remain fixed during scoring and feedback. Historical means and variances are D-dimensional vectors shared across patches. Below, all squares, maxima, and variance bounds act elementwise; $\mathbf{1}$ is the D-dimensional all-ones vector.

Residual scoring For the original query view (and analogously for each augmentation),

$$\mathbf{r}_{t,i} = \mathbf{q}_{t,i} - \boldsymbol{\mu}_{t,i}^{\mathrm{OT}}, \qquad \mathbf{v}_{t,i} = \boldsymbol{\sigma}_{t-1}^2 + \kappa \boldsymbol{\nu}_{t,i}^{\mathrm{OT}},$$
$$E_{t,i}^G = \frac{1}{2D} \sum_{d=1}^{D} \left[\log(2\pi v_{t,i,d}) + \frac{(r_{t,i,d} - \mu_{t-1,d})^2}{v_{t,i,d}} \right]. \tag{S1}$$

Predictive variances are floored at $\delta = 10^{-6}$. Median/scaled-MAD normalization and clipping give z^G; the second pass changes only rejection costs while retaining the first-pass moments in Eq. S1.

Association-corrected accumulation After prediction, let $\mathcal{L}_t$ and $w_{t,i}$ be the accepted patch set and weights from Algorithm S2. Store effective mass n_{t-1}, first-moment sum $\mathbf{u}_{t-1}$, and second-moment sum $\mathbf{h}_{t-1}$. With history momentum $\gamma = 0.99$, the candidate update is

$$\begin{aligned}
\widetilde{n}_t &= \gamma n_{t-1} + \sum_{i \in \mathcal{L}_t} w_{t,i}, \\
\widetilde{\mathbf{u}}_t &= \gamma \mathbf{u}_{t-1} + \sum_{i \in \mathcal{L}_t} w_{t,i} \mathbf{r}_{t,i}, \\
\widetilde{\mathbf{h}}_t &= \gamma \mathbf{h}_{t-1} + \sum_{i \in \mathcal{L}_t} w_{t,i} \left(\mathbf{r}_{t,i}^{\odot 2} - \kappa \boldsymbol{\nu}_{t,i}^{\mathrm{OT}} \right).
\end{aligned} \tag{S2}$$

The subtraction removes the modeled association contribution before estimating drift. Updates with fewer than two accepted residuals or nonpositive total weight are skipped.

Support anchoring and variance stabilization Let $(\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0^2)$ be the support-initialized statistics, $\xi = 0.05$ the anchoring weight, and $\chi = 0.10$ the shrinkage weight. Candidate moments are stabilized and mixed with support moments:

$$\begin{aligned}
\widehat{\boldsymbol{\mu}}_t &= \widetilde{\mathbf{u}}_t / \widetilde{n}_t, \\
\widehat{\mathbf{m}}_t &= \max\left(\widetilde{\mathbf{h}}_t / \widetilde{n}_t, \widehat{\boldsymbol{\mu}}_t^{\odot 2} + \delta \mathbf{1} \right), \\
\boldsymbol{\mu}_t &= (1 - \xi) \widehat{\boldsymbol{\mu}}_t + \xi \boldsymbol{\mu}_0, \\
\mathbf{m}_t &= (1 - \xi) \widehat{\mathbf{m}}_t + \xi (\boldsymbol{\sigma}_0^2 + \boldsymbol{\mu}_0^{\odot 2}), \\
\mathbf{s}_t &= \max(\mathbf{m}_t - \boldsymbol{\mu}_t^{\odot 2}, \delta \mathbf{1}), \\
\boldsymbol{\sigma}_t^2 &= \mathrm{clip}_{[0.25\boldsymbol{\sigma}_0^2,\, 4\boldsymbol{\sigma}_0^2]} \left((1 - \chi) \mathbf{s}_t + \chi \bar{s}_t \mathbf{1} \right),
\end{aligned} \tag{S3}$$

where $\mathbf{m}_t$ is the anchored second moment and $\bar{s}_t = D^{-1} \sum_d s_{t,d}$. A final floor of δ is applied to $\boldsymbol{\sigma}_t^2$. The implementation stores the stabilized moments for the next query:

$$n_t = \widetilde{n}_t, \qquad \mathbf{u}_t = n_t \boldsymbol{\mu}_t, \qquad \mathbf{h}_t = n_t (\boldsymbol{\sigma}_t^2 + \boldsymbol{\mu}_t^{\odot 2}). \tag{S4}$$

Initialization and normalization Support-only leave-one-view-out queries initialize the mean and corrected second moment using the same variance subtraction; the initial variance is floored and shrunk across channels. When feedback is enabled, support statistics are re-estimated with the two-pass procedure. The initial state uses the support residual count n_0 and the identities in Eq. S4. State is reset for each category and run. After an accepted update, NLLs of the selected residuals are recomputed with the updated Gaussian. Their median and scaled MAD update the Gaussian normalization by EMA with momentum γ; the MAD has floor 10^{-3}. Support-derived OT-gap and rejection normalization remain fixed.

K CORE ALGORITHMS

Algorithms S1 and S2 summarize association and calibration. Query and support patches are D-dimensional vectors; i, j, and v index query patches, support patches, and support views. All vector squares and variances are elementwise. Parameters $\rho, \varepsilon, \beta, \tau$ control rejection mass, entropy, spatial cost, and view weighting, respectively. The normalized gap, Gaussian NLL, and rejection evidence are denoted by $z^{\mathrm{OT}}, z^{G}, z^{R}$.

Algorithm S1. OT-GMA: per-view capacity-constrained association

Input: Query patches $\mathbf{Q}_t$, support views $\{\mathbf{S}^v\}$; rejection costs $c_{t,i,\perp}$ (zero on the first pass); retained-view count, $\rho, \varepsilon, \beta, \tau$.

Output: NN distance d^{NN}, transport cost d^{OT}, gap g^{OT}, normal moments $(\boldsymbol{\mu}^{\mathrm{OT}}, \boldsymbol{\nu}^{\mathrm{OT}})$, and rejection allocation a.

Normalize patches; select $\mathcal{V}_t$ by similarity of normalized global descriptors.
Compute $d^{\mathrm{NN}}_{t,i} = \min_{v,j}(1 - \mathbf{q}_{t,i}^{\top}\mathbf{s}_j^v)$ over all support views.
for each $v \in \mathcal{V}_t$ **do**
  Set $D^v_{i,j} = 1 - \mathbf{q}_{t,i}^{\top}\mathbf{s}_j^v$ and $C^v_{i,j} = D^v_{i,j} + \beta\|\mathbf{p}_i - \mathbf{p}_j\|_2^2$, with $\mathbf{p} \in [0,1]^2$.
  Set $a_i = 1/N$, $b_j = (1-\rho)/M$, and $b_\perp = \rho$ for N query and M support patches.
  Solve the entropic plan $\boldsymbol{\Pi}^v$ for $[\mathbf{C}^v, \mathbf{c}_{t,\perp}]$ with marginals $\mathbf{a}, \mathbf{b}$ using log-domain Sinkhorn.
  Condition on normal columns: $P^v_{i,j} = \Pi^v_{i,j} / \sum_{j=1}^{M} \Pi^v_{i,j}$; set $a^v_{t,i} = \Pi^v_{i,\perp} / \sum_{j\in\{1,\ldots,M,\perp\}} \Pi^v_{i,j}$.
  Compute $d^v_{t,i} = \sum_j P^v_{i,j} D^v_{i,j}$, $\boldsymbol{\mu}^v_{t,i} = \sum_j P^v_{i,j}\mathbf{s}^v_j$, $\mathbf{m}^v_{t,i} = \sum_j P^v_{i,j}(\mathbf{s}^v_j)^{\odot 2}$.
  Compute the normal-mass-weighted feature cost $\bar{d}^v_t = \sum_i \sum_{j=1}^{M} \Pi^v_{i,j} D^v_{i,j} / \sum_i \sum_{j=1}^{M} \Pi^v_{i,j}$.
end for; set $\omega^v_t = \exp(-\bar{d}^v_t/\tau) / \sum_{u\in\mathcal{V}_t} \exp(-\bar{d}^u_t/\tau)$.
Aggregate $(d^{\mathrm{OT}}_{t,i}, \boldsymbol{\mu}^{\mathrm{OT}}_{t,i}) = \sum_v \omega^v_t (d^v_{t,i}, \boldsymbol{\mu}^v_{t,i})$.
Set $\boldsymbol{\nu}^{\mathrm{OT}}_{t,i} = \sum_v \omega^v_t \mathbf{m}^v_{t,i} - (\boldsymbol{\mu}^{\mathrm{OT}}_{t,i})^{\odot 2}$, with a numerical floor of zero.
Return $a_{t,i} = \sum_v \omega^v_t a^v_{t,i}$, $g^{\mathrm{OT}}_{t,i} = d^{\mathrm{OT}}_{t,i} - d^{\mathrm{NN}}_{t,i}$, and the distances and moments.

Algorithm S2. TR-TGC: reference-preserving feedback and causal updates

Input: Frozen encoder and support memory; query stream; support-initialized statistics and settings in Table S9.

Output: Image/pixel anomaly scores and updated residual statistics.

for each query image t **do**
  for each query augmentation **do**
    Run Algorithm S1 with $c_{t,i,\perp} = 0$; retain all first-pass outputs.
    Compute residuals $\mathbf{r}_{t,i}$ and NLLs $E^G_{t,i}$ using history at $t-1$; normalize to $z^G_{t,i}$ (Sec. J).
    Set $c^{(2)}_{t,i,\perp} = -\eta \tanh([z^G_{t,i}]_+)$, where $[x]_+ = \max(x, 0)$.
    Rerun Algorithm S1 with these costs; retain only its rejection allocation a^*.
    Fuse first-pass distances and calibrated evidence with a^* for patch scoring.
  end for; align and combine query-view outputs for image/pixel predictions.
  Fix the current predictions before any historical update.
  if online adaptation is enabled and the original query passes the image guard **then**
    Select low-risk original-view patches $\mathcal{L}_t$ using evidence and view-consistency ranks.
    Set $w_{t,i} = \exp(-[z^{\mathrm{OT}}_{t,i}]_+ - [z^G_{t,i}]_+ - [z^R_{t,i}]_+)$ for $i \in \mathcal{L}_t$.
    Update residual moments and NLL normalization as in Sec. J.
  else retain the historical state.
  end if
end for