PARC: JOINT MODELING OF PATCH APPEARANCE AND REGIONAL COMPOSITION FOR TRAINING-FREE STRUCTURAL AND LOGICAL ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

In real-world industrial anomaly detection, production lines may contain products exhibiting both structural and logical anomalies. While conventional anomaly detection methods perform well in detecting structural anomalies, they often struggle with logical anomalies. Recent methods alleviate this limitation but often require category-specific training or auxiliary modules, which can limit generalizability and introduce substantial computational overhead. To overcome these limitations, we propose **PARC**, a training-free framework that jointly models **P**atch **A**ppearance and **R**egional **C**omposition using shared normal references. For appearance assessment, PARC evaluates local deviations by matching patch features against normal references. For composition assessment, it groups normal features into recurring patterns, summarizes each image by the proportions of these patterns within each spatial region, and scores how unusual these regional proportions are relative to normal images. Because structural and logical anomalies may occur independently, PARC calibrates the two scores using normal score distributions and fuses them for effective detection. Without requiring auxiliary modules or category-specific training, PARC achieves a combined image-level AUROC of 91.12% on MVTec-LOCO-AD using frozen features and raises the mean AUROC using five different backbones from 72.69% to 84.74%.

1 INTRODUCTION

In real-world industrial anomaly detection (Liu et al., 2024; Li et al., 2025; Wang et al., 2024), inspection systems need to identify two distinct types of defects: structural anomalies (Bergmann et al., 2019; Zou et al., 2022) and logical anomalies (Bergmann et al., 2022). Structural anomalies typically appear as local appearance defects, such as scratches or cracks, that alter the visual texture of product surfaces. In contrast, logical anomalies mainly arise when the overall composition of a product violates expected constraints, even if all individual components appear intact. For example, a package may be missing a required item, contain an incorrect quantity, or have components placed in incorrect positions. Consequently, reliable industrial inspection requires assessing not only the local appearance of individual parts but also the overall composition of the product.

Because anomalous products are scarce, only normal images of the target category are available for building the detector in many practical settings (Bergmann et al., 2020). A common approach to exploiting these normal examples is patch matching. Methods such as PatchCore (Roth et al., 2022) utilize pretrained features to compare each image patch with normal reference patches (Bae et al., 2023). Large matching distances provide evidence of structural anomalies. However, when an item is missing, the remaining items and the exposed background may still match normal examples. Thus, finding a normal match for every visible patch does not necessarily indicate that the overall product composition is normal.

To assess product composition, recent methods (Zhang et al., 2026; Hsieh & Lai, 2024; Zhang et al., 2025) often introduce additional mechanisms beyond local patch matching. Some methods (You et al., 2022; Guo et al., 2025; Zhang et al., 2026) rely on category-specific network training to learn compositional constraints, which can limit generalizability because the network must be retrained for each new product category. Other methods (Jeong et al., 2023; Hsieh & Lai, 2024; Zhang et al.,

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

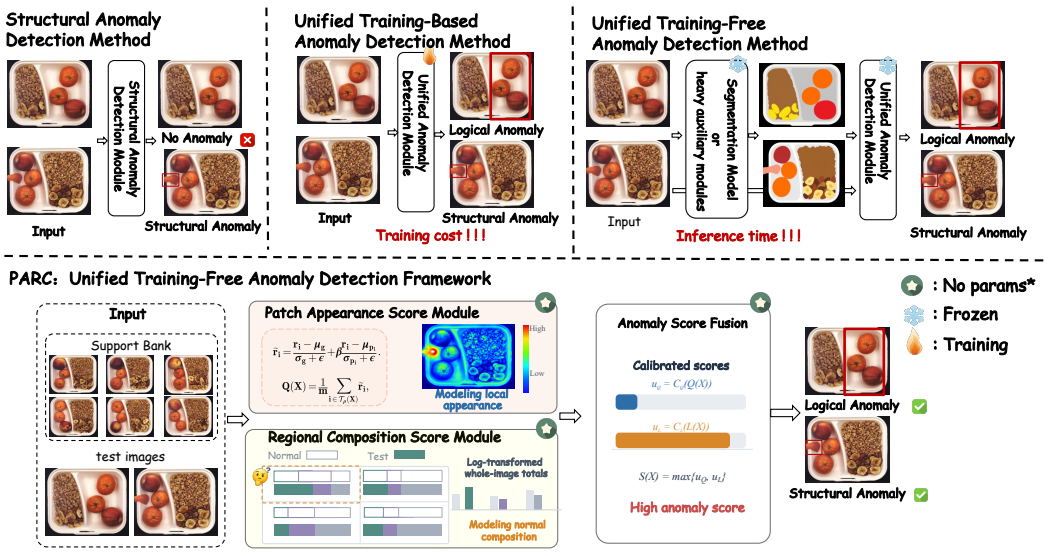


Figure 1: Motivation and overview of PARC. Top: representative detection approaches. Bottom: PARC combines patch appearance and regional composition scores through normal-data calibration and fusion. "No params*" denotes no trainable neural-network parameters in the marked modules; normal-data statistics are still fitted.

2025) rely on auxiliary modules, *e.g.,* component segmentors, which introduce substantial computational overhead. Therefore, our goal is to capture compositional information without category-specific network training or heavy auxiliary modules.

A key observation is that logical anomalies can alter the amount and spatial distribution of recurring visual patterns even when every visible patch still finds a similar normal match. For example, moving an intact component from one region to another changes regional proportions of visual patterns without necessarily producing large matching distances. Motivated by this observation, we represent each image not only by how well its patches match normal references but also by how recurring patterns are distributed across spatial regions. Therefore, we propose **PARC**, a *training-free* framework that realizes this idea through two complementary scores. The *patch appearance score* measures how well patches match normal references. The *regional composition score* describes each image using the proportions of recurring patterns within fixed spatial regions and jointly models how these proportions vary together across normal images. Because a combination of proportions can be unusual even when each individual proportion appears normal, the regional composition score captures evidence that the patch appearance alone may miss. Figure 1 provides an overview of PARC and illustrates how its two complementary branches assess patch appearance and regional composition using shared normal references.

Structural and logical anomalies can occur independently, so the two scores provide complementary evidence. However, because they measure different quantities, their raw values are not directly comparable. We therefore calibrate each score against its distribution over normal images to place both responses on a common scale, and use the larger calibrated response as the final anomaly score. Either source of evidence can thus trigger detection without requiring the anomaly type to be specified in advance. Both scores are computed from the same fixed feature extractor and normal reference images, without category-specific network training or heavy auxiliary modules.

We evaluate PARC using fixed features from five existing detectors (Damm et al., 2025; Roth et al., 2022; Lendering et al., 2026; Fučka et al., 2026a; Cohen & Hoshen, 2021). On MVTec-LOCO-AD (Bergmann et al., 2022), PARC improves the combined image-level AUROC from all five backbones compared with their original scoring procedures, increasing the mean from 72.69% to 84.74%. This metric jointly evaluates structural and logical anomalies against normal images. Using the AnomalyVFM (Fučka et al., 2026a) backbone, PARC achieves an AUROC of 91.12%. We further evaluate PARC on MVTec AD (Bergmann et al., 2019), VisA (Zou et al., 2022), and

CAD-SD (Ishida et al., 2023), and conduct ablation studies to examine the contributions of regional statistics and joint modeling.

Our contributions are summarized as follows:

- We introduce a composition representation based on regional proportions of recurring normal patterns, providing anomaly evidence beyond patch matching without requiring component segmentation.
- We propose PARC, a training-free framework that calibrates and combines patch appearance and regional composition scores to detect both structural and logical anomalies without category-specific network training.
- We demonstrate consistent improvements over the original scoring procedures across five feature sources on MVTec-LOCO-AD, with ablation studies validating the contributions of the proposed components.

2 Related Work

Feature-based Industrial Anomaly Detection Feature-based detectors use pretrained representations to characterize normal appearance. Reference-matching approaches compare test patches with features from normal images (Cohen & Hoshen, 2021; Roth et al., 2022). PatchCore, for example, stores representative normal patches in a memory bank and uses nearest-neighbor distances as anomaly evidence. Other methods describe normal features through position-specific Gaussian distributions or low-dimensional subspaces (Defard et al., 2021; Lendering et al., 2026). Recent work also explores stronger representations, using frozen DINOv2 features or adapting foundation models to anomaly detection (Damm et al., 2025; Fučka et al., 2026a; Oquab et al., 2024). PARC uses normal-reference matching to measure local appearance deviations. However, small local distances do not establish whether normal-looking elements occur in the expected quantities and locations, motivating the use of composition information.

Logical Anomaly Detection Logical anomalies violate constraints on the presence, quantity, or arrangement of otherwise normal elements (Bergmann et al., 2022). To capture these constraints, learned models incorporate dependencies across image regions (Yao et al., 2023; Zhang et al., 2024; Batzner et al., 2024; SAKAI et al., 2026). Other approaches compare extracted objects and their relations (Peng et al., 2025; Fučka et al., 2026b; Gu et al., 2025) or assess composition through language (Jin et al., 2025; Kwon et al., 2025; Zhang et al., 2025). Component-based methods also describe composition through statistical summaries (Kim et al., 2024) or learned models of component maps (Fučka et al., 2025; Zhang et al., 2026). Among these, ComAD obtains component maps by clustering normal features and measures component areas, colors, and counts (Liu et al., 2023). CSAD trains a component segmenter using pseudo-labels and concatenates component histograms from spatial grid cells (Hsieh & Lai, 2024). PARC instead describes composition through regional proportions of recurring patterns, directly aggregating feature similarities without extracting component masks.

Feature Distributions for Anomaly Detection Feature distributions offer a way to describe composition without component extraction. Globally pooled feature statistics provide image-level evidence complementary to local detection (Sugawara & Imamura, 2024; Li et al., 2026a). SINBAD represents feature sets using random-projection histograms (Cohen et al., 2024). Its image detector scores crops at multiple locations and scales, then averages their anomaly scores. PARC instead uses soft assignments to normal feature clusters and jointly models regional pattern proportions and whole-image statistics, retaining regional information before scoring. Normal-data calibration is also used to combine local and composition evidence (Kim et al., 2024; Sugawara & Imamura, 2024; Li et al., 2026b). LogSAD uses normal-validation calibration followed by maximum fusion (Zhang et al., 2025). PARC calibrates its patch appearance and regional composition scores using out-of-fold normal scores and takes their maximum, without target-category network training.

3 Method

PARC describes an image using recurring visual patterns, namely feature patterns that repeatedly appear in normal products: matching distances describe how well normal references explain local

appearance, while spatial pattern distributions describe the composition of the image. We derive a patch appearance score Q and a regional composition score L using these normal references, then calibrate and combine their responses using score distributions obtained from normal images.

The framework requires local features with spatial locations, but does not depend on a particular feature extractor. For each product category, we construct a support bank and fit normal statistics while keeping the feature extractor fixed. No network training or fine-tuning on the target category is required.

3.1 Normal References and Matching

For each product category, we partition the normal training images $\mathcal{D}$ into a support set $\mathcal{D}_s$ and a disjoint estimation set $\mathcal{D}_n$. The former supplies features for matching and clustering; the latter is used to estimate score statistics and normal score distributions. Keeping estimation images outside the support bank prevents their features from matching themselves when estimating normal distances.

Given an image X, a fixed feature extractor ϕ produces $N = HW$ local features $\{\mathbf{x}_i\}_{i=1}^{N}$ on an $H \times W$ grid. Let $\mathbf{p}_i \in [0,1]^2$ denote the normalized spatial coordinates of feature i. Images processed by the same extractor use a consistent feature grid. We collect all local features from $\mathcal{D}_s$ into a support bank $\mathcal{B} = \{\mathbf{s}_j\}_{j=1}^{M}$. Each query feature is matched to its nearest normal reference:

$$j_i^* = \arg\min_j d(\mathbf{x}_i, \mathbf{s}_j), \qquad r_i = d(\mathbf{x}_i, \mathbf{s}_{j_i^*}), \tag{1}$$

where d is cosine distance. The search spans the entire bank without restricting the reference location.

A small matching distance indicates similar local appearance but does not establish normal composition: missing or misplaced components may still find similar normal references while altering the proportions and locations of recurring patterns.

3.2 Patch Appearance Score

The patch appearance score measures how well normal references explain local appearance. The matching distance can be typical at one position and unusual at another. We use globally standardized matching distances as the base response and add a position-specific term weighted by β:

$$\widetilde{r}_i = \frac{r_i - \mu_g}{\sigma_g + \epsilon} + \beta \frac{r_i - \mu_{\mathbf{p}_i}}{\sigma_{\mathbf{p}_i} + \epsilon}. \tag{2}$$

The global mean μ_g and standard deviation σ_g are estimated over all query locations in normal estimation images. The position-specific statistics $\mu_{\mathbf{p}_i}$ and $\sigma_{\mathbf{p}_i}$ are estimated across those images at the same feature-grid location. The global term provides a common distance scale, while the position-specific term measures deviations relative to normal variation at that location. The constant ϵ ensures numerical stability.

Since a local defect may affect only a small part of an image, averaging all responses can dilute its signal. We retain a fraction $\rho \in (0,1]$ of the strongest responses and average the largest $m = \max(1, \lceil \rho N \rceil)$ values:

$$Q(X) = \frac{1}{m} \sum_{i \in \mathcal{T}_\rho(X)} \widetilde{r}_i, \tag{3}$$

where $\mathcal{T}_\rho(X)$ is the index set of the m largest standardized responses. A larger $Q(X)$ indicates local regions that are less well explained by normal references.

3.3 Regional Composition Score

The patch appearance score measures appearance differences at individual locations. Composition also depends on which recurring patterns occur and where they occur. We describe this information through regional pattern proportions and compare them with those of normal images.

Grouping normal features into recurring patterns. Comparing these proportions across images requires a common set of patterns. The support bank can contain many features with similar appearances, which should contribute to the same pattern (Sivic & Zisserman, 2003). We group the support features into K clusters using Euclidean K-means. Each center $\mathbf{c}_k$ is the mean feature of its group and represents a recurring pattern. The groups require no semantic part annotations.

A query feature may resemble several patterns, especially when it spans a component boundary. Soft assignment lets it contribute to these patterns according to its similarity to their centers. We use cosine similarity, as in local matching, and normalize the contributions into weights:

$$a_{ik}(X) = \frac{\exp(\cos(\mathbf{x}_i, \mathbf{c}_k)/\tau)}{\sum_{\ell=1}^{K} \exp(\cos(\mathbf{x}_i, \mathbf{c}_\ell)/\tau)}. \tag{4}$$

Here, $a_{ik}(X)$ is feature i's contribution to pattern k, with $\sum_k a_{ik}(X) = 1$. The temperature $\tau > 0$ controls weight concentration: smaller values favor the most similar centers.

Regional pattern distributions. Image-wide pattern proportions describe composition but discard location (Lazebnik et al., 2006). For example, moving an intact component between regions can preserve image-wide proportions while changing where its recurring patterns occur.

We partition normalized image coordinates into a fixed $R \times R$ grid, with the same boundaries for every image. For region r and recurring pattern k, we average the assignment weights to measure the proportion of the region associated with k:

$$h_{r,k}(X) = \frac{1}{N_r} \sum_{i:\mathbf{p}_i \in r} a_{ik}(X), \tag{5}$$

where N_r is the number of query features in region r. Each regional histogram sums to one, so regions are described on a common proportional scale.

The regional histograms retain location information. We also represent whole-image pattern totals on a logarithmic scale. On this scale, the same absolute change has less effect when the pattern's total is larger. Specifically, we sum the weights over the image and define $g_k(X) = \log(1 + \sum_{i=1}^{N} a_{ik}(X))$. The sum inside the logarithm also equals $\sum_{r=1}^{R^2} N_r h_{r,k}(X)$. Concatenating the regional histograms and these K global values in a fixed order gives $\mathbf{h}(X) \in \mathbb{R}^{(R^2+1)K}$, which describes regional pattern proportions together with log-transformed whole-image totals.

Modeling normal composition. Pattern proportions vary even among normal images, and some proportions vary together. We estimate the descriptor mean $\boldsymbol{\mu}_h$ and covariance $\boldsymbol{\Sigma}_h$ from normal estimation images, using Ledoit–Wolf shrinkage (Ledoit & Wolf, 2004) to stabilize covariance estimation. The regional composition score is

$$L(X) = \big(\mathbf{h}(X) - \boldsymbol{\mu}_h\big)^\top \boldsymbol{\Sigma}_h^{-1} \big(\mathbf{h}(X) - \boldsymbol{\mu}_h\big). \tag{6}$$

This squared Mahalanobis distance measures composition deviation relative to normal variation, revealing unusual combinations of pattern proportions even when each proportion is plausible.

3.4 Normal Score Calibration and Fusion

The patch appearance and regional composition scores measure different quantities and have different numerical scales. To compare their responses, we express each score relative to scores obtained from normal images.

Calibration should reflect how the detector scores a new normal image. We use cross-fitting on $\mathcal{D}_n$ to obtain a score for every estimation image using statistics fitted on other images. For each fold, we fit the patch appearance statistics and composition model on the remaining folds, then score the held-out images. We pool the held-out scores to form the normal reference distributions. The support bank and cluster centers remain fixed throughout.

For score $A \in \{Q, L\}$, let $\{z_m^A\}_{m=1}^n$ denote its normal out-of-fold scores, where $n = |\mathcal{D}_n|$. We express a score t by its smoothed rank among these references:

$$\widehat{F}_A(t) = \frac{0.5 + \sum_{m=1}^{n} \mathbf{1}[z_m^A \le t]}{n+1}. \tag{7}$$

Algorithm 1: PARC: patch appearance and regional composition scores from Normal References

Input: Normal images $\mathcal{D}$; frozen feature extractor ϕ; test image X.
Output: Anomaly score $S(X)$.

Normal references
  Construct support bank $\mathcal{B}$ and cluster centers $\{\mathbf{c}_k\}$
  Estimate score statistics and calibration mappings $\mathcal{C}_Q, \mathcal{C}_L$ from disjoint estimation images
Normal matching
  Extract local features $\{\mathbf{x}_i\} \leftarrow \phi(X)$
  Match each feature to a normal reference: $j_i^* \leftarrow \arg\min_j d(\mathbf{x}_i, \mathbf{s}_j)$
Appearance: how well do patches match?
  Read matching distances $r_i \leftarrow d(\mathbf{x}_i, \mathbf{s}_{j_i^*})$
  Compute $Q(X)$ from the largest standardized distances (Eq. 3)
Composition: how are normal patterns distributed?
  Compute soft weights $a_{ik}(X)$ from similarities to cluster centers (Eq. 4)
  Form $\mathbf{h}(X)$ from regional mean weights and log-transformed global sums; compute its deviation $L(X)$ from normal composition (Eq. 6)
Combine complementary evidence
  $u_Q \leftarrow \mathcal{C}_Q(Q(X)), \quad u_L \leftarrow \mathcal{C}_L(L(X))$
  return $S(X) = \max\{u_Q, u_L\}$

Here, $\mathbf{1}[\cdot]$ equals one when the condition holds and zero otherwise. Thus, the sum counts normal scores no greater than t, and smoothing keeps the rank strictly between zero and one.

Scores above the largest normal reference have the same rank. To distinguish these responses, we extend the mapping according to how far they exceed that maximum:

$$\mathcal{C}_A(t) = \mathrm{clip}_{[\delta,1-\delta]}\left(1 - \frac{1 - \widehat{F}_A(t)}{1 + \max(0, t - b_A)/v_A}\right), \tag{8}$$

where $b_A = \max_m z_m^A$ is the largest normal reference score and $v_A = \max(b_A - \min_m z_m^A, \epsilon)$ is the normal score range, floored at ϵ. Before clipping, the mapping equals $\widehat{F}_A(t)$ for $t \leq b_A$ and increases toward one for $t > b_A$, with the excess scaled by v_A. The operator $\mathrm{clip}_{[\delta,1-\delta]}$ restricts the output to $[\delta, 1-\delta]$.

We then refit the patch appearance statistics and composition model on all of $\mathcal{D}_n$ for test-time scoring. The calibrated scores are $u_Q(X) = \mathcal{C}_Q(Q(X))$ and $u_L(X) = \mathcal{C}_L(L(X))$.

Local appearance defects and changes in the composition or arrangement of normal elements need not occur together. We therefore retain the stronger calibrated response:

$$S(X) = \max\{u_Q(X), u_L(X)\}. \tag{9}$$

This rule allows either score to contribute a high anomaly score without specifying the anomaly type in advance.

Cross-fitting is used only to construct the normal score references. At inference, each image undergoes a single feature extraction, scored against the fixed support bank and cluster centers using statistics fitted on the full estimation set.

4 Experiments

We examine whether the regional composition of recurring patterns reveals anomalies that receive low local matching scores. After that, we compare PARC with existing methods, evaluate it across feature sources, and analyze how the composition readout complements local matching.

Datasets and metrics. We primarily evaluate on MVTec-LOCO-AD (Bergmann et al., 2022), which distinguishes structural anomalies, such as appearance defects, from logical anomalies, such

Table 1: Detection performance and method requirements on MVTec-LOCO-AD. ✓/× indicate whether a method is training-free or does not require heavy auxiliary modules. Inference time is the end-to-end latency for one image, reported in ms/img.

	Method requirements			AUROC (%)		
Method	Training free	Without heavy auxiliary modules	Inference time (ms/img)	Combined	Structural	Logical
LogiCo	×	×	2336.65	95.37	95.46	95.50
CSAD	×	✓	96.28	95.85	93.69	97.67
LogSAD	✓	×	2138.66	90.24	91.71	89.37
ComAD	✓	×	168.39	82.01	74.55	88.00
PatchCore	✓	✓	91.34	73.41	79.29	69.53
AnomalyDINO	✓	✓	110.57	78.73	87.05	72.04
SubspaceAD	✓	✓	127.34	79.42	90.71	70.94
PARC	✓	✓	114.35	91.12	93.59	89.42

LogiCo and CSAD achieve strong detection performance but require a separately trained model for each object category, limiting their applicability across categories with a single model.

as incorrect quantities or arrangements of otherwise normal elements. We additionally evaluate on MVTec AD (Bergmann et al., 2019), VisA (Zou et al., 2022), and CAD-SD (Ishida et al., 2023). We report category-averaged image-level AUROC (%), which measures how well a score ranks anomalous images above normal images; higher is better. On LOCO, Combined evaluates all anomalies against normal images, whereas Structural and Logical evaluate each anomaly type separately against normal images. PARC results are averaged over three seeds; $\pm$ denotes the sample standard deviation across seeds.

Features and normal data. Unless otherwise specified, we use AnomalyVFM features (Fučka et al., 2026a). For each category, normal training images are split evenly into a support bank for patch matching and a reference set providing matching-distance, pattern-proportion, and score-distribution statistics, obtained via five-fold cross-fitting with statistics fitted on the full pool. We use $K = 80$ pattern groups, softmax membership at $\tau = 0.30$, and a composition descriptor concatenating 2×2 regional membership proportions with each pattern's image-wide log total membership. The patch appearance score uses $\beta = 0.25$, averaging the largest 1% of standardized patch distances. Both scores are mapped with normal-reference rank calibration and fused by $\max(u_Q, u_L)$. PARC requires no target-category network training or fine-tuning, but constructs a support bank and fits normal-reference statistics; implementation and comparison settings appear in Appendix B.

4.1 Main results

Comparison with existing methods. Table 1 compares PARC with seven representative methods, including PatchCore (Roth et al., 2022), AnomalyDINO (Damm et al., 2025), SubspaceAD (Lendering et al., 2026), ComAD (Liu et al., 2023), LogSAD (Zhang et al., 2025), CSAD (Hsieh & Lai, 2024), and LogiCo (Zhang et al., 2026).

PARC achieves 91.12 Combined AUROC, with 93.59 Structural and 89.42 Logical AUROC, using only frozen features and normal-reference statistics. This is higher than LogSAD's 90.24, the strongest training-free baseline, while running nearly 20× faster , and comes within 4–5 points of CSAD and LogiCo, which require per-category network training .

Evaluation across feature sources. To examine whether PARC's gains extend beyond AnomalyVFM, we apply the readout to features extracted by four existing detectors—AnomalyDINO (Damm et al., 2025), PatchCore (Roth et al., 2022), SPADE (Cohen & Hoshen, 2021), and SubspaceAD (Lendering et al., 2026)—and compare against their original scoring procedures (Table 2). Each original scorer uses the full training set, which PARC divides into support and reference images. PARC improves Combined, Structural, and Logical AUROC for all five sources, raising averages from 72.69 to 84.74, from 79.98 to 86.07, and from 67.36 to 84.38.

Table 2: Evaluation across feature sources on MVTec-LOCO-AD (AUROC, %). Original: each method's original scoring procedure; +PARC: default PARC readout applied to its features (AnomalyVFM is our default source), averaged over three seeds. Bold marks the higher mean per pair.

	Combined		Structural		Logical	
Method / feature source	Original	+PARC	Original	+PARC	Original	+PARC
AnomalyDINO	78.73	**84.45**	87.05	**87.15**	72.04	**82.96**
PatchCore	73.41	**80.22**	79.29	**79.33**	69.53	**81.75**
SPADE	68.68	**80.23**	64.45	**79.43**	72.60	**81.68**
SubspaceAD	79.42	**87.68**	90.71	**90.87**	70.94	**86.06**
AnomalyVFM (default)	63.22	**91.12**	78.38	**93.59**	51.71	**89.42**
Mean	72.69	**84.74**	79.98	**86.07**	67.36	**84.38**

Table 3: AUROC (%) on additional datasets with AnomalyVFM features. PARC values are means ± standard deviations over three seeds.

Dataset	Original	+PARC
MVTec AD	96.44	$\mathbf{99.41}_{\pm 0.05}$
VisA	92.88	$\mathbf{96.33}_{\pm 0.10}$
CAD-SD	55.10	$\mathbf{99.81}_{\pm 0.04}$

Table 4: AUROC (%) on MVTec-LOCO-AD for the patch-appearance and regional-composition scores. Each score uses its normal-reference rank mapping; fusion takes their maximum.

Score	Combined	Structural	Logical
u_Q	$81.93_{\pm 0.10}$	$92.60_{\pm 0.13}$	$73.58_{\pm 0.24}$
u_L	$87.42_{\pm 0.36}$	$83.82_{\pm 0.65}$	$\mathbf{90.88}_{\pm 0.10}$
$\max(u_Q, u_L)$	$\mathbf{91.12}_{\pm 0.17}$	$\mathbf{93.59}_{\pm 0.15}$	$89.42_{\pm 0.44}$

Evaluation on additional datasets. Beyond LOCO, we use AnomalyVFM features and keep the PARC configuration unchanged on MVTec AD, VisA, and CAD-SD. PARC achieves 99.41, 96.33, and 99.81 AUROC, respectively, improving over the original scorer on all three datasets (Table 3).

4.2 Analysis of PARC

Complementarity of the two scores. To test whether composition adds evidence beyond local matching, we evaluate each score separately and their fusion (Table 4). Among the individual scores, the patch appearance score performs better on structural anomalies, while the regional composition score performs better on logical anomalies. Fusion reaches 91.12 Combined AUROC, exceeding the patch appearance and regional composition scores by 9.19 and 3.70 percentage points, respectively, although composition alone retains the highest Logical AUROC. Figure 2 illustrates this complementarity: the structural example produces a high patch appearance score, whereas the logical example produces a high regional composition score despite a low patch appearance score.

Joint modeling of composition statistics. To test the contribution of correlations across composition statistics, we replace the fitted covariance with a block-diagonal approximation: entries within each region's own vector and within the global vector are kept, while all cross-region and region-to-global entries are set to zero. This reduces fused Combined AUROC from 91.12 to 90.57 (Table 5, row 1), showing that the correlations between regional and global statistics carry anomaly evidence.

Regional proportions and soft membership. We next test each remaining design choice (Table 5). Replacing regional proportions with whole-image proportions reduces Combined AUROC to 90.05 (row 2), and permuting patch locations while preserving global membership totals reduces it to 90.36 (last row); together these show that spatial partitioning and correct spatial arrangement both contribute to detection. Removing log totals reduces Combined AUROC to 89.93 (row 3), indicating whole-image totals provide complementary evidence to regional proportions. Replacing soft membership with nearest-center assignment produces the largest drop, from 91.12 to 86.90 (row 4), supporting soft membership when forming the composition descriptor.

Computational cost. We report the average of the five category medians at batch size 1. Feature processing dominates at 108.63 ms, while appearance matching and calibration adds 3.85 ms, the

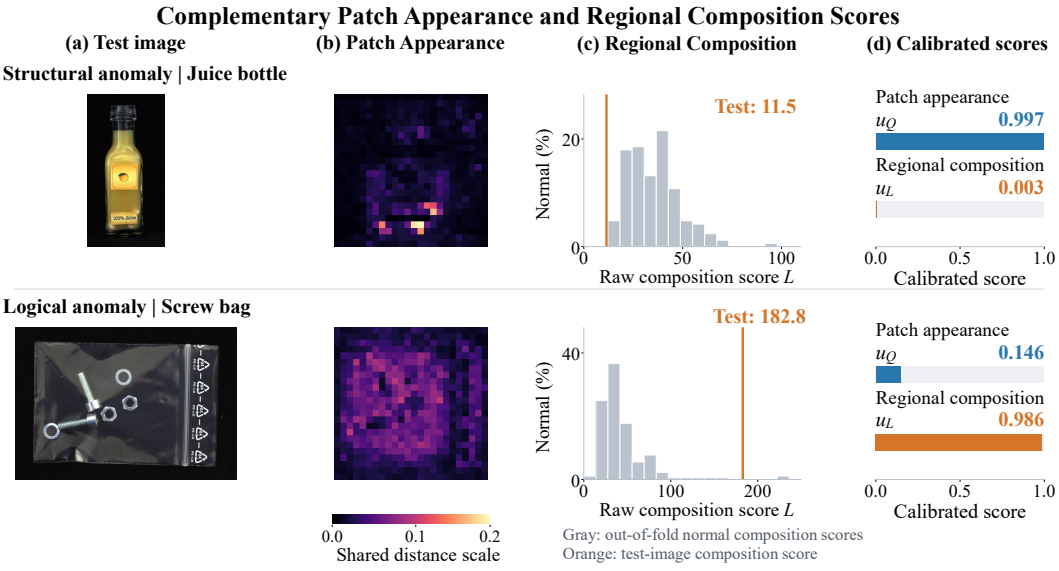


Figure 2: Complementary responses to a structural juice-bottle anomaly and a logical screw-bag anomaly (seed 0). In (c), gray histograms show out-of-fold normal composition scores and orange lines mark test scores. Panel (d) shows u_Q and u_L as normal-reference ranks. Distances in (b) share a color scale; the patch appearance score uses category-specific reference statistics.

Table 5: Composition ablations on LOCO with the appearance score fixed; mean AUROC (%) over three seeds. Each × replaces one default: whole-image proportions, omitted log totals, nearest-center assignment, or block-diagonal covariance. Last row permutes patch locations.

Composition design				AUROC (%)	
Regional proportions	Log totals	Soft membership	Full covariance	Logical AUROC composition only	Combined AUROC fused with appearance
✓	✓	✓	×	89.90	90.57
×	✓	✓	✓	89.09	90.05
✓	×	✓	✓	90.02	89.93
✓	✓	×	✓	82.80	86.90
✓	✓	✓	✓	**90.88**	**91.12**
Default with patch locations permuted				89.60	90.36

regional composition score adds 1.82 ms, and fusion adds 0.01 ms, for a total of 114.35 ms per image. Beyond the appearance matching required for structural detection, the composition score and fusion add only 1.83 ms (1.6%) to the total.

5 Conclusion

We present PARC, a training-free framework that detects both structural and logical anomalies through complementary patch appearance and regional composition scores calibrated against normal references. On MVTec-LOCO-AD, PARC reaches 91.12 Combined AUROC and raises the mean across five feature sources from 72.69 to 84.74, with the largest gain on logical anomalies; the same readout also improves scoring on MVTec AD, VisA, and CAD-SD. Ablations confirm the contributions of regional proportions, soft membership, and joint covariance, with only 1.83 ms added beyond appearance matching. Pixel-level localization and few-shot regimes are left for future work.

DECLARATION ON GENERATIVE AI

Generative AI tools were used solely to assist with language polishing, grammar correction, and improving the readability of this manuscript. The scientific content, methodology, experiments, analysis, and conclusions were developed entirely by the authors. All AI-assisted edits were reviewed and approved by the authors.

REFERENCES

Jaehyeok Bae, Jae-Han Lee, and Seyun Kim. Pni : Industrial anomaly detection using position and neighborhood information. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6373–6383, October 2023.

Kilian Batzner, Lars Heckler, and Rebecca König. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 128–138, January 2024.

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 (CVPR)*, June 2019.

Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond Dents and Scratches: Logical Constraints in Unsupervised Anomaly Detection and Localization. *International Journal of Computer Vision*, 130(4):947–969, April 2022. ISSN 1573-1405. doi: 10.1007/s11263-022-01578-9. URL https://doi.org/10.1007/s11263-022-01578-9.

Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences, 2021. URL https://arxiv.org/abs/2005.02357.

Niv Cohen, Issar Tzachor, and Yedid Hoshen. Set features for anomaly detection. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=aukOnn7J4M.

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 (WACV)*, pp. 1319–1329, February 2025.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: A patch distribution modeling framework for anomaly detection and localization. In Alberto Del Bimbo, Rita Cucchiara, Stan Sclaroff, Giovanni Maria Farinella, Tao Mei, Marco Bertini, Hugo Jair Escalante, and Roberto Vezzani (eds.), *Pattern Recognition. ICPR International Workshops and Challenges*, pp. 475–489, Cham, 2021. Springer International Publishing. ISBN 978-3-030-68799-1.

Matic Fučka, Vitjan Zavrtanik, and Danijel Skočaj. Salad – semantics-aware logical anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 21843–21852, October 2025.

Matic Fučka, Vitjan Zavrtanik, and Danijel Skočaj. Anomalyvfm – transforming vision foundation models into zero-shot anomaly detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 35555–35566, June 2026a.

Matic Fučka, Vitjan Zavrtanik, and Danijel Skočaj. Objectcore - efficient few-shot logical anomaly detection using object representations. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3857–3867, March 2026b.

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 (CVPR)*, pp. 15194–15203, June 2025.

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

Yu-Hsuan Hsieh and Shang-Hong Lai. Csad: Unsupervised component segmentation for logical anomaly detection. In *35th British Machine Vision Conference 2024, BMVC 2024, Glasgow, UK, November 25-28, 2024*. BMVA, 2024. URL https://papers.bmvc2024.org/0854.pdf.

Kengo Ishida, Yuki Takena, Yoshiki Nota, Rinpei Mochizuki, Itaru Matsumura, and Gosuke Ohashi. Sa-patchcore: Anomaly detection in dataset with co-occurrence relationships using self-attention. *IEEE Access*, 11:3232–3240, 2023. doi: 10.1109/ACCESS.2023.3234745.

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 (CVPR)*, pp. 19606–19616, June 2023.

Er Jin, Qihui Feng, Yongli Mou, Gerhard Lakemeyer, Stefan Decker, Oliver Simons, and Johannes Stegmaier. Logicad: Explainable anomaly detection via vlm-based text feature extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 4129–4137, 2025.

Soopil Kim, Sion An, Philip Chikontwe, Myeongkyun Kang, Ehsan Adeli, Kilian M Pohl, and Sang Hyun Park. Few shot part segmentation reveals compositional logic for industrial anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 8591–8599, 2024.

Yejin Kwon, Daeun Moon, Youngje Oh, and Hyunsoo Yoon. LogicQA: Logical anomaly detection with vision language model generated questions. In Georg Rehm and Yunyao Li (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 6: Industry Track)*, pp. 411–432, Vienna, Austria, July 2025. Association for Computational Linguistics. ISBN 979-8-89176-288-6. doi: 10.18653/v1/2025.acl-industry.29. URL https://aclanthology.org/2025.acl-industry.29/.

S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pp. 2169–2178, 2006. doi: 10.1109/CVPR.2006.68.

Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

Camile Lendering, Erkut Akdag, and Egor Bondarau. Subspacead: Training-free few-shot anomaly detection via subspace modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28557–28566, June 2026.

Changyi Li, Chao Yang, Yu Xiao, and Kari Tammi. Unislad: A unified framework for structural and logical industrial visual anomaly detection, 2026a. URL https://arxiv.org/abs/2606.20768.

Changyi Li, Miao Yu, Kai Dong, and Yu Xiao. Training-free logical and structural anomaly detection via calibrated fusion, 2026b. URL https://arxiv.org/abs/2609.05091.

Zhuo Li, Yuhao Yan, Xiangheng Wang, Yifei Ge, and Lin Meng. A survey of deep learning for industrial visual anomaly detection. *Artificial Intelligence Review*, 58(9):279, Jun 2025. ISSN 1573-7462. doi: 10.1007/s10462-025-11287-7. URL https://doi.org/10.1007/s10462-025-11287-7.

Jiaqi Liu, Guoyang Xie, Jinbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. Deep Industrial Image Anomaly Detection: A Survey. *Mach. Intell. Res.*, 21(1):104–135, February 2024. ISSN 2731-538X, 2731-5398. doi: 10.1007/s11633-023-1459-z. URL https://link.springer.com/10.1007/s11633-023-1459-z.

Tongkun Liu, Bing Li, Xiao Du, Bingke Jiang, Xiao Jin, Liuyi Jin, and Zhuo Zhao. Component-aware anomaly detection framework for adjustable and logical industrial visual inspection. *Advanced Engineering Informatics*, 58:102161, 2023. ISSN 1474-0346. doi: https://doi.org/10.1016/j.aei.2023.102161. URL https://www.sciencedirect.com/science/article/pii/S1474034623002896.

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, Herve Jegou, 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. Featured Certification.

Yun Peng, Xiao Lin, Nachuan Ma, Jiayuan Du, Chuangwei Liu, Chengju Liu, and Qijun Chen. Sam-lad: Segment anything model meets zero-shot logic anomaly detection. *Knowledge-Based Systems*, 314:113176, 2025.

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 (CVPR)*, pp. 14318–14328, June 2022.

Shunsuke SAKAI, Tatsuhito Hasegawa, and Makoto Koshino. Logical anomaly detection with masked image modeling. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856. URL https://openreview.net/forum?id=uuuaRCMYE3.

Josef Sivic and Andrew Zisserman. Video Google: A text retrieval approach to object matching in videos. In *IEEE International Conference on Computer Vision*, volume 2, pp. 1470–1477, 2003.

Shota Sugawara and Ryuji Imamura. Puad: Frustratingly simple method for robust anomaly detection. In *2024 IEEE international conference on image processing (ICIP)*, pp. 842–848. IEEE, 2024.

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 (CVPR)*, pp. 22883–22892, June 2024.

Haiming Yao, Wenyong Yu, Wei Luo, Zhenfeng Qiang, Donghao Luo, and Xiaotian Zhang. Learning global-local correspondence with semantic bottleneck for logical anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(5):3589–3605, 2023.

Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 4571–4584. Curran Associates, Inc., 2022. doi: 10.52202/068431-0330. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/1d774c112926348c3e25ea47d87c835b-Paper-Conference.pdf.

Jie Zhang, Masanori Suganuma, and Takayuki Okatani. Contextual affinity distillation for image anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 149–158, January 2024.

Jinjin Zhang, Guodong Wang, Yizhou Jin, and Di Huang. Towards training-free anomaly detection with vision and language foundation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15204–15213, June 2025.

Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Logico: A unified framework for logical and structural anomaly detection. In Paolo Favaro, Zuzana Kukelova, Atsuto Maki, Anna Rohrbach, Konrad Schindler, and Federico Tombari (eds.), *Computer Vision – ECCV 2026*, pp. 244–261, Cham, 2026. Springer Nature Switzerland. ISBN 978-3-032-37550-6.

Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner (eds.), *Computer Vision – ECCV 2022*, pp. 392–408, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-20056-4.

A Appendix Overview

Appendix B specifies the feature grid, clustering, cross-fitting, and numerical settings used by every PARC result. Appendix C records how each baseline is reproduced. Appendix D varies the composition and appearance hyperparameters. Appendix E reports category-level results and the two scores by category. Appendix F varies the support bank, fusion rule, and normal-reference size. Appendix G gives the stage-level timing behind the main-text summary. Appendix H provides additional detection examples.

B Implementation Details

B.1 Feature Grid Standardization

PARC consumes one feature map per image and does not resize features from different extractors to a common grid. AnomalyVFM RADIO features are extracted at 768×768, pooled by 2×2, and scored on a 24×24 grid. The same readout is applied without changing K, τ, or R to the native grids of AnomalyDINO (28×28), PatchCore and SPADE (28×28), and SubspaceAD (32×32). Regional boundaries follow the integer division in Section 3, so every region is nonempty on these grids.

B.2 Clustering and Cross-Fitting

For each category and seed, normal training images are split evenly into a support set and a disjoint estimation set. K-means uses Euclidean distance on support features, with 10 initializations and at most 300 iterations. Centers fitted for the main setting are reused rather than refit. The estimation images receive a fixed five-fold partition. Each fold fits the appearance statistics and Ledoit–Wolf composition model on the other folds and scores the held-out fold. Test images are scored by statistics fitted on the full estimation set. The support bank and cluster centers stay fixed during cross-fitting.

B.3 Numerical Stability

Patch features are ℓ_2-normalized before cosine matching. Soft membership uses temperature $\tau = 0.30$. Ledoit–Wolf shrinkage estimates the composition covariance, and its precision matrix is checked against the inverse identity. The normal-rank mapping uses the smoothed rank in Eq. 7, extends scores above the largest normal reference by Eq. 8, and clips the result to $[10^{-6}, 1 - 10^{-6}]$. The normal score range is floored at 10^{-8}. Composition descriptors and covariance statistics use float64.

C Comparison Settings

C.1 Baseline Reproduction

All image-level AUROC values use the same MVTec-LOCO-AD test split. PARC uses frozen AnomalyVFM features and averages three seeds. PatchCore, AnomalyDINO, and SubspaceAD use their released full-normal-set scoring procedures; SPADE is included only in the feature-source comparison. ComAD repeats its released segmentation script with process seeds 0–2. LogSAD uses one released full-data run per category. CSAD and LogiCo evaluate one existing checkpoint per category and are not retrained. LogiCo reads precomputed component masks, whereas its latency in Table 1 includes online mask generation. Latency uses batch size 1, the same 30 images per category, 20 warm-up calls, and 100 measured calls, with four CPU threads and one RTX 4090.

D HYPERPARAMETER SENSITIVITY

D.1 EFFECT OF K, τ, AND R

Table 6 changes one of K, τ, and R at a time around the default $K = 80$, $\tau = 0.30$, and $R = 2$. Combined AUROC stays within 2.06 points of the default for $K \in \{20, 40, 60\}$, $\tau \in \{0.10, 0.20, 0.40\}$, and $R \in \{1, 3\}$.

Table 6: Composition sensitivity on MVTec-LOCO-AD. Combined AUROC (%) is the mean and sample standard deviation over three seeds. The final row is the default.

Setting	Combined	Logical	Structural
$K = 20$	89.06±0.12	86.85±0.26	92.38±0.05
$K = 40$	90.22±0.21	88.42±0.33	92.89±0.28
$K = 60$	90.66±0.10	88.80±0.11	93.37±0.32
$\tau = 0.10$	90.62±0.25	88.63±0.48	93.49±0.09
$\tau = 0.20$	91.06±0.24	89.20±0.53	93.72±0.10
$\tau = 0.40$	90.92±0.10	89.25±0.35	93.39±0.30
$R = 1$	90.40±0.10	88.52±0.40	93.07±0.50
$R = 3$	90.92±0.45	89.09±0.78	93.65±0.17
$K = 80$, $\tau = 0.30$, $R = 2$	91.12±0.17	89.42±0.44	93.59±0.15

D.2 EFFECT OF β AND ρ

The default $\beta = 0.25$ and $\rho = 0.01$ are determined on the normal estimation set and kept identical across all categories and backbones. Table 7 varies one of them at a time while keeping the composition score fixed. The reported default reaches 91.12 Combined AUROC. Removing the position term, or averaging 2% rather than 1% of the patches, changes it by at most 0.37 points. Increasing β from 0.25 to 0.50 or 1.00 reduces the fused Structural AUROC.

Table 7: Appearance-score sensitivity on MVTec-LOCO-AD. Values are fused AUROC (%) over three seeds. The third row is the default.

β	ρ	Combined	Logical	Structural
0	0.010	90.87±0.36	89.08±0.59	93.43±0.14
0.25	0.005	91.07±0.29	89.29±0.56	93.63±0.23
0.25	0.010	91.12±0.17	89.42±0.44	93.59±0.15
0.25	0.020	90.75±0.25	89.39±0.45	92.82±0.17
0.50	0.010	90.43±0.30	89.23±0.57	92.40±0.41
1.00	0.010	90.02±0.41	89.28±0.72	91.52±0.42

E PER-CATEGORY RESULTS

E.1 MVTEC-LOCO-AD

Table 8 breaks down the default PARC score. Pushpins is the weakest category, at 81.71 Combined AUROC, and accounts for most of the gap between the category mean and the stronger categories.

E.2 MVTEC AD, VISA, AND CAD-SD

On MVTec AD, the category mean is 99.41 Combined AUROC. Fourteen of the fifteen categories are at least 98.57. Screw is the lowest, at 96.60 ± 1.35. On VisA, the mean is 96.33 over twelve categories. Macaroni2 is the outlier at 81.42 ± 1.01; the other eleven categories are between 93.07 and 100.00. CAD-SD comprises only one product category, screw; PARC achieves 99.81 ± 0.04 AUROC on it.

Table 8: Per-category PARC AUROC (%) on MVTec-LOCO-AD, mean over three seeds.

Category	Combined	Structural	Logical
breakfast_box	95.47±0.63	93.28±0.51	97.84±0.88
juice_bottle	98.94±0.24	99.58±0.36	98.53±0.18
pushpins	81.71±0.89	87.08±1.26	76.94±1.09
screw_bag	89.36±0.67	95.08±0.53	85.94±0.78
splicing_connectors	90.10±1.13	92.93±0.96	87.87±1.26
Mean	91.12±0.17	93.59±0.15	89.42±0.44

E.3 Score Complementarity by Category

Table 4 shows that the two scores specialize on different anomaly types. Table 9 shows where this holds within MVTec-LOCO-AD. Composition exceeds appearance on logical anomalies in every category. The margin is largest for screw bag, 88.30 against 58.56, and smallest for splicing connectors, 86.65 against 82.90. Appearance exceeds composition on structural anomalies in four of the five categories. Fusion achieves higher Combined AUROC than appearance alone in every category, but it is below composition on logical anomalies for breakfast box, juice bottle, pushpins, and screw bag. The maximum therefore retains structural evidence while giving up part of the logical response in four of the five categories.

Table 9: Per-category AUROC (%) of the calibrated scores on MVTec-LOCO-AD, mean over three seeds.

	Combined			Structural			Logical		
Category	u_Q	u_L	Fused	u_Q	u_L	Fused	u_Q	u_L	Fused
breakfast_box	83.48	88.80	95.47	92.74	79.29	93.28	73.44	99.11	97.84
juice_bottle	94.24	98.03	98.94	96.22	97.10	99.58	92.93	98.65	98.53
pushpins	72.59	77.82	81.71	86.64	73.46	87.08	60.07	81.70	76.94
screw_bag	71.80	88.24	89.36	93.91	88.14	95.08	58.56	88.30	85.94
splicing_connectors	87.57	84.22	90.10	93.50	81.13	92.93	82.90	86.65	87.87

F Additional Ablations

F.1 Support Bank Size

Table 10 keeps a nested subset of the normal support images and refits both nearest-neighbor matching and the $K = 80$ centers. Using one quarter of the support bank lowers Combined AUROC from 91.12 to 90.22. The composition score remains within 0.39 points of its full-bank Logical AUROC at every tested size.

Table 10: Support-bank fraction on MVTec-LOCO-AD. Combined is the fused score and Logical is the composition score. Means are over three seeds.

Support fraction	Combined	Logical
0.25	90.22±0.34	90.55±0.48
0.50	90.87±0.22	90.83±0.44
0.75	90.83±0.23	90.49±0.44
1.00	91.12±0.17	90.88±0.10

F.2 Fusion Strategy

Table 11 compares calibration and fusion rules on the same two raw scores. Maximum fusion of the default calibrated scores reaches 91.12 Combined AUROC. Replacing the maximum with the mean

drops it to 85.69. Using uncalibrated scores reverses the two branches: Logical AUROC rises to 90.93, while Structural AUROC falls to 83.89.

Table 11: Calibration and fusion on MVTec-LOCO-AD, AUROC (%) over three seeds.

Rule	Combined	Logical	Structural
Default calibration, max	91.12±0.17	89.42±0.44	93.59±0.15
ECDF, max	89.36±1.81	87.56±1.99	91.93±1.63
z-score, max	91.00±0.28	89.05±0.52	93.74±0.05
Default calibration, mean	85.69±0.42	83.36±0.26	89.18±0.53
Raw scores, max	87.47±0.37	90.93±0.15	83.89±0.63

F.3 Normal-Data Size and Covariance Stability

Table 12 fits both scores on a nested subset of the estimation images. One quarter of the estimation pool gives 88.51 Combined AUROC, 2.61 points below the full pool. The Ledoit–Wolf covariance becomes more ill-conditioned as more estimation images are added, from condition number 9923 to 30631, but detection improves over the same range. Shrinkage therefore keeps the larger sample usable despite the higher condition number.

Table 12: Estimation-pool fraction on MVTec-LOCO-AD. Condition number is the mean condition of the fitted composition covariance.

Estimation fraction	Combined AUROC	Condition number
0.25	88.51±0.40	9923
0.50	89.94±0.40	16385
0.75	90.94±0.17	23784
1.00	91.12±0.17	30631

G Additional Timing Details

The main text reports four stages for one image at batch size 1. Table 13 gives the category medians. Feature processing includes decoding, preprocessing, host-to-device transfer, the backbone, the image head, pooling, and normalization. Each stage is median-aggregated separately, so a row need not equal the sum of its stage columns. Across categories, feature processing accounts for the variation: juice bottle takes 88.72 ms, while breakfast box takes 134.06 ms. The composition score stays between 1.78 and 1.90 ms.

Table 13: Per-category median latency (ms/image) at batch size 1.

Category	Feature	Appearance	Composition	Fusion	Total
breakfast_box	134.06	3.81	1.85	0.01	139.76
juice_bottle	88.72	3.74	1.78	0.01	94.26
pushpins	108.08	3.98	1.79	0.01	113.97
screw_bag	108.81	3.87	1.90	0.01	114.60
splicing_connectors	103.46	3.87	1.79	0.01	109.14
Mean of medians	108.63	3.85	1.82	0.01	114.35

H Additional Qualitative Examples

Figure 2 in the main text illustrates the complementary responses of the two scores on one structural and one logical example. We further provide two sets of qualitative visualizations covering all five MVTec-LOCO-AD categories (three logical: Breakfast Box, Juice Bottle, Splicing Connectors; two structural: Pushpins, Screw Bag). Figure 3 shows full-resolution anomaly-score heatmaps alongside ground-truth masks. Figure 4 presents cropped views of the same cases for finer-grained inspection.

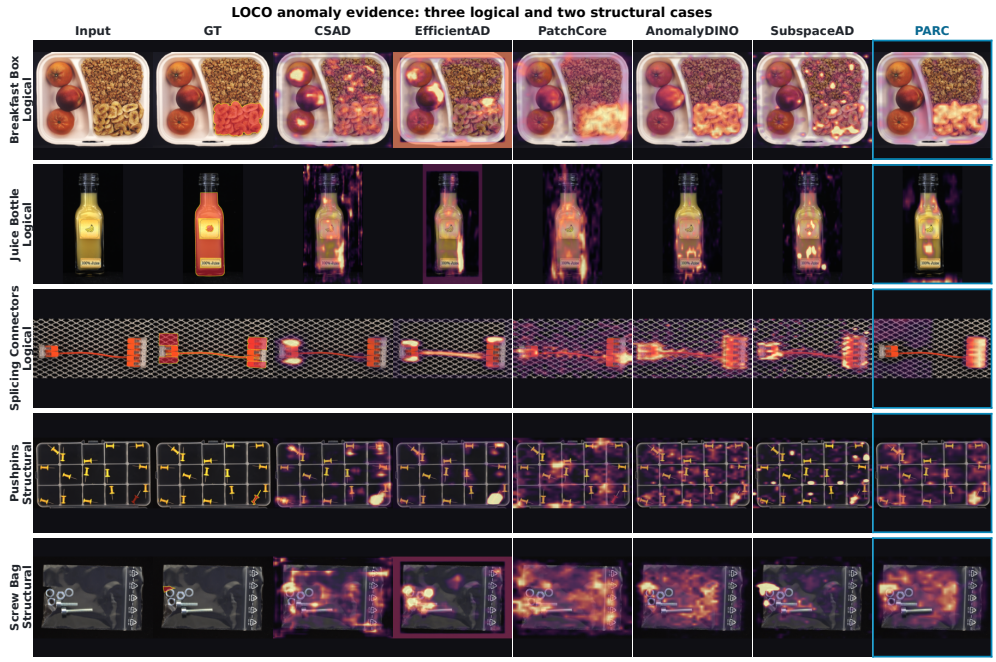


Figure 3: Full-resolution anomaly-score heatmaps for all five MVTec-LOCO-AD categories.

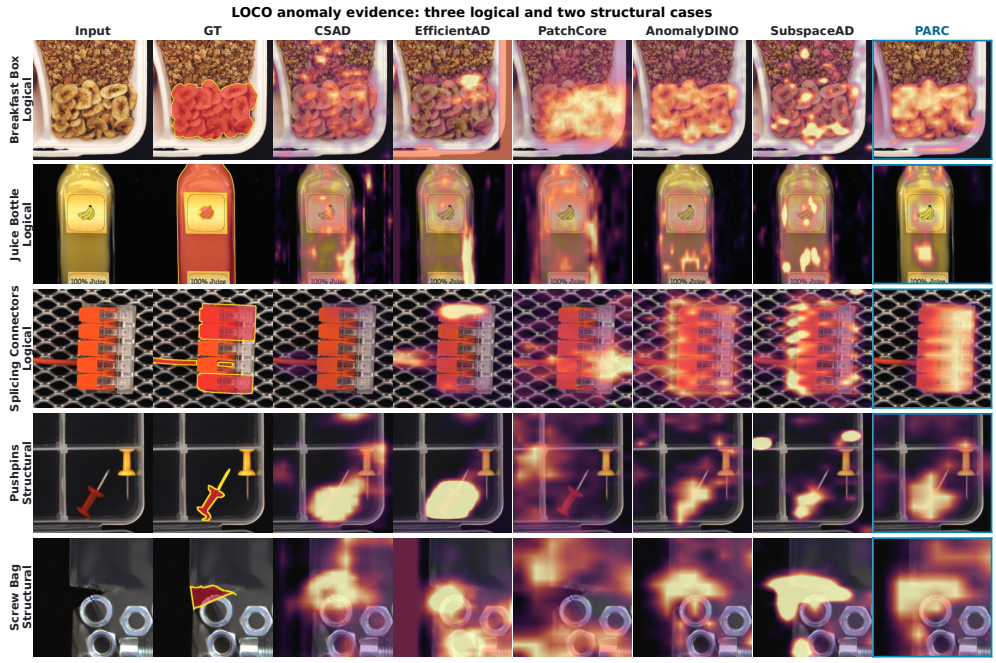


Figure 4: Cropped views of the same five cases as Figure 3, facilitating fine-grained comparison of anomaly localization.