RISK-CONTROLLED EFFICIENT TEMPLATE MUTUAL MATCHING FOR ANOMALY DETECTION IN TEXTURE-COMPLEX INDUSTRIAL INSPECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Industrial chip inspection must jointly limit missed defects, over-inspection of normal products, and computational cost. This balance is difficult when normal texture appears primarily on the carrier strip, which also belongs to the valid inspection region and therefore cannot simply be masked out. We propose Texture Reweighting and Risk Control for HETMM (TRRC-HETMM), combining texture-disentangled mutual matching, miss-risk calibration, and safe dynamic computation. The method estimates position-conditioned variation subspaces from normal templates and reweights residuals within and orthogonal to those subspaces. Independently calibrated release thresholds and guard rules govern fast decisions and fallback to full matching. Controlled high-frequency suppression and an equal-energy residual-shuffling control further support the contribution of normal strip texture to over-inspection. On a sealed industrial test set with 1,000 defective and 19,000 normal images, the frozen system misses 4 defective samples and over-inspects 2,765 normal samples, corresponding to rates of 0.40% and 14.55%, respectively. At the same four-miss operating point, HETMM over-inspects 6,718 normal samples; TRRC-HETMM reduces this number by 58.84%. Experiments on public benchmarks further assess generalization across object categories.

1 INTRODUCTION

Visual inspection on high-technology chip production lines has strongly asymmetric costs. A defective sample that escapes inspection may proceed to packaging or delivery, whereas a normal sample flagged as defective repeatedly occupies reinspection capacity, slows the line, and increases labor cost. A deployable system must therefore do more than maximize average AUROC: it must control over-inspection and computation at a stringent miss operating point. This work originates from that production requirement rather than from post hoc optimization of a public benchmark score.

The central difficulty is the strip that carries the chip. Upstream manufacturing and transportation controls produce normal variations in strip directionality, brightness, and local roughness. Yet the strip itself must remain inside the inspection region because breakage, scratches, and missing material can occur on the strip or at the chip-strip interface. Cropping away the "background" would shrink the safety-relevant region, while global smoothing or response suppression could weaken genuine defect evidence. This setting leads to a broader scientific question: when normal variation and anomaly evidence coexist in a region that cannot be masked, how can a detector suppress directions of normal variation while preserving evidence that is orthogonal to those directions or exceeds their normal magnitude?

We use template mutual matching rather than closed-category object detection as the technical starting point. Object detectors typically require predefined defect categories and corresponding annotations. Template matching instead screens for departures from normal appearance and is therefore, in principle, better suited to precede category-specific detection when separating previously unspecified defects. Our experiments assess detection and generalization on the defects currently available; they do not establish coverage of every unknown defect. This motivation defines the intended role of the system rather than replacing an open-set evaluation.

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.

Building on HETMM's bidirectional mutual matching (Chen et al., 2025), we introduce Texture Reweighting and Risk Control for HETMM (TRRC-HETMM). A position-conditioned subspace reorganizes anomaly evidence so that high-variance normal texture does not dominate the matching distance. Independent risk calibration determines which samples may be released, instead of selecting thresholds on the test set. Finally, safe dynamic computation admits only unambiguous samples to the fast path and sends gray-zone or guard-triggered samples to full PTS-60 matching. These components answer, respectively, what should be compared, when a decision is trustworthy, and where computation should be allocated.

Our contributions are fourfold. (1) We introduce a position-conditioned texture-disentangled metric for bidirectional query-template residuals. It separates normal texture directions from orthogonal structural residuals without removing the strip from inspection. (2) We place fast binary decisions, mandatory fallback, and full matching under branchwise miss-risk budgets, making efficiency subordinate to the safety operating point. (3) We construct a mechanism study that progresses from statistical association to controlled intervention, including residual-direction statistics, suppression of normal strip high frequencies, an equal-energy residual-shuffling control, and region-specific defect detection. (4) On an entity-disjoint industrial chip dataset and three public benchmarks, we report frozen operating points, equal-miss comparisons, cross-object generalization, statistical tests, and end-to-end efficiency.

2 RELATED WORK

Normality modeling and synthetic anomalies. Industrial anomaly detection commonly learns from normal images alone. Deep SVDD compresses normal samples into a feature hypersphere, while FCDD extends one-class description to localizable fully convolutional maps (Ruff et al., 2018; Liznerski et al., 2021). PaDiM models position-wise feature distributions, PatchCore performs nearest-neighbor retrieval from a representative patch memory, and Patch SVDD brings one-class description to local patches (Defard et al., 2021; Roth et al., 2022; Yi & Yoon, 2020). A second line of work learns decision boundaries from proxy anomalies or reconstruction: CutPaste creates pasted self-supervision, DRAEM jointly learns normal reconstruction and anomaly segmentation, and RealNet and DeSTSeg use realistic synthetic anomalies and denoising student-teacher constraints, respectively (Li et al., 2021; Zavrtanik et al., 2021; Zhang et al., 2024; 2023). These methods improve general anomaly representations, but synthetic distributions can still differ from production texture. We generate no defect category; instead, we estimate permissible variation directions from normal templates themselves.

Distillation, reconstruction, and efficient detection. Uninformed Students, STFPM, Reverse Distillation, EfficientAD, and Dinomaly model normality through predictive uncertainty, multi-level feature distillation, reverse knowledge transfer, lightweight teacher-student learning, and Transformer reconstruction, respectively (Bergmann et al., 2020; Wang et al., 2021; Deng & Li, 2022; Batzner et al., 2024; Guo et al., 2025). SimpleNet trains a discriminator through simple feature adaptation and anomalous feature generation, whereas DifferNet, FastFlow, and CFLOW-AD use normalizing flows to balance localization and speed (Liu et al., 2023; Rudolph et al., 2021; Yu et al., 2021; Gudovskiy et al., 2022). These methods mainly optimize the accuracy-efficiency trade-off of a single model. Our dynamic path instead retains full HETMM as a safety fallback and lets the fast branch handle only samples permitted by calibration and guard rules.

Reference matching and cross-category generalization. SPADE, PatchCore, and HETMM rely on local query-reference correspondences. HETMM further uses bidirectional matching and PTS center/boundary prototypes to detect local structural changes and missing content (Cohen & Hoshen, 2020; Roth et al., 2022; Chen et al., 2025). RegAD learns registration-based reference comparison across categories, while WinCLIP, AnomalyCLIP, AnomalyDINO, and UniVAD study zero- or few-shot generalization with vision-language or foundation-model representations (Huang et al., 2022; Jeong et al., 2023; Zhou et al., 2024; Damm et al., 2025; Gu et al., 2025). AnomalyVFM adapts vision foundation models for zero-shot detection using synthetic training data, while PTAD uses texture-free 3D references to address pose and texture variation (Fučka et al., 2026; Zhuo et al., 2026). We retain explicit normal templates because fixed production stations provide stable geome-

try. Our contribution is not to eliminate references, but to explain and reweight residuals caused by normal texture at the same canonical position.

Datasets, risk control, and dynamic inference. MVTec AD, MVTec LOCO, VisA, and Real-IAD cover conventional objects and textures, logical constraints, larger multi-object collections, and real multi-view inspection, respectively (Bergmann et al., 2021; 2022; Zou et al., 2022; Wang et al., 2024). Prior work has also studied how texture can mislead industrial anomaly detectors (Nakkina et al., 2025). These benchmarks are usually compared through ranking metrics, whereas production deployment additionally requires control at a fixed threshold. Learn-then-Test and conformal risk control (CRC) provide finite-sample tools for multiple risks and bounded monotone losses (Angelopoulos et al., 2025; 2024). BranchyNet, MSDNet, and DynamicViT allocate computation through early exits or dynamic tokens (Teerapittayanon et al., 2016; Huang et al., 2018; Rao et al., 2021). We do not present CRC or early exit as a new theory. Instead, we use branchwise risk budgets to constrain fast release, which can cause misses, and embed gray-zone fallback into a bidirectional template-matching system.

3 PRELIMINARIES

Let $\mathcal{Z}$ be a bank of normal templates, x a query image, and $y \in \{0, 1\}$ its label, where $y = 1$ denotes a defective sample. The valid region Ω_{inspect} includes the chip, carrier strip, and interface. The fixed ROI excludes only pixels outside the process-defined inspection range; it does not use test labels to remove difficult regions. An encoder produces $F_l(x) \in \mathbb{R}^{C_l \times H_l \times W_l}$. Full HETMM returns an anomaly map $S(x)$ and aggregates the largest K smoothed pixel scores inside the ROI into a continuous image score $s(x)$. The fast path produces only an image-level decision and no pixel map that can be interpreted as localization.

For threshold τ, a miss and an over-inspection event are $y = 1, s(x) \leq \tau$ and $y = 0, s(x) > \tau$, respectively. AUROC measures ranking by continuous scores, whereas FN and FP characterize one frozen operating point; neither substitutes for the other. The dynamic system also introduces a route $r(x) \in \{\text{fast}, \text{full}\}$. Score validity, threshold risk, and computational routing are therefore distinct questions. We minimize normal-sample over-inspection risk R_{over} subject to a defective-sample miss-risk constraint:

$$\min_{\tau, \Theta} R_{\text{over}}(\tau; \Theta) \quad \text{s.t.} \quad R_{\text{miss}}(\tau; \Theta) \leq \epsilon, \tag{1}$$

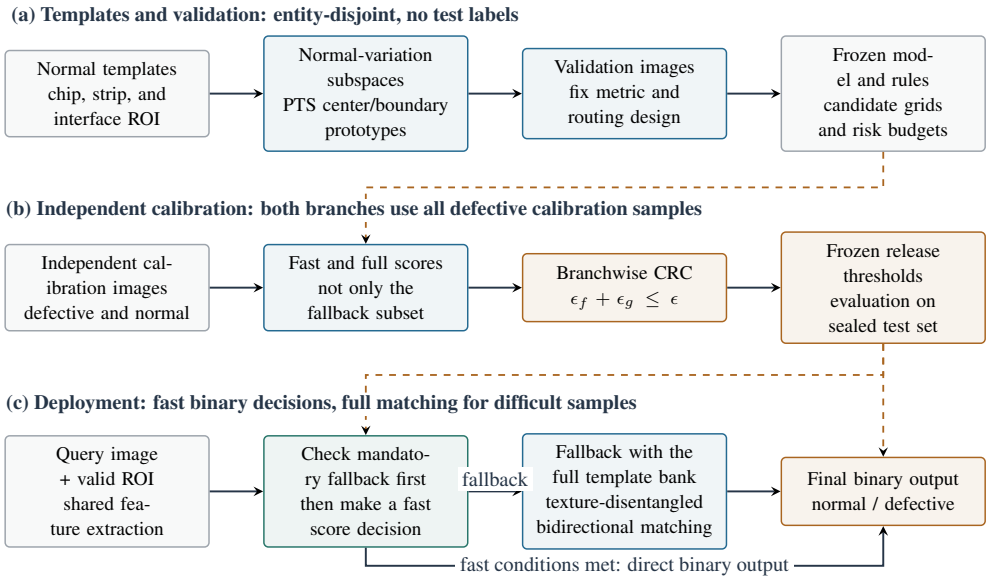
where Θ includes the metric, template bank, aggregation, and routing rules. The system always returns a binary normal/defective output; fallback is an internal computational path, not a referral to manual review. Deployment initially uses exhaustive manual verification and switches to sampling-based audits only after sufficient operational validation. All model metrics reported here are measured before any manual correction.

4 METHOD

4.1 UNIFIED FRAMEWORK

TRRC-HETMM combines texture-disentangled matching, branchwise risk calibration, and safe dynamic computation while preserving the complete valid ROI on every path. Normal templates construct position-conditioned variation subspaces and the PTS-60 bank. The validation set selects feature layers, the maximum subspace dimension, score aggregation, routing intervals, and guard rules; these choices are then frozen. An independent calibration set selects only the release thresholds of the fast and full branches, and the sealed test set is used once for evaluation. Separating these roles prevents a favorable test-set miss count from being fed back into the method.

The three components are connected in causal order. Texture disentanglement first changes the evidence representation between query and template. Risk calibration then converts continuous scores into budget-constrained operating points. Dynamic computation finally uses the frozen scores and thresholds to decide whether fallback is required. Dynamic routing without texture modeling would inherit the original false positives, while texture modeling without independent calibration could still benefit from test-set threshold selection. Full definitions and the algorithm appear in Appendices A–C.



No candidate-template filtering, no template augmentation, and no third prediction label.
The fast path returns no localization map; deployment uses full manual verification before switching to sampling-based audits.

Figure 1: The unified TRRC-HETMM framework. Texture disentanglement changes the anomaly evidence, risk calibration constrains release, and dynamic computation sends difficult samples to full template matching.

4.2 Position-conditioned texture-disentangled mutual matching

At a fixed inspection station, the same canonical position has a consistent physical meaning. For layer l and position p, we collect features from local neighborhoods of normal templates. We first clip outlying observations with a coordinate-wise median and median absolute deviation (MAD), then estimate a shrinkage covariance $\Sigma_{l,p} = U_{l,p}\Lambda_{l,p}U_{l,p}^\top + E_{l,p}$. The subspace dimension is determined by cumulative explained variance γ and cap $k_{\max}$. Here, U is not a defect dictionary; it contains variation directions repeatedly observed at the same position in normal templates.

For residual $\delta = q - t$ between query feature q and template feature t, we define

$$d_{\mathrm{TD}} = w^{\perp}\|(I - UU^\top)\delta\|_2^2 + w^{\parallel}\|\bar{\Lambda}^{-1/2}U^\top\delta\|_2^2, \tag{2}$$

where $\bar{\Lambda}_{jj} = \max(\Lambda_{jj}, 10^{-6})$. The first term measures residuals orthogonal to normal variation; the second reweights subspace residuals by normal variance. Each subspace direction has coefficient $w^{\parallel}/\bar{\Lambda}_{jj}$: higher-variance directions receive smaller coefficients, but these need not be below $w^{\perp}$. Large deviations within the subspace still produce responses. Defects may contribute to either component; anomaly preservation is assessed empirically.

We replace the cosine distance in both directions of HETMM local matching with this metric. For query position q and normal-template neighborhood $\mathcal{N}(p)$, the forward term finds the closest template, whereas the backward term requires every critical template prototype to be explained by the query:

$$d_l^{\rightarrow}(q) = \min_{t\in\mathcal{N}(p)} d_{\mathrm{TD}}(q,t), \qquad d_l^{\leftarrow}(t) = \min_{q\in\mathcal{N}(p)} d_{\mathrm{TD}}(q,t). \tag{3}$$

The forward term responds to local appearance changes; the backward term adds evidence for missing or incomplete structures. Multi-level maps are resampled to canonical coordinates, smoothed only within the valid ROI, and aggregated by the mean of the Top-20 responses. We use 3.5-MAD clipping and 0.10 covariance shrinkage, with no additional position-reliability formula.

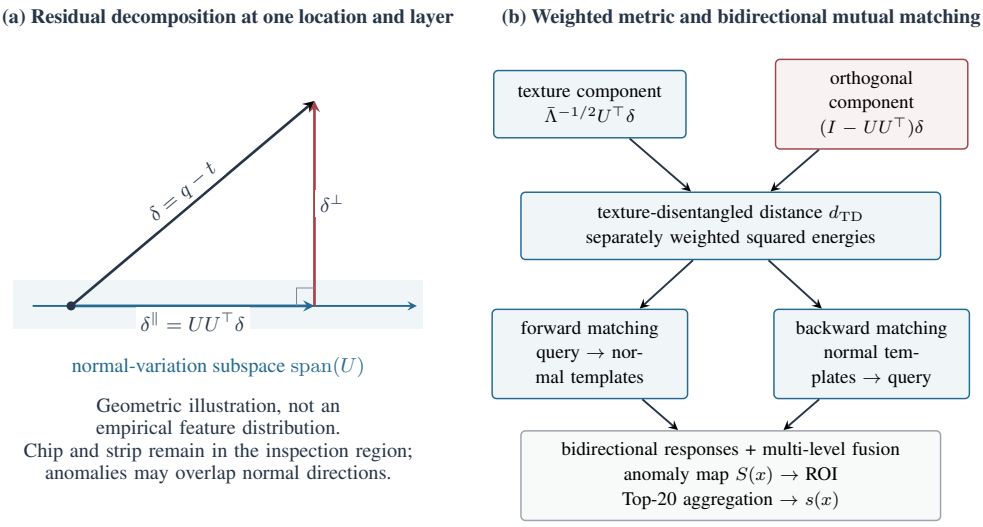


Figure 2: The position-conditioned texture-disentangled distance used in forward and backward local template matching.

4.3 Independent Risk Calibration

For a defective sample, define the monotone loss of branch j as $\ell_j(x, y; \tau) = \mathbb{I}[y = 1, s_j(x) \le \tau]$. The fast and full branches are each run on all $n_1 = 600$ defective calibration images. Models, templates, normalization, routing, fallback rules, budgets, and candidate grids are frozen before the calibration outcomes are inspected. Following CRC (Angelopoulos et al., 2024), branch j selects the largest feasible threshold:

$$\hat{\tau}_j = \max\{\tau \in \mathcal{G}_j : (n_1 \widehat{R}_j(\tau) + 1)/(n_1 + 1) \le \epsilon_j\}. \tag{4}$$

As the candidate threshold increases, the miss loss is nondecreasing; choosing the largest feasible value reduces over-inspection as much as the budget permits. We split the total 0.40% budget into $\epsilon_f = 0.18\%$ for fast release and $\epsilon_g = 0.22\%$ for the full branch. Both calibration branches observe zero misses, and $1/601 = 0.166\%$ makes each budget feasible at this sample size. The split is not an empirical addition of confidence values. It follows a pathwise decomposition of the system miss event and a union bound. The resulting statement controls expectation jointly over calibration and a new exchangeable sample; it is not a zero-miss guarantee for a fixed model under arbitrary future distributions and does not replace drift monitoring.

4.4 Safe Dynamic Computation

The fast route uses L3 features and 30 normal templates, avoiding full bidirectional matching against PTS-60 for every sample. A Sigmoid maps the raw anomaly value to s_{route}. Query and prototype descriptors are spatially averaged within the ROI and L2-normalized. Because a single early-exit score can mistake a difficult anomaly for a normal sample, we add two uncertainty checks. The cosine-similarity margin between the best and second-best prototypes measures reference ambiguity, while the relative forward-backward discrepancy d_{sym} measures whether the query can be explained in one direction but not the other. A margin below 0.02 or $d_{\mathrm{sym}} > 0.20$ forces fallback before any fast decision. The frozen rule is

$$\hat{y}(x) = \begin{cases} 0, & \neg b(x),\ s_{\mathrm{route}} < 0.58, \\ 1, & \neg b(x),\ s_{\mathrm{route}} > 0.72, \\ \mathbb{I}[s_{\mathrm{full}} > 0.68], & \text{otherwise}, \end{cases} \tag{5}$$

where b denotes either guard. Samples in the gray zone $[0.58, 0.72]$ and all guard-triggered samples use the full PTS-60 bank. The fast path returns no localization map, while the full branch returns an anomaly map. Here, "safe" means that fast release is constrained by an independent budget and

mandatory fallback; it does not mean that the system cannot miss a defect under distribution shift. Both routes map to the same normal/defective production interface, so dynamic computation does not change downstream semantics.

5 EXPERIMENTS

5.1 SETUP

Industrial data and splits. The sealed test set contains 20,000 images: 1,000 defective and 19,000 normal. The validation set contains 200 defective and 1,000 normal images, and the independent risk-calibration set contains 600 defective and 3,000 normal images. An additional 500 normal images form the templates. All sets are disjoint by physical chip identity, preventing repeated captures of the same chip from crossing splits; templates come from training-stage reference images and are never selected from the test set. Six cameras and three illumination stations acquire 5496×3672 RGB images. After cropping the valid region, images are resized to 640×640. The chip, strip, and interface all belong to the ROI, and the task is image-level normal/defective classification. Inference uses no test annotations except for the strip labels used only in the mechanism study.

Public benchmarks and comparisons. We use the official splits and all categories of MVTec AD, VisA, and Real-IAD. A fixed 10% of normal training images is used for validation, and Real-IAD is separated by physical entity. All listed methods were rerun in one unified hardware/software environment from their official implementations or released configurations, with the same preprocessing and pooled evaluation code. All public benchmarks run the full matching path. Image- and pixel-level AUROC are pooled over all test samples. UniVAD uses four randomly drawn normal reference images per category; the references are excluded from that draw's evaluation, and the reported score is the mean of four independent draws. No CRC threshold is selected on public data, so these experiments evaluate only continuous anomaly scores and make no claim of miss-risk control. Comparisons include the memory-bank method PatchCore, lightweight student-teacher model EfficientAD, bidirectional template method HETMM, Transformer reconstruction model Dinomaly, and few-shot unified model UniVAD.

Implementation and efficiency protocol. We set $k_{\max} = 16$, $\gamma = 0.90$, $w^{\parallel} = 0.25$, $w^{\perp} = 1$, and $K = 20$. A single frozen configuration produces every chip result, with no tuning for a defect region or test subset. Efficiency is measured on one RTX 4090 D at batch size 1. After 100 warm-up iterations, the complete 20,000-image test set is processed in acquisition-time order for 50 runs. FPS, P50, P95, and P99 are computed within each run and then averaged across runs. Timing includes image preprocessing, CPU-GPU transfer, model inference, postprocessing, and GPU synchronization at the measurement boundaries. We therefore report end-to-end system speed rather than network-forward time alone. Classification statistics still use 20,000 independent images; repeated runs measure time only and do not enlarge the statistical sample.

Metrics and statistics. Image AUROC is computed from continuous scores over all normal and defective samples. Pixel AUROC is used only on public datasets with pixel masks. The miss rate is FN/1,000 and the over-inspection rate is FP/19,000. Because the industrial objective emphasizes a low-miss operating point, we report each method's frozen threshold and a descriptive equal-FN point along the HETMM test-score curve; the latter is not interpreted as an independently calibrated result. We give a Wilson interval for the miss rate and use an exact McNemar test for paired decisions on normal samples.

5.2 MAIN CHIP RESULTS

For the main table, AUROC uses the common continuous score obtained by running the full branch offline on every sample. FN, FP, and speed come from the frozen dynamic system. Relative to the frozen HETMM operating point, our system produces 44 fewer misses and 1,130 fewer false positives. The Wilson 95% interval for the miss rate is $[0.16\%, 1.02\%]$. At FN=4 along the HETMM test-score curve, HETMM has 6,718 false positives (35.36%); our method removes 3,953 of them, a relative reduction of 58.84%. This equal-FN comparison is descriptive, not an independent calibration guarantee. At the main operating points, 1,219 normal images are wrong under HETMM but

Table 1: Sealed-test results. Rates and AUROC are percentages; P95 is in ms. Best values are bold; venue/year identify the original methods.

Method	Venue/year	AUROC↑	FN↓	FP↓	Miss↓	Over-inspection↓	FPS↑/P95↓
PatchCore	CVPR 2022	69.28	289	6,392	28.90	33.64	19.8/50.3
EfficientAD-M	WACV 2024	82.44	147	5,890	14.70	31.00	**78.5**/14.3
HETMM	IJCV 2025	93.18	48	3,895	4.80	20.50	72.9/13.8
TRRC-HETMM	Ours	**98.40**	**4**	**2,765**	**0.40**	**14.55**	78.3/**13.7**

Table 2: Unified-environment reruns of full-branch AUROC (image/pixel, %), pooled over test samples. Best image and pixel values are bold separately.

Method	Venue/year	MVTec AD↑	VisA↑	Real-IAD↑
HETMM	IJCV 2025	99.0/98.1	96.1/97.2	84.7/92.2
EfficientAD-M	WACV 2024	99.1/98.6	98.1/97.9	77.5/88.7
Dinomaly	CVPR 2025	98.5/98.2	98.6/**98.7**	88.5/**97.4**
UniVAD (4-shot)	CVPR 2025	98.5/97.3	94.8/98.3	89.3/93.5
TRRC-HETMM	Ours	**99.4/98.7**	**98.7**/97.9	**92.0**/96.9

correct under our method, while 89 change in the opposite direction. The exact McNemar test gives $p \approx 2.61 \times 10^{-254}$.

Table 1 also distinguishes score ranking from operating-point behavior. HETMM reaches 93.18% AUROC but still misses 48 samples at its frozen threshold. Texture disentanglement raises full-branch AUROC to 98.40%, showing that the gain is not merely a threshold shift. CRC primarily changes false positives at a fixed operating point without altering score ranking, and dynamic computation then restores throughput while preserving four misses. The final gain therefore arises from evidence representation, threshold selection, and computational routing rather than from any single module.

5.3 Public datasets

Our method has the highest image AUROC among the tabulated results on all three benchmarks, although its pixel AUROC is below Dinomaly on VisA and Real-IAD. These results support cross-object detection applicability; they do not establish universal superiority in localization or a public-data risk guarantee.

The purpose of the public experiments is to test whether texture disentanglement is overly specific to chip carrier strips, not to extrapolate the industrial risk claim to unrelated datasets. MVTec AD is near saturation and leaves limited room for improvement. VisA and Real-IAD contain more complex objects and viewpoints, yet image-level performance remains strong, suggesting that position-conditioned normal-variation modeling transfers beyond chips. The absence of uniform pixel-level superiority also clarifies that Top-K image aggregation and the risk operating point are optimized for screening, not as a replacement for a dedicated fine-segmentation model.

5.4 Testing the core hypothesis

Relative to the normal mean, define the subspace explanation ratio $\eta = \|UU^\top \Delta\|_2^2/(\|\Delta\|_2^2 + \varepsilon)$. At the main HETMM operating point, η is $0.28/0.43/0.37/0.62$ for TN/FP/TP/FN, respectively, with Cliff's $\delta = 0.61$ between FP and TN. The strip-response ratio is 0.62 for FP and 0.29 for TN. These quantities establish a descriptive association; we do not equate the residual-to-normal-mean proxy with a query-template residual.

The targeted intervention includes all 1,489 normal images in the main test set that satisfy the eligibility criteria specified before evaluation; selection does not use FP/TN outcomes. Let M be the intersection of the strip label and valid ROI. On the actual processing channel, a Gaussian filter

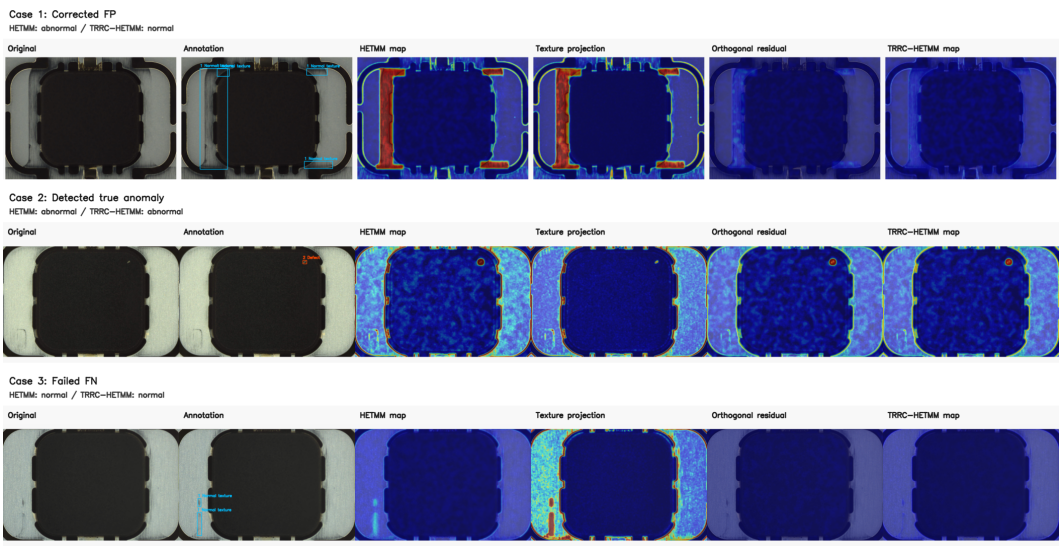


Figure 3: Real cases: a corrected normal-texture false positive, a detected true anomaly, and a strip anomaly missed by both methods. Each row shows the original image, annotation, HETMM map, texture projection, orthogonal residual, and our map.

extracts the low-frequency component $L = G_{k,\sigma}(x)$, and a fixed retention coefficient α reconstructs

$$x^{(T)} = (1 - M) \odot x + M \odot [L + \alpha(x - L)]. \tag{6}$$

Here $\alpha = 1$ means no suppression and $\alpha = 0$ removes the selected high-frequency component completely. A fixed boundary treatment is used, and pixels outside the mask remain unchanged. The processing channel, k, σ, α, and boundary settings cannot be disclosed because of industrial process restrictions, but they were fixed before inspecting the intervention results and applied identically to every sample. The equal-energy control block-shuffles the targeted intervention residual within the same region using a fixed seed and block size. Its perturbation is matched by RMS magnitude over masked pixels, added back to the original image, and clipped to the legal pixel range. The block size and seed are subject to the same restriction. Model, templates, routing, and thresholds remain frozen.

Before intervention, the subset contains 1,098 FP and 391 TN images. The targeted intervention changes 428 FP to TN and 15 TN to FP, reducing the FP rate from 73.74% to 46.00%. The residual-shuffling control changes 79 FP to TN and 38 TN to FP, leaving a 70.99% FP rate. Region-specific detection then checks whether strip suppression broadly removes true anomaly evidence. Among 268 chip-defect images, HETMM and our method detect 255 and 268; among 770 strip-defect images, they detect 735 and 766. All four remaining misses of our method are strip defects. The 38 images carrying both regional labels enter both counts, and no interface-defect sample is available, so we make no preservation claim for that region.

Together, the study provides three layers of evidence. First, higher subspace explanation and strip response among FP than TN images establish an observational association with normal texture directions. Second, selectively weakening the targeted high-frequency component only where the strip intersects the ROI corrects many false positives. Third, an RMS-matched control that disrupts spatial structure is much weaker, ruling out the simple explanation that any perturbation of equal magnitude lowers the score. Region-specific detection simultaneously tests whether the gain comes from uniformly suppressing anomaly responses. Because all four remaining misses occur on the strip, the evidence supports the proposed mechanism but does not imply that strip defects have been fully solved.

5.5 Dynamic paths, efficiency, and ablation

The fast path directly decides 17,716 images (TP/FN/TN/FP = 528/0/15,240/1,948), while 2,284 images fall back to full matching (468/4/995/817). Fast normal release covers 76.20% of all images,

Table 3: Main-module ablation. Over-inspection is in percent; best values (including ties) are bold.

Configuration	AUROC	FN	FP	Over-inspection	FPS
HETMM	93.18	48	3,895	20.50	72.9
TD	**98.40**	**4**	3,218	16.94	69.0
TD+CRC	**98.40**	**4**	2,839	14.94	69.0
TD+CRC+SDC	**98.40**	**4**	**2,765**	**14.55**	**78.3**

and full fallback covers 11.42%. The fast path therefore has no observed miss on this fixed test set, but this observation is not a guarantee for future distributions.

End-to-end throughput rises from 72.9 to 78.3 FPS (+7.41%). P95 changes from 13.8 to 13.7 ms, while P99 increases from 15.8 to 19.9 ms. Dynamic computation improves average throughput but not tail latency. The main ablation appears below; the complete combinations and γ sensitivity are reported in Appendix E.

The fast path is valuable because it handles many unambiguous normal samples, not because it exits early on every image. Deployment-log review confirms that the four missed samples in the fallback group did enter the full branch. They are therefore difficult strip defects that full matching itself failed to recognize, rather than routing errors. The higher P99 comes from a small number of images that incur both routing and full matching, motivating future work on batched or asynchronous fallback. We do not use the gain in average FPS to conceal the degradation in tail latency.

6 DISCUSSION AND LIMITATIONS

Scope of the conclusions. The industrial dataset provides a stringent and relatively large fixed test, but it still comes from six cameras, three illumination stations, and the current production process. CRC relies on exchangeability between calibration and future samples. A systematic shift in upstream material, camera, illumination, or defect composition invalidates automatic reuse of the original budget. Deployment should continuously monitor score distributions, fallback rates, and audit errors, and must refreeze the configuration and recalibrate thresholds after a process change.

Anomaly coverage and localization. Template matching does not require every defect category to be enumerated in advance, but the test set can still cover only anomalies that were actually collected. Four strip anomalies remain missed by the full branch, and no interface anomaly is available. We therefore cannot claim zero misses on the strip or interface. Because the fast path produces no heatmap, its localization ability cannot be inferred from image-level classification. Applications requiring defect contours should run the full branch or a dedicated pixel-level model.

Limits to reproducibility. The raw chip images and some intervention parameters are restricted by company authorization and process confidentiality. Explicit data roles, frozen protocols, public benchmarks, and an equal-energy control reduce the verification difficulty caused by unavailable data, but they cannot make the industrial intervention fully reproducible by third parties. Public-data experiments provide independent cross-object evidence; they do not replace external verification in the restricted production setting.

7 CONCLUSION

TRRC-HETMM unifies texture-aware residual reweighting, independent miss-risk calibration, and dynamic computation within template mutual matching. Experiments support reduced over-inspection and cross-object applicability; routing improves average throughput but not tail latency. The limitations above preclude claims of zero misses or robustness to arbitrary drift. Coverage of unknown defects remains future work.

REPRODUCIBILITY STATEMENT

We specify the entity-disjoint splits, frozen strategy, calibration protocol, and timing protocol. The supplementary archive provides the implementation used on the experimental server for this study. Company confidentiality and authorization restrictions prevent release of the raw chip dataset and the identified intervention parameters, so the code release alone does not enable complete independent reproduction of the industrial experiments. Accuracy and timing comparisons also require the corresponding data, configurations, and hardware/software environment described in the paper.

ETHICS STATEMENT

The company authorized the data solely for this research. The dataset contains no human subjects, personally identifiable information, or personal data. During initial deployment validation, every model output is manually verified. Routine operation switches to sampling-based manual audits only after validation over a sufficient period and data volume. Model evaluation is kept separate from production outcomes after manual correction, and sampling-based audits do not guarantee correctness for every unaudited sample. Statistical risk control is not an unconditional production-safety guarantee.

AI USE STATEMENT

Artificial-intelligence tools were used for language polishing, typographical checks, and consistency review. They were not used to generate, modify, or replace experimental data, nor did they replace the authors' verification of the experimental design, result analysis, or conclusions. The authors reviewed and take responsibility for the final content.

REFERENCES

Anastasios N. Angelopoulos, Stephen Bates, Adam Fisch, Lihua Lei, and Tal Schuster. Conformal risk control. In *International Conference on Learning Representations*, 2024.

Anastasios N. Angelopoulos, Stephen Bates, Emmanuel J. Candès, Michael I. Jordan, and Lihua Lei. Learn then test: Calibrating predictive algorithms to achieve risk control. *The Annals of Applied Statistics*, 19(2):1641–1662, 2025. doi: 10.1214/24-AOAS1998.

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*, pp. 128–138, 2024.

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*, pp. 4183–4192, 2020.

Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 129:1038–1059, 2021. doi: 10.1007/s11263-020-01400-4.

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, 2022. doi: 10.1007/s11263-022-01578-9.

Zixuan Chen, Xiaohua Xie, Lingxiao Yang, and Jian-Huang Lai. Hard-normal example-aware template mutual matching for industrial anomaly detection. *International Journal of Computer Vision*, 133:2927–2949, 2025. doi: 10.1007/s11263-024-02323-0.

Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020.

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.

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*, pp. 475–489. Springer, 2021.

Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9737–9746, 2022.

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*, pp. 35555–35566, 2026.

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, 2025.

Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. CFLOW-AD: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 98–107, 2022.

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*, pp. 20405–20415, 2025.

Chaoqin Huang, Haoyan Guan, Aofan Jiang, Ya Zhang, Michael Spratling, and Yan-Feng Wang. Registration based few-shot anomaly detection. In *Computer Vision – ECCV 2022*, pp. 303–319. Springer, 2022. doi: 10.1007/978-3-031-20053-3_18.

Gao Huang, Danlu Chen, Tianhong Li, Felix Wu, Laurens van der Maaten, and Kilian Q. Weinberger. Multi-scale dense networks for resource efficient image classification. In *International Conference on Learning Representations*, 2018.

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.

Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. CutPaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9664–9674, 2021.

Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. SimpleNet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20402–20411, 2023.

Philipp Liznerski, Lukas Ruff, Robert A. Vandermeulen, Billy Joe Franks, Marius Kloft, and Klaus-Robert Müller. Explainable deep one-class classification. In *International Conference on Learning Representations*, 2021.

Tapan Ganatma Nakkina, Yuhao Zhong, Pete Sumethasorn, Haopeng Tian, and Satish Bukkapatnam. When textures deceive: Weakly supervised industrial anomaly detection with adapted-loss CycleGAN. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 4107–4116, 2025.

Yongming Rao, Wenliang Zhao, Benlin Liu, Jiwen Lu, Jie Zhou, and Cho-Jui Hsieh. DynamicViT: Efficient vision transformers with dynamic token sparsification. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

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.

Marco Rudolph, Bastian Wandt, and Bodo Rosenhahn. Same same but DifferNet: Semi-supervised defect detection with normalizing flows. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1906–1915, 2021. doi: 10.1109/WACV48630.2021.00195.

Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4393–4402, 2018.

Surat Teerapittayanon, Bradley McDanel, and H. T. Kung. BranchyNet: Fast inference via early exiting from deep neural networks. In *23rd International Conference on Pattern Recognition*, pp. 2464–2469, 2016. doi: 10.1109/ICPR.2016.7900006.

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.

Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for anomaly detection. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2021. doi: 10.5244/C.35.349.

Jihun Yi and Sungroh Yoon. Patch SVDD: Patch-level SVDD for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision*, 2020.

Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. FastFlow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677*, 2021.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. DRAEM: A discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8330–8339, 2021.

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.

Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. DeSTSeg: Segmentation guided denoising student-teacher for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3914–3923, 2023.

Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. AnomalyCLIP: Object-agnostic prompt learning for zero-shot anomaly detection. In *International Conference on Learning Representations*, 2024.

Wei Zhuo, Jianen Xiang, Miaomiao Liu, and Huajun Lu. PTAD: Pose and texture agnostic anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Findings*, pp. 6779–6788, 2026.

Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. SPot-the-Difference self-supervised pre-training for anomaly detection and segmentation. In *European Conference on Computer Vision*, pp. 392–408, 2022.

A METHOD DETAILS

For each layer l and canonical position p, we collect $\mathcal{V}_{l,p} = \{F_l(z_i)_q : q \in \mathcal{N}_{r_l}(p)\}$ from neighborhoods of normal templates and perform eigendecomposition after MAD clipping and covariance shrinkage. The subspace dimension is

$$k_{l,p} = \min\left(k_{\max}, \min\left\{k : \frac{\sum_{j=1}^{k}\lambda_j}{\sum_j \lambda_j} \geq \gamma\right\}\right). \tag{7}$$

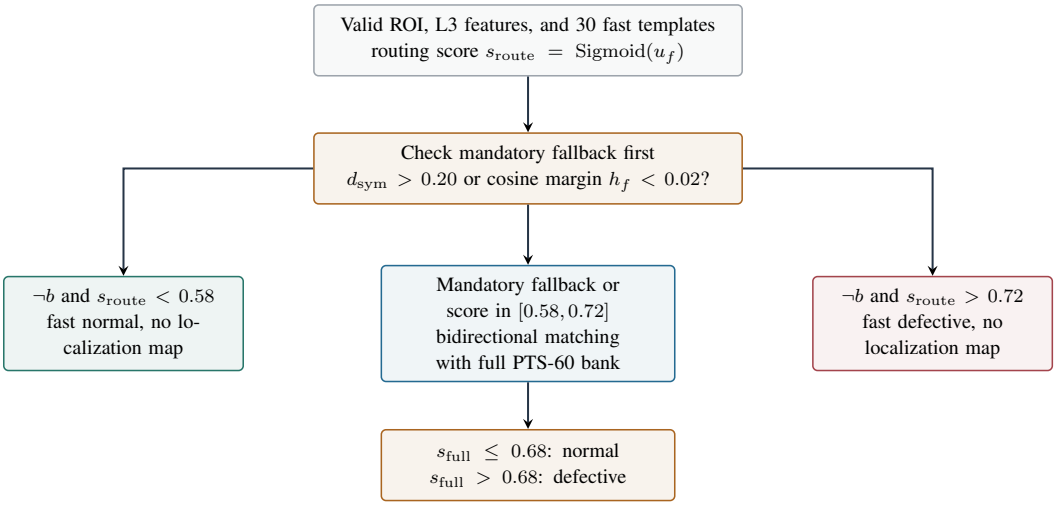
Let $P = UU^\top$. Then Equation (2) is equivalent to $d_{\mathrm{TD}} = \delta^\top M\delta$, where $M = w^\perp(I-P) + w^\parallel U\bar{\Lambda}^{-1}U^\top$. After eigenvalue clipping, nonnegative weights make M positive semidefinite. This is a position-conditioned quadratic metric, not a new PCA or whitening theory, and the implementation uses no additional position-reliability weight.

Forward matching fixes the query anchor, whereas backward matching fixes the template anchor:

$$S^{\rightarrow}_{l,p} = \min_{i,q\in\mathcal{N}_{r_l}(p)} d_{\mathrm{TD}}(F_l(x)_p, t_{i,q}; p, l), \tag{8}$$

$$S^{\leftarrow}_{l,p} = \max_i \min_{q\in\mathcal{N}_{r_l}(p)} d_{\mathrm{TD}}(F_l(x)_q, t_{i,p}; p, l). \tag{9}$$

The two directions and multiple levels are combined with frozen weights. The image score is the Top-20 mean of the smoothed anomaly map inside the valid ROI. Fast-route prototype descriptors are spatially averaged over the ROI from L3 features and L2-normalized. The raw routing score uses a standard Sigmoid. The prototype-margin guard is the difference between the first and second cosine similarities, and its threshold 0.02 is measured on the original cosine scale. The full branch directly accesses PTS-60 without candidate-template preselection or template augmentation.



Risk event: $A_{\mathrm{sys}} = A_f \cup (B_g \cap A_g) \subseteq A_f \cup A_g$
The full branch is calibrated on every defective sample; a path-conditional miss rate is not a branch risk budget.

Figure 4: Frozen dynamic inference. Mandatory fallback is checked before either fast-normal or fast-defective decisions.

B FINITE-SAMPLE INTERPRETATION OF RISK CALIBRATION

Let A_f denote fast normal release, B_g execution of full matching, and A_g normal release by the full branch. For a defective sample,

$$A_{\mathrm{sys}} = A_f \cup (B_g \cap A_g) \subseteq A_f \cup A_g, \qquad \mathbb{E}_{\mathrm{cal}} R_{\mathrm{sys}} \leq \epsilon_f + \epsilon_g. \tag{10}$$

For calibration, the fast-policy score is $+\infty$ when a guard forces fallback and otherwise equals s_{route}. Both calibrators count $s_j \le \tau$ as a miss. Fast normal release requires $s_{\text{route}} < \tau_f$, so its miss event is contained in the calibrated event; equality causes fallback. Full normal release uses $s_{\text{full}} \le \tau_g$, exactly matching its calibration loss. Even with zero calibration misses, the CRC correction is $1/(n_1+1)$. Here $n_1 = 600$ supports both prespecified branch budgets. Zero misses after threshold selection cannot simultaneously be treated as an ordinary binomial confidence statement for an unselected threshold on the same data. Updating the model, routing, or templates invalidates automatic reuse of the original calibration result. Test intervals for an independently frozen strategy and expectation-over-calibration control are distinct claims. Correlated images cannot be treated as independent samples to inflate the effective sample size.

C COMPLETE INFERENCE PROCEDURE

Algorithm 1 Frozen binary policy of TRRC-HETMM

Require: Image x, valid ROI, frozen features and subspaces, PTS-60 templates, frozen thresholds
Ensure: Binary output $\hat{y}$, route label, and full anomaly map when available

```
Extract shared features; compute s_route, d_sym, and guard quantities
b ← (d_sym > 0.20) or any other guard is triggered
if ¬b and s_route < 0.58 then
    return normal, fast path, empty localization map
else if ¬b and s_route > 0.72 then
    return defective, fast path, empty localization map
end if
Run texture-disentangled bidirectional matching over every PTS-60 template
Fuse anomaly map S and aggregate its Top-20 responses inside the valid ROI into s_full
return 𝕀[s_full > 0.68], full path, S
```

The fast branch produces no complete localization map; a fast-normal decision must not be interpreted as pixelwise localization correctness. Public pixel-level results run full matching on every image. If full matching is run separately for an early-exit sample to obtain a localization map, that extra computation belongs to offline evaluation and is not counted as a realized fast pixel-localization capability.

D REGION STATISTICS AND INTERVENTION DETAILS

In the fixed H1–H8 masks, label 4 denotes the valid ROI, label 1 the chip, label 3 the strip, and label 2 plus ROI pixels covered by neither chip nor strip the interface. These regions occupy 45.81%, 28.23%, and 25.96% of the ROI. Every image uses the same crop, alignment, resizing, and fixed spatial masks; pixels outside the ROI are ignored. The strip-response ratio is

$$\rho_{\text{strip}} = \frac{\sum_{p\in\Omega_{\text{strip}}} S_p}{\sum_{p\in\Omega_{\text{inspect}}} S_p + \varepsilon}. \tag{11}$$

A high strip-response ratio alone does not prove a false positive because a real strip defect can also produce a high ratio. The normal-image intervention changes only the confirmed-normal strip region while preserving the chip and geometry, and includes an equal-energy residual block-shuffling control. Intervention parameters were fixed before outcomes were inspected. The processing channel, filter scale, suppression coefficient, boundary settings, block size, and random seed cannot be disclosed because of industrial process restrictions.

E COMPLETE ABLATIONS AND SENSITIVITY

For TD, $\gamma = 0.80/0.85/0.90/0.95/0.98$ gives over-inspection rates of $18.8/17.5/16.9/17.7/18.1\%$ and FN counts of $6/5/4/6/11$. The validation set selects $\gamma = 0.90$.

Table 4: Definitions and outcomes of the documented H1–H8 mechanism checks. Region counts are non-exclusive because 38 images carry both chip and strip labels.

ID	Question and statistic	Observed result
H1	Does targeted normal-strip suppression selectively correct HETMM false positives?	Among 1,489 eligible normal images, baseline FP/TN=1,098/391. Targeted suppression changes 428 FP→TN and 15 TN→FP, reducing FP rate from 73.74% to 46.00%. The RMS-matched control changes only 79 FP→TN and 38 TN→FP (70.99% FP rate).
H2	Are regional anomaly responses preserved?	For 268 chip-defect images, HETMM/ours detect 255/268 and response retention is 98.9%. For 770 strip-defect images, they detect 735/766 and retention is 98.3%. No interface-defect sample is available.
H3	Is texture explanation associated with false positives?	The FP–TN texture-explanation effect is Cliff's $\delta = 0.61$. This is descriptive; regression on the thresholding score is omitted because it induces complete or quasi-complete separation.
H4	TN residual profile ($n = 15{,}105$)	Texture explanation 0.28 ± 0.11; normalized orthogonal residual 0.36 ± 0.14; strip-response ratio 0.29 ± 0.17.
H5	FP residual profile ($n = 3{,}895$)	0.43 ± 0.13; 0.32 ± 0.13; 0.62 ± 0.16, respectively. FP have higher texture explanation and strip response without higher orthogonal residual.
H6	TP residual profile ($n = 952$)	0.37 ± 0.19; 1.13 ± 0.54; 0.40 ± 0.24, respectively. The large orthogonal residual is consistent with detected evidence outside normal texture directions.
H7	FN residual profile ($n = 48$)	0.62 ± 0.09; 0.29 ± 0.08; 0.67 ± 0.11, respectively, consistent with difficult anomalies aligned with strip texture directions.
H8	What regions do the fixed masks cover?	Chip/strip/interface occupy 45.81/28.23/25.96% of the ROI; non-exclusive abnormal-image counts are 268/770/0. Pixels outside the ROI are ignored.

Algorithm 2 Restricted-parameter texture intervention and control

Require: Confirmed-normal image x, strip mask M_s, valid ROI M_r, frozen HETMM, fixed confidential parameters

1: $M \leftarrow M_s \cap M_r$; keep every pixel outside M unchanged
2: Extract low-frequency component $L \leftarrow G_{k,\sigma}(x)$ on the fixed processing channel
3: Form targeted image $x^{(T)} \leftarrow (1 - M) \odot x + M \odot [L + \alpha(x - L)]$
4: Block-shuffle the signed intervention residual $x^{(T)} - x$ inside M with the fixed seed
5: Scale the shuffled perturbation to match the masked RMS magnitude of $x^{(T)} - x$
6: Add the control perturbation to x, clip to the legal range, and keep the same mask boundary rule
7: Run the frozen model on x, $x^{(T)}$, and the control; record scores, maps, and paired decision flips

Table 5: Complete module ablation; best values (including ties) are bold. TD: texture disentanglement; CRC: conformal risk control; SDC: safe dynamic computation.

Configuration	AUROC	FN	FP	Over-inspection/%	FPS
HETMM	93.18	48	3,895	20.50	72.9
TD	**98.40**	**4**	3,218	16.94	69.0
TD+CRC	**98.40**	**4**	2,839	14.94	69.0
TD+CRC+SDC	**98.40**	**4**	**2,765**	**14.55**	**78.3**
CRC	93.18	8	4,560	24.00	–
SDC	93.18	10	4,408	23.20	–
CRC+SDC	93.18	10	4,294	22.60	–
TD+SDC	**98.40**	**4**	3,249	17.10	77.0
Random projection	94.20	8	4,085	21.50	–

F VISUAL SUMMARY OF EXPERIMENTS

Figure 5 plots the reported subspace sensitivity results. Markers denote evaluated configurations; connecting segments are visual guides. The two metrics use separate panels rather than a dual-axis scale.

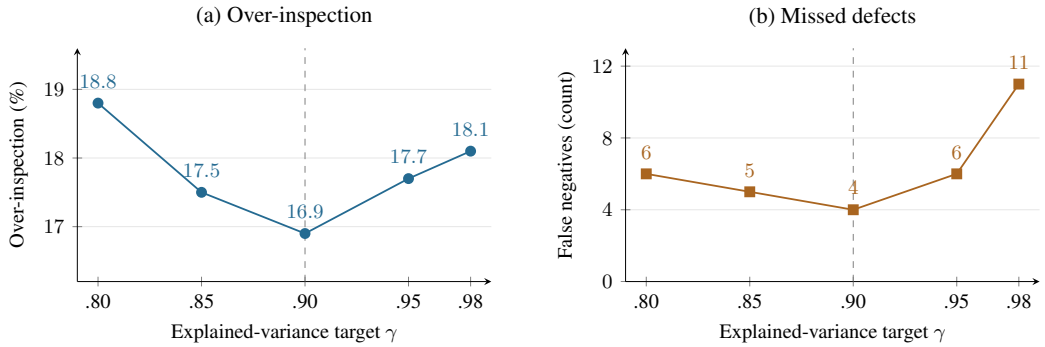


Figure 5: Effect of γ on over-inspection and missed defects, using the values in Appendix E. The dashed line marks $\gamma = 0.90$, selected during validation.

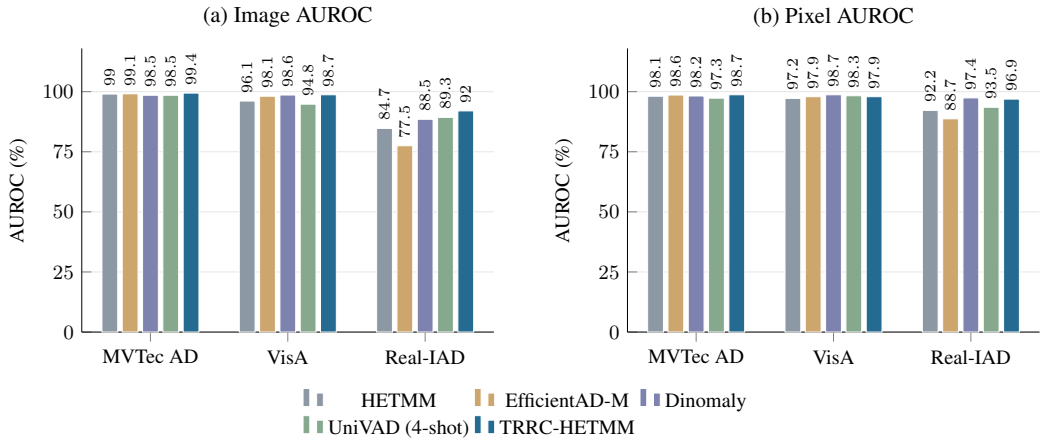


Figure 6: Public-benchmark results from Table 2, with image and pixel AUROC shown separately. Both bar-chart axes start at zero. All values are unified-environment reruns reported in this paper.

G NOTES ON THE REAL QUALITATIVE CASES

The first case in Figure 3 shows suppression of a normal strip-texture response; the second preserves a local true-anomaly response; and the third is missed by both methods, demonstrating that normal-texture modeling cannot eliminate every miss. Heatmap colors are not used to compare absolute scores across images. N5 follows an independent annotation scheme: label 1 is normal texture, label 2 is a defect, and label 3 is the valid ROI. These numbers do not share semantics with the region-statistics masks.