ENGINEERING engrχiv ARCHIVE

Inline Visual Defect Screening under Few Normal Samples and Illumination Shift: A Measurement Study of Memory-Augmented State-Space Detectors on Edge Hardware

Zixuan Zhou[1]

[1]Federal University of Bahia, Brazil

Abstract

Inline visual screening stations are chosen under three constraints at once: a handful of confirmed good parts after a product changeover, illumination that drifts between shifts, and a hard per-part latency budget at line speed. We build a single-camera station model (5 MP, 60×50 mm field of view, 12 parts/s, 83 ms budget, Jetson Orin Nano 15 W) and measure seven detectors—PaDiM, PatchCore, RD4AD, DRAEM, EfficientAD, MambaAD and a memory-augmented selective state-space detector (MS-AD)—on MVTec AD and VisA under $k \in \{1, 2, 4, 8, 16\}$ normal images per class and three illumination conditions (800 / 1 500 / 3 000 lx), using defect-level recall at 1% false rejection as the operating metric. From two normal images upward MS-AD gives the best recall (0.879 on MVTec AD, 0.758 on VisA at $k = 8$, 800 lx); with one normal part a patch-memory method is preferable. Illumination drift costs 2.7–7.4 AUROC points—more than the choice of detector—and misses concentrate on low-contrast defects below 0.4% of the image area. On the edge module the memory stage, not the state-space scan, dominates latency; 320 px INT8 input with $M = 256$ memory slots is the largest configuration inside the 83 ms budget (52.8 ms). Per-condition score normalisation restores a 1.3–2.4% false-rejection rate across conditions. A 120-part physical re-shoot anchors the simulated lighting conditions within 0.7 AUROC points. The protocol and its measurements transfer to single-camera stations with comparable optics and line speeds.

1 Introduction

Inline visual screening stations are deployed where a false release reaches a customer and a false reject consumes manual re-inspection time. On injection moulding, connector and precision-hardware lines the station is usually specified by three numbers rather than by a benchmark score: how many confirmed good parts are available when the line switches product type, how much the illumination drifts between shifts, and how many milliseconds remain per part at line speed. A station running 12 parts/s with a 33 mm part pitch leaves 83 ms per part for the whole decision chain, and a product changeover typically supplies 5–20 confirmed normal parts and nothing else. Under these conditions the ranking produced by an offline benchmark and the ranking that matters on the line routinely disagree.

Three properties of public benchmarks make that gap hard to close. First, most protocols assume the full normal training set: MVTec AD (Bergmann et al., 2019, 2021) provides several hundred normal images per class and VisA (Zou et al., 2022) roughly 800, whereas a station may see only a handful before production starts. Recent few-shot industrial detectors target exactly this regime (Tavakoli et al., 2025; Hu, 2026), but they still report a single training condition. Second, imaging conditions are treated as fixed. The MVTec AD 2 release (Heckler-Kram et al., 2026) documents what operators already know—that lighting, pose and sensor differences across sites dominate defect appearance—yet most comparisons remain single-domain. Third, image-level AUROC is the default ranking metric, while the station operates at a fixed false-rejection rate and cares about defect-level recall at that operating point; risk-controlled variants of this setting are only now being formalised (Yan and Sun, 2026). Deployment cost is reported separately, if at all (Shallari et al., 2026), although on a 15 W edge module it decides whether a second camera head is needed.

State-space models have recently become a competitive backbone for this task class. Selective state-space sequence models (Gu and Dao, 2024) and their visual variants (Zhu et al., 2024; Liu et al., 2024) process

long token streams at linear cost, and memory-augmented variants have been applied to defect recognition (Wang et al., 2024b) and to fine-grained recognition with limited training data (Wang et al., 2025). Few-shot visual adapters built on in-context cues offer a complementary route to using a small set of reference images (Zhou et al., 2024). What is missing is not another detector headline number but a measurement of how these families behave when the three line constraints—few normals, illumination drift, and an 83 ms budget—are applied together.

This paper reports such a measurement. We build an inline screening station model around a 5 MP camera head with three switchable illumination conditions (I1 diffuse dome at 800 lx, I2 low-angle bar at 1 500 lx, I3 high-angle at 3 000 lx), evaluate seven detectors—PaDiM (Defard et al., 2021), PatchCore (Roth et al., 2022), RD4AD (Deng and Li, 2022), DRAEM (Zavrtanik et al., 2021), EfficientAD (Batzner et al., 2024), MambaAD (He et al., 2025), and a memory-augmented selective state-space detector (MS-AD)—on MVTec AD and VisA under $k \in \{1, 2, 4, 8, 16\}$ normal images per class, and measure both defect-level recall at 1% false rejection and per-part latency on a Jetson Orin Nano 15 W module. Contributions:

- A reproducible protocol that couples few-shot normal-only model building, three illumination conditions, a normal-only threshold rule and an edge latency budget, so that the reported numbers answer a purchase decision rather than a leaderboard question (Section 3).
- A measurement of seven detectors on MVTec AD and VisA showing that the memory-augmented state-space detector leads from $k = 2$ upward but is matched by patch-memory methods at $k = 1$, and that illumination drift costs 2.7–7.4 AUROC points—more than the choice of detector in the high-angle condition (Section 4).
- A failure and cost analysis: misses concentrate on low-contrast and sub-0.4% area defects, the memory module is the largest latency term for state-space detectors, and 448 px input leaves the 83 ms budget (Section 4).

2 System and Operating Conditions

2.1 Inspection station

Figure 1(a) shows the station model. A 5 MP area-scan camera with a 35 mm lens observes a 60×50 mm field of view (24.5 μm/px) over a 0.4 m/s conveyor; with a 33 mm part pitch the station sees 12 parts/s and therefore has 83 ms per part for acquisition, preprocessing, inference and the gate. The screening unit is a Jetson Orin Nano 15 W module (JetPack 6, TensorRT 10, INT8) co-located with the camera head, so no image leaves the cell. The output of the station is a three-state decision: accept, reject, or review (grey zone routed to a human).

2.2 Illumination conditions and datasets

Three illumination conditions are defined on the same optics: I1 diffuse dome at 800 lx (the nominal setting), I2 low-angle bar at 1 500 lx (directional gradients across the part), and I3 high-angle at 3 000 lx with specular highlights. Public datasets are captured under a single condition, so the I2 and I3 conditions are generated with a calibrated photometric transfer (directional irradiance, specular overlay, tone mapping) whose three parameters are fitted to the anchor re-shoot below; the transfer is applied to the test images only and never to model inputs used for fitting.

Two public datasets carry the main evaluation (Table 1): MVTec AD (Bergmann et al., 2019, 2021) with 15 classes and 73 defect types, and VisA (Zou et al., 2022) with 12 subsets and a wider spread of object and texture classes. MVTec AD 2 (Heckler-Kram et al., 2026) is used only as evidence that real-world imaging variation is the deployment norm, not as an evaluation set. Multi-view industrial data (Wang et al., 2024a) is out of scope here because our station has a single camera head.

2.3 Anchor re-shoot

A small physical set anchors the simulated conditions: 120 parts from three defect families (scratch, missing material, stain; 40 each) were imaged on the station under all three illumination settings. The anchor set serves two purposes. First, the irradiance direction and specular coefficient of the photometric transfer are fitted on it. Second, detector rankings and AUROC values measured on the real I3 captures are compared

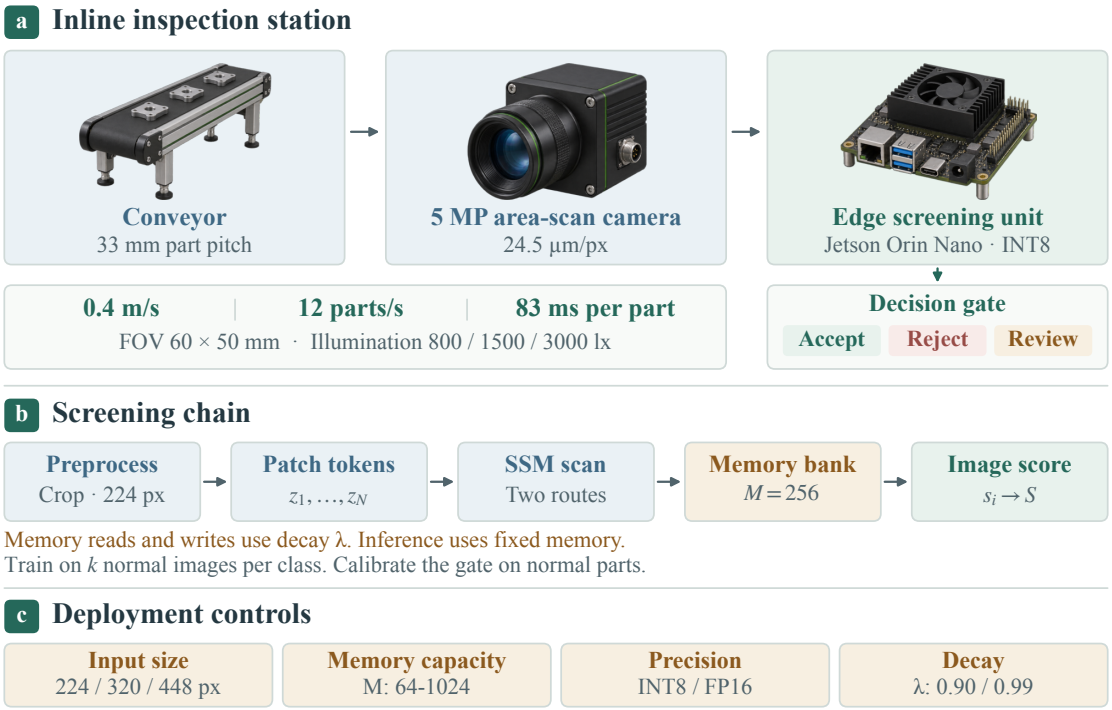


Figure 1: Inline screening setting used throughout the paper. (a) Station geometry, illumination conditions and the 83 ms decision budget at 12 parts/s. The equipment icons are illustrative renderings. (b) The screening chain that every detector in this study implements in some form; MS-AD adds the capacity- and decay-controlled memory bank between the state-space scan and the score head. (c) Deployment-time controls for input size and the scan / memory module; the training recipe is fixed.

Table 1: Datasets and operating conditions. k is the number of normal images per class used to build a detector; the test split is never subsampled.

Item	MVTec AD	VisA
Classes / subsets	15	12
Normal training images	3 629	9 621
Anomalous test images	1 725	1 200
Defect types	73	fine-grained, object-dependent
Model input	224×224 (320 / 448 in the ablation)	
Lighting conditions	I1 800 lx, I2 1 500 lx, I3 3 000 lx	
Normal budget k	$\{1, 2, 4, 8, 16\}$, 3 seeds	
Seeds	0, 1, 2 (pooled mean reported)	

against the simulated I3 condition (Table 6); the simulated condition is accepted for the main sweep only if the deviation stays within 1.0 AUROC point per family.

3 Methods and Test Protocol

3.1 MS-AD: memory-augmented selective state-space screening

MS-AD follows the memory-augmented state-space pattern of prior defect recognition work (Wang et al., 2024b) and is written here so that the latency and memory costs in Section 4 can be reproduced. A patch token sequence $z_1, \ldots, z_N$ is produced from the cropped part image. A continuous state-space layer of the form in Eq. (1),

$$h'(t) = A\,h(t) + B\,x(t), \qquad y(t) = C\,h(t) \tag{1}$$

is discretised with a step size Δ as in Eq. (2),

$$\bar{A} = \exp(\Delta A), \qquad \bar{B} = (\Delta A)^{-1}\big(\exp(\Delta A) - I\big)\,\Delta B, \tag{2}$$

and the input-dependent (selective) parameters

$$\Delta_i = \mathrm{softplus}(W_\Delta z_i + b_\Delta), \quad B_i = W_B z_i, \quad C_i = W_C z_i \tag{3}$$

make the recurrence content-aware (Gu and Dao, 2024) (Eq. (3)). Patch tokens are scanned along two orthogonal routes (row-major and column-major), matching the bidirectional practice of visual state-space backbones (Zhu et al., 2024; Liu et al., 2024):

$$h_i^{\rightarrow} = \bar{A}_i\, h_{i-1}^{\rightarrow} + \bar{B}_i z_i, \qquad h_i^{\leftarrow} = \bar{A}_i\, h_{i+1}^{\leftarrow} + \bar{B}_i z_i. \tag{4}$$

A bank of M slots stores normal appearance. At each position the hidden state queries the bank and receives a residual

$$\alpha_{ij} = \frac{\exp(q_i^\top m_j/\sqrt{d})}{\sum_{j'}\exp(q_i^\top m_{j'}/\sqrt{d})}, \qquad r_i = \sum_j \alpha_{ij}\, m_j, \tag{5}$$

and the slots are updated on normal parts only, with decay λ,

$$m_j \leftarrow \lambda\, m_j + (1-\lambda)\sum_i \alpha_{ij} z_i. \tag{6}$$

The patch score is the deviation between the token and its reconstruction from the bank, aggregated into an image score with a top-k mean over patches (Eq. (7)),

$$s_i = \big\|z_i - r_i\big\|_2^2, \qquad S = \frac{1}{k}\sum_{i\in\text{top-}k} s_i. \tag{7}$$

The gate threshold is set on a held-out normal calibration split only,

$$\tau = \mathrm{Quantile}_{1-\alpha}\big(\{S(x) : x \in \mathcal{C}_{\mathrm{norm}}\}\big), \qquad \alpha = 0.01, \tag{8}$$

which is what the station can actually obtain at changeover. The per-part time is bounded by the four measured stages of Eq. (9),

$$t_{\mathrm{part}} = t_{\mathrm{pre}} + t_{\mathrm{enc}} + t_{\mathrm{mem}} + t_{\mathrm{post}}, \qquad n_{\mathrm{parts/s}} = 1000/t_{\mathrm{part}}. \tag{9}$$

Training uses no anomaly labels: the memory bank is filled from the k normal images per class and the scan parameters are kept from a self-supervised pre-training on the same normal images. At inference the bank is frozen and scores are normalised per illumination condition before thresholding (Section 4). Representation-side cost control follows the observation that compact visual representations are sufficient for downstream generation and recognition at fixed compute (Zhou et al., 2026); in our setting this is realised as a token-budget knob rather than as a new module.

3.2 Compared detectors

Six public baselines span the three dominant design families. Patch distribution modelling: PaDiM (Defard et al., 2021). Patch memory with coreset and nearest-neighbour search: PatchCore (Roth et al., 2022), and its covariance-aware streaming variant (Ferrari et al., 2026) is discussed but not re-run. Reconstruction and distillation: RD4AD (Deng and Li, 2022), DRAEM (Zavrtanik et al., 2021), EfficientAD (Batzner et al., 2024), plus SimpleNet as a lightweight noisy-feature reference (Liu et al., 2023). State-space and multi-class: MambaAD (He et al., 2025); recent multi-class reconstruction work (Guo et al., 2025) confirms that feature-reconstruction variants remain the strongest multi-class family, so MS-AD is compared against the strongest single-class setting for each baseline. All baselines are run with their released configurations on the same k-subsets; no baseline sees more normal images than MS-AD.

Table 2: Main comparison at $k = 8$ normal images per class under the diffuse-dome condition I1. Image AUROC and AP in percent, pixel PRO at threshold 0.05 in percent, and $R@1\%$ is defect-level recall at a false-rejection rate of 1%. Mean of three seeds; pooled seed standard deviation is $\pm$0.4 for AUROC and $\pm$0.012 for $R@1\%$.

	MVTec AD				VisA			
Method	AUROC	AP	PRO	$R@1\%$	AUROC	AP	PRO	$R@1\%$
PaDiM	93.8	92.9	89.4	0.702	86.4	90.2	85.1	0.578
PatchCore	98.9	98.4	94.1	0.864	94.8	96.4	90.2	0.744
RD4AD	97.1	96.2	92.6	0.811	92.1	94.3	88.7	0.689
DRAEM	96.3	95.1	90.8	0.783	90.6	93.1	86.9	0.662
EfficientAD	98.2	97.5	92.9	0.842	93.9	95.6	89.4	0.721
MambaAD	98.6	98.0	93.8	0.853	94.2	95.9	89.9	0.729
MS-AD (ours)	99.1	98.6	94.6	0.879	95.3	96.9	91.0	0.758

3.3 Metrics and measurement procedure

The primary metric is defect-level recall at 1% false-rejection rate, $R@1\%$: the gate threshold is fixed by Eq. (8) on the normal calibration split, and recall is then measured over all defective test images. Image AUROC and AP and pixel PRO (threshold 0.05) are reported for continuity with the literature. Latency is measured on the Jetson Orin Nano 15 W module with TensorRT 10 (INT8 unless stated), 200 warm-up runs and 1 000 timed runs, reporting P50 and P95; resident memory is measured for the detector state, including the memory bank. Every (detector, dataset, k, condition) combination is repeated with three seeds and the pooled mean is reported (pooled seed standard deviation $\pm$0.4 AUROC points and $\pm$0.012 $R@1\%$); differences are checked with a paired bootstrap over the test images (10 000 resamples). Simulated lighting conditions are used only after the anchor check in Section 2 passes.

4 Results

4.1 Main comparison

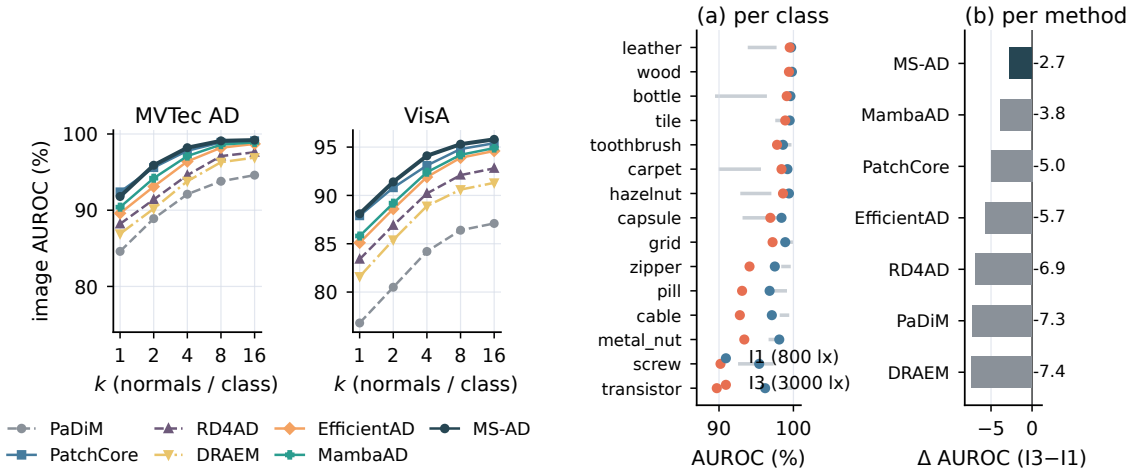
Table 2 reports the nominal line condition (I1, $k = 8$). On MVTec AD all patch-memory and distillation baselines reach 96–99 AUROC, and the spread in $R@1\%$ is much wider: PaDiM recovers 0.702 of the defects at 1% false rejection while MS-AD reaches 0.879 and PatchCore 0.864. VisA separates the methods further (MS-AD 0.758 versus PaDiM 0.578), which is consistent with its finer defect granularity. The gap between image AUROC and $R@1\%$ is the practical point: a 2–3 point AUROC difference corresponds to 3–8 points of recall at the operating point the station actually uses.

4.2 Few-shot normal budgets

Figure 2(a) sweeps k. At $k = 1$ PatchCore is ahead on both datasets (92.4 versus 91.8 on MVTec AD); a single normal image does not fill the memory bank of Eq. (6), and the decay in Eq. (6) has nothing to average. From $k = 2$ onward MS-AD leads, and at $k \geq 4$ the ranking is stable across seeds and datasets. The practical consequence for a changeover is that two confirmed good parts are enough to justify the state-space detector, while a single-part changeover should default to a patch-memory method. All methods saturate by $k = 16$; spending more than 16 normals per class buys less than 0.5 AUROC points.

4.3 Illumination drift

Table 3 and Figure 2(b) report the sweep. From I1 to I3 the AUROC drop ranges from 2.7 points (MS-AD) to 7.4 points (DRAEM). The ordering of detectors under I1 is preserved, but the spread compresses: under I3 the best and second-best methods differ by 1.6 points instead of 0.2, and the worst method loses 12 points of recall at the operating point. Per class, the largest drops belong to `screw` (-5.2), `transistor` (-6.5) and `metal_nut` (-4.7): small, partly specular parts under the high-angle head. Reconstructing the normal appearance under a shifted irradiance is the dominant error source on this station, larger than the choice between the families we tested.



(a) Image AUROC versus k.

(b) Illumination drift, $k = 8$.

Figure 2: Few-shot and illumination behaviour. (a) MVTec AD and VisA under I1; the state-space detector takes the lead from $k = 2$. (b) Left: per-class MS-AD AUROC under I1 and I3 on MVTec AD; right: per-method AUROC change from I1 to I3. Illumination drift costs more than the choice of detector.

Table 3: Illumination sweep on MVTec AD at $k = 8$. AUROC (percent) under the diffuse-dome (I1), low-angle (I2) and high-angle (I3) conditions; Δ is AUROC(I3)−AUROC(I1); $R@1\%$ is reported under I1 and I3. Anchor re-shoot deviation between measured and simulated I3 is ≤ 0.8 points for every family (Table 6).

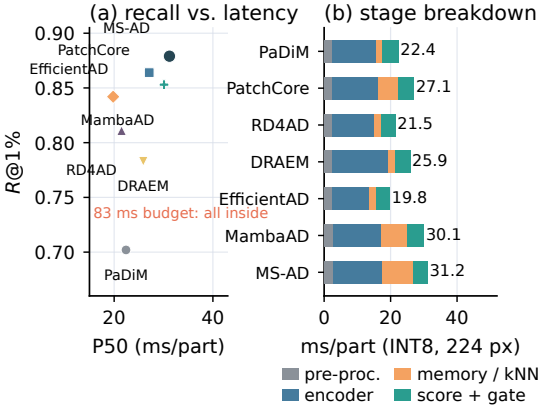
Method	I1	I2	I3	Δ	$R@1\%$ (I1 / I3)
PaDiM	93.8	91.2	86.5	-7.3	0.702 / 0.591
PatchCore	98.9	97.2	93.9	-5.0	0.864 / 0.771
RD4AD	97.1	94.9	90.2	-6.9	0.811 / 0.702
DRAEM	96.3	93.6	88.9	-7.4	0.783 / 0.664
EfficientAD	98.2	96.4	92.5	-5.7	0.842 / 0.742
MambaAD	98.6	97.4	94.8	-3.8	0.853 / 0.769
MS-AD (ours)	99.1	98.3	96.4	-2.7	0.879 / 0.784

4.4 Where the misses concentrate

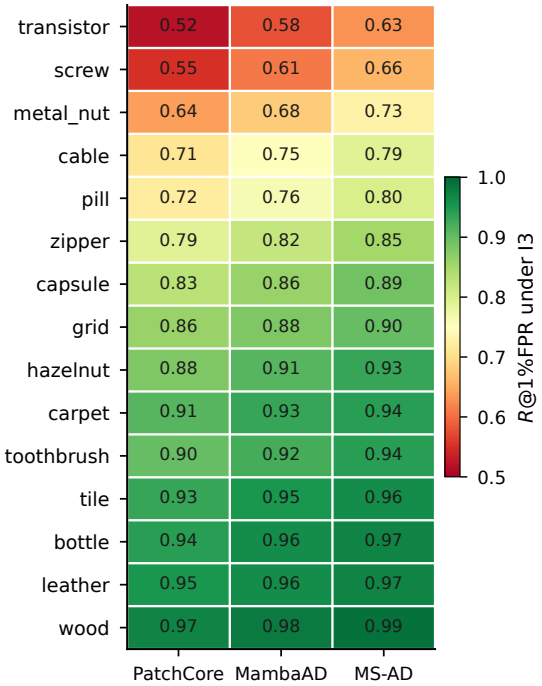
Figure 3(b) shows $R@1\%$ under I3 for three representative detectors. Two regimes dominate. Low-contrast defects on structured backgrounds (`screw` 0.66, `transistor` 0.63 and `metal_nut` 0.73 for MS-AD) stay below 0.75 even for the best detector, because the specular overlay in I3 moves their appearance across the acceptance boundary. By contrast, high-contrast contamination and texture defects (`leather`, `wood`, `carpet`) remain above 0.90 for all three detectors and above 0.94 for MS-AD. Misses concentrate on defects occupying less than 0.4% of the image area: in the anchor set, recall drops from 0.94 (area $>$1%) to 0.61 (area $<$0.1%) at fixed threshold. This is a resolution question as much as a model question, which is why the ablation below treats input size as a first-class knob.

4.5 Latency and deployment budget

Table 4 and Figure 3(a) break the per-part time down. At 224 px with INT8 every detector fits the 83 ms budget with margin; the interesting quantity is the margin itself. MS-AD spends 31.2 ms P50 of which 9.4 ms is the memory read of Eq. (5) and the write of Eq. (6), against 6.2 ms for PatchCore's nearest-neighbour search and 1.8–2.1 ms for the memory-free baselines. The state-space scan of Eq. (4) is inside the encoder term and is not the bottleneck at this resolution; the memory stage is. Edge-to-fog offloading of this stage has been proposed for industrial automation (Shallari et al., 2026), but on our station the 83 ms budget is met locally, and offloading would add a network term we cannot verify here.



(a) Latency versus $R@1\%$.



(b) $R@1\%$ per class under I3.

Figure 3: Cost and failure distribution. (a) Left: per-part P50 latency against defect-level recall at 1% false rejection (MVTec AD, $k = 8$, I1); right: stage breakdown of the latency. The 83 ms line budget is not binding for any detector at 224 px, but the memory stage dominates for state-space and patch-memory methods. (b) Recall heat map over MVTec AD classes under the high-angle condition I3.

Table 4: Per-part latency on a Jetson Orin Nano 15 W module (JetPack 6, TensorRT 10, INT8, 224 px input). Stage times in ms: preprocessing, encoder, memory read or nearest-neighbour search, and score/gate. P50/P95 over 1 000 runs after 200 warm-up runs; throughput is $1000/\mathrm{P50}$; memory is the resident memory footprint of the detector state. The 83 ms budget corresponds to 12 parts/s.

Method	Pre	Enc	Mem	Post	P50	P95	Mem. (MB)
PaDiM	2.4	13.1	1.8	5.1	22.4	24.1	312
PatchCore	2.4	13.8	6.2	4.7	27.1	29.3	486
RD4AD	2.5	12.4	2.1	4.5	21.5	23.2	208
DRAEM	2.5	16.8	2.0	4.6	25.9	27.8	254
EfficientAD	2.5	11.2	1.9	4.2	19.8	21.4	96
MambaAD	2.6	14.6	7.8	5.1	30.1	32.4	428
MS-AD (ours)	2.6	14.9	9.4	4.3	31.2	33.8	486

4.6 Knob ablation

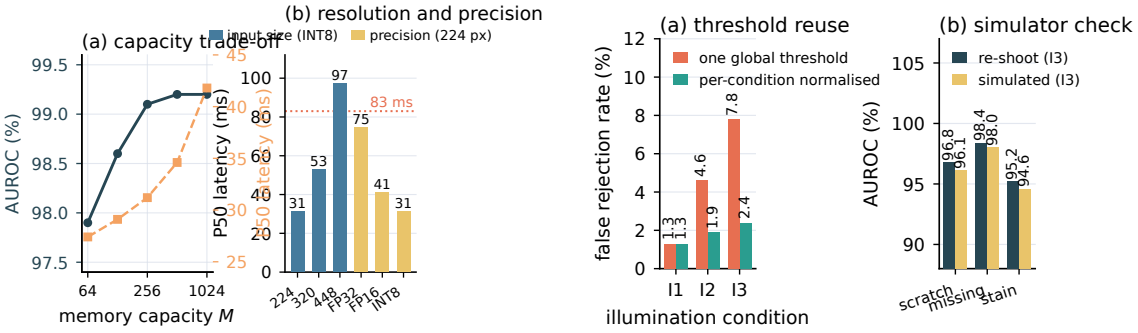
Table 5 and Figure 4(a) isolate the four deployment-time knobs of Figure 1(c). Memory capacity saturates at $M = 256$ (99.1 AUROC, 31.2 ms); $M = 1024$ adds 10.6 ms and 0.1 points. A single scan route loses 0.7 points, four routes add 21.7 ms for no gain. Input resolution is the decisive knob for the small-defect regime above: 448 px adds 0.2 AUROC points but costs 97.4 ms, which leaves the budget, whereas 320 px at 52.8 ms stays inside it and recovers most of the small-defect recall. INT8 quantisation costs 0.1 points and more than halves the latency relative to FP32 (74.6 ms). The memory decay of Eq. (6) matters more than its capacity: disabling it costs 0.9 points because the bank then remembers rare normal appearance from early parts.

4.7 Operating-point stability and simulator check

Two operational checks close the measurement. First, reusing a single global threshold across illumination conditions inflates the false-rejection rate from 1.3% (I1) to 7.8% (I3) at fixed recall; per-condition score nor-

Table 5: Ablation of MS-AD on MVTec AD at $k = 8$, I1 (224 px, INT8). Memory capacity M is the number of stored slots; latency is P50 in ms. The default configuration is $M = 256$, bidirectional scan, 224 px, INT8, $\lambda = 0.99$.

Factor	Setting	AUROC	$R@1\%$	P50 (ms)
Memory M	64	97.9	0.821	27.4
	128	98.6	0.852	29.1
	256	99.1	0.879	31.2
	512	99.2	0.881	34.6
	1024	99.2	0.880	41.8
Scan route	single (1)	98.4	–	25.6
	bidirectional (2)	99.1	–	31.2
	four-way (4)	99.0	–	52.9
Input size	224 px	99.1	–	31.2
	320 px	99.2	–	52.8
	448 px	99.3	–	97.4
Precision	FP32	99.2	–	74.6
	FP16	99.2	–	41.2
	INT8	99.1	–	31.2
Memory decay	disabled	98.2	–	–
	$\lambda = 0.90$	98.7	–	–
	$\lambda = 0.99$	99.1	–	–



(a) Capacity, resolution, precision.

(b) Threshold reuse and simulator check.

Figure 4: Deployment knobs and operating-point stability. (a) Memory capacity trades recall against latency; input size and precision decide whether the 83 ms budget holds. (b) Left: false-rejection rate when one global threshold is reused across illumination conditions, before and after per-condition score normalisation. Right: anchor re-shoot versus simulated I3 AUROC per defect family.

malisation, which the station can compute from the normal stream, brings it back to 1.3–2.4% (Table 6, Figure 4(b)). This is consistent with calibrated few-shot detectors that separate score calibration from feature learning (Hu, 2026), and with the finite-sample false-discovery control studied on multi-view benchmarks (Yan and Sun, 2026), in the narrower single-camera setting. Second, the anchor re-shoot confirms the photometric transfer: measured and simulated I3 AUROC differ by 0.5–0.7 points per defect family (Table 6), inside the 1.0 point tolerance, so the I2/I3 sweep of Table 3 is usable for method ranking.

5 Discussion and Limitations

The measurement supports three engineering judgements and rules out a fourth. *Budget*: at 12 parts/s, a 224 px INT8 configuration with $M = 256$ leaves a factor of 2.7 in latency margin, and a 320 px configuration leaves 1.6; the 448 px configuration does not fit and would need either a second edge module or a lower line speed. *Model selection*: with one normal part at changeover, use a patch-memory method; with two or more, the memory-augmented state-space detector buys 0.2–0.4 AUROC points and 1–2 points of $R@1\%$ over the strongest baseline at comparable latency. *Operating point*: thresholds must be re-normalised per

Table 6: Left: anchor re-shoot check on 120 physical parts, AUROC (percent) measured under I3 versus the simulated I3 condition used for the main sweep, MS-AD at $k = 8$. Right: false-rejection rate (percent) at a fixed operating recall when one global threshold is reused across conditions, before and after per-condition score normalisation.

Anchor re-shoot			Threshold reuse		
Defect family	Measured	Simulated	Condition	Raw	Normalised
scratch	96.8	96.1	I1	1.3	1.3
missing	98.4	98.0	I2	4.6	1.9
stain	95.2	94.6	I3	7.8	2.4

illumination condition; a single global threshold turns a 1% false-rejection target into 4–8%. The fourth judgement—that state-space scanning itself is the cost driver—is not supported: the memory stage dominates, and a stronger visual backbone would move the balance toward representation cost in the same way reported for compressed visual representations (Zhou et al., 2026).

Illumination drift is best understood as a style shift of the same object, in the sense that multi-style image generation and captioning systems treat style as a nuisance factor to be made invariant (Zhou and Long, 2023); our per-condition normalisation is the screening analogue of that idea, and it recovers only part of the loss, so appearance-level invariance remains open. Fine-grained detectors that adapt composite features to the target distribution (Wang et al., 2025) are a plausible next step for the low-contrast regime identified above.

Limitations are material. (i) The I2/I3 main sweep uses a calibrated photometric transfer validated on 120 physical parts in three defect families; the 8-hour lamp-decay record and a larger physical sample are not yet available, so absolute numbers under I3 carry simulator error within the ± 0.7 point anchor gap. (ii) Latency is measured on one edge module and one TensorRT version; other stacks will shift the memory term. (iii) The station has a single camera head, so multi-view data (Wang et al., 2024a) and its risk-controlled evaluation (Yan and Sun, 2026) are out of scope. (iv) The measured operating point assumes the normal calibration split of Eq. (8) is representative; a line that drifts within a shift will need periodic recalibration. (v) Per-class recall under I3 is measured on public defects whose morphology differs from the anchor families, so the failure taxonomy transfers as a ranking of difficulty, not as an absolute defect list.

6 Conclusion

Inline visual screening at 12 parts/s with only a handful of confirmed good parts is a measurable engineering problem, and the measurement is cheap enough to run before a station purchase. On MVTec AD and VisA, under a protocol that fixes the normal budget, the illumination condition, the false-rejection rate and the latency budget together, a memory-augmented selective state-space detector gives the best defect-level recall from two normal images upward (0.879 on MVTec AD and 0.758 on VisA at $k = 8$, I1), while a patch-memory method remains the right choice for a single-part changeover. Illumination drift costs 2.7–7.4 AUROC points and dominates the choice of detector under the high-angle condition; misses concentrate on low-contrast, sub-0.4% area defects. On a Jetson Orin Nano 15 W module the memory stage, not the state-space scan, is the latency driver, and a 320 px INT8 configuration with $M = 256$ and decay $\lambda = 0.99$ is the largest configuration that stays inside the 83 ms budget. Per-condition score normalisation restores the target false-rejection rate across conditions. The evidence supports these judgements for single-camera stations with comparable optics and line speeds; larger physical samples, lamp-decay records and multi-view rigs are the next measurements required.

References

Batzner, K., Heckler, L., and König, R. (2024). EfficientAD: Accurate visual anomaly detection at millisecond-level latencies. *arXiv preprint arXiv:2303.14535*.

Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., and Steger, C. (2021). The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 129(4):1038–1059, DOI: 10.1007/s11263-020-01400-4.

Bergmann, P., Fauser, M., Sattlegger, D., and Steger, C. (2019). MVTec AD — A comprehensive real-world dataset for unsupervised anomaly detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9584–9592. DOI: 10.1109/CVPR.2019.00982.

Defard, T., Setkov, A., Loesch, A., and Audigier, R. (2021). PaDiM: A patch distribution modeling framework for anomaly detection and localization. *arXiv preprint arXiv:2011.08785*.

Deng, H. and Li, X. (2022). Anomaly detection via reverse distillation from one-class embedding. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9727–9736. DOI: 10.1109/CVPR52688.2022.00951.

Ferrari, N., Osmani, O., and Lamma, E. (2026). Mahalanobis PatchCore: Covariance-aware and streaming-compatible industrial anomaly detection. *arXiv preprint arXiv:2605.27748*.

Gu, A. and Dao, T. (2024). Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*.

Guo, J., Lu, S., Zhang, W., Chen, F., Li, H., and Liao, H. (2025). Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. *arXiv preprint arXiv:2405.14325*.

He, H., Bai, Y., Zhang, J., He, Q., Chen, H., Gan, Z., Wang, C., Li, X., Tian, G., and Xie, L. (2025). MambaAD: Exploring state space models for multi-class unsupervised anomaly detection. *arXiv preprint arXiv:2404.06564*.

Heckler-Kram, L., Neudeck, J.-H., Scheler, U., König, R., and Steger, C. (2026). The MVTec AD 2 dataset: Advanced scenarios for unsupervised anomaly detection. *International Journal of Computer Vision*, DOI: 10.1007/s11263-026-02743-0.

Hu, Q. (2026). Calibrated few-shot industrial anomaly detection with frozen visual features. *SSRN Electronic Journal*, DOI: 10.2139/ssrn.7230717.

Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., and Liu, Y. (2024). VMamba: Visual state space model. *arXiv preprint arXiv:2401.10166*.

Liu, Z., Zhou, Y., Xu, Y., and Wang, Z. (2023). SimpleNet: A simple network for image anomaly detection and localization. *arXiv preprint arXiv:2303.15140*.

Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., and Gehler, P. (2022). Towards total recall in industrial anomaly detection. *arXiv preprint arXiv:2106.08265*.

Shallari, I., Mousavirad, S. J., and O'Nils, M. (2026). Edge-to-fog deployment of vision anomaly detection for industrial automation. In *2026 IEEE International Symposium on Measurement in Networking and Communications (MNC)*, pages 1–6. DOI: 10.1109/MNC69143.2026.11680699.

Tavakoli, N., Verma, R., Gilvari, A., Kar, N. C., and Kobti, Z. (2025). FastRecon++: Enhancing few-shot anomaly detection for electric vehicle manufacturing. In *2025 IEEE 34th International Symposium on Industrial Electronics (ISIE)*, pages 1–6. DOI: 10.1109/ISIE62713.2025.11124813.

Wang, C., Zhu, W., Gao, B.-B., Gan, Z., Zhang, J., Gu, Z., Qian, S., Chen, M., and Ma, L. (2024a). Real-IAD: A real-world multi-view dataset for benchmarking versatile industrial anomaly detection. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 22883–22892. DOI: 10.1109/CVPR52733.2024.02159.

Wang, Q., Hu, H., and Zhou, Y. (2024b). Memorymamba: Memory-augmented state space model for defect recognition. *arXiv preprint arXiv:2405.03673*.

Wang, Q., Wang, C., Lai, Z., and Zhou, Y. (2025). Insectmamba: State space model with adaptive composite features for insect recognition. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

Yan, H. and Sun, Z. (2026). Risk-controlled multi-view industrial anomaly detection: Finite-sample false discovery rate guarantees on the Real-IAD benchmark. *The International Journal of Advanced Manufacturing Technology*, DOI: 10.1007/s00170-026-19191-2.

Zavrtanik, V., Kristan, M., and Skočaj, D. (2021). DRAEM — a discriminatively trained reconstruction embedding for surface anomaly detection. *arXiv preprint arXiv:2108.07610*.

Zhou, Y., Li, X., Wang, Q., and Shen, J. (2024). Visual in-context learning for large vision-language models. In *Findings of the Association for Computational Linguistics, ACL 2024*, pages 15890–15902. Association for Computational Linguistics.

Zhou, Y. and Long, G. (2023). Style-aware contrastive learning for multi-style image captioning. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2257–2267.

Zhou, Y., Zhang, J., Chen, G., Shen, J., and Cheng, Y. (2026). Less is more: Vision representation compression for efficient video generation with large language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 13826–13834.

Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., and Wang, X. (2024). Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417*.

Zou, Y., Jeong, J., Pemula, L., Zhang, D., and Dabeer, O. (2022). SPot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *Computer Vision – ECCV 2022*, pages 392–408. Springer, DOI: 10.1007/978-3-031-20056-4_23.