applied sciences

MDPI

Article

AROMA: Adaptive ROI-Based Morphology-Aware Augmentation

Ho-Jun Han and Il-Young Moon *

School of Computer Science and Engineering, Korea University of Technology and Education, Cheonan 31253, Republic of Korea; hhjun321@koreatech.ac.kr

* Correspondence: iymoon@koreatech.ac.kr; Tel.: +82-41-560-1493

Abstract

Anomaly detection in industrial automated visual inspection (AVI) is constrained by severe data sparsity and class imbalance, as defect samples are both scarce and morphologically heterogeneous in real production environments. Existing context-aware synthesis frameworks, such as CASDA, alleviate this limitation by modeling defect–background relationships. However, they depend on domain-specific handcrafted rules and manually constructed compatibility matrices that must be redesigned for each new dataset, limiting scalability. We propose AROMA (adaptive ROI-based morphology-aware augmentation), a framework whose central contribution is the automatic, data-driven re-derivation of the defect–background compatibility matrix that prior work constructed by hand. AROMA profiles defect-morphology and background-context distributions and, from their patch-level co-occurrence, learns a symmetric compatibility gate that scores placement locations without any per-dataset hand-tuning, automatically replacing the handcrafted compatibility matrices required by previous approaches. It further summarizes each dataset's difficulty with a morphology complexity index (MCI) and a context complexity index (CCI); the CCI, in particular, characterizes when context-aware placement can help. To evaluate the placement mechanism independently of generative-model quality, we employ a training-free copy–paste compositing engine with gradient-domain or alpha blending, driven by this data-driven symmetric compatibility gate. Experiments on five industrial datasets spanning textiles (AITeX), metal commutators (Kolektor), steel (Severstal), magnetic tiles (MTD), and leather (MVTec) show that the effectiveness of context-aware ROI selection is not universal but jointly depends on two conditions: sufficient baseline headroom and adequate background-context complexity (CCI). Performance gains are consistently observed on datasets satisfying both conditions (AITeX: +9.4 pp mAP@0.5 over baseline and +2.8 pp over random placement; Severstal: +1.6 pp and +1.3 pp), whereas the advantage narrows to parity—without reversing—on datasets with near-ceiling baseline accuracy (Kolektor, MTD) or largely homogeneous backgrounds that make placement position inconsequential (MVTec Leather, where both placement strategies gain equally, about +9 pp). These findings indicate that context-aware ROI selection is beneficial when both improvement headroom and contextual diversity are present and is otherwise neutral rather than harmful, providing practical guidance on when adaptive ROI optimization is expected to outperform random placement.

Keywords: defect detection; data augmentation; deep learning; industrial visual inspection; context-aware data augmentation

Academic Editor: Angelo Marcelo Tusset

Received: 27 July 2026
Revised: 31 August 2026
Accepted: 3 September 2026
Published: 8 September 2026

1. Introduction

With the advancement of smart manufacturing, defect inspection has increasingly shifted towards deep learning-based automated visual inspection (AVI) systems [1]. Among these, one-class anomaly detection models have gained attention because they require only normal samples for training, reducing the need for costly defect annotation [2]. However, their performance remains highly dependent on the diversity and representativeness of the training data, limiting their effectiveness in data-scarce industrial environments [3].

In real-world manufacturing, defect samples are inherently scarce, resulting in a severe class imbalance that degrades model performance [4,5]. While conventional augmentation methods increase data diversity, they often fail to preserve realistic defect structures and contextual consistency [4]. Generative approaches, such as GANs and diffusion models, can produce more realistic defects but require substantial computational resources and may generate physically inconsistent results [6,7].

To address these limitations, context-aware synthesis explicitly models the relationship between defect morphology and background characteristics. For example, CASDA generates contextually consistent defect samples using a defect–background compatibility model, producing more discriminative training data than context-agnostic placement [8].

Despite this progress, existing context-aware frameworks rely on manually designed, domain-specific conditioning mechanisms, limiting their scalability across industrial datasets. Consequently, how to automatically derive ROI modeling and placement policies from intrinsic dataset characteristics remains an open challenge.

To address this challenge, we propose AROMA (adaptive ROI-based morphology-aware augmentation), a framework that automatically re-derives the defect–background compatibility model that prior context-aware methods built by hand. AROMA learns this model directly from patch-level defect–background co-occurrence statistics and characterizes each dataset's difficulty with complexity indices (MCI, CCI) that indicate when such context-aware placement is expected to help, enabling context-aware defect synthesis without domain-specific engineering. The principal contributions of this work are summarized as follows:

1.1. Automatic Re-Derivation of the Defect–Background Compatibility Model

We show that the handcrafted, per-dataset compatibility matrix of the prior context-aware synthesis can be replaced by a symmetric compatibility gate computed directly from patch-level defect–background co-occurrence statistics. This eliminates domain-specific hand-tuning and provides a unified, data-driven placement mechanism applicable across diverse industrial datasets.

1.2. A Training-Free Protocol That Isolates Placement from Generation

We evaluate the placement mechanism with a training-free copy–paste compositing engine (gradient-domain or alpha blending) rather than a learned generative model, so that the measured effect is attributable to context-aware placement rather than to generative-model quality. The same engine and synthesis budget are applied to both the AROMA and random arms, making the comparison a controlled ablation of the compatibility gate.

1.3. A Characterization of When Context-Aware ROI Selection Helps

We identify the conditions under which context-aware ROI selection is effective. Its benefits depend jointly on sufficient baseline headroom and adequate context complexity (CCI). When both conditions are satisfied, context-aware placement consistently outperforms random selection; otherwise, it converges with random placement.

1.4. CCI-Conditional Downstream Evaluation

We evaluate AROMA on five heterogeneous industrial datasets using a single training-free synthesis engine. The observed gains and convergence follow the two conditions above, empirically validating this characterization and highlighting the regimes in which AROMA is effective.

2. Related Work

2.1. Industrial Defect Detection

Automated industrial defect detection has traditionally been formulated as a supervised learning problem. In this setting, object detection and segmentation networks are trained on densely annotated defect images to localize surface anomalies [1,2]. Deployed systems demonstrate the practical maturity of this paradigm; for example, Stavropoulos et al. [9] present a vision-based system for real-time defect detection of rubber compound parts on the production line. Although such fully supervised approaches achieve strong performance when abundant labeled defects are available [10], their reliance on large annotated datasets is impractical in real manufacturing settings, where defects are rare and labeling is costly [3]. To address this constraint, one-class anomaly detection methods have emerged as the dominant paradigm [3], learning a model of normality exclusively from defect-free images and flagging deviations at inference time. Representative methods such as PatchCore [11], EfficientAD [12], SimpleNet [13], and Reverse Distillation [14] achieve high detection accuracy by modeling the feature distribution of normal samples. However, their performance degrades when the available normal data is limited or when subtle, structurally diverse defects must be distinguished from benign variation [3]. Consequently, augmenting the limited pool of defect data has become a central concern for robust industrial anomaly detection. This scarcity is compounded by data decentralization: defect samples are fragmented across production sites and machine groups, and privacy or operational constraints often preclude central pooling. Federated fault-diagnosis approaches address this by collaboratively adapting models across inconsistent machine nodes while recovering balanced decision boundaries locally [15]; data-side augmentation methods that operate purely on local statistics, such as the one proposed here, are complementary to this direction.

2.2. Data Augmentation

Conventional data augmentation expands training sets through geometric transformations (rotation, flipping, scaling) and photometric perturbations (brightness, contrast, color jitter) [4]. Although computationally inexpensive and widely adopted, these label-preserving operations only resample the existing appearance of defects and fail to introduce the structural variability that characterizes real industrial anomalies [4]. Cut–paste strategies partially address this gap by extracting defect patches and inserting them into normal images to synthesize new defective samples [16,17]. However, naive cut–paste frequently produces visually implausible composites, since achieving natural alignment between the pasted defect and the surrounding background texture, illumination, and geometry remains a challenging and largely unsolved problem [18]. As a result, augmentation realism—rather than mere quantity—has become the key bottleneck for downstream detection gains.

2.3. Generative Models for Augmentation

Generative models have been increasingly explored as a means of synthesizing diverse and realistic defect imagery. Generative adversarial networks (GANs) have been applied to defect synthesis [19,20], but they are prone to mode collapse and training instability, which limit the diversity of the generated defects and demand substantial per-dataset

tuning. Diffusion-based approaches, particularly denoising diffusion probabilistic models (DDPMs), produce markedly higher sample diversity and fidelity [7,21], establishing themselves as the current state of the art for image synthesis. Recent diffusion-based systems push this direction further: AnomalyDiffusion [22] generates anomalies from only a few samples by disentangling anomaly appearance from spatial location, and RealNet [23] shows that controlling the realism and strength of synthesized anomalies directly improves downstream detection. Nevertheless, most generative pipelines model the defect appearance in isolation and do not explicitly account for the background context in which a defect must plausibly appear [2]. This common gap—generating defects without principled background modeling—motivates synthesis methods that jointly reason about defect and context.

2.4. Context-Aware Defect Synthesis

A growing line of work seeks to incorporate spatial and contextual constraints into the defect generation process. ControlNet-based conditioning, for example, has been used to guide diffusion models with masks or layout cues so that synthesized defects respect the structure of the host image [21,24]. Other methods explicitly model defect–background compatibility, selecting or generating defects that are consistent with the local texture and semantics of the target region. CASDA [8] is a representative example of this direction, employing a handcrafted compatibility matrix to match defect types with appropriate background regions during synthesis. While effective on its target domain, CASDA's reliance on manually engineered, domain-specific rules limits its transferability, requiring substantial re-engineering for each new dataset.

2.5. Copy–Paste and ROI-Based Synthesis

Copy–paste augmentation, popularized by Dwibedi et al. [25] and later refined for instance segmentation [26], remains an attractive baseline because it preserves authentic defect pixels while avoiding the training cost and artifacts of generative models. The effectiveness of such methods, however, depends heavily on where defects are placed: the selection of regions of interest (ROIs) for placement strongly influences the realism and the downstream utility of the synthesized data [27]. Context-aware placement has been explored for natural-scene object detection: Dvornik et al. [27] select paste locations by modeling the visual context around candidate boxes, and InstaBoost [28] relocates each annotated instance within its own image using a probability-map-guided appearance-consistency prior. Both operate on annotated object instances in natural scenes; the industrial defect setting addressed here differs in that defects drawn from a scarce cross-image defect pool must be composited onto clean background images of textured surfaces, where no per-instance relocation prior exists. Most existing industrial approaches therefore adopt random or uniform ROI selection, which is domain-agnostic but suboptimal, as it ignores the morphological and contextual compatibility between a defect and its candidate location [27]. To date, the principled, data-driven learning of ROI placement policies in this setting has received little attention, leaving the choice of placement strategy largely heuristic.

2.6. Data-Driven Pipeline Adaptation

Beyond individual augmentation operators, recent research has investigated automatically adapting pipeline configurations to the characteristics of the data. Meta-learning and AutoML techniques optimize augmentation policies and architectural choices directly from data rather than relying on manual design [29,30], while complexity-guided feature learning leverages quantitative descriptors of dataset difficulty to steer model behavior [31]. These efforts demonstrate that measurable dataset statistics can serve as effective signals for configuring learning pipelines without handcrafted rules. Building on this insight, AROMA

introduces a measure-then-derive mechanism: a profiling stage measures each dataset's per-feature defect-morphology and background-context distributions and computes a defect–background compatibility model directly from their patch-level co-occurrence, while complexity indices (MCI, CCI) summarize each dataset's difficulty and characterize when context-aware placement can help. In contrast to CASDA's handcrafted, single-domain compatibility matrix, this data-driven formulation enables the synthesis pipeline to generalize across heterogeneous industrial datasets without per-dataset re-engineering.

3. Methodology

3.1. Experimental Setup

AROMA was evaluated on five publicly available industrial anomaly detection datasets spanning diverse surface characteristics: AITeX [32] (textile), KolektorSDD [33] (metal), Severstal [34] (steel), MTD [35] (ceramic), and MVTec Leather [3] (organic). These datasets cover a broad range of morphological and contextual complexities, enabling evaluation across heterogeneous industrial domains without per-dataset re-engineering. Their statistics are summarized in Table 1.

Table 1. Per-dataset statistics for the five-dataset canonical evaluation roster.

Dataset	Domain	MCI	CCI
AITeX	Textile	0.392	0.44
Kolektor	Metal commutator	0.328	0.224
Severstal	Steel	0.455	0.306
MTD	Magnetic tile	0.561	0.246
MVTec Leather	Leather	0.595	0.2

As shown in Figure 1, the five datasets span a broad range of complexity (MCI: 0.33–0.60, CCI: 0.20–0.44), placing them in distinct operating regimes rather than clustering in a single visual regime. The CCI axis (0.20–0.44), in particular, captures the background-context diversity across the five datasets.

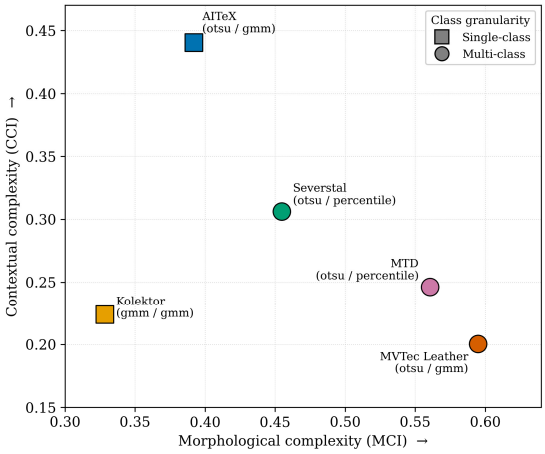

Figure 1. Dataset complexity landscape (MCI vs. CCI).

Three experimental conditions were evaluated for each dataset: Baseline (real data only), Random-ROI, and AROMA-ROI. Both augmentation methods used the same training-free copy–paste pipeline and the same synthetic data budget, differing only in the ROI placement mechanism: random placement versus AROMA's data-driven symmetric

compatibility gate. All models were evaluated on the same real test set, ensuring that performance differences were attributable solely to the placement mechanism.

3.2. AROMA Pipeline

The AROMA pipeline (Figure 2) progresses from dataset profiling and complexity analysis to defect synthesis. Unlike CASDA, which relies on a handcrafted, domain-specific compatibility matrix and morphological rules, AROMA re-derives its defect–background compatibility model directly from dataset statistics. By profiling defect morphology and background context, AROMA builds this compatibility model from patch-level co-occurrence and summarizes each dataset's difficulty with the morphology complexity index (MCI) and context complexity index (CCI), without manual tuning.

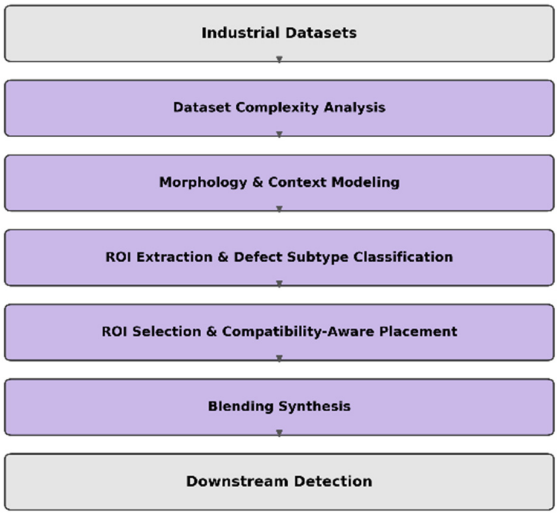

Figure 2. AROMA pipeline.

3.2.1. Dataset Complexity Analysis

The pipeline first quantifies dataset complexity with two complementary indices. The morphology complexity index (MCI) summarizes the structural variability of defect regions as the mean of four normalized statistics capturing intensity variation, distribution multimodality, defect-type diversity, and cluster separability. The context complexity index (CCI) measures background heterogeneity as the mean of four normalized components:

$$\mathrm{CCI} = \mathrm{Mean}(\mathrm{TextureEntropy}, \mathrm{ContextClusterCount}, \mathrm{FreqComplexity}, \mathrm{OrientVarian}) \quad (1)$$

where the components quantify texture disorder, the number of background context clusters, frequency-domain complexity, and gradient-orientation variability, estimated from 20,000 sampled image patches per dataset. Both indices characterize the evaluation roster; the CCI additionally conditions the downstream analysis of when context-aware placement is beneficial (Sections 4.2 and 4.3).

As shown in Table 2, decomposing the CCI into its four components reveals the background characteristics of each dataset. Orientation variance and frequency complexity provide the strongest discrimination, with AITeX exhibiting the highest values due to its directional, periodic texture, whereas Kolektor and MVTec Leather show consistently low values characteristic of smooth, uniform surfaces. In contrast, texture entropy varies only modestly (2.58–3.22), indicating that structural rather than intensity-based features dominate background differentiation. These measurements provide the quantitative basis for the background categories used during ROI extraction (Section 3.2.3).

Table 2. Per-dataset decomposition of the context complexity index (CCI) into its four measured components.

Dataset	Surface	CCI	Texture Entropy	Clusters	Frequency Complexity	Orientation Variance
AITeX	Textile	0.440	2.848	5	0.0158	1.3085
Kolektor	Metal	0.224	2.578	5	0.0002	0.0035
Severstal	Steel	0.306	3.008	5	0.0073	0.4326
MTD	Ceramic	0.246	3.215	5	0.0065	0.0093
MVTec Leather	Organic	0.200	2.981	4	0.0002	0.0005

3.2.2. Morphology and Context Modeling

AROMA models defect morphology and background context with per-dataset, data-driven partitions that require no hand-set constants. Defect instances are grouped by a Gaussian mixture whose number of clusters is selected per dataset by the Bayesian information criterion (BIC), yielding morphology clusters that adapt to each dataset's defect population. Background context is discretized into cells by per-feature tertile (P33/P66) binning of the profiled context features. These morphology clusters and context cells form the row and column indices of the symmetric compatibility model used for placement (Section 3.2.4); because both partitions are estimated from each dataset's own statistics, no dataset-specific constants are introduced.

As shown in Figure 3, defect shape varies mostly along an elongation continuum—from compact, low-aspect blobs to thin, near-linear scratches—so the resulting morphology clusters are essentially segments of this spectrum; each cluster k carries a prior P(k) = n_k/N, its share of the defect population, that serves as the morph_prior when ranking placements (Section 3.2.4).

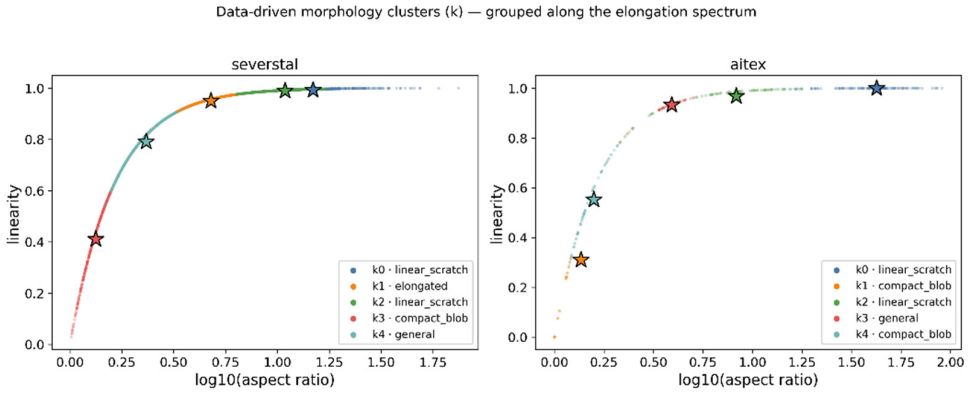

Figure 3. Data-driven morphology clusters for Severstal and AITeX.

Stars denote the per-cluster centroids, colour-matched to their cluster k. As shown in Figures 4 and 5, the five profiled context features and their P33/P66 tertile boundaries differ markedly across datasets, so the same discretization yields dataset-specific context cells; the structure-sensitive features (frequency energy, orientation entropy) vary the most across the roster, consistent with the CCI decomposition.

Red dashed vertical lines mark the P33/P66 tertile boundaries that discretize each feature into the three-level context cells indexing the compatibility model.

3.2.3. Background Categories and Defect Subtypes

Regions of interest (ROIs) were extracted from normal images to identify plausible defect placement locations: Otsu thresholding produces a foreground mask, connected-component analysis yields candidate ROIs, and components below a minimum connected-

component area are discarded as noise. The texture surrounding each retained ROI is classified into one of five categories—smooth, directional, periodic, organic, or complex—by the per-dataset percentile cascade of Table 3, and a continuity score and a stability score quantify each ROI's spatial coherence and its consistency across the normal-image population.

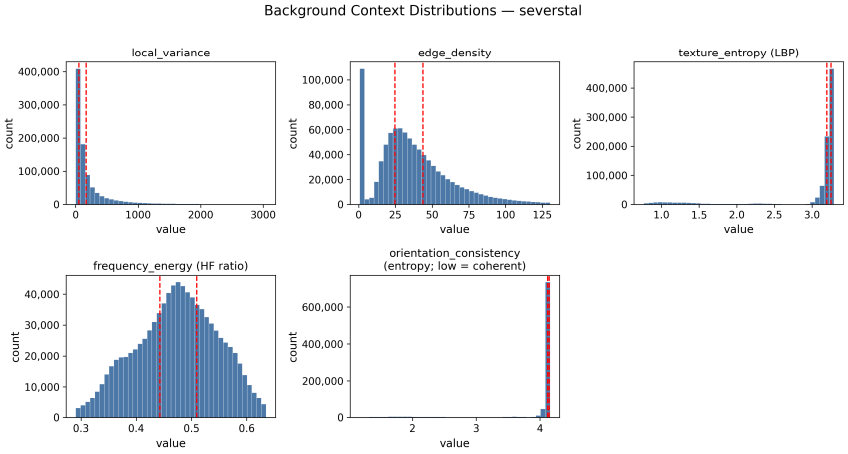

Figure 4. Per-dataset distributions of the five profiled background context features (Severstal).

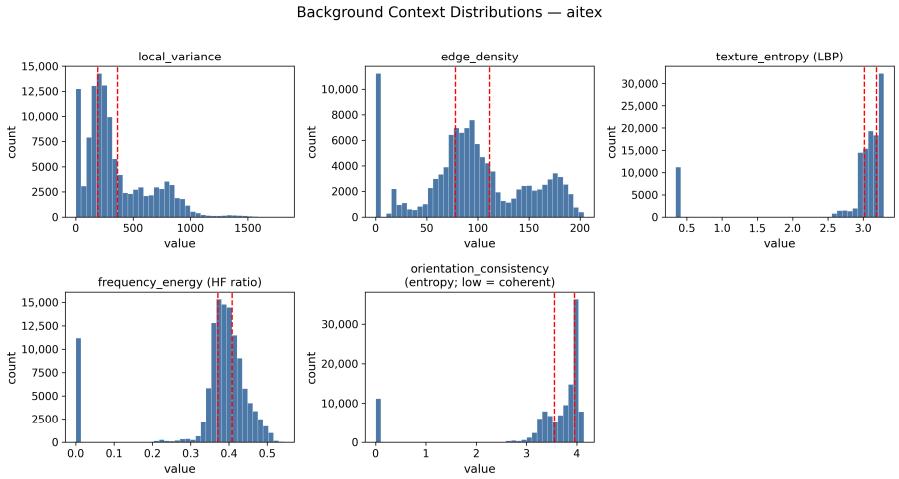

Figure 5. Per-dataset distributions of the five profiled background context features (AITeX).

Seed defects were extracted from defective images to provide morphological templates for synthesis. The segment anything model (SAM) [36] generates precise defect masks, with Otsu thresholding as a fallback when SAM segmentation is unavailable or unreliable. Each defect is classified into one of four generic morphological subtypes—linear_scratch, compact_blob, irregular, or general—using the threshold criteria from Table 4 on aspect ratio and solidity, whose numeric values are derived from each dataset's own defect population (Table 5); these subtype labels parameterize the subtype-specific elastic-warping variants generated before composition, without relying on domain-specific defect categories.

The thresholds are the P33 and P66 tertiles of each dataset's own defect population—the same discretization already applied to the background context features (Section 3.2.2)—so that every partition in the pipeline is anchored to the dataset it describes. The rule is therefore dataset-invariant in form, while its numeric values are estimated per dataset; the derived values are reported in Table 5.

The criteria involve only two features, both standard region shape descriptors [37]. The solidity of the defect region is the ratio of its actual area to the area of its convex hull, and therefore falls as a boundary becomes more ragged or strongly concave. For the aspect ratio, an ellipse is fitted to the defect region and the ratio of its major to minor axes is taken;

the fit is obtained from the region's second central moments [38], which makes the aspect ratio a function of the same eigenvalues that define linearity.

Table 3. Background texture categories from ROI extraction.

Category	Computational Criteria	Example Surface
Smooth	LocalVariance $\leq$ P25	bare metal, uniform coating
Directional	OrientationEntropy $\leq$ P25 AND FreqComplexity $\leq$ P25.	textile weave
Periodic	FreqComplexity $\geq$ P75.	tiles, checker patterns
Organic	TextureEntropy $\geq$ P50 AND LocalVariance $\geq$ P75.	leather, wood, fabric
Complex	else	Boundaries, composite surfaces

Table 4. Defect subtype classification criteria.

Subtype	Classification Criteria	Description
linear_scratch	AspectRatio > P66(AspectRatio)	Long, thin, linear defect
compact_blob	AspectRatio < P33(AspectRatio) AND Solidity > P66(Solidity)	Compact, nearly isotropic defect
irregular	Solidity < P33(Solidity)	Irregular, medium-linearity defect
general	Default label	Fallback category for ambiguous or mixed-morphology defects

Table 5. Per-dataset subtype thresholds derived from each dataset's defect population.

Dataset	Defects	Aspect Ratio P33	Aspect Ratio P66	Solidity P33	Solidity P66
AITeX	352	3.13	16.43	0.648	0.9
Kolektor	52	4.39	5.73	0.7	0.9
Severstal	3620	2.8	7.8	0.9	0.965
MTD	388	1.7	3.62	0.846	0.934
MVTec Leather	92	1.58	4.03	0.874	0.954

Every linearity threshold reduces to an aspect-ratio threshold, and the two features order the defect population identically. Eccentricity, which equals the square root of linearity, is redundant in the same sense; all three are profiled and reported, but only aspect ratio enters the classification.

The derived thresholds vary by a factor of 4.5 in aspect ratio (3.62 for MTD to 16.43 for AITeX) and span from 0.648 to 0.954 in solidity, so the same rule expresses a markedly different shape boundary in each domain. This is the morphological variation that a fixed constant would absorb silently.

As shown in Figures 6 and 7, the per-dataset distributions of the six morphology features confirm that the Table 4 decision boundaries fall between populated regions of the feature space rather than cutting through dense modes and that each dataset occupies a distinct morphological regime.

Red dashed vertical lines mark the Table 5 subtype thresholds on the two features that carry criteria. Aspect ratio is shown on a logarithmic axis.

3.2.4. ROI Selection and Compatibility-Aware Placement

Synthesis placement decomposes into three sequential decisions: (i) which defect to use as the source (ROI selection), (ii) onto which normal background image it should

be placed (background assignment), and (iii) at which position on that background (site resolution). All three stages consult a single compatibility model, defined first. Figure 8 presents the overall flow before each stage is described in turn.

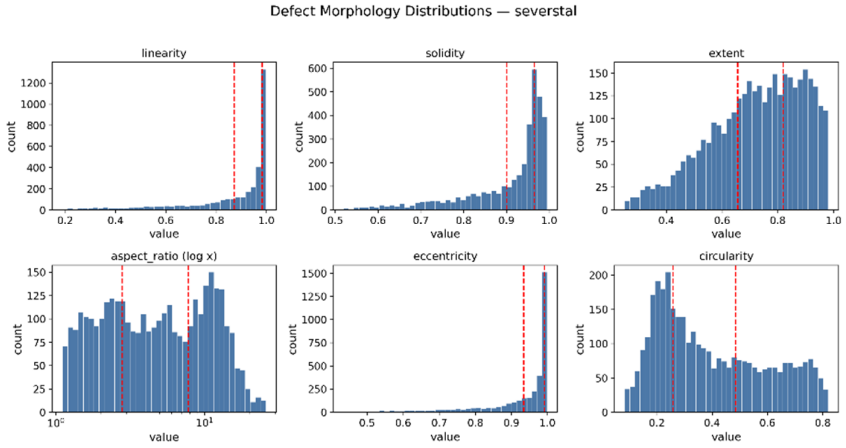

Figure 6. Per-dataset distributions of the six profiled defect morphology features (Severstal).

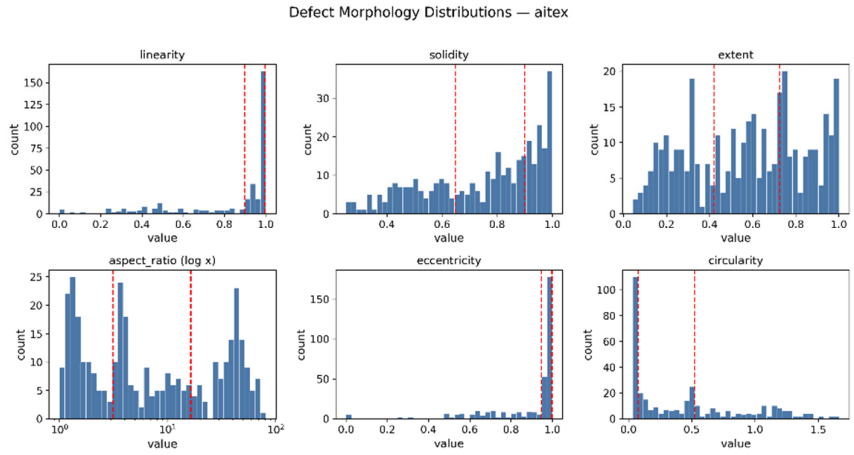

Figure 7. Per-dataset distributions of the six profiled defect morphology features (AITeX).

- Compatibility Model: matrix_symmetric

matrix_symmetric models the compatibility between a morphology cluster (k) and a 64 px discrete context cell (c). It is built from two observed distributions over the same context-cell space combined by their geometric mean:

$$\text{ctx\_prior}(k, c) \propto \sqrt{(\text{P\_def}(k, c) \cdot \text{P\_clean}(c))} \quad (2)$$

so that a cell scores highly only when defects of that cluster are actually observed near it and the cell exists in normal images. Each row is then normalized per cluster, so that a row expresses the relative compatibility of every context cell with cluster (k), peaking at 1 for the most compatible cell.

As shown in Figure 9, symmetric compatibility (ctx_prior) across the five datasets are shown as morphology-cluster × context-cell heatmaps (top-20 most-compatible cells, row-normalized so each cluster peaks at 1; each cluster's peak cell is boxed). Clusters prefer different cells on heterogeneous surfaces (AITeX) but share a dominant cell on near-uniform ones (MVTec Leather).

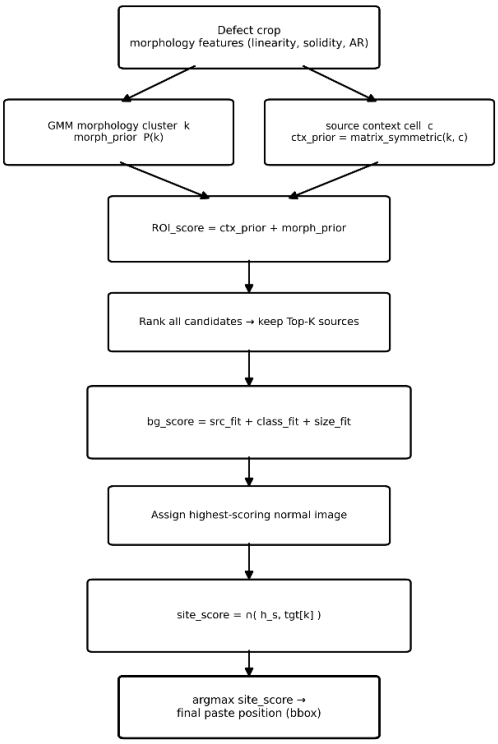

Figure 8. ROI selection and compatibility-aware placement flow.

In Figure 9, The red box in each row marks that morphology cluster's most compatible context cell.

- Defect Crop Selection

This stage determines which defects to use as sources. Each candidate is scored by combining its context-based compatibility with the cluster prior,

$$ROI\ score = \text{ctx\_prior} + \text{morph\_prior} \tag{3}$$

where ctx_prior is read from matrix_symmetric and morph_prior is the cluster prior P(k) in Section 3.2.2. The top (K) candidates by ROI_score become the sources for the placement stages that follow.

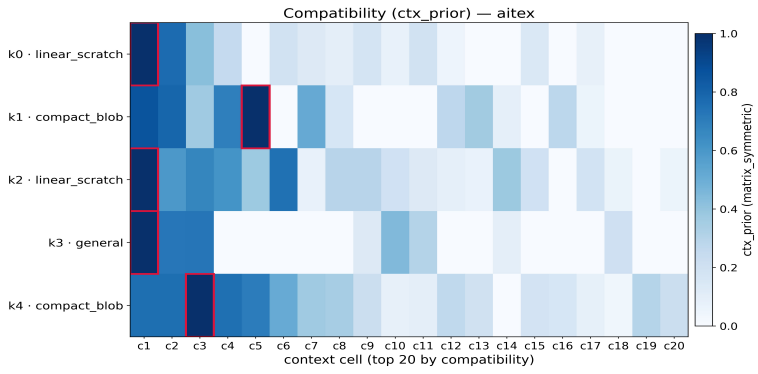

Figure 9. Symmetric compatibility (ctx_prior) across the five datasets as morphology-cluster × context-cell heatmaps.

- Background Assignment

This stage determines the normal background image for each selected ROI. Candidates are ranked by three fitness measures, all built on the histogram intersection ∩(a, b) = Σ_c min(a(c), b(c)):

$$\begin{aligned} \text{src\_fit}(g) &= \cap(h\_g, p\_src(v)) \\ \text{class\_fit}(g) &= \cap(h\_g, p\_cls) \\ \text{size\_fit}(g) &= \min(1, 0.95 \cdot W\_g/w, 0.95 \cdot H\_g/h) \end{aligned} \tag{4}$$

$$\text{bg\_score} = \text{src\_fit} + \text{class\_fit} + \text{size\_fit} \tag{5}$$

and the normal image with the highest score is assigned as the background.

As shown in Figure 10, the two histogram measures respond to different aspects of the pool and the final bg_score ranking combines them into the assigned background.

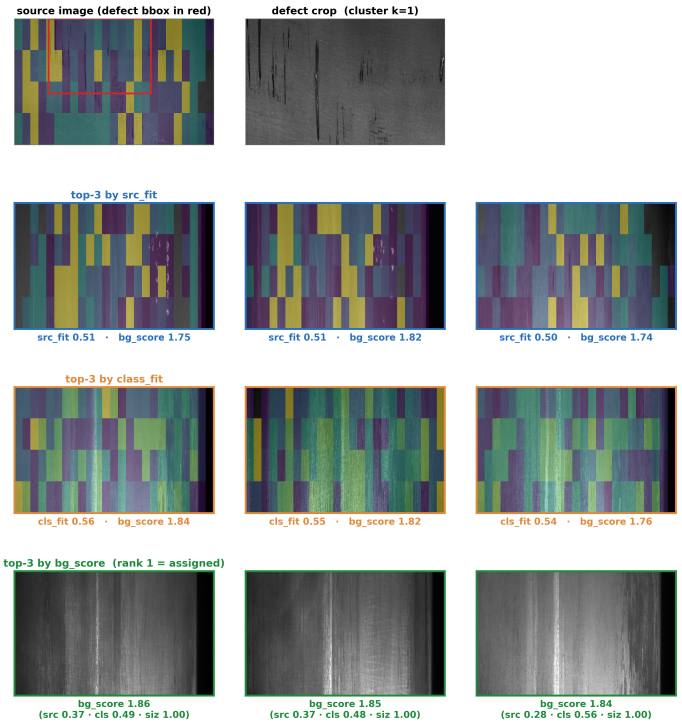

Figure 10. Background assignment on real images.

- Site Resolution

This stage determines where on the assigned background the defect is placed. A candidate position (s) is the top-left coordinate of the defect crop; the ring region around the crop represents the local context that would surround the defect after insertion. The ring of each candidate is converted into a context-cell histogram (h_s) and compared against the matrix_symmetric row of the defect's cluster (k) as the target profile:

$$\text{site\_score} = \cap(h\_s, \text{tgt}[k]) \tag{6}$$

The position with the highest site_score is selected. Placement, therefore, relies neither on geometric heuristics nor on random selection but on the position whose surroundings most closely resemble the context in which that defect morphology was observed in the real data.

As shown in Figure 11, the resolved site reproduces the ring context of the real defect: the tiles ringing the defect in its source image and the tiles ringing the selected position in the assigned background are tinted by the same target mass tgt[k] under one color scale, so

that the resemblance rewarded by the intersection score is directly visible. Here, the best site (best s) is the admissible position with the highest site score, and its measured ring histogram is used to characterize the local context.

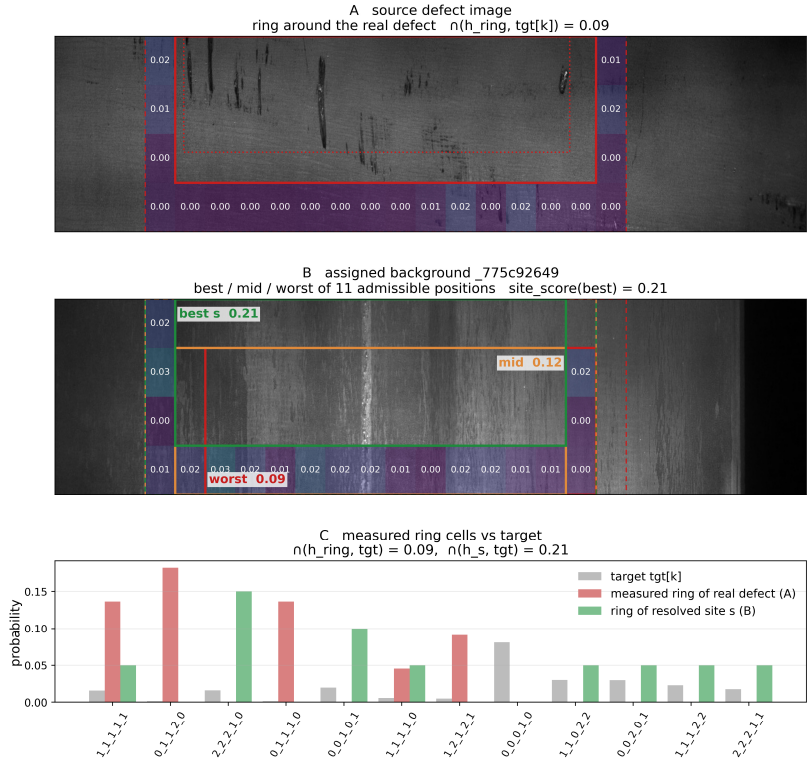

Figure 11. Ring context of the source defect versus the resolved site.

3.2.5. Blending Synthesis

Each selected defect crop is placed at the designated bounding box on the target normal image. The same blend operator is applied identically to the AROMA and random arms, ensuring that blending is not a confounding factor. For each synthesized image, the corresponding defect mask and bounding box are co-saved, providing pixel-level localization for downstream detection.

3.3. Experimental Procedure

3.3.1. Dataset Split

Train–test partitioning followed the on-disk structure provided by each source dataset (AITeX, Kolektor, Severstal, MTD, MVTec Leather). Where a predefined split was unavailable, a fixed-seed partition was applied for reproducibility. In all cases, synthetic images were added exclusively to the training partition, while the test partition consisted solely of real images, enforcing strict isolation between synthetic data and evaluation. The resulting splits and the per-condition synthesis budget are summarized in Table 6.

Table 6. Dataset splits and augmentation scale.

Dataset	Clean Images	Defect Instances	Train/Test	Synth per Condition	Classes
AITeX	7169	352	5018/2151	600	1
Kolektor	347	52	243/104	600	2
Severstal	5902	3620	4131/1771	600	4
MTD	956	388	669/287	600	5
MVTec Leather	245	92	172/73	600	5

All downstream results are reported over three random seeds (1, 2, 42) to establish statistical reliability. For each seed, three augmentation conditions were evaluated on an identical real test set under paired testing: (i) baseline: original training data only; (ii) random-ROI: synthetic defects placed uniformly at random locations; (iii) AROMA: synthetic defects placed by the data-driven compatibility-aware policy (Section 3.2.4). Paired testing ensures that all differences in downstream detection performance are attributable to the augmentation strategy, not to randomness in the train–test split or model initialization.

3.3.2. Downstream Detection Protocol

Downstream utility was assessed via supervised YOLOv8n object detection [39]. Each of the three augmentation conditions (baseline/random/AROMA) trained an independent detector and evaluated it by mean average precision at IoU 0.5 (mAP@0.5) on the shared real test set.

Multi-class datasets (Severstal, MVTec Leather, MTD) were evaluated with per-dataset class enumeration, imgsz 640, and rectangular training enabled. Single-class datasets (AITeX, Kolektor) were trained with a single synthetic defect class; AITeX used imgsz 256 with rectangular training disabled (the AITeX tiles are square), while Kolektor used imgsz 640.

All training runs used identical hyperparameters across all three conditions: 100 epochs, a validation fraction of 0.3, and a batch size of 64. These hyperparameters and the supervised-detection protocol were fixed before any runs and are the source of all downstream mAP@0.5 numbers reported in Sections 4.2 and 4.3.

3.3.3. ROI Coverage Evaluation

The ability of the ROI selection policy to explore the defect-placement space was evaluated by comparing two ROI sets generated from the same candidate pool: AROMA's compatibility-aware selection and a uniform-random selection (Section 3.2.4). Because coverage metrics increase with the number of selected ROIs, the larger set was randomly subsampled (using a fixed seed) to match the size of the smaller set before evaluation, ensuring that the comparison reflected selection quality rather than selection quantity. Three metrics were computed on the size-matched sets: (1) morphology coverage; (2) context coverage, the fractions of morphology clusters and context cells represented in the selection; and (3) rare-pair coverage, the fraction of rarely co-occurring morphology–context pairs (pairs under-represented in the candidate pool) that were selected. Higher coverage indicates a more balanced and comprehensive exploration of the placement space.

3.3.4. Background-Selection Compatibility Evaluation

A second experiment isolates the background-selection component of the placement policy (Section 3.2.4) from ROI selection. Both AROMA and Random use identical ROI set and defect instances, differing only in the assignment of defect-free background images. AROMA ranks candidate backgrounds by histogram-intersection similarity to the ROI's context signature, whereas Random samples uniformly from the same valid background pool. Because both methods share the same ROIs, defect pixels, and synthesis pipeline, any observed differences can be attributed solely to background selection. Background-selection quality is evaluated for each ROI by computing the histogram intersection between the assigned background texture (intensity, gradient magnitude, and local variance) and a dataset-level reference histogram pooled from the background regions of real defect images. The resulting scores are compared using a one-sided Mann–Whitney U test.

3.3.5. Augmentation Budget

For each dataset, both augmented conditions (Random and AROMA) add the same number of synthetic defect images to the real training partition under an identical synthesis

budget (Section 3.3.1), so that the two conditions differ only in the placement mechanism and not in the amount of synthetic data. The compatibility and clean-background gates are applied only to the AROMA arm; the Random arm is a deliberately naive baseline with no placement gating, so the comparison measures the full data-driven placement framework rather than any single component.

4. Experimental Results and Analysis

This section evaluates AROMA across the five industrial datasets of Section 3.1 (AITeX, Kolektor, Severstal, MTD, MVTec Leather). Section 4.1 isolates the ROI selection and background-selection mechanisms via coverage and background-compatibility statistics. All conditions are compared against a real-only Baseline. AROMA is also compared against Random, but with Random instantiated differently per experiment—such as an independently re-selected ROI set, the same ROIs with only background assignment randomized, or an ungated re-selected ROI set—so that a Random result in one figure does not carry over to another.

4.1. ROI Quality Evaluation

The compatibility-driven selection of Section 3.2.4 deliberately concentrates placements on compatible morphology–context combinations rather than spreading them uniformly. To characterize the resulting placement distribution independently of synthesis quality, three coverage statistics (Section 3.3.3) were computed over AROMA's final placements and an independently and uniformly re-selected Random set drawn from the same candidate pool, at an equalized selection size: morphology coverage, context coverage, and rare-pair coverage—the fractions of available morphology clusters, context cells, and rarely co-occurring morphology–context pairs received at least one placement.

As shown in Figure 12, both arms cover every morphology cluster and the uniform Random arm attains equal or higher context coverage on all five datasets (most visibly MTD, 0.85 vs. 0.48), an expected consequence of the design since AROMA concentrates placements on compatible cells rather than spreading them. The per-pair coverage quota nevertheless keeps AROMA's rare-pair coverage at or above Random's on AITeX (1.00 vs. 0.87), Severstal (0.99 vs. 0.92), and MTD (1.00 vs. 0.68); Kolektor is the exception (0.60 vs. 1.00), and MVTec Leather's candidate pool contains no rare pairs. Placement breadth is thus traded for compatibility, and the downstream consequence of this trade is measured in Section 4.3.

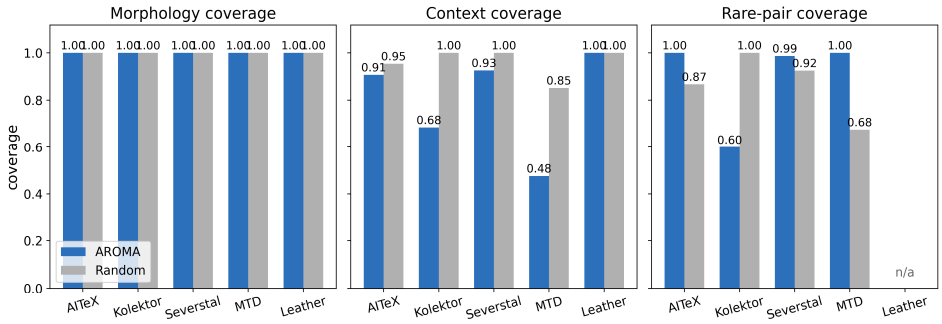

Figure 12. ROI placement coverage (morphology, context, and rare-pair) for AROMA versus an equal-size uniform Random selection from the same candidate pool.

Table 7 complements the coverage statistics with two distributional-evenness measures of the selected set, the normalized Shannon entropy (H/log n) and the Gini coefficient of the morphology-cluster frequency distribution, computed for both arms at the same equalized budget.

Table 7. Distributional evenness of the selected ROIs.

Dataset	Entropy (AROMA)	Entropy (Random)	Gini (AROMA)	Gini (Random)
Severstal	0.984	0.956	0.122	0.207
MTD	0.956	0.867	0.21	0.338
MVTec Leather	0.896	0.899	0.243	0.243
AITeX	0.815	0.859	0.406	0.336
Kolektor	0.883	0.802	0.296	0.39

Two observations follow. First, AROMA's selection is as even as or more even than the uniform-random selection on four of the five datasets (higher entropy and lower Gini on Severstal, MTD, and Kolektor; a tie on MVTec Leather). This is a direct consequence of the per-pair coverage quotas in the allocation (Section 3.2.4): uniform sampling inherits the candidate pool's cluster imbalance, whereas the quota actively spreads selections across morphology–context pairs. Second, on AITeX—the most heterogeneous surface—AROMA deliberately trades distributional evenness for compatibility (entropy 0.815 vs. 0.859; Gini 0.406 vs. 0.336), concentrating placements on compatible pairs; this breadth-for-compatibility trade is the intended behavior, and its downstream consequence is measured in Section 4.2, where AITeX shows AROMA's largest gain over random placement. Entropy and Gini measure distributional evenness rather than placement quality; they are reported alongside the coverage statistics above.

Figure 13 illustrates this behavior qualitatively by directly invoking the same compatibility-scan and uniform-random placement functions on one fixed representative image per dataset—a constructed illustration of the placement mechanism itself, distinct from the independently re-selected Random arm of Figure 10. The compatibility-scored candidates concentrate away from void or texture-degenerate regions, whereas the uniformly placed candidates disregard local context and land on such regions.

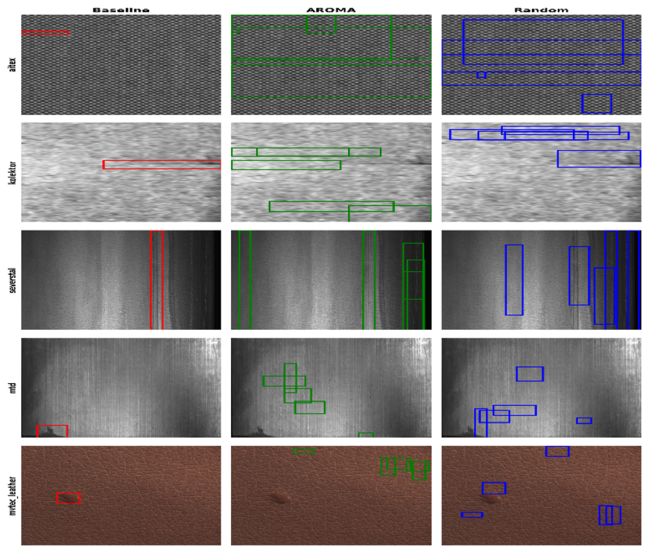

Figure 13. Qualitative ROI placement comparison.

To test the background-selection mechanism independently of the copy–paste engine, AROMA's ROI set is given a symmetric random-background control that keeps the same ROIs and randomizes only the background assignment. Each ROI's assigned normal background was scored by its texture–histogram intersection with the dataset's pooled real defect background (0–1, higher is more similar); then, the two arms were compared with a

one-sided Mann–Whitney U test. As shown in Figure 14, AROMA's assigned backgrounds score higher than Random's on four of the five datasets.

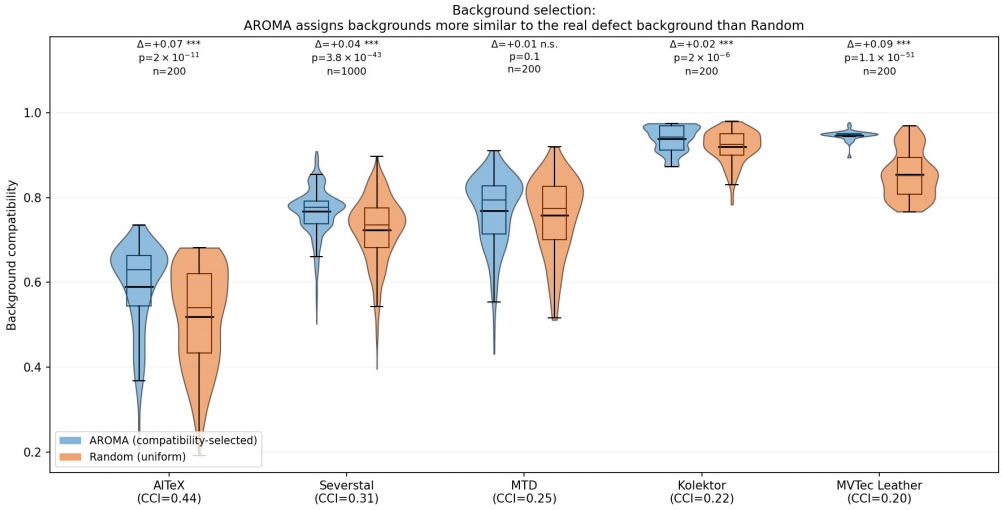

Figure 14. Background-selection compatibility (AROMA vs. Random) across the five datasets.

In Figure 14, Δ is the mean difference (AROMA—Random) and p is from a one-sided Mann–Whitney U test; *** denotes $p < 0.001$ and n.s. not significant.

4.2. Downstream Detection Performance—Single-Class

On AITeX (Table 8), AROMA achieves 0.4533 mean mAP@0.5, outperforming both Baseline (0.3591) and Random (0.4249). The improvement is robust across seeds: both augmented methods raise the mean while reducing seed variance (Baseline std 0.0426 versus 0.0261/0.0300 for Random/AROMA). The single-class, morphologically homogeneous nature of AITeX—combined with its tiled repetitive structure—appears amenable to context-aware placement: AROMA's ability to identify and prioritize compatible background contexts within the tile geometry yields measurable gains.

Table 8. YOLOv8n detection performance on the AITeX test set (mean ± std, n = 3 seeds).

Method	mAP@0.5
Baseline	0.3591 ± 0.0426
Random	0.4249 ± 0.0261
AROMA	0.4533 ± 0.0300
Random vs. Baseline	+6.57 pp
AROMA vs. Baseline	+9.42 pp
AROMA vs. Random	+2.84 pp

In Figure 15 presents sample images from each dataset at each stage of the proposed process.

On Kolektor (Table 9), baseline performance is already near-ceiling (0.9503 mAP@0.5); synthetic augmentation provides only marginal gains. Random achieves 0.9837 (+3.34 pp), and AROMA achieves 0.9866 (+3.63 pp vs. Baseline, +0.29 pp vs. Random). Both augmented methods narrow this spread (AROMA std 0.0080). AROMA's performance parity with Random indicates that in near-ceiling, low-headroom regimes, the placement policy contributes less than baseline dataset characteristics.

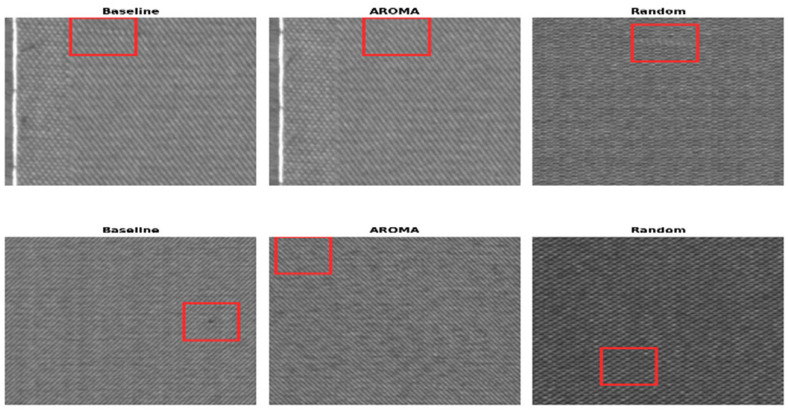

Figure 15. AITeX ROI comparison.

Table 9. YOLOv8n detection performance on the Kolektor test set (mean ± std, n = 3 seeds).

Method	mAP@0.5
Baseline	0.9503 ± 0.0398
Random	0.9837 ± 0.0196
AROMA	0.9866 ± 0.0080
Random vs. Baseline	+3.34 pp
AROMA vs. Baseline	+3.63 pp
AROMA vs. Random	+0.29 pp

In Figure 16 presents sample images from each dataset at each stage of the proposed process.

The contrast between AITeX and Kolektor validates the conditional nature of AROMA's benefit: meaningful gains emerge when baseline headroom exists and the defect–background compatibility structure is informative. No claim of universal improvement is asserted; instead, the results demonstrate that AROMA recovers value in constrained regimes where random placement leaves coverage gaps, while remaining neutral in high-headroom or data-rich scenarios.

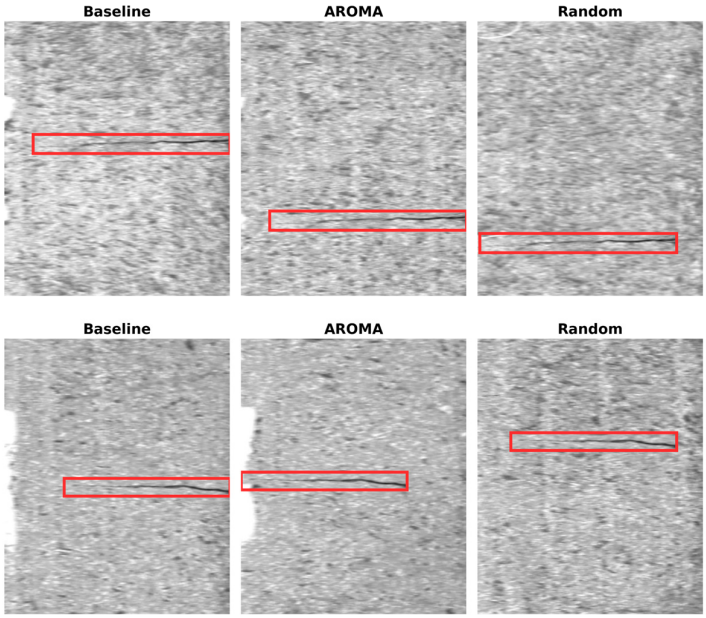

Figure 16. Kolektor ROI comparison.

4.3. Downstream Detection Performance—Multi-Class

Severstal (Table 10) shows that AROMA achieves +1.32 pp over Random (0.5197 vs. 0.5065 mAP@0.5) from a low baseline (0.5033) that retains substantial headroom. Class-wise, AROMA gains over Random are broad: c1 +2.46 pp, c2 +2.56 pp, c4 +1.55 pp, with only c3 slightly negative (−1.29 pp). Class c2 (under-represented) remains the weakest across all methods (0.3812 AROMA), though AROMA recovers the ground Random loses there (Random −2.63 pp vs. Baseline).

Table 10. YOLOv8n supervised detection performance on Severstal (mean ± std, n = 3 seeds).

Method	mAP@0.5	Class 1 AP	Class 2 AP	Class 3 AP	Class 4 AP
Baseline	0.5033 ± 0.0221	0.4664 ± 0.0400	0.3819 ± 0.0829	0.5951 ± 0.0083	0.5698 ± 0.0314
Random	0.5065 ± 0.0171	0.4500 ± 0.0436	0.3556 ± 0.0220	0.6141 ± 0.0028	0.6063 ± 0.0301
AROMA	0.5197 ± 0.0152	0.4746 ± 0.0061	0.3812 ± 0.0355	0.6012 ± 0.0075	0.6218 ± 0.0395
Random vs. Baseline	+0.32 pp	−1.64 pp	−2.63 pp	+1.90 pp	+3.66 pp
AROMA vs. Baseline	+1.64 pp	+0.82 pp	−0.07 pp	+0.61 pp	+5.20 pp
AROMA vs. Random	+1.32 pp	+2.46 pp	+2.56 pp	−1.29 pp	+1.55 pp

In Figure 17 presents sample images from each dataset at each stage of the proposed process.

MTD (Table 11) shows modest augmentation gains from a near-ceiling baseline: Random adds +1.38 pp over Baseline, and AROMA adds +1.79 pp, edging Random by +0.41 pp (0.9293 vs. 0.9252)—effectively convergent at n = 3. Per-class divergence is notable: Fray dominates both methods' gains over Baseline (+4.07 pp Random, +3.86 pp AROMA), while AROMA additionally lifts Crack (+3.28 pp vs. Random) but concedes Uneven (−3.02 pp vs. Random). The offsetting class-level movements suggest that context-aware placement offers limited net leverage in the near-ceiling regime.

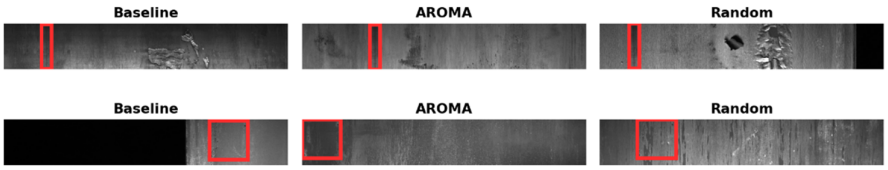

Figure 17. Severstal ROI comparison.

Table 11. YOLOv8n supervised detection performance on MTD (mean ± std, n = 3 seeds).

Method	mAP@0.5	Blowhole	Break	Crack	Fray	Uneven
Baseline	0.9113 ± 0.0227	0.9929 ± 0.0036	0.9377 ± 0.0412	0.9423 ± 0.0466	0.8773 ± 0.0708	0.8064 ± 0.0389
Random	0.9252 ± 0.0101	0.9865 ± 0.0054	0.9502 ± 0.0256	0.9402 ± 0.0418	0.9179 ± 0.0715	0.8311 ± 0.0740
AROMA	0.9293 ± 0.0191	0.9936 ± 0.0019	0.9630 ± 0.0185	0.9731 ± 0.0137	0.9159 ± 0.0731	0.8009 ± 0.0641
Random vs. Baseline	+1.38 pp	−0.64 pp	+1.25 pp	−0.20 pp	+4.07 pp	+2.46 pp
AROMA vs. Baseline	+1.79 pp	+0.07 pp	+2.53 pp	+3.08 pp	+3.86 pp	−0.56 pp
AROMA vs. Random	+0.41 pp	+0.71 pp	+1.28 pp	+3.28 pp	−0.20 pp	−3.02 pp

In Figure 18 presents sample images from each dataset at each stage of the proposed process.

MVTec Leather (Table 12): Both methods add roughly +9 pp over Baseline (Random +9.03 pp, AROMA +8.92 pp), yet the AROMA−Random gap is null (−0.11 pp)—the two placement strategies are statistically indistinguishable, with seed-level differences swinging between −10.1 pp and +11.1 pp. Per-class movements likewise offset one another (Cut +3.66 pp and Poke +1.55 pp for AROMA vs. Random, against Glue −3.43 pp and Fold −2.30 pp), and the dominant gain—Fold, +27–29 pp for both methods—is shared rather than method-specific. The parity has a concrete mechanism: on this near-uniform surface, the similarity-ranked background assignment collapses onto a small background pool, so compatibility ranking carries no positional signal and AROMA's selection reduces to a diversity-restricted variant of random placement. Augmentation itself is highly effective on Leather; which backgrounds are chosen is not the operative variable.

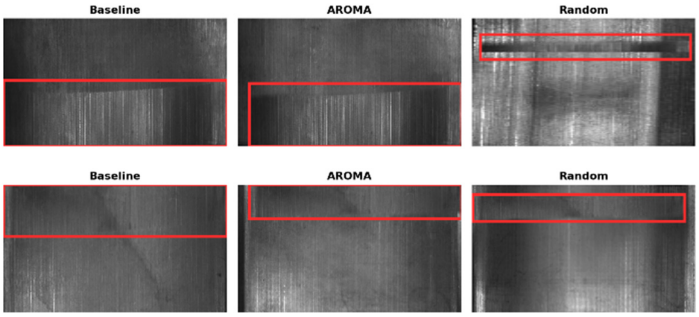

Figure 18. MTD ROI comparison.

In Figure 19 presents sample images from each dataset at each stage of the proposed process.

Across the three multi-class datasets, the AROMA−random gap responds to baseline headroom and surface heterogeneity: Severstal (heterogeneous, low-baseline, +1.32 pp) shows a positive direction; MTD (near-ceiling, limited leverage, +0.41 pp) converges; and MVTec Leather (monotone background, uninformative compatibility signal, −0.11 pp) is neutral. Combined with single-class results in Section 4.2 (AITeX +2.84 pp; Kolektor +0.29 pp), the pattern confirms that context-aware placement is never harmful and delivers its gains conditionally, where baseline headroom and informative surface heterogeneity coincide.

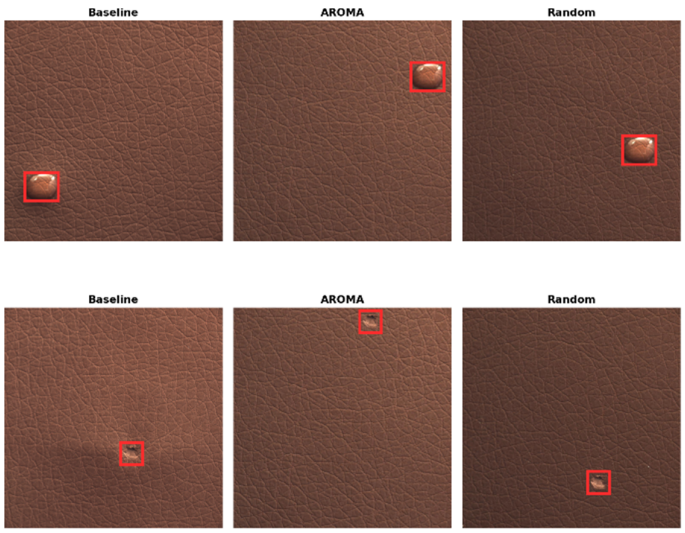

Figure 19. MVTec leather ROI comparison.

To verify that the observed gains are not an artifact of a single detector architecture, we repeated the Severstal and AITeX experiments with YOLOv11n, a more recent de-

tector generation, under an identical protocol (same splits, same synthetic budget, same three seeds).

Table 12. YOLOv8n supervised detection performance on MVTec Leather (mean ± std, n = 3 seeds).

Method	mAP@0.5	Color	Cut	Fold	Glue	Poke
Baseline	0.7481 ± 0.0581	0.8559 ± 0.1244	0.6985 ± 0.1277	0.4537 ± 0.1234	0.8693 ± 0.1360	0.8629 ± 0.0736
Random	0.8384 ± 0.0857	0.9317 ± 0.1097	0.8310 ± 0.1854	0.7470 ± 0.2198	0.9393 ± 0.0537	0.7432 ± 0.1947
AROMA	0.8373 ± 0.0568	0.9313 ± 0.1103	0.8677 ± 0.1697	0.7240 ± 0.2424	0.9050 ± 0.1559	0.7587 ± 0.1040
Random vs. Baseline	+9.03 pp	+7.58 pp	+13.25 pp	+29.33 pp	+7.00 pp	−11.97 pp
AROMA vs. Baseline	+8.92 pp	+7.54 pp	+16.92 pp	+27.03 pp	+3.57 pp	−10.42 pp
AROMA vs. Random	−0.11 pp	−0.03 pp	+3.66 pp	−2.30 pp	−3.43 pp	+1.55 pp

YOLOv11n (Tables 13 and 14) reproduces the YOLOv8n pattern. On Severstal (Table 13), AROMA outperforms Random in all three seeds (+1.44, +0.65, +2.34 pp; mean +1.48 pp, 0.5080 vs. 0.4932 mAP@0.5), with the gain again concentrated in the classes where Random's frequency-following synthesis under-serves the minority: c2 (+2.10 pp vs. Random) and c4 (+4.81 pp vs. Random), while c1 and c3 remain at baseline levels. AROMA is also the only condition whose recall exceeds Baseline (0.515 vs. 0.503), whereas Random trades recall (0.481) for precision. On AITeX (Table 14), both augmentation arms clearly exceed Baseline (Random +3.30 pp, AROMA +4.27 pp on mAP@0.5; mAP@0.5:0.95 rises from 0.159 to 0.207–0.217), but the AROMA−Random gap (+0.97 pp) sits within the seed-level variance of this small dataset (244–246 training images), and we therefore read the two placement strategies as comparable, both exceeding Baseline. The class-conditional gain structure and its magnitude are thus stable across detector generations, indicating that the compatibility-gated placement effect is a property of the augmented data rather than of a specific detector architecture.

Table 13. YOLOv11n supervised detection performance on Severstal (mean ± std, n = 3 seeds).

Method	mAP@0.5	Class 1 AP	Class 2 AP	Class 3 AP	Class 4 AP
Baseline	0.4864 ± 0.0289	0.4465 ± 0.0284	0.3027 ± 0.0372	0.6116 ± 0.0062	0.5848 ± 0.0467
Random	0.4932 ± 0.0028	0.4434 ± 0.0261	0.3110 ± 0.0416	0.6243 ± 0.0012	0.5941 ± 0.0311
AROMA	0.5080 ± 0.0067	0.4461 ± 0.0353	0.3320 ± 0.0255	0.6117 ± 0.0039	0.6422 ± 0.0370
Random vs. Baseline	+0.68 pp	−0.30 pp	+0.83 pp	+1.27 pp	+0.93 pp
AROMA vs. Baseline	+2.16 pp	−0.04 pp	+2.93 pp	+0.00 pp	+5.74 pp
AROMA vs. Random	+1.48 pp	+0.27 pp	+2.10 pp	−1.26 pp	+4.81 pp

Table 14. YOLOv11n detection performance on the AITeX test set (mean ± std, n = 3 seeds).

Method	mAP@0.5
Baseline	0.3637 ± 0.0874
Random	0.3967 ± 0.0271
AROMA	0.4064 ± 0.0487
Random vs. Baseline	+3.30 pp
AROMA vs. Baseline	+4.27 pp
AROMA vs. Random	+0.97 pp

4.4. Ablation Study—Stage-Wise Contribution of the Placement Pipeline

The placement pipeline in Section 3.2.4 makes three sequential decisions for each synthetic sample: (i) defect ROI selection by morphology–context compatibility, (ii) clean-background assignment by fitness ranking, and (iii) intra-image site resolution by ring-based context scoring. To examine how these stages jointly contribute to the Severstal gain in Table 8, we perform leave-one-out ablations by replacing one stage with its random counterpart while keeping the others unchanged: ROI-random (random crop selection), Background-random (random background assignment), and Site-random (random paste position). Because randomly assigned backgrounds provide no ring statistics, Background-random also uses random site placement; thus, it jointly removes stages (ii) and (iii), rather than (ii) alone. All variants follow the Table 8 protocol (YOLOv8n, identical hyperparameters, and n = 3 seeds) with an equal synthetic budget of 2534 images, matching the real training-set size. The Baseline and Random results are carried over from Table 8 for reference.

In Table 15, Every partial variant underperforms the full pipeline. Removing ROI selection causes the largest drop (−3.80 pp), followed by site resolution (−2.06 pp) and background assignment with its cascaded site loss (−2.00 pp). The direction is consistent across all three seeds for ROI and site selection and in two of the three seeds for background assignment. With only three seeds, these results are directional rather than inferential; notably, the ROI-selection deficit exceeds seed-level variation in every seed.

The stage contributions are complementary rather than additive. Random is the strongest variant after the full pipeline (0.5065), while every single-stage ablation falls below both Random and the real-only Baseline. Notably, ROI-random underperforms Random in all three seeds, showing that AROMA-selected crops do not benefit from random placement. This is consistent with Section 4.3: similarity-based background assignment narrows the background pool, which is beneficial only when combined with compatibility-aware crop selection. Likewise, Background-random and Site-random perform similarly (0.4997 vs. 0.4991), suggesting that background fitness is realized primarily through site resolution.

Table 15. Leave-one-out ablation of the AROMA placement pipeline on Severstal (YOLOv8n mAP@0.5, mean ± std, n = 3 seeds).

Method	Stages (ROI·BG·Site)	mAP@0.5	vs. AROMA (Full)
AROMA (full)	A·A·A	0.5197 ± 0.0152	—
Random	R·R·R	0.5065 ± 0.0171	−1.32 pp
Baseline	— (real only)	0.5033 ± 0.0221	−1.64 pp
Background-random	A·R·R	0.4997 ± 0.0298	−2.00 pp
Site-random	A·A·R	0.4991 ± 0.0311	−2.06 pp
ROI-random	R·A·A	0.4817 ± 0.0091	−3.80 pp

Implication: The ablation indicates that the gain arises not from independent stage contributions but from their coherent interaction. Crop compatibility supports background selection, while background fitness is realized through site resolution. Removing any stage therefore breaks this chain, supporting the integrated design of Section 3.2.4 over reduced variants.

4.5. Sensitivity of the Tertile Partition Boundaries

The only dataset-independent convention is the use of P33/P66 to define the tertile partitions. We therefore verify that these fixed percentile positions do not act as a hidden tuning knob.

Tertile-boundary sensitivity (P33/P66): The subtype labels in Section 3.2.3 are used only for interpretation and reporting; they are not used by ROI ranking in Equation (2), so boundary shifts cannot change the generated placements. We therefore test the stability of the labels by perturbing P33 and P66 by ±5 percentile points individually and jointly (P33 → 28–38; P66 → 61–71), yielding eight perturbed configurations per dataset, while retaining the degeneracy safeguard of Section 3.2.3. The response is proportional, with no instability cliff: single-boundary shifts relabel 3.1–9.1% of defects and joint shifts relabel 7.7–14.5% (Table 16), consistent with the distribution mass directly affected by the boundary shifts rather than an amplified response.

Table 16. Stability of the subtype labels under tertile-boundary perturbation (P33/P66 shifted by ±5 percentile points).

Perturbation	Subtype Reassignment
Single boundary ±5 pt	3.1–9.1%
Both boundaries ±5 pt	7.7–14.5%

Together with the ablation in Section 4.4, these results identify the source of AROMA's effect: downstream gains depend on the context-compatibility mechanism. Removing a placement stage changes the outcome, while replacing compatibility-scored ROI selection with random selection reduces Severstal mAP by 3.8. In contrast, the only remaining fixed convention, the tertile boundaries, can be perturbed without affecting placement and causes only proportional changes in the reported labels.

5. Discussion

The principal contribution of AROMA is methodological. The defect–background compatibility model that the prior context-aware synthesis (CASDA) constructed by hand is re-derived automatically from patch-level co-occurrence statistics: a symmetric compatibility gate scores placement locations directly from each dataset's own profiling, removing the per-domain handcrafted compatibility matrix and morphological rules that CASDA required. This advance in design holds independently of downstream accuracy, and it was applied across five heterogeneous industrial surfaces (textile, metal, steel, machined ceramic, leather) without per-dataset re-engineering. The complexity indices (MCI, CCI) serve as descriptive statistics of each dataset's difficulty; the CCI additionally identifies where context-aware placement can matter.

In Section 4, AROMA's context-aware ROI selection shows a consistent positive direction over random placement on datasets with sufficient baseline headroom, improving mAP@0.5 by +2.84 pp on AITeX and +1.32 pp on Severstal, although the n = 3 paired comparison does not reach statistical significance (paired t, df = 2). By contrast, MVTec Leather presents a near-homogeneous background in which defect placement is largely background-agnostic, providing little informative compatibility signal: both placement strategies gain equally from augmentation (≈+9 pp over baseline) and their difference is null (−0.11 pp).

Across the dataset roster, the AROMA–random gap closely follows the context complexity index (CCI), decreasing from heterogeneous surfaces (AITeX, CCI 0.440, +2.84 pp) to moderately structured steel (Severstal, 0.306, +1.32 pp), approaching zero on magnetic tiles (MTD, 0.246, +0.41 pp) and the near-saturated commutator set (Kolektor, 0.224, +0.29 pp), and vanishing on the nearly uniform leather surface (CCI 0.200, −0.11 pp). These results suggest that AROMA's benefit is not universal but depends jointly on two conditions: sufficient baseline detection headroom and meaningful contextual diversity. When either condition is absent, context-aware ROI selection converges with random placement. This

CCI–gain relationship is observed on five datasets and is therefore offered as a mechanistically grounded hypothesis and practical guidance, not as a statistically validated predictive law.

The comparison between the two single-class datasets further supports this interpretation, while illustrating its limits. AITeX combines a high CCI (0.440) with substantial baseline headroom (baseline 0.359), producing the largest observed improvement (+2.84 pp). In contrast, Kolektor exhibits lower context complexity (CCI 0.224) and an almost saturated baseline (0.950), leaving little opportunity for ROI optimization; consequently, AROMA and random placement become statistically indistinguishable (+0.29 pp). Rather than contradicting the proposed mechanism, Kolektor represents the limiting case in which baseline headroom is effectively exhausted.

Finally, we do not rely on full-image distribution metrics such as FID [40], KID [41], or LPIPS [42] as evidence for placement quality. Because both methods composite identical defect pixels, these metrics remain nearly unchanged by construction and primarily measure global image distributions rather than localized placement decisions. For this reason, downstream supervised defect detection (Sections 4.2 and 4.3) is treated as the principal evaluation criterion. For the same reason, the evaluation isolates the placement policy and does not compare against methods that alter defect appearance, such as generative defect synthesis; since such synthesizers are orthogonal to the placement gate, combining them with AROMA's placement pipeline is a natural follow-up that we are pursuing.

AROMA's compatibility gate reasons solely about surface texture, evaluating whether a defect's morphology is compatible with the surrounding local texture. This design is effective for the predominantly surface-centric datasets considered in this study, where the inspected material occupies most of the image (e.g., textile, steel, ceramic, and leather). A natural extension is to augment the ROI selection stage with foreground-aware object reasoning by segmenting the inspected product and modeling defect placement relative to its geometry, in addition to local texture. Such an extension would allow the same data-driven policy framework to generalize from surface-centric inspection tasks to a broader class of object-centric industrial datasets, where a discrete manufactured object occupies the image against a surrounding background.

6. Conclusions

We present AROMA, a framework whose principal contribution is the automatic, data-driven derivation of the defect–background compatibility model. Unlike prior context-aware synthesis methods such as CASDA, which relied on manually engineered compatibility rules, AROMA computes a symmetric compatibility gate directly from patch-level defect–background co-occurrence statistics for each dataset, enabling context-aware placement without per-dataset handcrafting. In addition, dataset difficulty is characterized through the proposed morphology complexity index (MCI) and context complexity index (CCI), providing measurable criteria for determining when context-aware placement is likely to be beneficial. This methodological contribution is independent of downstream accuracy and has been successfully applied across heterogeneous industrial surfaces without domain-specific re-engineering.

Under a training-free copy–paste engine, AROMA consistently outperformed random ROI selection on datasets with both sufficient baseline headroom and adequate context complexity, including AITeX (+2.84 pp) and Severstal (+1.32 pp mAP@0.5). By contrast, performance converged to parity—without ever falling below random placement—on datasets with limited headroom or nearly homogeneous backgrounds, such as MTD and MVTec Leather. The comparison between the two single-class datasets further illustrates this behavior: despite using the same framework, AROMA produced positive gains on

AITeX, which combines high context complexity with substantial baseline headroom, whereas Kolektor, whose baseline is already near ceiling (0.950), showed no meaningful advantage over random placement—the limiting case in which exhausted headroom leaves placement strategy inconsequential regardless of surface heterogeneity. These observations suggest that the effectiveness of context-aware ROI selection depends jointly on baseline headroom and contextual complexity, rather than on either factor alone.

Overall, this work both clarifies the practical limits of context-aware ROI selection and establishes a principled alternative to manually engineered placement policies by grounding compatibility modeling in measurable dataset statistics. More broadly, AROMA provides a general methodology for automatically adapting defect placement strategies to previously unseen industrial datasets, offering a scalable foundation for future data augmentation systems.

Author Contributions: Conceptualization, H.-J.H. and I.-Y.M.; methodology, H.-J.H. and I.-Y.M.; software, H.-J.H. and I.-Y.M.; data curation, H.-J.H. and I.-Y.M.; investigation, H.-J.H. and I.-Y.M.; writing—original draft preparation, H.-J.H. and I.-Y.M.; writing—review and editing, H.-J.H. and I.-Y.M.; supervision, I.-Y.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in AITEX AFID: https://www.aitex.es/afid/; KolektorSDD: https://www.vicos.si/resources/kolektorsdd/; Severstal Steel Defect Detection: https://www.kaggle.com/c/severstal-steel-defect-detection; Magnetic Tile Surface Defect (MTD): https://datasetninja.com/magnetic-tile-surface-defect; MVTec AD (leather category): https://www.mvtec.com/research-teaching/datasets/mvtec-ad—(accessed on 2 September 2026).

Acknowledgments: The authors would like to thank the editors and reviewers for their constructive suggestions and comments that helped improve the quality of the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rippel, O.; Merhof, D. Anomaly Detection for Automated Visual Inspection: A Review. In *Bildverarbeitung in der Automation*; Technologien für die Intelligente Automation; Springer: Berlin/Heidelberg, Germany, 2023; Volume 17. [CrossRef]
2. Hütten, N.; Alves Gomes, M.; Hölken, F.; Andricevic, K.; Meyes, R.; Meisen, T. Deep Learning for Automated Visual Inspection in Manufacturing and Maintenance: A Survey of Open-Access Papers. *Appl. Syst. Innov.* **2024**, *7*, 11. [CrossRef]
3. Bergmann, P.; Batzner, K.; Fauser, M.; Sattlegger, D.; Steger, C. The MVTec Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. *Int. J. Comput. Vis.* **2021**, *129*, 1038–1059. [CrossRef]
4. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
5. Buda, M.; Maki, A.; Mazurowski, M.A. A Systematic Study of the Class Imbalance Problem in Convolutional Neural Networks. *Neural Netw.* **2018**, *106*, 249–259. [CrossRef]
6. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *arXiv* **2014**. [CrossRef]
7. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. *arXiv* **2020**. [CrossRef]
8. Han, H.J.; Moon, I.Y. CASDA: Enhancing Steel Defect Detection Through Context-Aware Data Augmentation Framework. *Appl. Sci.* **2026**, *16*, 6137. [CrossRef]
9. Stavropoulos, P.; Papacharalampopoulos, A.; Petridis, D. A Vision-Based System for Real-Time Defect Detection: A Rubber Compound Part Case Study. *Procedia CIRP* **2020**, *93*, 1230–1235. [CrossRef]
10. Han, H.J.; Moon, I.Y. Research on Object Detection Model-Based Equipment Defect Detection Method. *J. Adv. Navig. Technol.* **2025**, *29*, 859–866. [CrossRef]

11. Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; Gehler, P. Towards Total Recall in Industrial Anomaly Detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 14318–14328. [CrossRef]
12. Batzner, K.; Heckler, L.; König, R. EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; IEEE: New York, NY, USA, 2024; pp. 128–138. [CrossRef]
13. Liu, Z.; Zhou, Y.; Xu, Y.; Wang, Z. SimpleNet: A Simple Network for Image Anomaly Detection and Localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2023; pp. 20402–20411. [CrossRef]
14. Deng, H.; Li, X. Anomaly Detection via Reverse Distillation from One-Class Embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 9737–9746. [CrossRef]
15. Yang, B.; Lei, Y.; Li, N.; Li, X.; Si, X.; Chen, C. Balance Recovery and Collaborative Adaptation Approach for Federated Fault Diagnosis of Inconsistent Machine Groups. *Knowl. Based Syst.* **2025**, *317*, 113480. [CrossRef]
16. Li, C.-L.; Sohn, K.; Yoon, J.; Pfister, T. CutPaste: Self-Supervised Learning for Anomaly Detection and Localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2021; pp. 9664–9674. [CrossRef]
17. Zavrtanik, V.; Kristan, M.; Skočaj, D. DRAEM—A Discriminatively Trained Reconstruction Embedding for Surface Anomaly Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2021; pp. 8310–8319. [CrossRef]
18. Pérez, P.; Gangnet, M.; Blake, A. Poisson Image Editing. *ACM Trans. Graph.* **2003**, *22*, 313–318. [CrossRef]
19. Zhang, G.; Cui, K.; Hung, T.-Y.; Lu, S. Defect-GAN: High-Fidelity Defect Synthesis for Automated Defect Inspection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*; IEEE: New York, NY, USA, 2021; pp. 2524–2534. [CrossRef]
20. Niu, S.; Li, B.; Wang, X.; Lin, H. Defect Image Sample Generation with GAN for Improving Defect Recognition. *IEEE Trans. Autom. Sci. Eng.* **2020**, *17*, 1611–1622. [CrossRef]
21. Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. High-Resolution Image Synthesis with Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2022; pp. 10684–10695. [CrossRef]
22. Hu, T.; Zhang, J.; Yi, R.; Du, Y.; Chen, X.; Liu, L.; Wang, Y.; Wang, C. AnomalyDiffusion: Few-Shot Anomaly Image Generation with Diffusion Model. *Proc. AAAI Conf. Artif. Intell.* **2024**, *38*, 8526–8534. [CrossRef]
23. Zhang, X.; Xu, M.; Zhou, X. 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 (CVPR)*; IEEE: New York, NY, USA, 2024; pp. 16699–16708.
24. Zhang, L.; Rao, A.; Agrawala, M. Adding Conditional Control to Text-to-Image Diffusion Models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2023; pp. 3836–3847. [CrossRef]
25. Dwibedi, D.; Misra, I.; Hebert, M. Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2017; pp. 1310–1319. [CrossRef]
26. Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.-Y.; Cubuk, E.D.; Le, Q.V.; Zoph, B. Simple Copy-Paste Is a Strong Data Augmentation Method for Instance Segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2021; pp. 2918–2928. [CrossRef]
27. Dvornik, N.; Mairal, J.; Schmid, C. Modeling Visual Context Is Key to Augmenting Object Detection Datasets. In *Proceedings of the European Conference on Computer Vision (ECCV)*; Springer: Cham, Switzerland, 2018; pp. 364–380. [CrossRef]
28. Fang, H.-S.; Sun, J.; Wang, R.; Gou, M.; Li, Y.-L.; Lu, C. InstaBoost: Boosting Instance Segmentation via Probability Map Guided Copy-Pasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2019; pp. 682–691. [CrossRef]
29. Cubuk, E.D.; Zoph, B.; Mané, D.; Vasudevan, V.; Le, Q.V. AutoAugment: Learning Augmentation Strategies from Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2019; pp. 113–123. [CrossRef]
30. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. RandAugment: Practical Automated Data Augmentation with a Reduced Search Space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*; IEEE: New York, NY, USA, 2020; pp. 702–703. [CrossRef]
31. Ho, T.K.; Basu, M. Complexity Measures of Supervised Classification Problems. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 289–300. [CrossRef]
32. Silvestre-Blanes, J.; Albero-Albero, T.; Miralles, I.; Pérez-Llorens, R.; Moreno, J. A Public Fabric Database for Defect Detection Methods and Results. *Autex Res. J.* **2019**, *19*, 363–374. [CrossRef]

33. Tabernik, D.; Šela, S.; Skvarč, J.; Skočaj, D. Segmentation-Based Deep-Learning Approach for Surface-Defect Detection. *J. Intell. Manuf.* **2020**, *31*, 759–776. [CrossRef]
34. Severstal: Steel Defect Detection. Kaggle Competition. 2019. Available online: https://www.kaggle.com/c/severstal-steel-defect-detection (accessed on 1 January 2026).
35. Huang, Y.; Qiu, C.; Yuan, K. Surface Defect Saliency of Magnetic Tile. *Vis. Comput.* **2020**, *36*, 85–96. [CrossRef]
36. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y.; et al. Segment Anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2023; pp. 4015–4026. [CrossRef]
37. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 4th ed.; Pearson: New York, NY, USA, 2018.
38. Hu, M.-K. Visual Pattern Recognition by Moment Invariants. *IRE Trans. Inf. Theory* **1962**, *8*, 179–187. [CrossRef]
39. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2016; pp. 779–788. [CrossRef]
40. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2017**, *30*, 6626–6637.
41. Bińkowski, M.; Sutherland, D.J.; Arbel, M.; Gretton, A. Demystifying MMD GANs. In Proceedings of the International Conference on Learning Representations (ICLR), Vancouver, BC, Canada, 30 April–3 May 2018.
42. Zhang, R.; Isola, P.; Efros, A.A.; Shechtman, E.; Wang, O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2018; pp. 586–595. [CrossRef]