BEYOND FEATURE DIRECTION: MAGNITUDE-AWARE MATCHING OF PATCHES AND SPATIAL CONTEXTS FOR TRAINING-FREE ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Training-free industrial anomaly detection has recently benefited from pretrained vision foundation models, where query patches are typically compared with a small set of normal reference patches using normalized features. However, feature normalization emphasizes directional similarity while discarding the magnitude information present before normalization, and independent patch matching overlooks anomalies expressed through abnormal local spatial configurations. We address these limitations with a magnitude-aware patch and local-context matching framework built on frozen DINOv2 features. For each patch, we preserve the unit direction of the post-LayerNorm representation while restoring its pre-LayerNorm magnitude relative to normal reference statistics, yielding a representation that retains both directional and radial information. We further construct an ordered 3×3 context representation by concatenating each patch with its eight spatial neighbors, enabling direct comparison of local configurations without introducing learnable parameters. Patch-level and context-level nearest-neighbor distances to normal memory banks are then combined to obtain the final anomaly score. The proposed method requires neither additional training nor anomalous validation data. Experiments on MVTec AD and VisA under 1-, 2-, and 4-shot settings show consistent improvements over the AnomalyDINO baseline at both image and pixel levels, while repeated experiments with randomly sampled normal references confirm that the gains are robust to reference selection. These results indicate that pre-normalization feature magnitude and local spatial context provide complementary cues for training-free industrial anomaly detection.

1 INTRODUCTION

Industrial anomaly detection aims to identify defective regions or abnormal samples from visual observations, often under limited supervision. In many practical settings, collecting large-scale anomalous data is expensive or unrealistic, making training-free or few-shot anomaly detection particularly attractive. Recent approaches have shown that pretrained vision foundation models, especially DINOv2 (Oquab et al., 2023), provide strong generic visual representations that can be directly exploited for anomaly detection without task-specific training. Among them, nearest-neighbor matching between query patches and a small set of normal reference patches has emerged as a simple yet effective paradigm (Roth et al., 2022; Damm et al., 2025).

Despite their effectiveness, existing training-free patch-matching methods typically rely on normalized patch features (Damm et al., 2025). This design emphasizes directional similarity in feature space, but largely discards the magnitude information contained in the original representation before normalization. As a result, two patches may appear similar under direction-only matching even when their pre-normalization feature magnitudes differ substantially. Moreover, recent few-shot anomaly detection methods commonly score patches independently against normal patch representations (Damm et al., 2025; Lendering et al., 2026), which can overlook anomalies that are better characterized by abnormal local spatial configurations rather than by the center patch alone. Figure 1 illustrates both issues: direction-only matching leaves substantial overlap between normal and anomalous patches, while incorporating magnitude and local context helps suppress false anomaly evidence for normal patches and strengthen anomaly evidence for truly anomalous ones.

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

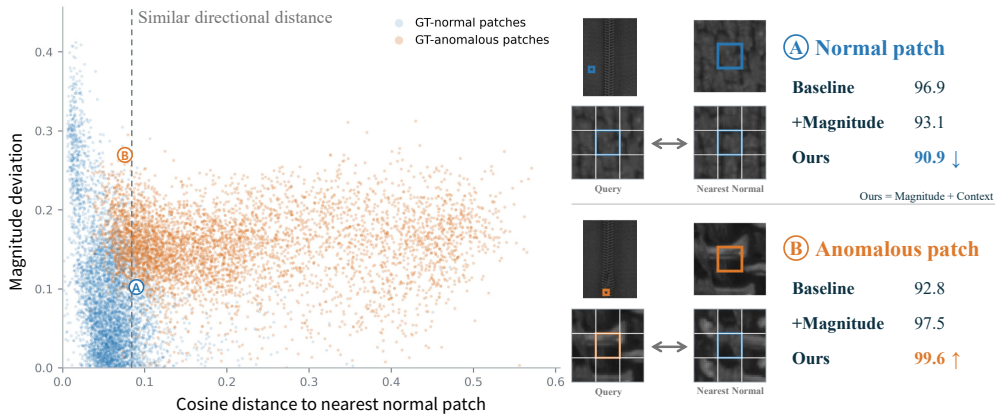

Figure 1: **Motivation for magnitude- and context-aware anomaly matching.** Direction-only matching with normalized DINO features leaves substantial overlap between normal and anomalous patches. The left panel visualizes query patches by their directional distance to the nearest normal patch and their pre-normalization magnitude deviation. The right panel reports normal-score percentiles for two representative patches: magnitude and context suppress anomaly evidence for a normal patch (A) while strengthening it for an anomalous patch (B). These examples illustrate that pre-normalization feature magnitude and ordered local spatial context provide complementary cues for training-free anomaly detection.

In this work, we revisit training-free anomaly detection from these two perspectives. First, we ask whether the magnitude of pretrained features discarded by normalization contains useful anomaly cues. To this end, we construct a magnitude-aware patch representation that preserves the unit direction of the post-LayerNorm feature while reintroducing the relative magnitude measured from its pre-LayerNorm counterpart. Second, we ask whether comparing a patch in isolation is sufficient for anomaly detection. To capture local structural inconsistency, we further represent each patch together with its ordered local neighborhood and compare such local contexts against normal reference contexts. These two complementary signals are combined in a fully training-free manner without modifying the frozen DINOv2 backbone or introducing any learnable parameters.

Based on these ideas, we propose a training-free anomaly detection framework that jointly exploits feature direction, pre-normalization magnitude, and local spatial context. Our method constructs separate memory banks of patch representations and ordered local-context representations extracted from normal reference images, computes nearest-neighbor distances in the two spaces, and combines them to obtain the final anomaly score. Because the method only changes the matching representation and scoring stage, it preserves the simplicity and practicality of existing training-free pipelines.

We evaluate the proposed method on MVTec AD (Bergmann et al., 2019) and VisA (Zou et al., 2022) under 1-, 2-, and 4-shot settings. Using the fixed-reference three-run protocol for the main comparison, our method consistently improves upon the AnomalyDINO (Damm et al., 2025) baseline at both image and pixel levels. Additional repeated experiments with randomly sampled normal references further confirm that the gains are robust to reference selection. Ablation studies also show that magnitude-aware matching and local-context matching provide complementary benefits, supporting our central hypothesis that pre-normalization feature magnitude and local spatial configuration are both useful cues for training-free industrial anomaly detection.

Our contributions are summarized as follows:

- We identify pre-normalization feature magnitude as a complementary anomaly cue that is discarded by conventional direction-only matching with normalized DINO features.
- We propose a fully training-free anomaly detection framework that combines magnitude-aware patch matching with ordered local-context matching, without introducing any learnable parameters.

- We demonstrate consistent improvements over the AnomalyDINO baseline on MVTec AD and VisA across multiple shot settings, and provide robustness and ablation results that support the effectiveness of the proposed cues.

2 Related Work

2.1 Patch-based Industrial Anomaly Detection

Patch-level feature matching has been widely used for industrial anomaly detection (Cohen & Hoshen, 2020; Defard et al., 2021; Roth et al., 2022). PatchCore (Roth et al., 2022) constructs a memory bank of nominal patch features extracted from a pretrained network and detects anomalies through nearest-neighbor distances. It further aggregates features over local neighborhoods to retain spatial information and employs coreset subsampling to reduce the memory-bank size. Complementary approaches have explored one-class learning, reconstruction and distillation, flow-based density estimation, and discriminative modeling (Ruff et al., 2018; Zavrtanik et al., 2021; Deng & Li, 2022; Yu et al., 2021; Gudovskiy et al., 2022; Liu et al., 2023; Batzner et al., 2024). This line of work established patch-level nearest-neighbor matching as a strong paradigm for detecting and localizing industrial defects.

Several approaches have extended patch-based anomaly detection beyond direct nearest-neighbor matching. PaDiM (Defard et al., 2021) models the distribution of pretrained patch features at each spatial location, while DRAEM (Zavrtanik et al., 2021) combines reconstruction with discriminative segmentation for anomaly localization. Reverse Distillation (Deng & Li, 2022) detects anomalies through discrepancies between teacher and student representations, whereas FastFlow (Yu et al., 2021) and CFlow-AD (Gudovskiy et al., 2022) use normalizing flows to model the distribution of normal features. More recent approaches such as SimpleNet (Liu et al., 2023) and EfficientAD (Batzner et al., 2024) further explore discriminative and distillation-based formulations for efficient anomaly detection. Together, these methods demonstrate that the representation and comparison of local patch features are central to industrial anomaly detection.

2.2 Training-Free Anomaly Detection with Foundation Models

Foundation models have also enabled anomaly detection with little or no task-specific optimization (Jeong et al., 2023; Li et al., 2024b; Damm et al., 2025). WinCLIP (Jeong et al., 2023) exploits CLIP (Radford et al., 2021) representations and textual prompts for zero- and few-shot anomaly classification and segmentation, while MuSc (Li et al., 2024b) performs training-free anomaly detection by mutually comparing patches across unlabeled test images. In the vision-only setting, AnomalyDINO (Damm et al., 2025) uses normalized DINOv2 patch features and normal-reference nearest-neighbor matching. SuperAD (Zhang et al., 2025) follows a related DINOv2 memory-bank formulation for training-free anomaly detection under more challenging industrial conditions.

Several recent methods further investigate how foundation-model features should be exploited (Zhou et al., 2024; Cao et al., 2024; Gu et al., 2024; Li et al., 2024a; Lv et al., 2025). PatchEAD (Huang et al., 2026) proposes a unified patch-focused visual prompting framework with feature alignment and foreground masking, while SubspaceAD (Lendering et al., 2026) replaces nearest-neighbor memory matching with a low-dimensional normal subspace estimated from frozen DINOv2 patch features. These approaches primarily improve how normal visual representations are selected, aligned, or modeled. In contrast, we focus on two properties of the matching representation itself: the radial magnitude discarded by normalized DINO features and the spatial configuration discarded when patches are compared independently. Recent analyses have also revealed nontrivial feature-norm behavior in vision transformers (Darcet et al., 2024), motivating closer examination of feature magnitude rather than direction alone. We recover the former from pre-LayerNorm features and incorporate the latter through ordered local-context matching, while keeping the backbone frozen and introducing no learnable parameters.

3 METHOD

3.1 OVERVIEW

We consider a training-free few-shot anomaly detection setting in which only a small number of normal reference images are available. Given a frozen DINOv2 backbone (Oquab et al., 2023), our goal is to measure how far each query patch deviates from the normal reference set without introducing additional training or learnable parameters.

Existing patch-matching approaches primarily compare normalized feature directions (Roth et al., 2022; Damm et al., 2025). We extend this formulation with two complementary cues: pre-LayerNorm feature magnitude and ordered local spatial context. As illustrated in Figure 2, a single frozen DINOv2 forward pass produces magnitude-aware patch features, which are matched against normal references at both the patch level and the ordered $n \times n$ context level. The two nearest-neighbor distances are then combined to obtain the final anomaly score.

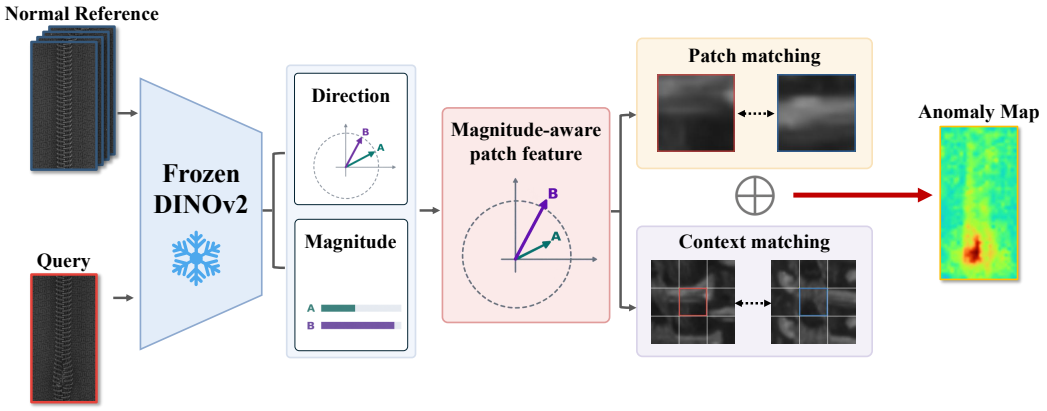

Figure 2: **Overview of the proposed magnitude- and context-aware framework.** Normal reference images and a query image are each processed once by a frozen DINOv2 backbone. The post-LayerNorm representation provides feature direction, while the corresponding pre-LayerNorm representation provides relative feature magnitude. These cues are combined into a magnitude-aware patch representation, which is used for both patch-level nearest-neighbor matching and ordered 3×3 local-context matching. The two distances are fused to obtain the final anomaly score.

3.2 MAGNITUDE-AWARE PATCH REPRESENTATION

Let $h_i^{\text{pre}} \in \mathbb{R}^D$ be the feature of patch i immediately before the final DINOv2 LayerNorm (Oquab et al., 2023; Ba et al., 2016), and let

$$h_i^{\text{post}} = \text{LN}\left(h_i^{\text{pre}}\right). \tag{1}$$

We extract its unit direction $u_i \in \mathbb{R}^D$ as

$$u_i = \frac{h_i^{\text{post}}}{\left\|h_i^{\text{post}}\right\|_2}. \tag{2}$$

To recover scale information discarded by normalized matching, we define the centered pre-LayerNorm magnitude $r_i \in \mathbb{R}_{\geq 0}$ as

$$r_i = \left\| h_i^{\text{pre}} - \overline{h_i^{\text{pre}}} \mathbf{1}_D \right\|_2, \tag{3}$$

where $\overline{h_i^{\text{pre}}} \in \mathbb{R}$ denotes the mean over the D feature dimensions and $\mathbf{1}_D \in \mathbb{R}^D$ is the all-ones vector. We normalize it by the median magnitude $m \in \mathbb{R}_{>0}$ over normal reference patches,

$$m = \text{median}_{j \in \mathcal{R}} r_j, \tag{4}$$

where $\mathcal{R}$ denotes the index set of normal reference patches.

We then construct the magnitude-aware representation $z_i \in \mathbb{R}^D$ as

$$z_i = \left(\frac{r_i}{m}\right)^{\alpha} u_i. \tag{5}$$

We use $\alpha = 0.5$ by default. When $\alpha = 0$, z_i reduces to the direction-only representation. The patch-level anomaly distance $d_i^p \in \mathbb{R}_{\geq 0}$ can be calculated as follows:

$$d_i^p = \min_{z_j \in \mathcal{M}_p} \|z_i - z_j\|_2 . \tag{6}$$

Thus, a query patch receives a high patch-level anomaly score when even its closest normal reference patch is distant in the magnitude-aware feature space.

3.3 Ordered Local-Context Matching

Patch-wise matching can miss anomalies expressed through abnormal neighboring configurations. We therefore concatenate the magnitude-aware features in an ordered $n \times n$ neighborhood:

$$c_i = \frac{1}{n} \text{Concat}\left(z_i^{(1)}, \ldots, z_i^{(n^2)}\right), \tag{7}$$

where $z_i^{(q)}$ denotes the q-th feature in the local neighborhood centered at patch i, ordered in raster scan with replicate padding at boundaries. The factor $1/n$ compensates for the scale increase from concatenating n^2 features. Using the normal context bank $\mathcal{M}_c$, we define

$$d_i^c = \min_{c_j \in \mathcal{M}_c} \|c_i - c_j\|_2 . \tag{8}$$

This score measures whether the local configuration surrounding the query patch has a close counterpart among normal reference regions.

3.4 Final Anomaly Score Calculation

The patch-level and context-level distances capture complementary aspects of normality. We combine them using a weighted sum,

$$s_i = w d_i^p + (1 - w) d_i^c, \tag{9}$$

where we use $w = 0.5$ by default. Therefore, the final patch anomaly score is simply the average of the magnitude-aware patch distance and the ordered context distance.

Both branches are computed from a single forward pass through the frozen DINOv2 backbone, without additional learnable parameters or task-specific training. The main additional cost comes from constructing and searching the local-context memory bank. The image-level score is the average of the top 1% patch scores, while patch scores are resized for localization.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate on MVTec AD (Bergmann et al., 2019) and VisA (Zou et al., 2022), containing 15 and 12 industrial anomaly categories, respectively. Following the few-shot setting of AnomalyDINO (Damm et al., 2025), we use 1, 2, or 4 normal reference images per category. No anomalous samples are used for training, parameter selection, or validation.

Backbone and implementation. We use a frozen DINOv2-S/14 (Oquab et al., 2023) at 672×672 resolution with a single forward pass per image. Unless otherwise specified, we use centered pre-LayerNorm magnitude, $\alpha = 0.5$, a 3×3 context, and $w = 0.5$. All comparisons use the same normal references.

Evaluation metrics. We report image-level AUROC and AUPR, and pixel-level AUROC and PRO. PRO is integrated up to a false-positive rate of 0.3, following standard practice (Bergmann et al., 2019; Damm et al., 2025). All metrics are percentages. Main results are averaged over three fixed reference sets; five-run results with randomly sampled references are provided in the appendix.

4.2 Main Results

Table 1 shows that our method achieves the best results among the compared methods across all metrics on MVTec AD and VisA for 1-, 2-, and 4-shot settings. The gains are particularly pronounced in MVTec localization and VisA image-level detection.

Table 1: Comparison with existing few-shot anomaly detection methods on MVTec AD and VisA. All values are reported in percentages. Published baseline results are taken from the corresponding papers. Bold indicates the best result.

			MVTec AD				**VisA**			
			Image-level		Pixel-level		Image-level		Pixel-level	
Shot	Method	Backbone	AUROC	AUPR	AUROC	PRO	AUROC	AUPR	AUROC	PRO
1	PatchCore	WRN50-2	83.4	92.2	92.0	79.7	79.9	82.8	95.4	80.5
	WinCLIP	CLIP-B/16	93.1	96.5	95.2	87.1	83.8	85.1	96.4	85.1
	PromptAD	CLIP-B/16	94.6	97.1	95.9	87.9	86.9	88.4	96.7	85.1
	IIPAD	CLIP-L/14	94.2	97.2	96.4	89.8	85.4	87.5	96.9	87.3
	AnomalyDINO	DINOv2-S	96.6	98.2	96.8	92.7	87.4	89.0	97.8	92.5
	PatchEAD$^+$	DINOv2-B	95.0	97.0	96.1	91.6	89.1	89.7	97.4	90.6
	SubspaceAD	DINOv2-S	93.2	96.5	96.7	91.9	87.4	87.4	96.5	87.6
	Ours	DINOv2-S	**97.3**	**98.5**	**97.4**	**93.8**	**89.8**	**91.2**	**97.9**	**93.2**
2	PatchCore	WRN50-2	86.3	93.8	93.3	82.3	81.6	84.8	96.1	82.6
	WinCLIP	CLIP-B/16	94.4	97.0	96.0	88.4	84.6	85.8	96.8	86.2
	PromptAD	CLIP-B/16	95.7	97.9	96.2	88.5	88.3	90.0	97.1	85.8
	IIPAD	CLIP-L/14	95.7	97.9	96.7	90.3	86.7	88.6	97.2	87.9
	AnomalyDINO	DINOv2-S	96.9	98.2	97.0	93.1	89.7	90.7	98.0	93.4
	PatchEAD$^+$	DINOv2-B	96.2	97.6	96.4	92.2	90.6	91.0	97.7	91.3
	SubspaceAD	DINOv2-S	94.1	97.1	96.8	92.0	89.0	88.6	96.6	87.8
	Ours	DINOv2-S	**97.4**	**98.5**	**97.5**	**94.1**	**91.7**	**92.5**	**98.1**	**94.1**
4	PatchCore	WRN50-2	88.8	94.5	94.3	84.3	85.3	87.5	96.8	84.9
	WinCLIP	CLIP-B/16	95.2	97.3	96.2	89.0	87.3	88.8	97.2	87.6
	PromptAD	CLIP-B/16	96.6	98.5	96.5	90.5	89.1	90.8	97.4	86.2
	IIPAD	CLIP-L/14	96.1	98.1	97.0	91.2	88.3	89.6	97.4	88.3
	AnomalyDINO	DINOv2-S	97.7	98.7	97.2	93.4	92.6	92.9	98.2	94.1
	PatchEAD$^+$	DINOv2-B	97.0	98.3	96.9	92.9	91.9	92.3	98.0	92.1
	SubspaceAD	DINOv2-S	94.9	97.6	96.8	92.0	90.1	89.9	96.7	87.9
	Ours	DINOv2-S	**98.2**	**98.9**	**97.6**	**94.4**	**93.7**	**94.1**	**98.3**	**94.7**

4.3 Component Analysis

Table 2 evaluates magnitude and context under the 1-shot fixed-reference setting. Magnitude-aware matching yields higher mean performance than the baseline on both datasets, while context mainly affects image-level detection. Combining both achieves the highest mean performance in most metrics, except for VisA Pixel AUROC (97.93 vs. 97.92). The baseline and context-only variants use

cosine matching, whereas magnitude-aware variants use Euclidean distance. This choice is necessary because cosine similarity is invariant to feature magnitude. Thus, the table represents a practical component ablation of the full pipeline.

Table 2: Component ablation on MVTec AD and VisA under the 1-shot setting. Results are averaged over three fixed normal reference sets and reported as mean ± standard deviation. The baseline and context-only variants use cosine matching, while magnitude-aware variants use Euclidean matching.

Dataset	Variant	Img. AUROC	Img. AUPR	Pix. AUROC	PRO
MVTec	AnomalyDINO	96.65±0.46	98.23±0.24	96.84±0.18	92.73±0.16
	+ Magnitude	96.81±0.55	98.31±0.35	97.19±0.15	93.42±0.15
	+ Context	96.90±0.68	98.38±0.33	96.81±0.16	92.79±0.14
	Magnitude + Context	**97.25±0.70**	**98.54±0.36**	**97.40±0.14**	**93.81±0.13**
VisA	AnomalyDINO	87.43±1.52	89.05±1.27	97.80±0.12	92.51±0.65
	+ Magnitude	88.47±1.35	90.01±1.14	**97.93±0.12**	93.05±0.65
	+ Context	88.40±0.92	89.89±0.78	97.69±0.14	92.00±0.84
	Magnitude + Context	**89.76±0.83**	**91.15±0.77**	97.92±0.13	**93.22±0.76**

4.4 What Does Pre-LayerNorm Magnitude Capture?

We analyze whether the relative pre-LayerNorm magnitude r/m itself contains anomaly-related information. Using ground-truth masks only for this diagnostic analysis, we select patches with zero anomalous occupancy as normal and those with at least 80% anomalous occupancy as anomalous; boundary and background patches are excluded.

Figure 3(a) shows that anomalous patches tend to exhibit larger r/m, although the normal and anomalous distributions remain overlapping. Figure 3(b) further shows that the mean r/m of anomalous patches is higher than that of normal patches in all 27 categories across MVTec AD and VisA. These results indicate that pre-LayerNorm magnitude provides a complementary anomaly cue, but is not sufficient as a standalone anomaly score.

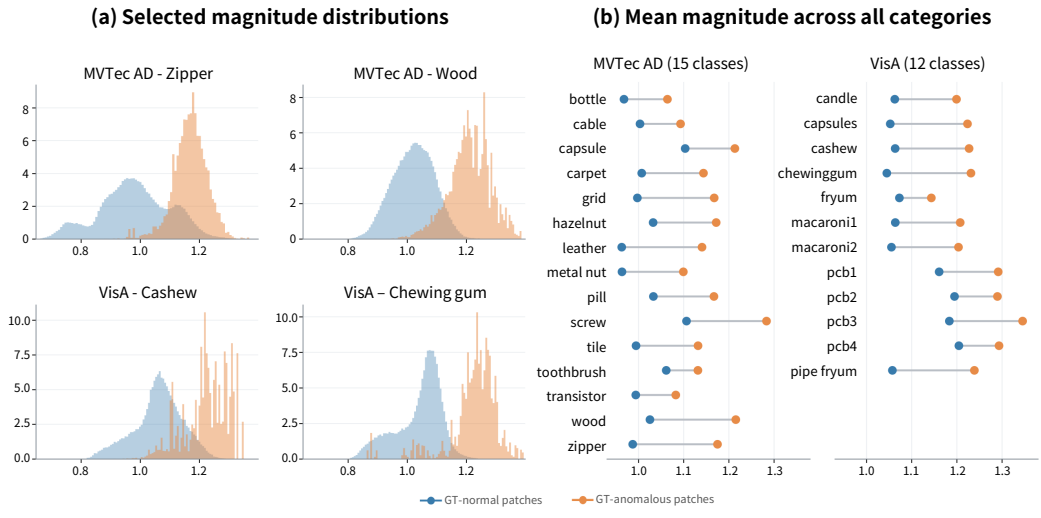

Figure 3: **Pre-LayerNorm magnitude contains anomaly-related information.** (a) Representative distributions of the relative pre-LayerNorm magnitude r/m for ground-truth normal and anomalous patches on MVTec AD and VisA. Anomalous patches generally exhibit a shift toward larger relative magnitudes, while the two distributions remain overlapping. (b) Class-wise mean r/m across all 15 MVTec AD and 12 VisA categories. Anomalous patches have a higher mean relative magnitude than normal patches in all 27 categories. These results support using pre-LayerNorm magnitude as a complementary cue rather than as a standalone anomaly score.

4.5 Design Ablations

Figure 4 analyzes four design choices under the 1-shot setting using three fixed reference sets. We plot the gap to 100 for Image AUROC and PRO as mean ± standard deviation, where lower is better.

Magnitude exponent. Introducing magnitude ($\alpha > 0$) generally improves performance over direction-only matching. We use $\alpha = 0.5$ as a stable square-root scaling rather than a universally optimal setting.

Patch–context fusion. Neither context-only ($w = 0$) nor patch-only ($w = 1$) consistently dominates across datasets and metrics. We therefore use $w = 0.5$ as a simple balanced fusion.

Magnitude source. Centered pre-LayerNorm magnitude consistently outperforms its post-LayerNorm counterpart across both datasets, supporting the use of radial information before the final LayerNorm.

Context size. A 3×3 context improves over isolated 1×1 matching, while 5×5 provides no consistent additional gain and increases matching cost. We therefore use 3×3 by default.

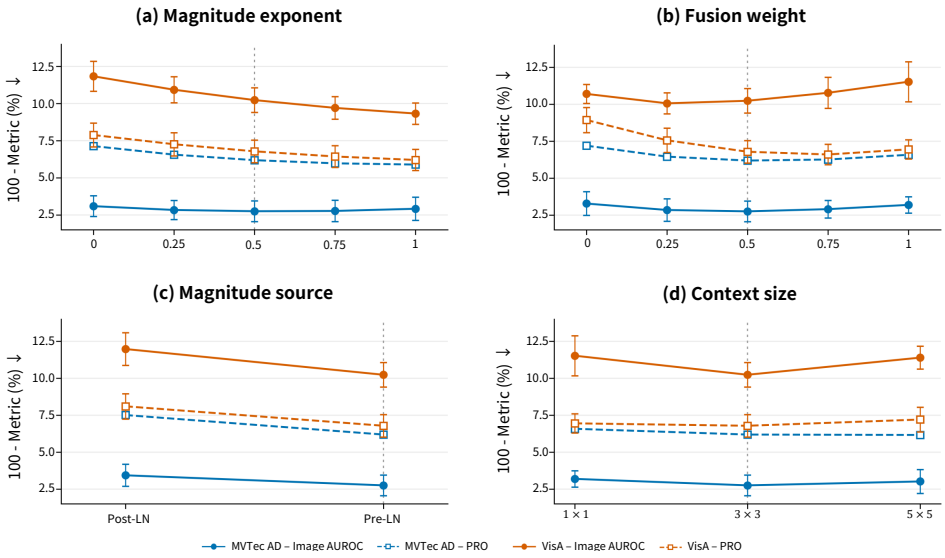

Figure 4: **Sensitivity to key design choices.** We analyze (a) the magnitude exponent α, (b) the patch–context fusion weight w, (c) the source of feature magnitude, and (d) the local context size. Results are obtained under the 1-shot setting using three paired fixed reference sets and are reported as mean ± sample standard deviation. Here, Metric denotes Image AUROC or PRO; we report $100 - \text{AUROC}$ and $100 - \text{PRO}$, so lower is better. The vertical dashed lines indicate our default configuration. Our choices provide a stable trade-off across datasets and metrics rather than being universally optimal for every individual metric.

4.6 Qualitative Results

Figure 5 compares the baseline, magnitude-only, context-only, and full-model anomaly maps using the same 1-shot references on MVTec AD and VisA. Across representative examples, the full model generally produces anomaly responses that are more spatially concentrated around the annotated defect regions while reducing distracting activations in surrounding normal areas.

These qualitative differences are consistent with the motivation behind our magnitude- and context-aware matching. In particular, incorporating pre-LayerNorm magnitude can alter patch-level anomaly evidence even when the normalized feature directions are similar, while ordered local context provides additional cues when the center patch alone is ambiguous.

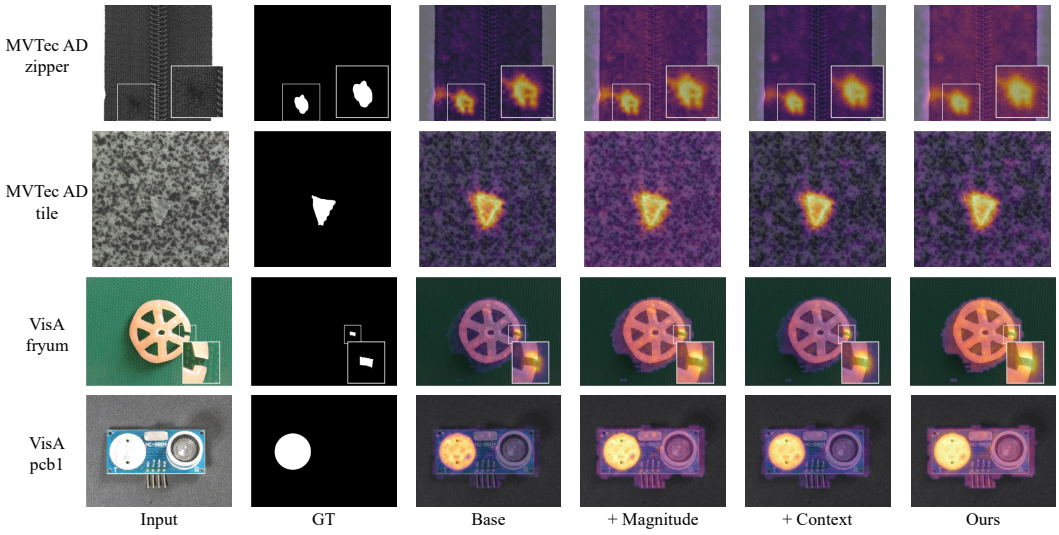

Figure 5: **Qualitative comparison of anomaly localization.** Representative 1-shot examples from MVTec AD and VisA show the input, ground truth, AnomalyDINO baseline, component variants, and full method. Zoomed regions are included for small defects where appropriate. Each anomaly map is independently min–max normalized for visualization; colors therefore indicate relative spatial anomaly evidence within each map and should not be interpreted as directly comparable absolute scores.

Together, these effects produce anomaly maps that better align with the spatial extent of several defects, complementing the improvements observed in pixel-level AUROC and PRO.

5 Discussion and Limitations

Our analysis shows that pre-LayerNorm magnitude provides anomaly-related information complementary to normalized feature direction, although the normal and anomalous distributions partially overlap. Likewise, ordered $n \times n$ context captures local structural inconsistency, while adaptive or multi-scale context may further improve modeling of more complex defects.

Our method keeps the DINOv2 backbone frozen and requires only a single forward pass per image. The main additional cost arises from constructing and searching the local-context memory bank. Developing more efficient context representations is a promising direction for future work.

6 Conclusion

We presented a training-free anomaly detection framework that revisits two sources of information underutilized in conventional normalized patch matching: pre-normalization feature magnitude and local spatial context.

Our method preserves the normalized semantic direction of DINOv2 features while reintroducing relative pre-LayerNorm magnitude and comparing ordered 3×3 patch neighborhoods against normal reference contexts. The resulting patch- and context-level distances are combined without introducing learnable parameters or requiring anomalous validation data.

Experiments on MVTec AD and VisA across 1-, 2-, and 4-shot settings demonstrate consistent improvements over the AnomalyDINO baseline at both image and pixel levels. Ablations and repeated reference sampling further support the complementary benefits of magnitude and context and the robustness of the gains. Anomalous patches also show higher mean relative pre-LayerNorm magnitude across all 27 evaluated categories, highlighting the value of radial and local structural cues beyond normalized feature direction.

AI USE STATEMENT

Generative AI tools were used as a writing and formatting aid to help structure initial manuscript drafts and generate LaTeX code for tables. The authors subsequently revised and verified all AI-assisted content and take full responsibility for the final manuscript.

ETHICS STATEMENT

This work uses publicly available industrial anomaly detection benchmarks and does not involve human subjects, personally identifiable information, or sensitive personal data. We do not foresee significant ethical concerns arising from the proposed methodology.

REPRODUCIBILITY STATEMENT

We provide the complete formulation of our method and all key implementation details in Sections 3 and 4.1, including the backbone, input resolution, magnitude scaling, context size, fusion weight, and evaluation protocol. Additional robustness experiments are provided in the appendix. We plan to release the implementation to facilitate reproduction of the reported results.

REFERENCES

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

Kilian Batzner, Lars Heckler, and Rebecca König. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 127–137. IEEE, 2024.

Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9584–9592. IEEE, 2019.

Yunkang Cao, Jiangning Zhang, Luca Frittoli, Yuqi Cheng, Weiming Shen, and Giacomo Boracchi. Adaclip: Adapting clip with hybrid learnable prompts for zero-shot anomaly detection. In *European conference on computer vision*, pp. 55–72. Springer, 2024.

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

Simon Damm, Mike Laszkiewicz, Johannes Lederer, and Asja Fischer. Anomalydino: Boosting patch-based few-shot anomaly detection with dinov2. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1319–1329. IEEE, 2025.

Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need registers. In *International conference on learning representations*, volume 2024, pp. 2632–2652, 2024.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International conference on pattern recognition*, pp. 475–489. Springer, 2021.

Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 9727–9736. IEEE, 2022.

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

Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *2022 IEEE/CVF winter conference on applications of computer vision (WACV)*, pp. 1819–1828. IEEE, 2022.

Po-Han Huang, Jeng-Lin Li, Po-Hsuan Huang, Ming-Ching Chang, and Wei-Chao Chen. Patchead: Unifying industrial visual prompting frameworks for patch-exclusive anomaly detection. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5531–5540. IEEE, 2026.

Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. Winclip: Zero-/few-shot anomaly classification and segmentation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19606–19616. IEEE, 2023.

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

Xiaofan Li, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Promptad: Learning prompts with only normal samples for few-shot anomaly detection. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16848–16858. IEEE, 2024a.

Xurui Li, Ziming Huang, Feng Xue, and Yu Zhou. Musc: Zero-shot industrial anomaly classification and segmentation with mutual scoring of the unlabeled images. In *International Conference on Learning Representations*, volume 2024, pp. 2172–2197, 2024b.

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

Wenxi Lv, Qinliang Su, and Wenchao Xu. One-for-all few-shot anomaly detection via instance-induced prompt learning. In *International Conference on Learning Representations*, volume 2025, pp. 63657–63677, 2025.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14298–14308. IEEE, 2022.

Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International conference on machine learning*, pp. 4393–4402. PMLR, 2018.

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

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dræm–a discriminatively trained reconstruction embedding for surface anomaly detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8310–8319. IEEE, 2021.

Huaiyuan Zhang, Hang Chen, Yu Cheng, Shunyi Wu, Linghao Sun, Linao Han, Zeyu Shi, and Lei Qi. Superad: A training-free anomaly classification and segmentation method for cvpr 2025 vand 3.0 workshop challenge track 1: Adapt & detect. *arXiv preprint arXiv:2505.19750*, 2025.

Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. In *International Conference on Learning Representations*, volume 2024, pp. 49705–49737, 2024.

Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European conference on computer vision*, pp. 392–408. Springer, 2022.

A PER-CLASS ANALYSIS

We further examine the proposed method at the category level under the 1-shot setting. Results are averaged over the same three fixed normal reference sets used in the main experiments, with AnomalyDINO and our method using identical references for every run.

A.1 MVTEC AD

Table 3 reports the per-class results on MVTec AD using image-level AUROC and AUPR and pixel-level AUROC and PRO. The proposed method improves a broad range of object and texture categories, showing that the aggregate gains are not dominated by a single defect type.

The improvements are particularly visible for transistor, wood, and zipper. For example, zipper exhibits a substantial gain in PRO, while transistor improves across both image- and pixel-level metrics. Several texture categories, including tile and leather, also benefit from the proposed matching scheme. A few categories show minor decreases in individual metrics, suggesting that the contribution of magnitude and local context can vary with defect characteristics.

Table 3: Per-class results on MVTec AD under the 1-shot setting, averaged over three fixed normal reference sets. All values are percentages.

	Image AUROC		Image AUPR		Pixel AUROC		PRO	
Category	Base	Ours	Base	Ours	Base	Ours	Base	Ours
Bottle	99.7	99.9	99.9	100.0	98.8	98.9	95.9	96.6
Cable	92.7	93.3	96.3	96.6	94.6	95.9	89.4	90.6
Capsule	90.2	89.4	97.7	97.4	98.4	98.6	97.1	97.1
Carpet	100.0	100.0	100.0	100.0	98.9	99.2	97.8	98.3
Grid	99.1	99.7	99.7	99.9	99.4	99.5	97.2	97.3
Hazelnut	97.5	97.5	98.7	98.6	99.4	99.4	97.4	97.1
Leather	100.0	100.0	100.0	100.0	99.1	99.3	97.9	98.5
Metal Nut	99.9	100.0	100.0	100.0	96.7	97.1	94.2	94.8
Pill	93.7	96.3	98.8	99.3	96.5	96.1	97.3	97.2
Screw	93.2	92.8	97.6	97.6	98.4	98.8	93.4	94.7
Tile	100.0	100.0	100.0	100.0	95.3	96.2	88.0	90.5
Toothbrush	97.4	99.0	99.1	99.6	99.1	99.3	94.0	95.0
Transistor	90.9	93.1	87.1	89.7	87.1	89.0	67.3	70.7
Wood	98.0	99.4	99.4	99.8	94.9	95.8	94.7	95.7
Zipper	97.4	98.5	99.3	99.6	96.0	97.7	89.2	93.3

A.2 VISA

Table 4 reports the corresponding per-class results on VisA under the same evaluation protocol. Compared with MVTec AD, VisA shows particularly pronounced gains in image-level detection for several challenging categories.

Substantial improvements are observed for cashew, PCB2, PCB3, and PCB4, with PCB4 showing the largest image-level gain among these categories. Localization performance also improves for several classes, including macaroni2, cashew, and PCB3. These results indicate that the proposed magnitude- and context-aware matching provides useful complementary cues across different object structures and defect patterns, although minor decreases remain for a small number of individual metrics.

Table 4: Per-class results on VisA under the 1-shot setting, averaged over three fixed normal reference sets. All values are percentages.

	Image AUROC		Image AUPR		Pixel AUROC		PRO	
Category	Base	Ours	Base	Ours	Base	Ours	Base	Ours
Candle	87.8	90.0	88.2	90.8	98.9	98.8	96.7	96.5
Capsules	98.3	98.7	99.0	99.3	96.7	96.7	95.0	95.5
Cashew	86.3	90.9	93.3	95.6	99.1	99.4	96.1	97.3
Chewing Gum	98.0	97.5	99.1	98.9	99.5	99.5	92.0	92.0
Fryum	94.7	95.8	97.7	98.2	96.6	96.9	93.1	93.7
Macaroni1	87.4	88.2	88.0	88.1	99.3	99.6	97.6	97.8
Macaroni2	62.7	62.8	59.4	59.6	98.8	99.3	92.0	95.2
PCB1	91.4	94.0	92.0	94.2	99.1	99.2	92.5	92.6
PCB2	84.7	88.9	87.6	91.4	95.6	95.8	90.0	90.2
PCB3	85.0	88.8	86.4	90.8	96.6	96.4	88.5	90.3
PCB4	80.3	88.0	81.7	89.9	94.6	94.7	78.5	79.2
Pipe Fryum	92.6	93.7	96.2	96.8	98.9	98.9	98.0	98.2

B Random-Reference Robustness

To evaluate whether the improvements depend on a particular choice of normal references, we repeat the 1-, 2-, and 4-shot experiments using five independently sampled reference sets on MVTec AD and VisA. For each run, the baseline and our method use exactly the same normal references, enabling a paired comparison and reducing the influence of variability caused by reference selection.

Table 5 reports the mean and standard deviation of the paired improvement, defined as $\Delta = \text{Ours} - \text{Baseline}$, over the five runs. Positive values therefore indicate improvements over AnomalyDINO. Our method achieves positive mean gains across all datasets, shot settings, and evaluation metrics. Importantly, this trend is maintained as the number of reference images increases from 1 to 4 shots, suggesting that the improvements are not specific to a particular few-shot configuration. The gains are particularly consistent for pixel-level metrics on MVTec AD and image-level metrics on VisA. Overall, these results show that positive mean gains are maintained across different sampled normal reference sets.

Table 5: Robustness to random normal-reference selection. Each entry reports the mean $\pm$ standard deviation of the paired improvement $\Delta = \text{Ours} - \text{AnomalyDINO}$ over five random reference sets. All values are reported in percentage points.

Dataset	Shot	Img. AUROC	Img. AUPR	Pix. AUROC	PRO
MVTec AD	1	$+0.381 \pm 0.427$	$+0.211 \pm 0.213$	$+0.566 \pm 0.027$	$+1.107 \pm 0.074$
	2	$+0.406 \pm 0.284$	$+0.226 \pm 0.167$	$+0.502 \pm 0.055$	$+1.082 \pm 0.072$
	4	$+0.284 \pm 0.145$	$+0.151 \pm 0.082$	$+0.431 \pm 0.014$	$+1.016 \pm 0.045$
VisA	1	$+1.934 \pm 0.424$	$+1.896 \pm 0.518$	$+0.138 \pm 0.036$	$+0.666 \pm 0.195$
	2	$+1.475 \pm 0.386$	$+1.391 \pm 0.324$	$+0.091 \pm 0.027$	$+0.574 \pm 0.131$
	4	$+0.988 \pm 0.316$	$+0.887 \pm 0.221$	$+0.043 \pm 0.013$	$+0.515 \pm 0.108$

C Controlled All-Euclidean Component Analysis

To disentangle the contribution of feature magnitude from the choice of distance metric, we repeat the 1-shot component ablation using Euclidean distance for all four variants. All variants use the

same three fixed normal reference sets. The direction-only baseline uses unit-normalized descriptors, while magnitude-aware variants use $z_i = (r_i/m)^{1/2}u_i$.

For unit-normalized descriptors, Euclidean distance is a monotonic transformation of cosine distance and therefore preserves the nearest-neighbor ranking at the matching stage. Using Euclidean distance for all variants provides a controlled comparison in which the distance metric is held fixed.

Table 6: Controlled component ablation under the 1-shot setting using three fixed normal reference sets. All four variants use Euclidean distance. Results are reported as mean ± sample standard deviation in percentages.

Dataset	Variant	Img. AUROC	Img. AUPR	Pix. AUROC	PRO
MVTec AD	Direction only	96.63 ± 0.46	98.23 ± 0.23	96.84 ± 0.18	92.68 ± 0.17
	+ Magnitude	96.81 ± 0.55	98.31 ± 0.35	97.19 ± 0.15	93.42 ± 0.15
	+ Context	96.91 ± 0.70	98.39 ± 0.34	96.90 ± 0.16	92.86 ± 0.15
	Magnitude + Context (Ours)	97.25 ± 0.70	98.54 ± 0.36	97.40 ± 0.14	93.81 ± 0.13
VisA	Direction only	87.22 ± 1.47	88.89 ± 1.24	97.75 ± 0.12	92.26 ± 0.66
	+ Magnitude	88.48 ± 1.35	90.01 ± 1.14	97.93 ± 0.12	93.05 ± 0.65
	+ Context	88.16 ± 1.01	89.70 ± 0.86	97.70 ± 0.13	92.12 ± 0.80
	Magnitude + Context (Ours)	89.76 ± 0.83	91.15 ± 0.77	97.93 ± 0.13	93.22 ± 0.76

Table 6 shows that adding magnitude increases the mean performance across all four metrics on both datasets when compared with the corresponding variant without magnitude. This trend holds both with and without the context branch. Therefore, the observed gains from magnitude are not solely attributable to changing the matching metric from cosine to Euclidean distance. These results are presented as descriptive evidence, and we do not claim statistical significance.

D Computational Efficiency

We compare the inference time of AnomalyDINO and our method using a frozen DINOv2-S/14 backbone at 672×672 resolution on a single RTX 4090 GPU. All experiments use float32 precision, a batch size of 1, and reference seed 0. For each image, we average three timing measurements and report the mean and standard deviation over all test images (1,725 images for MVTec AD and 2,162 images for VisA).

The reported time includes preprocessing, feature extraction, masking, nearest-neighbor search, and final anomaly-map generation, while file I/O and memory-bank construction are excluded. As shown in Table 7, our additional magnitude and local-context matching introduces an inference overhead ranging from $1.25\times$ to $1.67\times$ depending on the dataset and number of reference images.

Table 7: Inference-time comparison with AnomalyDINO. Times are reported in milliseconds per image as mean ± standard deviation.

Dataset	Shot	AnomalyDINO	Ours	Relative Time
MVTec AD	1	40.01 ± 5.75	52.65 ± 5.48	$1.32\times$
MVTec AD	2	41.77 ± 5.60	60.31 ± 6.07	$1.44\times$
MVTec AD	4	43.96 ± 5.07	73.43 ± 9.26	$1.67\times$
VisA	1	63.03 ± 5.96	78.70 ± 11.18	$1.25\times$
VisA	2	64.70 ± 6.28	84.72 ± 14.65	$1.31\times$
VisA	4	66.81 ± 6.75	97.91 ± 16.92	$1.47\times$