SubspaceAD: Training-Free Few-Shot Anomaly Detection via Subspace Modeling

Supplementary Material

A. Per-Category Few-Shot Results

Tables 5 and 6 provide per-category few-shot performance on the MVTec-AD and VisA datasets.

MVTec-AD (Table 5). *SubspaceAD* achieves consistently strong performance across nearly all MVTec-AD categories. Even with only one normal image, several categories (e.g., Bottle, Carpet, Grid, Leather, Tile, Toothbrush) reach perfect or near-perfect AUROC, indicating robust capture of normal appearance from one example. The *Transistor* category is a notable exception, with a 1-shot Pixel PRO score of 64.9%. This limitation is due to the characteristic of patch-based visual anomaly detection, which focuses on local appearance and therefore does not explicitly account for logical or structural anomalies, such as missing or misplaced components.

VisA (Table 6). *SubspaceAD* exhibits greater performance variability across VisA categories, which is expected given the dataset's higher visual diversity and complex backgrounds [43]. Categories, such as *Cashew* and *Chewing Gum*, achieve excellent results, with 1-shot image-level AUROC scores of 97.7% and 99.1%, respectively. Conversely, categories like *Macaroni2* (80.8%) and *PCB3* (86.4%) are more challenging. This degradation primarily stems from background artifacts that resemble true defects and from the high intra-class variability of normal samples. Both factors hinder patch-based methods from forming a compact subspace of normality from only a few shots.

B. Performance of DINOv3 Backbones

For completeness, we report the few-shot results using the DINOv3-7B backbone [36] in Table 7. These results are obtained using 4096-dimensional features, extracted from layers 22–28 with 16×16 patch tokens. SubspaceAD$_{448}$ and SubspaceAD$_{672}$ denote models evaluated at input image resolutions of 448×448 and 672×672 pixels, respectively. As shown in Table 7, the DINOv3 backbone is consistently outperformed by the DINOv2-G backbone (see main paper, Table 1), particularly on the VisA dataset.

C. Full-Shot Setting Analysis

A comparison in the full-shot setting, where all available normal training samples are used to build the model of normality, is provided in Table 8. On MVTec-AD, AnomalyDINO achieves slightly superior performance across most metrics (e.g., 99.5% vs 99.2% I-AUROC).

On the more challenging VisA dataset, *SubspaceAD* surpasses AnomalyDINO in both image-level and pixel-level AUROC (98.2% vs. 97.6% and 99.1% vs. 98.8%, respectively). However, AnomalyDINO retains a modest advantage in the pixel-level PRO score (96.1% vs. 94.9%). This gain comes at a substantial computational and memory cost, because AnomalyDINO stores dense patch-level features for every normal image and performs K-NN retrieval over these embeddings, typically on the order of 10^6 feature vectors per category on VisA. In contrast, *SubspaceAD* performs a single forward pass and a lightweight subspace projection, yielding an inference time of approximately 300 ms per image at 672px (see Sec. D), while requiring no feature memory banks.

D. Inference Time Analysis

An analysis of inference speed is presented in Table 10. To ensure an algorithmic comparison independent of hardware, we further evaluate the standalone *scoring head* latency (isolated from the backbone forward pass) in Table 9.

End-to-End Latency. Table 10 reports the end-to-end inference times for various backbone configurations. Because these measurements were obtained on different hardware (NVIDIA A40 for AnomalyDINO vs. NVIDIA H100 for SubspaceAD), direct wall-clock speed comparisons are not appropriate. For both methods, the total inference time is primarily dominated by the forward pass of the frozen backbone. The key distinction is architectural: SubspaceAD achieves its performance using only the backbone and a lightweight projection, eliminating the need to store or search through feature memory banks.

Algorithmic Fairness (Scoring Head Only). To isolate the scoring mechanism from the backbone's overhead, we measured the time required to compute anomaly scores from extracted features on an H100. As shown in Table 9, both methods exhibit comparable latency in the few-shot regime. SubspaceAD requires ≈74 ms per image, while AnomalyDINO requires ≈80 ms. While our subspace projection time is strictly invariant to the size of the support set once the PCA components are determined, the k-NN retrieval overhead of AnomalyDINO remains similarly marginal for the small values of K evaluated here.

E. Additional Qualitative Results

To complement the qualitative experiments in the main paper (Fig. 1 and Fig. 3), this section provides a more extensive set of qualitative results. We present detailed anomaly maps for representative samples from the VisA dataset in Fig. 6 and the MVTec-AD dataset in Fig. 7. These examples further illustrate the model's localization performance across diverse object categories and anomaly types.

F. Failure Cases and Limitations

While SubspaceAD demonstrates strong performance, it is important to acknowledge its limitations, particularly those inherent to patch-based, few-shot methodologies. Two primary modes of failure are identified, which are common challenges in this domain and are illustrated in Fig. 8.

Table 5. Detailed few-shot anomaly segmentation results of *SubspaceAD* on the MVTec-AD dataset. We report mean Image AUROC (%), Image AUPR (%), Pixel AUROC (%), and Pixel PRO (%) results.

Category	1-shot				2-shot				4-shot			
	Img AUROC	Img AUPR	Pxl AUROC	PRO	Img AUROC	Img AUPR	Pxl AUROC	PRO	Img AUROC	Img AUPR	Pxl AUROC	PRO
Bottle	99.9 ± 0.1	100.0 ± 0.0	98.6 ± 0.1	96.0 ± 0.2	100.0 ± 0.0	100.0 ± 0.0	98.7 ± 0.0	96.1 ± 0.2	100.0 ± 0.0	100.0 ± 0.0	98.7 ± 0.0	96.2 ± 0.2
Cable	91.9 ± 1.2	95.4 ± 0.7	95.6 ± 0.2	89.3 ± 0.7	92.5 ± 1.4	96.1 ± 0.7	95.5 ± 0.3	90.2 ± 0.5	93.3 ± 0.6	96.6 ± 0.4	95.7 ± 0.2	90.5 ± 0.5
Capsule	88.4 ± 11.3	96.8 ± 4.0	98.1 ± 0.2	96.6 ± 1.1	92.9 ± 3.1	98.5 ± 0.7	98.2 ± 0.1	97.2 ± 0.2	93.2 ± 5.1	98.5 ± 1.3	98.3 ± 0.1	97.4 ± 0.1
Carpet	100.0 ± 0.0	100.0 ± 0.0	99.2 ± 0.0	98.2 ± 0.1	100.0 ± 0.0	100.0 ± 0.0	99.2 ± 0.0	98.2 ± 0.1	100.0 ± 0.0	100.0 ± 0.0	99.2 ± 0.0	98.2 ± 0.1
Grid	100.0 ± 0.0	100.0 ± 0.0	99.5 ± 0.0	97.8 ± 0.3	100.0 ± 0.0	100.0 ± 0.0	99.5 ± 0.0	98.1 ± 0.2	100.0 ± 0.0	100.0 ± 0.0	99.5 ± 0.0	98.2 ± 0.2
Hazelnut	97.6 ± 3.6	98.6 ± 2.1	99.5 ± 0.2	97.0 ± 1.1	97.2 ± 6.1	98.0 ± 4.3	99.6 ± 0.1	97.6 ± 0.6	99.2 ± 1.8	99.5 ± 1.1	99.6 ± 0.0	97.8 ± 0.2
Leather	100.0 ± 0.0	100.0 ± 0.0	98.9 ± 0.0	98.3 ± 0.2	100.0 ± 0.0	100.0 ± 0.0	98.9 ± 0.0	98.4 ± 0.1	100.0 ± 0.0	100.0 ± 0.0	98.9 ± 0.0	98.4 ± 0.1
Metal Nut	100.0 ± 0.0	100.0 ± 0.0	97.2 ± 0.9	94.8 ± 0.8	100.0 ± 0.0	100.0 ± 0.0	97.7 ± 0.2	95.3 ± 0.2	100.0 ± 0.0	100.0 ± 0.0	97.8 ± 0.1	95.0 ± 0.2
Pill	95.8 ± 0.8	99.2 ± 0.2	95.4 ± 0.5	97.3 ± 0.1	96.6 ± 0.5	99.3 ± 0.1	95.8 ± 0.1	97.4 ± 0.2	97.1 ± 0.6	99.4 ± 0.1	96.2 ± 0.4	97.5 ± 0.2
Screw	89.0 ± 2.1	95.8 ± 0.9	99.1 ± 0.1	96.7 ± 0.3	90.7 ± 2.4	96.5 ± 1.0	99.2 ± 0.1	97.1 ± 0.2	91.5 ± 0.9	96.8 ± 0.5	99.3 ± 0.0	97.3 ± 0.1
Tile	100.0 ± 0.1	100.0 ± 0.0	97.7 ± 0.1	93.6 ± 0.6	100.0 ± 0.0	100.0 ± 0.0	97.7 ± 0.1	93.7 ± 0.1	100.0 ± 0.1	100.0 ± 0.0	97.6 ± 0.0	93.8 ± 0.1
Toothbrush	98.5 ± 1.5	99.4 ± 0.6	98.9 ± 0.6	96.3 ± 0.6	97.8 ± 2.3	99.1 ± 0.9	98.8 ± 0.5	96.2 ± 0.6	99.0 ± 0.6	99.6 ± 0.2	99.2 ± 0.2	96.8 ± 0.5
Transistor	96.6 ± 0.8	94.5 ± 1.4	90.4 ± 1.1	64.9 ± 1.1	96.7 ± 1.3	94.7 ± 2.3	92.1 ± 0.8	66.6 ± 0.6	96.3 ± 2.4	94.7 ± 3.0	92.6 ± 1.1	67.8 ± 0.5
Wood	99.6 ± 0.3	99.9 ± 0.1	96.1 ± 0.4	96.5 ± 0.1	99.6 ± 0.2	99.9 ± 0.1	96.2 ± 0.6	96.7 ± 0.1	99.7 ± 0.1	99.9 ± 0.0	96.4 ± 0.2	96.7 ± 0.2
Zipper	99.6 ± 0.1	99.9 ± 0.0	98.1 ± 0.2	92.3 ± 0.6	99.8 ± 0.1	99.9 ± 0.0	98.3 ± 0.2	92.7 ± 0.7	99.8 ± 0.1	99.9 ± 0.0	98.4 ± 0.2	93.1 ± 0.6
Mean	**97.1 ± 0.9**	**98.6 ± 0.4**	**97.5 ± 0.1**	**93.7 ± 0.2**	**97.6 ± 0.6**	**98.8 ± 0.4**	**97.7 ± 0.1**	**94.1 ± 0.1**	**97.9 ± 0.4**	**99.0 ± 0.2**	**97.8 ± 0.1**	**94.3 ± 0.0**

Table 6. Detailed few-shot anomaly segmentation results of *SubspaceAD* on the VisA dataset. We report mean Image AUROC (%), Image AUPR (%), Pixel AUROC (%), and Pixel PRO (%) results.

Category	1-shot				2-shot				4-shot			
	Img AUROC	Img AUPR	Pxl AUROC	PRO	Img AUROC	Img AUPR	Pxl AUROC	PRO	Img AUROC	Img AUPR	Pxl AUROC	PRO
Candle	94.5 ± 0.8	94.4 ± 0.6	99.4 ± 0.0	95.0 ± 0.4	94.1 ± 0.6	93.6 ± 0.7	99.4 ± 0.0	94.8 ± 0.6	93.9 ± 0.9	93.3 ± 1.1	99.4 ± 0.0	94.8 ± 0.4
Capsules	96.7 ± 0.9	98.0 ± 0.5	98.5 ± 0.2	97.7 ± 0.2	96.8 ± 0.6	98.0 ± 0.3	98.6 ± 0.1	97.8 ± 0.3	97.2 ± 0.8	98.3 ± 0.5	98.6 ± 0.1	98.0 ± 0.3
Cashew	97.7 ± 0.7	98.9 ± 0.3	99.0 ± 0.0	98.2 ± 0.2	97.9 ± 1.3	99.0 ± 0.6	99.2 ± 0.1	98.2 ± 0.2	98.4 ± 1.0	99.2 ± 0.5	99.2 ± 0.1	98.2 ± 0.2
Chewing Gum	99.1 ± 0.2	99.5 ± 0.1	99.6 ± 0.0	96.2 ± 0.3	99.1 ± 0.0	99.5 ± 0.0	99.6 ± 0.0	96.4 ± 0.2	99.1 ± 0.0	99.5 ± 0.0	99.6 ± 0.0	96.4 ± 0.2
Fryum	97.3 ± 0.6	98.8 ± 0.3	96.5 ± 0.3	93.6 ± 0.4	98.0 ± 0.2	99.1 ± 0.1	96.8 ± 0.2	93.5 ± 0.4	98.1 ± 0.4	99.2 ± 0.2	97.1 ± 0.1	93.7 ± 0.4
Macaroni1	92.4 ± 1.3	92.0 ± 1.6	99.7 ± 0.1	94.3 ± 0.9	92.7 ± 1.1	92.2 ± 1.3	99.7 ± 0.0	94.5 ± 0.7	94.3 ± 1.4	93.7 ± 1.3	99.8 ± 0.0	95.2 ± 0.3
Macaroni2	80.8 ± 5.2	76.5 ± 8.1	99.7 ± 0.0	96.6 ± 0.5	80.9 ± 5.6	76.2 ± 9.2	99.7 ± 0.1	97.0 ± 0.4	84.2 ± 3.7	81.0 ± 6.4	99.8 ± 0.0	97.2 ± 0.4
PCB1	92.5 ± 2.6	91.1 ± 1.9	99.3 ± 0.0	93.3 ± 0.7	92.9 ± 0.8	91.2 ± 0.6	99.3 ± 0.0	93.9 ± 0.4	93.5 ± 1.1	91.9 ± 1.0	99.4 ± 0.0	94.0 ± 0.4
PCB2	88.6 ± 2.3	86.6 ± 1.9	96.9 ± 0.3	89.3 ± 1.3	90.5 ± 0.9	87.5 ± 1.3	97.1 ± 0.1	89.6 ± 0.6	91.6 ± 1.4	88.7 ± 2.1	97.2 ± 0.1	90.1 ± 0.4
PCB3	86.4 ± 1.3	86.0 ± 1.7	94.9 ± 0.2	86.4 ± 0.5	88.0 ± 2.3	87.3 ± 2.8	95.2 ± 0.2	86.6 ± 0.2	90.1 ± 2.0	89.9 ± 1.8	95.4 ± 0.1	88.1 ± 0.8
PCB4	98.5 ± 0.6	98.3 ± 0.7	96.4 ± 0.4	83.9 ± 1.6	98.2 ± 0.7	97.9 ± 0.9	96.7 ± 0.4	83.2 ± 1.3	98.4 ± 0.4	98.1 ± 0.6	96.8 ± 0.1	82.0 ± 1.6
Pipe Fryum	95.8 ± 2.0	97.6 ± 1.1	98.7 ± 0.1	97.9 ± 0.1	95.9 ± 2.0	97.7 ± 1.1	98.8 ± 0.1	97.9 ± 0.1	97.8 ± 0.2	98.6 ± 0.2	99.0 ± 0.0	98.0 ± 0.1
Mean	**93.4 ± 0.7**	**93.1 ± 0.8**	**98.2 ± 0.1**	**93.5 ± 0.2**	**93.7 ± 0.5**	**93.3 ± 0.8**	**98.3 ± 0.0**	**93.6 ± 0.1**	**94.7 ± 0.2**	**94.3 ± 0.5**	**98.4 ± 0.0**	**93.8 ± 0.2**

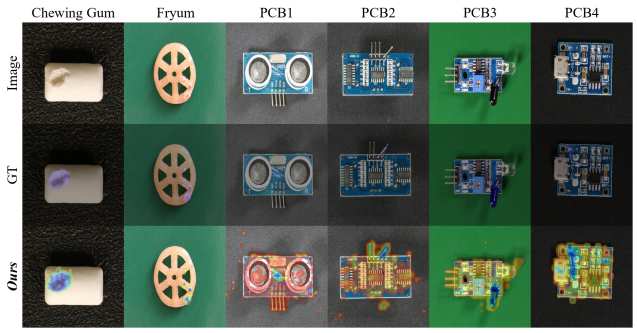

Figure 6. *Additional qualitative results on VisA.* Examples from six categories (*PCB1–4*, *Chewing Gum*, *Fryum*). Rows show the input image, ground-truth mask, and our prediction. *SubspaceAD* accurately localizes both subtle texture anomalies and fine structural defects across diverse VisA domains.

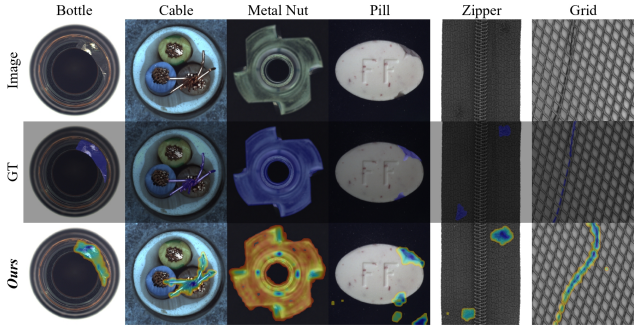

Figure 7. *Additional qualitative results on MVTec-AD.* Examples from six categories (*Bottle*, *Cable*, *Metal Nut*, *Pill*, *Zipper*, *Grid*). Rows show the input image, ground-truth mask, and our prediction. *SubspaceAD* effectively localizes structural, positional, and surface anomalies across varied MVTec categories.

Logical and Structural Anomalies. As a patch-based method, SubspaceAD excels at modeling the *local appearance* and *texture* of normal samples. However, it does not explicitly model global spatial relationships or semantic rules. Consequently, it struggles with logical anomalies, such as a missing component. For example, in the MVTec-AD *Transistor* category (discussed in Appendix A), when a transistor is missing, the exposed circuit-board background may be incorrectly identified as normal texture. The model lacks the semantic, object-level understanding to know that a component *should* be present in that specific location.

Background Artifacts and High Intra-Class Variance. The model struggles in categories with high normal variance or complex, cluttered backgrounds, as seen in some parts of the VisA dataset. As noted in Appendix A, categories Macaroni2 and PCB3

Table 7. Few-shot anomaly detection and localization results using the DINOv3-7B backbone. This backbone is consistently outperformed by DINOv2-G (see Table 1 in the main manuscript).

		MVTec-AD				VisA			
		Image-level		Pixel-level		Image-level		Pixel-level	
Setup	**Method**	AUROC	AUPR	AUROC	PRO	AUROC	AUPR	AUROC	PRO
1-shot	SubspaceAD$_{448}$ (DINOv3)	97.7	98.6	97.4	92.9	91.3	91.4	98.0	91.2
	SubspaceAD$_{672}$ (DINOv3)	97.0	97.8	97.1	93.2	92.8	92.8	98.0	92.7
2-shot	SubspaceAD$_{448}$ (DINOv3)	97.9	98.8	97.6	93.3	93.0	92.8	98.1	91.6
	SubspaceAD$_{672}$ (DINOv3)	97.2	98.2	97.2	93.4	93.8	93.8	98.1	93.0
4-shot	SubspaceAD$_{448}$ (DINOv3)	98.3	99.0	97.8	93.7	93.3	93.0	98.3	91.9
	SubspaceAD$_{672}$ (DINOv3)	98.1	98.7	97.4	93.7	94.2	94.0	98.3	93.3

Table 8. Comparison of anomaly detection and localization performance on MVTec-AD and VisA in the **full-shot** setting. Best results are in **bold**, and second-best are underlined.

		MVTec-AD				VisA			
		Image-level		Pixel-level		Image-level		Pixel-level	
Setup	**Method**	AUROC	AUPR	AUROC	PRO	AUROC	AUPR	AUROC	PRO
Full-shot	AnomalyDINO	**99.5**	**99.8**	**98.2**	**95.0**	97.6	98.0	98.8	**96.1**
Full-shot	*SubspaceAD* (Ours)	99.2	99.6	**98.2**	94.9	**98.2**	**98.3**	**99.1**	94.9

Table 9. Hardware-normalized scoring head latency (H100). We measure the time to process features *after* extraction. SubspaceAD latency is invariant to K, whereas memory-bank retrieval scales with the number of shots.

Method	**Setting (K)**	**Scoring Time (ms / img)**
AnomalyDINO	1-shot	80.1
AnomalyDINO	2-shot	80.3
AnomalyDINO	4-shot	80.5
SubspaceAD (Ours)	1–4 shot	**74.1**

are challenging because their normal samples exhibit significant variation. This makes it difficult to form a single, compact subspace of normality from only a few shots. Furthermore, benign background artifacts (e.g., shadows, debris) that are not present in the few-shot support set may be incorrectly flagged as anomalies, since the model has no mechanism to infer that such regions belong to the normal background.

Outlook. Despite these limitations, the overall results demonstrate that even a simple, training-free subspace model surpasses far more complex approaches in both detection accuracy and efficiency. The clarity of its statistical formulation, combined with its strong few-shot generalization, highlights the potential of foundation-model representations when paired with lightweight, interpretable modeling. Future work may extend this direction by incorporating geometric or semantic priors to better handle logical and structural anomalies.

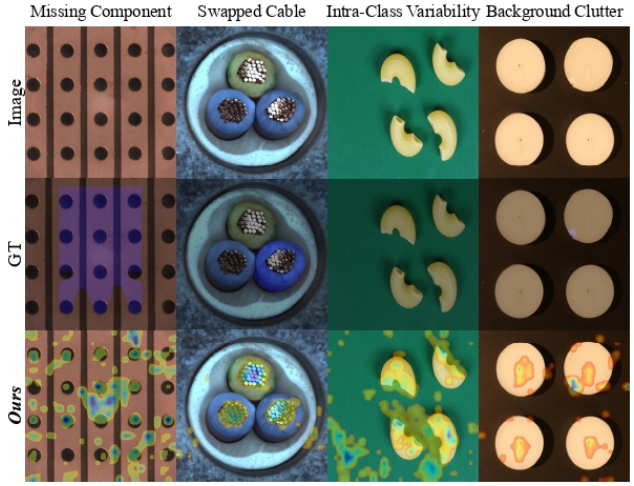

Figure 8. Qualitative failure cases. Each row shows the image, ground-truth mask, and our anomaly map. Examples include missed structural defects (e.g., missing transistor component), incorrect detection of cable swaps, and false positives from intra-class variability or background clutter.

G. Impact of Rotation-Agnostic Preprocessing

In the standard MVTec-AD dataset, the orientation of the *Transistor* object is strictly fixed, meaning misrotation explicitly constitutes an anomaly. To assess the effect of a uniform, rotation-agnostic preprocessing pipeline across all categories, we conducted an ablation where random rotations were applied to the normal support samples.

However, for orientation-dependent categories like *Transistor*,

Table 10. Inference time comparison with AnomalyDINO on MVTec-AD (1-shot, **448px**). Note that AnomalyDINO results are from the authors' paper [11], measured on an NVIDIA A40, while our measurements are obtained on an NVIDIA H100.

Method	Backbone	Params (M)	GPU	Time (ms / image)
WinCLIP [18]	CLIP ViT-B/16+	150.0	NVIDIA T4	389
AnomalyDINO [11]	DINOv2 ViT-S	21.7	NVIDIA A40	60
AnomalyDINO [11]	DINOv2 ViT-B	85.8	NVIDIA A40	84
AnomalyDINO [11]	DINOv2 ViT-L	303.3	NVIDIA A40	141
SubspaceAD (Ours) – DINOv2 Backbones (NVIDIA H100)				
SubspaceAD (Ours)	DINOv2 ViT-S	21.7	NVIDIA H100	36
SubspaceAD (Ours)	DINOv2 ViT-B	85.8	NVIDIA H100	56
SubspaceAD (Ours)	DINOv2 ViT-L	303.3	NVIDIA H100	112
SubspaceAD (Ours)	DINOv2 ViT-G	1100.0	NVIDIA H100	127
SubspaceAD (Ours) – DINOv3 Backbones (NVIDIA H100)				
SubspaceAD (Ours)	DINOv3 ViT-S	21.0	NVIDIA H100	16
SubspaceAD (Ours)	DINOv3 ViT-B	86.0	NVIDIA H100	21
SubspaceAD (Ours)	DINOv3 ViT-7B	7000.0	NVIDIA H100	330

rotating the normal samples during PCA fitting fundamentally alters the model's definition of normality. By incorporating rotated features into the normal subspace, the model inadvertently learns to accept rotational defects as normal variations, causing it to miss actual misrotation anomalies during inference.

Table 11 outlines the performance under this rotation-agnostic protocol. As expected, forcing the subspace to account for rotational variance leads to explicit performance deltas compared to the standard aligned baseline (cf. Table 5). In the 1-shot setting, Image AUROC decreases by 7.4% (from 96.6% to 89.2%) and Pixel PRO drops by 3.8% (from 64.9% to 61.1%). This degradation persists in the 4-shot setting, which exhibits a 5.0% delta in Pixel PRO (dropping from 67.8% to 62.8%). These results demonstrate that applying uniform rotational preprocessing is counterproductive for categories where orientation is a defining characteristic of the normal state.

Table 11. Few-shot anomaly segmentation results of *SubspaceAD* on the MVTec-AD *Transistor* category using a **rotation-agnostic** protocol. We report mean Image AUROC (%), Image AUPR (%), Pixel AUROC (%), and Pixel PRO (%) ± standard deviation across 5 seeds.

Shots	Img AUROC	Img AUPR	Pxl AUROC	Pxl PRO
1-shot	89.2 ± 7.6	84.9 ± 6.0	82.8 ± 2.8	61.1 ± 1.0
2-shot	92.8 ± 3.1	88.4 ± 3.9	83.1 ± 1.0	61.9 ± 1.1
4-shot	93.5 ± 3.8	90.0 ± 4.0	83.9 ± 1.5	62.8 ± 0.6