

DL-SubspaceAD: Training-Free Anomaly Segmentation via Deep Layer DINOv3 Features on Full-Resolution MVTec AD 2

VAND 4.0 Challenge: Industrial Track (Regular Setting)

Code: https://github.com/riips/VAND4.0_IndustrialTrack_Regular

Tenshi Ito
Yachiyo Engineering Co., Ltd.
tn-ito@yachiyo-eng.co.jp

Abstract

This technical report presents Deep Layer SubspaceAD (DL-SubspaceAD) for the CVPR 2026 VAND 4.0 Challenge Industrial Track (Regular setting). We focus on training-free unsupervised anomaly detection built on foundation models and adopt SubspaceAD, which has recently achieved strong performance. SubspaceAD detects anomalies using a low-dimensional subspace constructed from features extracted from arbitrary layers of a frozen foundation model. By selecting deeper layers for SubspaceAD than in the original formulation, we obtain substantially improved performance. For binarizing anomaly maps, we employ MEBin and SAMFiner, which were used in the top-performing RoBiS method at VAND 3.0. On the official leaderboard, our method achieves 8-category average SegF1 scores of 46.88% on $\text{TEST}_{\text{Private}}$ and 45.36% on $\text{TEST}_{\text{PrivateMixed}}$.

1. Introduction

1.1. Background

Anomaly detection and segmentation in industrial imagery aim to identify and localize subtle deviations from nominal product appearance. The problem is difficult because defects are rare and visually diverse, and sufficient defective examples are often unavailable during development. Under this setting, unsupervised learning from normal images alone is the dominant paradigm; on standard benchmarks such as MVTec AD [1] and VisA [13], many methods—including memory bank approaches such as PatchCore [9]—have reported strong performance.

By contrast, on MVTec AD 2 [4], a successor benchmark designed to reflect more realistic capture conditions, resolutions, and defect scales, fine defects and distribu-

tion shift due to illumination changes are key challenges, and methods that excel on earlier benchmarks can suffer large accuracy drops. MVTec AD 2 targets high-resolution industrial images and provides, in addition to the public test split with released ground truth ($\text{TEST}_{\text{Public}}$), private test splits ($\text{TEST}_{\text{Private}}$ and $\text{TEST}_{\text{PrivateMixed}}$) evaluated on undisclosed labels; the mixed private split emphasizes greater diversity of imaging conditions. In such settings, aggressive resizing can weaken defect cues, and how continuous score maps are converted into final binary masks strongly affects performance, so both representation density and downstream binarization matter.

At CVPR 2025, VAND 3.0 [3] elicited a wide range of solutions for MVTec AD 2. For example, ISVL [11] combines multi-scale feature embeddings with prototype-based reconstruction in a hybrid framework aimed at localization under difficult conditions involving transparent objects and illumination variation. RoBiS [8] trains detectors on high-resolution inputs with Swin-Cropping and pseudo-illumination augmentation, and achieves high SegF1 as an industrial pipeline that refines binary masks with MEBin [5] and SAMFiner. SuperAD [12] builds a memory bank from DINOv2 features and estimates anomalous pixels by nearest-neighbor search alone, demonstrating that a concise training-free design can still be competitive.

However, many strong VAND 3.0 methods rely in practice on category-dependent designs, such as switching pre-processing, binarization, or training procedures per object category.

1.2. Challenge description

The CVPR 2026 VAND 4.0 Industrial Track evaluates anomaly segmentation on MVTec AD 2 [4]. As with MVTec AD and VisA, the dataset assumes learning and validation from normal images, while incorporating scenarios closer to deployment—for example, by emphasizing dis-

tribution shift through imaging conditions that differ from training—to narrow the gap between research benchmarks and field requirements. In the Regular setting, only the defect-free Train set is used per object category; the Validation set may be used to estimate thresholds and related quantities, but images from $\text{TEST}_{\text{Public}}$ must not be used for training. Overall performance is based on pixel-level F1 (SegF1), measured on both $\text{TEST}_{\text{Private}}$ (normal and anomalous images captured under the same illumination as training) and $\text{TEST}_{\text{PrivateMixed}}$ (normal and anomalous images under diverse illumination, including conditions not seen during training); methods are ranked by the average rank over 8 categories on these two splits. The challenge further requires that model design be category-agnostic: hyperparameters and the overall architecture must be shared across all object categories or adapt automatically to the inspected object. Thus, VAND 4.0 uses the same primary metric as VAND 3.0—SegF1 on MVTec AD 2—while explicitly requiring category-agnostic designs, which is a major difference from the previous edition.

2. Methodology

2.1. Model design

2.1.1. Approach

SuperAD, a top-performing method at the VAND 3.0 Challenge, is a training-free approach that stores strong visual features extracted from DINOv2 in a memory bank and estimates anomalous regions by nearest-neighbor search against test image features. This result shows that, without additional training, appropriately leveraging foundation model representations can yield feature spaces effective for industrial anomaly detection. Motivated by this perspective, we focus on statistical modeling of normality that directly exploits foundation model features.

Our method adopts SubspaceAD [7], a recent approach proposed for CVPR 2026. SubspaceAD extracts patch features from normal images using a frozen foundation model and applies PCA to estimate a low-dimensional subspace in which normal patch features predominantly lie. At inference time, patch features from a test image are reconstructed in this subspace, and the reconstruction residual with respect to the original features serves as the anomaly score. This design builds a per-category normality reference from normal images alone, without additional neural network training or a large memory bank.

2.1.2. Architecture

Our pipeline comprises four stages: (i) DINOv3 feature extraction at native resolution, (ii) per-category construction of a PCA-based normal subspace, (iii) generation of a continuous anomaly map from PCA reconstruction residuals, and (iv) binarization with MEBin and SAMFiner.

As the feature extractor, we use DINOv3 ViT-7B/16 [10], which consists of 40 Transformer blocks. Because MVTec AD 2 contains anomalies that are very small relative to image size, we do not uniformly resize images to a low resolution; instead, we feed DINOv3 inputs at the original image resolution. Since a ViT processes images as fixed-size patches, we apply zero padding only to the right and bottom edges when necessary.

To construct the per-category PCA normal subspace, we convert all Train set images into patch features with DINOv3 and average features from layers 39, 40 (out of 40) at each patch location. We then apply PCA to all patch features obtained from normal training images and select the number of principal components per category so that the cumulative explained variance reaches 99%. This layer choice applies to ViT-7B/16 the trend observed in the layer ablation on DINOv3 ViT-H+/16 in Sec. 4.3.1, where deeper layers yielded the highest accuracy.

At inference time, we project patch features extracted from a test image into the PCA subspace of the corresponding category. We compute the squared norm of the difference between the original features and their reconstruction in the PCA subspace, and use it as the patch-wise anomaly score.

Finally, we convert the continuous anomaly map into a binary anomaly map. We apply MEBin and SAMFiner from RoBiS as post-processing. MEBin extracts candidate anomalous regions from the continuous map, and SAMFiner refines them using the segmentation capability of SAM [6]. Unlike the original RoBiS paper, we do not use test-set- or category-dependent processing; we apply a unified procedure across all settings. Specifically, we use ViT-H as the SAM backbone in SAMFiner and retain only the highest-confidence mask among the confidence-weighted masks produced by SAM.

2.1.3. Training

Our method requires no training. The frozen DINOv3 weights are fixed; the only quantities estimated per category are PCA statistics computed from patch features of all normal images in the Train set.

3. Dataset & Evaluation

3.1. Benchmark dataset utilization

We use the industrial dataset MVTec AD 2. Pre-processing follows Sec. 2.1.2: native resolution input with minimal zero padding only; we do not apply data augmentation.

3.2. Evaluation criteria

We measure anomaly detection performance with SegF1. Computing SegF1 requires a binary anomaly map derived from the continuous anomaly map; we generate it with MEBin and SAMFiner from RoBiS. For the binary anomaly

Table 1. SegF1 scores (in %) on $\text{TEST}_{\text{Private}}$ and $\text{TEST}_{\text{PrivateMixed}}$ on the VAND 4.0 Industrial Track leaderboard.

Category	$\text{TEST}_{\text{Private}}$	$\text{TEST}_{\text{PrivateMixed}}$
can	1.30	0.58
fabric	70.91	68.59
fruit jelly	47.88	45.97
rice	61.90	61.35
sheet metal	75.43	72.71
vial	44.90	44.98
wallplugs	3.38	3.31
walnuts	69.32	65.41
Average	46.88	45.36

map, SegF1 is computed from pixel-level precision and recall:

$$\text{SegF1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (1)$$

Following the challenge evaluation protocol, the final ranking is determined by the average of the ranks of mean SegF1 on $\text{TEST}_{\text{Private}}$ and $\text{TEST}_{\text{PrivateMixed}}$:

$$\text{rank}_{\text{final}} = \frac{\text{rank}(\text{SegF1}_{\text{Private}}) + \text{rank}(\text{SegF1}_{\text{PrivateMixed}})}{2}. \quad (2)$$

3.3. Evaluation setting

Our method follows the Regular setting. We do not use any images from $\text{TEST}_{\text{Public}}$ of MVTEC AD 2 for training; we build the SubspaceAD normal subspace using only the Train set of each target category. Hyperparameters and model architecture are shared across all categories; we do not employ category-dependent design.

4. Results

4.1. Performance metrics

Tab. 1 reports results on $\text{TEST}_{\text{Private}}$ and $\text{TEST}_{\text{PrivateMixed}}$. Our method achieves 46.88% on $\text{TEST}_{\text{Private}}$ and 45.36% on $\text{TEST}_{\text{PrivateMixed}}$. The gap between the two splits is 1.52 percentage points in SegF1.

4.2. Comparison

Tab. 2 compares our method with approaches from VAND 3.0. Our method achieves performance comparable to SuperAD, a VAND 3.0 method, while using a single design shared across all categories, whereas VAND 3.0 methods are tuned per category.

4.3. Ablation study

The official DINOv3 repository releases pretrained weights for multiple model sizes; the largest is ViT-7B/16 and the

next largest is ViT-H+/16. Although ViT-7B/16 is expected to yield better anomaly detection due to stronger generic representations, its large parameter count and slow inference make it impractical for rapid experimentation. We therefore iterate on ViT-H+/16 and transfer effective design choices to ViT-7B/16. This section reports layer ablations for PCA on DINOv3 ViT-H+/16 and an ablation of SAMFiner on DINOv3 ViT-7B/16.

4.3.1. Layer ablation

We evaluate 16 configurations on ViT-H+/16, selecting pairs of layers every two layers across all 32 layers. For each configuration, we average features from the two selected layers and apply PCA to build the normal subspace. Tab. 3 shows results on $\text{TEST}_{\text{Public}}$ when varying the selected layers (MEBin only). Performance varies substantially with layer choice. The best result is obtained with layers 31, 32, achieving mean SegF1 of 49.30%. In addition to 31, 32, deeper pairs such as 29, 30 also perform comparatively well, suggesting that features from deeper layers of foundation models such as DINOv3 are important for anomaly detection. Based on these results, the best configuration on ViT-H+/16 uses the last two of 32 layers (31, 32); we apply the same policy to ViT-7B/16 by averaging layers 39, 40 out of 40.

4.3.2. Post-processing ablation

Tab. 4 reports SegF1 on $\text{TEST}_{\text{Public}}$ for DINOv3 ViT-7B/16 with and without SAMFiner. Applying SAMFiner improves SegF1 by 3.8 percentage points (from 51.51% to 55.31%).

5. Discussion

5.1. Challenges & Solutions

The challenge disallows category-dependent design. Initially, following the official SubspaceAD implementation, we used intermediate-to-final layers (55%–70% depth) for subspace fitting and applied data augmentation such as rotation and exposure changes. While some categories improved, others degraded, lowering mean SegF1 overall. We concluded that improving model generality is more important than fixing individual failure modes, and focused on layer selection based on the observation that ViTs capture different representations at different depths [2]. As shown in Sec. 4.3.1, layer ablation confirmed that deeper features are effective; on ViT-7B/16, using the last two layers 39, 40 out of 40 yielded strong performance. This suggests that, on difficult settings such as MVTEC AD 2 where robustness is required, selecting layers that capture semantic representations of foundation models is important.

Table 2. SegF1 (in %) on $\text{TEST}_{\text{Private}}$ / $\text{TEST}_{\text{PrivateMixed}}$ for VAND 3.0 methods and ours. ISVL, RoBiS, and SuperAD use category-dependent designs; ours is category-agnostic.

Category	ISVL	RoBiS	SuperAD	Ours
can	17.03 / 9.21	1.86 / 0.84	17.3 / 1.9	1.30 / 0.58
fabric	84.95 / 81.76	87.46 / 73.37	77.4 / 65.3	70.91 / 68.59
fruit jelly	68.1 / 67.76	53.63 / 52.62	41.3 / 40.9	47.88 / 45.97
rice	66.34 / 66.93	63.86 / 63.23	60.9 / 61.2	61.90 / 61.35
sheet metal	69.69 / 69.74	70.98 / 70.92	59.5 / 59.7	75.43 / 72.71
vial	44.01 / 42.12	48.73 / 48.83	42.8 / 40.8	44.90 / 44.98
wallplugs	11.99 / 6.24	14.38 / 3.40	13.7 / 6.7	3.38 / 3.31
walnuts	68.34 / 67.65	67.13 / 58.94	69.1 / 69.1	69.32 / 65.41
Average	53.81 / 51.43	51.00 / 46.52	47.8 / 43.2	46.88 / 45.36

Table 3. Layer ablation on DINOv3 ViT-H+/16 on $\text{TEST}_{\text{Public}}$. Mean SegF1 (in %) with MEBin only.

Layer	$\text{TEST}_{\text{Public}}$
1, 2	9.06
3, 4	9.99
5, 6	10.37
7, 8	12.31
9, 10	12.76
11, 12	19.66
13, 14	25.36
15, 16	27.09
17, 18	30.91
19, 20	31.24
21, 22	33.94
23, 24	32.93
25, 26	34.47
27, 28	37.79
29, 30	42.82
31, 32	49.30

Table 4. Effect of SAMFiner on DINOv3 ViT-7B/16: SegF1 (in %) on $\text{TEST}_{\text{Public}}$.

	$\text{TEST}_{\text{Public}}$
MEBin	51.51
MEBin + SAMFiner	55.31

5.2. Model Robustness & Adaptability

We attribute robustness to illumination variation partly to using features from deep layers such as 39, 40 in DINOv3 ViT-7B/16. As shown in Tab. 1, the gap between $\text{TEST}_{\text{Private}}$ and $\text{TEST}_{\text{PrivateMixed}}$ is only 1.52 percentage points in SegF1, indicating strong robustness to imaging conditions. Deeper layers are believed to encode more

semantic representations, which may explain this behavior; detailed analysis remains future work.

5.3. Future Work

Our training-free approach leveraging foundation model representations achieves competitive accuracy. However, we use only a subset of the deepest layers. If each layer captures complementary representations, combining them may further improve performance. Finer analysis of per-layer features and methods that exploit foundation model representations more fully remain important directions.

6. Conclusion

This report summarizes our anomaly segmentation method for MVTEC AD 2 submitted to the CVPR 2026 VAND 4.0 Industrial Track (Regular setting). Under the class-agnostic constraint, we extract dense patch features at native resolution with frozen DINOv3 ViT-7B/16, estimate a per-category normal subspace by PCA from normal training images only, and form continuous anomaly maps from SubspaceAD-style reconstruction residuals without additional neural training. Layer selection is fixed to the average of the last two layers 39, 40 out of 40 on ViT-7B/16, guided by ablations on the lighter ViT-H+/16 backbone on $\text{TEST}_{\text{Public}}$. For binarization, we use MEBin and SAMFiner from RoBiS.

On the official private test leaderboard, we achieve 8-category average SegF1 scores of 46.88% on $\text{TEST}_{\text{Private}}$ and 45.36% on $\text{TEST}_{\text{PrivateMixed}}$ (Tab. 1). With the same model design, performance on $\text{TEST}_{\text{PrivateMixed}}$ degrades only modestly under changed illumination, indicating reasonable robustness under SegF1, while large score variation across categories remains; analyzing failure modes and improving low-scoring categories are key next steps. Richer integration of multi-layer features and further generalization of the binarization pipeline, as outlined in the Discussion, are directions for future work.

References

- [1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 1
- [2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 3
- [3] Lars Heckler-Kram, Ashwin Vaidya, Jan-Hendrik Neudeck, Ulla Scheler, Dick Ameln, Samet Akcay, and Paula Ramos. From benchmarks to reality: Advancing visual anomaly detection by the vand 3.0 challenge. *arXiv preprint arXiv:2509.17615*, 2025. 1
- [4] Lars Heckler-Kram, Jan-Hendrik Neudeck, Ulla Scheler, Rebecca König, and Carsten Steger. The mvtec ad 2 dataset: Advanced scenarios for unsupervised anomaly detection. *International Journal of Computer Vision*, 134(4):175, 2026. 1
- [5] Ziming Huang, Xurui Li, Haotian Liu, Feng Xue, Yuzhe Wang, and Yu Zhou. Anomalyncd: Towards novel anomaly class discovery in industrial scenarios. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 4755–4765, 2025. 1
- [6] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 2
- [7] Camile Lendering, Erkut Akdag, and Egor Bondarev. Subspacead: Training-free few-shot anomaly detection via subspace modeling. *arXiv preprint arXiv:2602.23013*, 2026. 2
- [8] Xurui Li, Zhongsheng Jiang, Tingxuan Ai, and Yu Zhou. Robis: Robust binary segmentation for high-resolution industrial images. *arXiv preprint arXiv:2505.21152*, 2025. 1
- [9] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328, 2022. 1
- [10] Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025. 2
- [11] Xingao Wang, Shuying Xia, Zhaohong Liao, Mengjie Xie, Handa Wang, and Zhi Gao. Accurate anomaly localization in challenging industrial settings via a hybrid detection framework. <https://github.com/ISVL119/isvl/blob/main/vand3.0.pdf>, 2025. VAND 3.0 Challenge, CVPR 2025 Workshop on Visual Anomaly and Novelty Detection. 1
- [12] 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. 1
- [13] 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*, pages 392–408. Springer, 2022. 1