TPFREEDAS: TRAINABLE-PARAMETER-FREE DINO-GUIDED ANOMALY SYNTHESIS FOR FEW-SHOT ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Few-shot anomaly detection relies on synthetic anomalies when real anomalies and annotations are scarce. Yet a spatially plausible CutPaste-style perturbation may still be unsuitable for the image structure or local material or functional region. We frame synthesis validity as three progressively finer compatibility conditions: foreground support, image-level synthesis suitability, and local-region compatibility. TPFreeDAS is a trainable-parameter-free DINO-guided framework that addresses these conditions by reusing frozen DINO representations already available in DINO-based detectors, without an extra feature backbone. **Foreground-Constrained Anomaly Synthesis (FCAS)** estimates placement support from DINO patch responses. **Structure-Adaptive Synthesis Routing (SASR)** selects a synthesis path using foreground coverage, feature dispersion, and boundary complexity as reliability cues. For selected structured categories, **Region-Aware and Quality-Gated Synthesis (RAQG)** uses representation-derived pseudo-regions and validates feature changes inside and outside the intended anomaly mask. TPFreeDAS changes only training-time synthesis, adding no learnable synthesis parameters or inference-time overhead. On PCBMarket, it reduces 4-shot per-image synthesis latency by 35.51% relative to our CutPaste implementation and exceeds FoundAD by up to 14.89 percentage points in I-AUROC and 12.10 in PRO; it improves most reported VisA and MVTec AD metrics. All training and inference code and model weights will be released.

1 INTRODUCTION

Anomaly detection aims to identify defective products and localize potential anomalous regions from manufacturing images (Bergmann et al., 2019). Because real anomalies are scarce, diverse, and costly to annotate, few-shot methods adapt detectors to new categories using only one or a few normal references (Jeong et al., 2023; Li et al., 2024; Fang et al., 2023; Lv et al., 2025). Vision foundation models further improve this setting through transferable patch-level representations (Caron et al., 2021; Oquab et al., 2023; Siméoni et al., 2025). In particular, FoundAD (Zhai et al., 2026) uses a frozen visual encoder and a lightweight predictor to recover normal features from synthetic anomalous contexts; feature residuals are then used for image-level detection and pixel-level localization. Since no real anomalies are available during training, it constructs contexts with CutPaste (Li et al., 2021). Consequently, synthesis quality determines whether the predictor learns meaningful anomalous deviations or artificial generation cues. Synthesis validity is hierarchical: plausible placement alone may not ensure compatibility with image structure or local regions. FoundAD's category-level heuristics and fixed CutPaste policy can lead to three progressively finer mismatches. First, a **foreground–placement mismatch** occurs when coarse masks retain backgrounds or exclude valid object regions, producing spatially implausible anomalies. Second, a **structure–strategy mismatch** arises because foreground constraints that suit compact objects may become fragmented or overly restrictive for dense textures, repetitive boundaries, or distributed components. Third, a **region–anomaly compatibility mismatch** remains within structured foregrounds containing different material or functional regions, where an anomaly may be placed on an incompatible local structure. Figure 1 illustrates these limitations on dense, smooth, and component-rich images.

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.

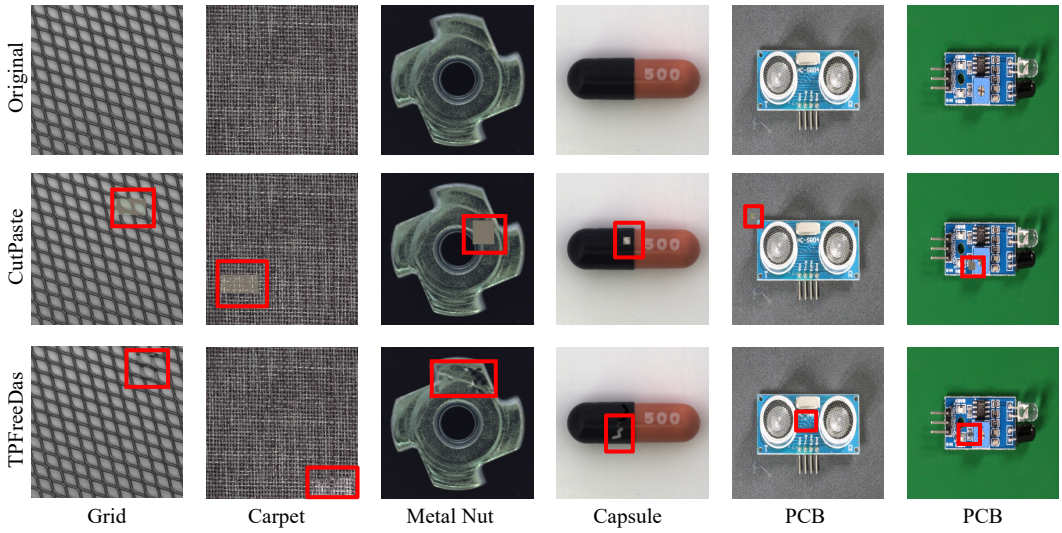

Figure 1: CutPaste (Li et al., 2021) and TPFreeDAS synthesis examples on grid, carpet, metal nut, capsule, and PCB images. The examples illustrate foreground-placement, image-structure, and local-region compatibility issues. The red boxes mark synthetic anomalies.

We propose **TPFreeDAS**, a trainable-parameter-free DINO-guided synthesis framework that addresses these levels while retaining the FoundAD encoder, predictor, objective, and inference pipeline. It reuses the detector's frozen DINO features without an additional synthesis backbone. **Foreground-Constrained Anomaly Synthesis** estimates spatial support for placement; **Structure-Adaptive Synthesis Routing** selects between foreground-constrained and CutPaste paths according to support reliability; and **Region-Aware and Quality-Gated Synthesis** constrains local placement and validates feature changes for selected structured categories.

We contribute: 1) a three-level decomposition of CutPaste-based synthesis compatibility into spatial validity, image-level suitability, and local-region compatibility; 2) TPFreeDAS, which reuses frozen DINO features in FCAS, SASR, and RAQG without a synthesis backbone, learnable synthesis parameters, auxiliary losses, or test-time computation; and 3) 1-, 2-, and 4-shot results on VisA, MVTec AD, and PCBMarket, including CutPaste replacement in OneNIP, SuperSimpleNet, and FoundAD, with PCBMarket gains over FoundAD of up to 14.89 and 12.10 points in I-AUROC and PRO, respectively.

2 Related Work

2.1 Foundation Model-Based Anomaly Detection

Visual foundation models provide transferable representations for anomaly detection: DINO learns object-centric spatial patch features (Caron et al., 2021; Oquab et al., 2023; Siméoni et al., 2025), whereas CLIP aligns visual and textual semantics (Radford et al., 2021). Earlier feature-based methods, including SPADE (Cohen & Hoshen, 2020) and PatchCore (Roth et al., 2022), compare test features with normal references or memory-bank embeddings. Recent few-shot methods exploit stronger pretrained priors through compositional vision-language prompts (Jeong et al., 2023), one-class prompt learning (Li et al., 2024), support-feature reconstruction (Fang et al., 2023), instance-induced prompts (Lv et al., 2025), or diffusion-based anomaly reconstruction (Yan et al., 2026). FoundAD (Zhai et al., 2026) instead learns a predictor that maps anomalous context features toward their normal counterparts. TPFreeDAS retains this backbone and residual-based inference, focusing solely on better synthetic context construction.

2.2 Synthetic Anomaly Generation

Synthetic anomalies provide training signals without requiring real anomaly annotations. CutPaste (Li et al., 2021) pastes transformed normal regions to learn anomaly-sensitive representations, while Spot-the-Difference (Zou et al., 2022) learns from synthetically altered image pairs. FoundAD (Zhai et al., 2026) uses CutPaste-style images as anomalous contexts and normal-image features as prediction targets. Diffusion-based methods have also been explored for anomaly reconstruction using pretrained generative priors (Rombach et al., 2022); for example, AnomalySD (Yan et al., 2026) performs text-guided reconstruction and detects anomalies from reconstruction discrepancies. TPFreeDAS reuses frozen DINO representations for spatial support, synthesis-path selection, and local quality validation.

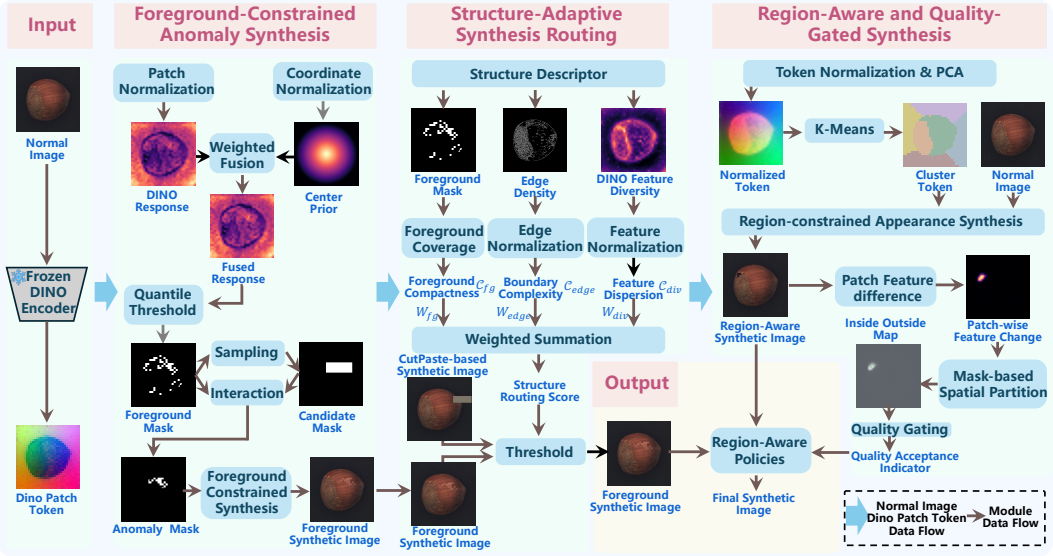

Figure 2: TPFreeDAS uses frozen DINO features at three levels of synthesis compatibility: FCAS estimates spatial support, SASR selects a path according to foreground-constraint reliability, and RAQG constrains local placement and validates feature changes for selected structured categories. The selected image is used only during training.

3 Methodology

3.1 Problem Formulation and Framework Overview

We consider few-shot anomaly detection, where only a few normal images are available for adapting an anomaly detector to a new category. For category c, the normal training set is denoted by $\mathcal{D}_{\mathrm{tr}}^{c} = \{I_i^c\}_{i=1}^{K_s}$, where $K_s \in \{1, 2, 4\}$. No real anomaly images or pixel-level anomaly annotations are used during training.

Given a normal image $I \in \mathbb{R}^{3\times H\times W}$, we employ DINOv3 ViT-B/16 (Siméoni et al., 2025) as a frozen visual encoder Φ^* to extract spatial patch tokens:

$$Z = \Phi^*(I) = [z_1, \ldots, z_{N_p}]^\top \in \mathbb{R}^{N_p\times D}, \tag{1}$$

where N_p and D denote the number and dimension of patch tokens, respectively.

Following FoundAD Zhai et al. (2026), a lightweight predictor Ψ reconstructs the normal representation from a synthetic anomalous context, using features extracted from the corresponding normal image as the target. FreeDAS modifies only the construction of this context while retaining the frozen DINOv3 encoder, predictor, objective, and inference pipeline. FreeDAS addresses the foreground–placement, structure–strategy, and region–anomaly compatibility mismatches through three modules. Foreground-Constrained Anomaly Synthesis (FCAS) restricts anomaly placement

using an instance-level foreground mask. Structure-Adaptive Synthesis Routing (SASR) combines foreground compactness, feature dispersion, and boundary complexity to select between FCAS and the original CutPaste path. For selected structured categories, Region-Aware and Quality-Gated Synthesis (RAQG) constructs pseudo-structural regions, generates region-compatible anomalies, and retains candidates with localized feature changes. We denote the outputs of the foreground-constrained, CutPaste-based, and region-aware paths by I_s^{fg}, I_s^{base}, and I_s^{reg}, respectively, and the final selected synthetic image by I_s.

3.2 Foreground-Constrained Anomaly Synthesis

Foreground-Constrained Anomaly Synthesis (FCAS) estimates placement support, not semantic segmentation. It reuses frozen DINO patch magnitudes as the primary spatial cue and normalized coordinates only for a weak center prior. For each patch token z_i, its ℓ_2 magnitude is normalized within the current image to obtain the DINO response $\widehat{s}_i$. Independently, the normalized patch coordinate $(x_i, y_i) \in [-1, 1]^2$ is converted into the center prior c_i. The two signals are combined through weighted fusion:

$$\begin{aligned} \widehat{s}_i &= \frac{\|z_i\|_2 - \min_j \|z_j\|_2}{\max_j \|z_j\|_2 - \min_j \|z_j\|_2 + \epsilon}, \\ c_i &= 1 - \mathrm{clip}\left(\sqrt{x_i^2 + y_i^2}, 0, 1\right), \\ r_i &= \lambda_r \widehat{s}_i + \lambda_c c_i, \end{aligned} \tag{2}$$

where r_i is the fused response and λ_r and λ_c weight the DINO response and center prior, respectively. With $\lambda_r > \lambda_c$, the latter serves only as a weak spatial regularizer.

The fused response is converted into a patch-level foreground estimate using an image-adaptive quantile threshold:

$$\begin{aligned} \tau_{\mathrm{fg}} &= \mathrm{Quantile}\left(\{r_i\}_{i=1}^{N_p}, q\right), \\ M_{\mathrm{fg}}^p(i) &= \mathbb{I}\left[r_i \geq \tau_{\mathrm{fg}}\right], \end{aligned} \tag{3}$$

where $q = 0.55$. The patch mask is reshaped and mapped to the original resolution of the normal image. Standard morphological closing and opening are subsequently applied to suppress isolated responses and fill small discontinuities, producing the foreground mask M_{fg}.

The foreground mask specifies where an anomaly may occur but does not determine its shape, scale, or local position. We therefore sample an initial candidate mask C from a CutPaste-style mask distribution $\mathcal{P}_{\mathrm{mask}}$. The sampled candidate is then restricted through mask interaction:

$$C \sim \mathcal{P}_{\mathrm{mask}}(I),\ M = \mathrm{Interact}\left(C, M_{\mathrm{fg}}\right) = C \cap M_{\mathrm{fg}}, \tag{4}$$

where C is the candidate mask before foreground restriction and M is the final anomaly mask. This interaction removes candidate pixels located outside the estimated foreground while retaining the sampled local anomaly shape inside the estimated spatial support.

Given an anomaly mask M, the normalized image is mapped back to the RGB range as $\bar{I} = \mathrm{Denorm}(I)$. Foreground-constrained appearance synthesis then blends a CutPaste-style transformation $T(\bar{I})$ into the selected anomaly region:

$$I_s^{\mathrm{fg}} = \mathrm{Norm}\left[\bar{I} \odot (1 - \mathcal{G}_\sigma(M)) + T(\bar{I}) \odot \mathcal{G}_\sigma(M)\right], \tag{5}$$

where $\mathcal{G}_\sigma(\cdot)$ smooths the anomaly-mask boundary and I_s^{fg} is the foreground-constrained synthetic image. The anomaly mask is used only for sample construction and does not serve as an additional supervision target.

3.3 Structure-Adaptive Synthesis Routing

SASR estimates when foreground-constrained synthesis may be unreliable, rather than classifying image structure. Dense boundaries, repetitive textures, or distributed components can fragment

FCAS support. The descriptors C_{fg}, C_{div}, and C_{edge} provide complementary reliability cues: foreground coverage, token dispersion, and boundary density:

$$\begin{aligned} \widetilde{z}_i &= \frac{z_i}{\|z_i\|_2 + \epsilon}, \qquad E = \mathrm{EdgeNorm}(\widetilde{E}), \\ \big(C_{\mathrm{fg}}, C_{\mathrm{div}}, C_{\mathrm{edge}}\big) &= \Big(\langle M_{\mathrm{fg}}\rangle_{xy}, \mathrm{clip}\big(\sum_{d=1}^{D} \mathrm{Var}_i[\widetilde{z}_{i,d}], 0, 1\big), \mathrm{clip}\big(\langle E\rangle_{xy}/\kappa, 0, 1\big)\Big), \end{aligned} \tag{6}$$

where $\langle A\rangle_{xy} = (HW)^{-1}\sum_{x=1}^{W}\sum_{y=1}^{H} A(x,y)$ denotes the spatial mean. The maps $\widetilde{E}$ and E are the raw and normalized edge responses, respectively. The operator $\mathrm{Var}_i[\cdot]$ computes variance over all patch tokens indexed by i, and $\kappa = 0.20$ is the edge-density normalization factor.

A fixed weighted sum combines these cues for synthesis selection:

$$S_{\mathrm{route}}(I) = w_{\mathrm{fg}} C_{\mathrm{fg}} + w_{\mathrm{div}} C_{\mathrm{div}} + w_{\mathrm{edge}} C_{\mathrm{edge}}, \tag{7}$$

where w_{fg}, w_{div}, and w_{edge} weight foreground coverage, feature dispersion, and boundary complexity, respectively. The score is a fixed heuristic for synthesis selection. With $\delta = 0.50$, $I_s^{\mathrm{gen}} = I_s^{\mathrm{fg}}$ when $S_{\mathrm{route}}(I) < \delta$ and $I_s^{\mathrm{gen}} = I_s^{\mathrm{base}}$ otherwise. The latter conservatively recovers the baseline CutPaste path when the foreground constraint may be unreliable.

3.4 Region-Aware and Quality-Gated Synthesis

Even within estimated foreground support, an anomaly may cross incompatible local structures. For selected structured categories, RAQG forms a coarse representation-derived partition to constrain placement, not to identify semantic parts. Normalized frozen tokens, low-rank PCA, and K-Means implement it:

$$\begin{aligned} X &= \mathrm{TokenNorm}(Z), \qquad U = \mathrm{PCA}_{d_{\mathrm{pca}}}\left(X - \frac{1}{N_p}\mathbf{1}\mathbf{1}^{\top}X\right), \\ L &= \mathrm{KMeans}(U, K), \qquad C_{\mathrm{reg}} = \mathrm{Upsample}\left[\mathrm{Reshape}(L)\right]. \end{aligned} \tag{8}$$

where X is the normalized token representation, U is the compact cluster-token representation, d_{pca} is the retained PCA dimension, L contains the cluster assignments, and C_{reg} is the pseudo-region map at the input-image resolution.

The pseudo-region map provides structural support for anomaly placement. Given a sampled anomaly type a, TPFreeDAS selects a structurally compatible region and performs region-constrained synthesis as $(M_{\mathrm{reg}}, I_s^{\mathrm{reg}}) = \mathcal{S}_{\mathrm{reg}}(I, C_{\mathrm{reg}}, a)$, where M_{reg} and I_s^{reg} denote the anomaly mask and synthetic image, respectively. This jointly constrains anomaly location and local structural compatibility. Since a compatible placement may still induce weak or diffuse changes, TPFreeDAS further compares I_s^{reg} with I in the frozen DINO feature space using the patch-wise change $d_i = \|\Phi^*(I_s^{\mathrm{reg}})_i - \Phi^*(I)_i\|_2^2$. The region-constrained mask is then mapped to the DINO patch grid for mask-based spatial partition. It separates the feature-change map into inside and outside patch sets and aggregates their average responses:

$$\begin{aligned} \Omega_{\mathrm{in}} &= \left\{i \mid M_{\mathrm{reg}}^{p}(i) = 1\right\}, \qquad \Omega_{\mathrm{out}} = \left\{i \mid M_{\mathrm{reg}}^{p}(i) = 0\right\}, \\ D_{\mathrm{in}} &= \frac{1}{|\Omega_{\mathrm{in}}|}\sum_{i\in\Omega_{\mathrm{in}}} d_i, \quad D_{\mathrm{out}} = \frac{1}{|\Omega_{\mathrm{out}}|}\sum_{i\in\Omega_{\mathrm{out}}} d_i, \end{aligned} \tag{9}$$

where D_{in} measures the feature change inside the intended anomaly region and D_{out} measures outside feature leakage.

Post-synthesis quality validation favors a localized representation change: strong change inside the intended mask with limited outside leakage and an optional relative-leakage penalty:

$$\begin{aligned} Q &= D_{\mathrm{in}} - \lambda_{\mathrm{out}} D_{\mathrm{out}} - \lambda_{\mathrm{leak}}\frac{D_{\mathrm{out}}}{D_{\mathrm{in}} + \epsilon}, \\ g_{\mathrm{acc}} &= \mathrm{Gate}\left(D_{\mathrm{in}}, D_{\mathrm{out}}, Q\right) \in \{0, 1\}, \end{aligned} \tag{10}$$

where Q is the quality score and g_{acc} is the quality acceptance indicator. The weights λ_{out} and λ_{leak} control the absolute outside penalty and optional relative-leakage penalty, respectively. A candidate

Table 1: Cross-framework comparison in VisA between CutPaste and TPFreeDAS. The best results are highlighted in blue.

Framework	Synthesis	Setting	I-AUROC↑	AUPR↑	P-AUROC↑	PRO↑
OneNIP ECCV′24	CutPaste	1-shot	92.80	–	98.70	42.50
	TPFreeDAS	1-shot	93.36	95.01	98.80	43.11
SuperSimpleNet ICPR′24	CutPaste	2-shot	93.40	–	–	87.40
	TPFreeDAS	2-shot	95.20	–	–	92.20
FoundAD ICLR′26	CutPaste	1-shot	92.60	92.00	99.70	98.00
	TPFreeDAS	1-shot	93.40	93.18	99.71	98.06
	CutPaste	2-shot	93.80	93.30	99.70	98.20
	TPFreeDAS	2-shot	95.39	95.00	99.77	98.45
	CutPaste	4-shot	94.40	94.00	99.70	98.40
	TPFreeDAS	4-shot	96.06	95.66	99.78	98.37

is retained when $g_{\text{acc}} = 1$ and rejected otherwise. Let $\mathcal{C}_{\text{str}}$ denote the predefined structured-category set, which is fixed across all shot settings. The final synthetic image is selected as

$$I_s = \begin{cases} I_s^{\text{reg}}, & c \in \mathcal{C}_{\text{str}} \text{ and } g_{\text{acc}} = 1, \\ I_s^{\text{base}}, & c \in \mathcal{C}_{\text{str}} \text{ and } g_{\text{acc}} = 0, \\ I_s^{\text{gen}}, & c \notin \mathcal{C}_{\text{str}}. \end{cases} \tag{11}$$

Thus, an accepted region-aware image becomes the final synthetic context for a structured category, while a rejected candidate is replaced by the conservative CutPaste-based result.

3.5 Training Objective

TPFreeDAS only constructs the synthetic context I_s, while training follows FoundAD (Zhai et al., 2026). The frozen encoder extracts the normal target H from I, and the predictor estimates it from I_s:

$$H = \Phi^*(I), \qquad \widehat{H} = \Psi(\Phi^*(I_s)),$$
$$\mathcal{L}_{\text{pred}} = \frac{1}{BN_pD} \sum_{b=1}^{B} \sum_{i=1}^{N_p} \sum_{d=1}^{D} \left(H_{b,i,d} - \widehat{H}_{b,i,d} \right)^2. \tag{12}$$

Only Ψ is optimized; Φ^* remains frozen and TPFreeDAS adds no learnable synthesis parameters. During inference, TPFreeDAS is removed and the original FoundAD pipeline computes image- and pixel-level scores from feature residuals.

4 Experiments

4.1 Datasets and Implementation

We evaluate TPFreeDAS on MVTec AD (Bergmann et al., 2019), VisA (Zou et al., 2022), and PCB-Market (Huang et al., 2020), which contain 15, 12, and six categories, respectively. The PCBMarket categories are missing hole, mouse bite, open circuit, short, spur, and spurious copper. For each category, we construct 1-shot, 2-shot, and 4-shot training sets by sampling only normal images. No real anomaly images or annotations are used for training, while the original test images and pixel-level masks are retained for evaluation. All images are resized to 512×512 and encoded by a frozen DINOv3 ViT-B/16, yielding a 32×32 grid of 768-dimensional patch features. The predictor contains six Transformer blocks with a hidden dimension of 384 and uses a dropout rate of 0.2. We train only the predictor for 2000 epochs using AdamW with a learning rate of 1×10^{-3}, weight decay of 1×10^{-4}, and the MSE feature-prediction loss. The batch size is 8 on a single NVIDIA RTX 3090 GPU, except for 1-shot PCBMarket, where it is set to 1. TPFreeDAS is removed during inference. We report image-level AUROC (I-AUROC) and AUPR for anomaly detection, and pixel-level AUROC (P-AUROC) and per-region overlap (PRO) for localization. Following common practice (Bergmann et al., 2019), PRO is integrated over false-positive rates from 0 to 0.3. Higher values indicate better performance for all metrics.

Table 2: Comparison with few-shot anomaly detection methods on VisA and MVTec AD. FoundAD and TPFreeDAS use the generalized multi-class-one-model setting. Best and second-best results are marked in blue and green.

Shot	Method	w/o Text	VisA				MVTec AD				Reference
			I-AUROC	AUPR	P-AUROC	PRO	I-AUROC	AUPR	P-AUROC	PRO	
1	PatchCore (Roth et al., 2022)	✓	79.90	82.80	95.40	80.50	83.40	92.20	92.00	79.70	CVPR′22
	WinCLIP (Jeong et al., 2023)	✗	83.80	85.10	96.40	85.10	93.10	96.50	95.20	87.10	CVPR′23
	InCTRL (Zhu & Pang, 2024)	✗	83.20	84.10	89.00	66.70	91.30	95.20	94.60	87.80	CVPR′24
	AnomalyCLIP (Zhou et al., 2024)	✗	87.70	87.70	83.20	90.10	95.20	97.20	94.60	87.60	ICLR′24
	PromptAD (Li et al., 2024)	✗	86.90	88.40	96.70	85.10	94.60	97.10	95.90	87.90	CVPR′24
	LogSAD (Zhang et al., 2025)	✗	89.80	90.30	97.50	88.20	95.50	97.30	97.00	92.50	CVPR′25
	FoundAD (Zhai et al., 2026)	✓	92.60	92.00	99.70	98.00	96.10	97.90	96.80	92.80	ICLR′26
	TPFreeDAS	✓	93.40	93.18	99.71	98.06	96.32	97.98	96.87	92.91	
2	PatchCore (Roth et al., 2022)	✓	81.60	84.80	96.10	82.60	86.30	93.80	93.30	82.30	CVPR′22
	WinCLIP (Jeong et al., 2023)	✗	84.60	85.80	96.80	86.20	94.40	97.00	96.00	88.40	CVPR′23
	InCTRL (Zhu & Pang, 2024)	✗	86.30	86.80	89.80	68.10	91.80	95.50	95.20	88.30	CVPR′24
	AnomalyCLIP (Zhou et al., 2024)	✗	87.80	89.10	84.50	90.80	95.40	97.30	94.90	87.80	ICLR′24
	PromptAD (Li et al., 2024)	✗	88.30	90.00	97.10	85.80	95.70	97.90	96.20	88.50	CVPR′24
	LogSAD (Zhang et al., 2025)	✗	91.80	92.00	97.80	89.70	96.30	97.60	97.30	93.10	CVPR′25
	FoundAD (Zhai et al., 2026)	✓	93.80	93.30	99.70	98.20	96.90	98.30	97.00	93.20	ICLR′26
	TPFreeDAS	✓	95.39	95.00	99.77	98.45	96.96	98.31	96.92	93.37	
4	PatchCore (Roth et al., 2022)	✓	85.30	87.50	96.80	84.90	88.80	94.50	94.30	84.30	CVPR′22
	WinCLIP (Jeong et al., 2023)	✗	87.30	88.80	97.20	87.60	95.20	97.30	96.20	89.00	CVPR′23
	InCTRL (Zhu & Pang, 2024)	✗	87.80	88.00	90.20	68.80	93.10	96.30	95.80	89.50	CVPR′24
	AnomalyCLIP (Zhou et al., 2024)	✗	88.80	90.10	85.20	91.40	96.10	97.80	95.50	88.20	ICLR′24
	PromptAD (Li et al., 2024)	✗	89.10	90.80	97.40	86.20	96.60	98.50	96.50	90.50	CVPR′24
	LogSAD (Zhang et al., 2025)	✗	93.20	93.40	98.10	90.50	96.60	97.70	97.50	93.50	CVPR′25
	FoundAD (Zhai et al., 2026)	✓	94.40	94.00	99.70	98.40	97.10	98.60	97.20	93.60	ICLR′26
	TPFreeDAS	✓	96.06	95.66	99.78	98.37	97.22	98.55	97.57	93.61	

Table 3: Comparison between FoundAD and TPFreeDAS on PCBMarket (Huang et al., 2020). The three rightmost columns report 4-shot synthesis efficiency; FoundAD uses CutPaste synthesis. Dashes indicate unmeasured settings. Best values are highlighted in blue.

Shot	Method	I-AUROC↑	AUPR↑	P-AUROC↑	PRO↑	Batch latency (ms)↓	Per-image latency (ms)↓	Throughput (images/s)↑
1	FoundAD	62.15	59.36	88.17	65.85	–	–	–
	TPFreeDAS	68.85	62.19	92.28	77.95	–	–	–
2	FoundAD	59.73	61.00	88.12	67.80	–	–	–
	TPFreeDAS	70.95	68.15	92.26	79.55	–	–	–
4	FoundAD	60.28	61.33	89.77	72.02	3061.60	382.70	2.613
	TPFreeDAS	75.17	75.12	92.82	80.82	1974.32	246.79	4.052

4.2 Experimental Results

For each few-shot setting, we average results over three normal support sets sampled with independent seeds. We evaluate TPFreeDAS through cross-framework replacement, comparisons with few-shot baselines, and module and hyperparameter ablations.

Cross-Framework Synthesis Comparison. Table 1 tests whether TPFreeDAS gains extend beyond FoundAD by replacing CutPaste in OneNIP (Gao, 2024), SuperSimpleNet (Rolih et al., 2024), and FoundAD (Zhai et al., 2026) across their listed VisA shot settings. It improves 16 of 17 comparable metrics. Gains reach 1.80 points in I-AUROC and 4.80 in PRO on SuperSimpleNet, and up to 1.66 in I-AUROC and 1.70 in AUPR on FoundAD. The only decrease is 0.03 points in FoundAD 4-shot PRO, suggesting the benefit is not confined to FoundAD.

Quantitative Comparison with Few-Shot Baselines. Table 2 compares TPFreeDAS with representative few-shot methods under 1-shot, 2-shot, and 4-shot settings. TPFreeDAS achieves the best result on 20 of the 24 dataset–shot–metric combinations under the reported protocols, using a multi-class-one-model setting without textual prompts. On VisA, TPFreeDAS obtains the best I-AUROC, AUPR, and P-AUROC under all three shot settings. In the 1-shot setting, it surpasses the strong

Table 4: Module omission ablation study.

Configuration	I-AUROC↑	AUPR↑	P-AUROC↑	PRO↑
w/o FCAS	93.610	92.980	99.720	98.250
w/o SASR	94.890	94.550	99.710	98.410
w/o RAQG	94.440	93.940	99.710	98.270
Full TPFreeDAS	**95.391**	**95.003**	**99.770**	**98.454**

Table 5: Ablation study of the response-fusion weights in FCAS.

λ_r	λ_c	I-AUROC↑	AUPR↑	P-AUROC↑	PRO↑
1.00	0.00	94.118	93.479	99.744	98.282
0.90	0.10	94.407	93.861	99.735	98.284
0.75	0.25	94.176	93.356	99.729	98.200
0.50	0.50	94.782	94.721	99.728	98.201
0.85	**0.15**	**95.391**	**95.003**	**99.770**	**98.454**

Table 6: Ablation study of the structure-descriptor weights in SASR.

w_{fg}	w_{div}	w_{edge}	I-AUROC↑	AUPR↑	P-AUROC↑	PRO↑
0.35	0.35	0.30	94.617	93.728	99.747	98.305
0.25	0.25	0.50	93.904	92.952	99.732	98.301
0.20	0.25	0.50	94.104	93.046	99.735	98.311
0.40	0.30	0.30	94.747	94.239	99.741	98.272
0.30	**0.30**	**0.40**	**95.391**	**95.003**	**99.770**	**98.454**

Table 7: Ablation study of the quality-gating weights in RAQG.

λ_{out}	λ_{leak}	I-AUROC↑	AUPR↑	P-AUROC↑	PRO↑
0	0	95.339	94.818	99.756	98.348
1	0	95.000	94.665	99.759	**98.473**
1	500	95.230	94.776	99.756	98.424
2	**0**	**95.391**	**95.003**	**99.770**	98.454

LogSAD baseline (Zhang et al., 2025) by 2.21 points in P-AUROC and 9.86 points in PRO. The corresponding margins remain 1.68 and 7.87 points under the 4-shot setting, supporting the gains in pixel-level localization.

On MVTec AD, TPFreeDAS achieves the best I-AUROC and PRO under all three shot settings and reaches the best P-AUROC in the 4-shot setting. It also remains competitive with prompt-based methods such as PromptAD (Li et al., 2024) and LogSAD (Zhang et al., 2025) in the remaining metrics. Compared with FoundAD (Zhai et al., 2026), most MVTec AD gains are modest and two metrics decline slightly; larger gains occur on VisA and PCBMarket.

PCBMarket Comparison. On PCBMarket (Huang et al., 2020), whose examples include repeated circuits and components, TPFreeDAS exceeds FoundAD (Zhai et al., 2026) across all reported metrics and shots (Table 3), with gains up to 14.89, 13.79, and 12.10 points in I-AUROC, AUPR, and PRO, respectively. This pattern is consistent with local compatibility control. For 4-shot synthesis, it achieves $1.551\times$ our CutPaste throughput and 35.51% lower per-image latency.

Component Ablation. The omission pattern in Table 4 is consistent with the three controls: removing FCAS gives the largest I-AUROC and AUPR drops (1.781 and 2.023 points), supporting spatial placement; removing RAQG lowers them by 0.951 and 1.063 points, consistent with local refinement and quality validation. SASR gives smaller but consistent gains, supporting its role as a safeguard against unreliable foreground constraints.

Hyperparameter Analysis. Tables 5–7 favor DINO-dominant FCAS fusion $(\lambda_r, \lambda_c) = (0.85, 0.15)$ over omitting the center prior or assigning equal weights. Among tested SASR mixtures, $(w_{\text{fg}}, w_{\text{div}}, w_{\text{edge}}) = (0.30, 0.30, 0.40)$ leads all four metrics with modest boundary emphasis. For RAQG, $(\lambda_{\text{out}}, \lambda_{\text{leak}}) = (2, 0)$ leads three metrics and trails the highest PRO by 0.019 points.

Qualitative Comparison. Figure 3 compares the anomaly localization heatmaps produced by FoundAD (Zhai et al., 2026) and TPFreeDAS on representative samples from Candle, Cashew, Macaroni, and PCB. FoundAD occasionally produces diffuse responses over normal structures or background regions. In contrast, the detector trained with TPFreeDAS yields more concentrated responses within the ground-truth anomaly regions and suppresses unrelated activations more effectively. These maps are consistent with the intended effect of structurally constrained synthesis.

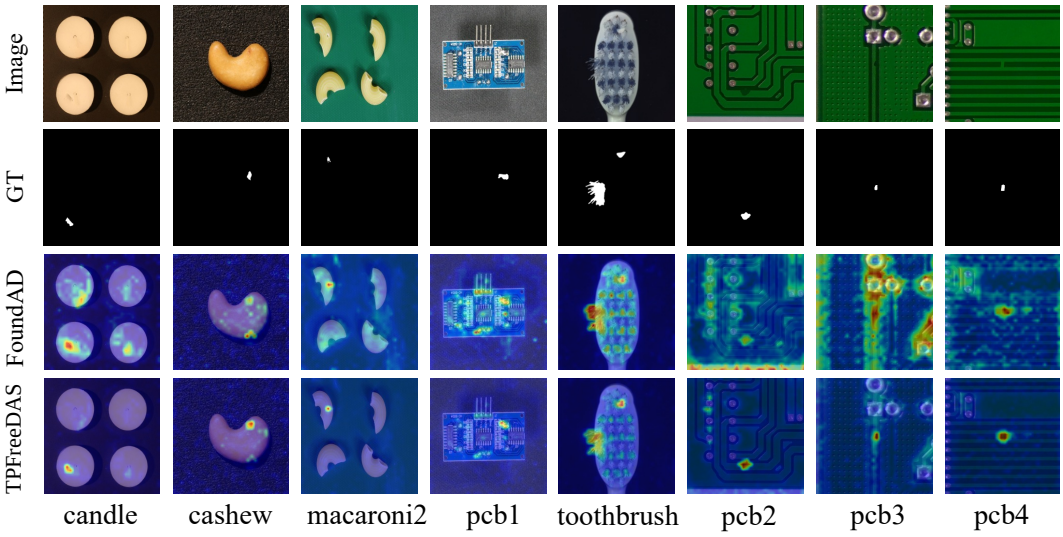

Figure 3: Anomaly localization heatmaps of FoundAD (Zhai et al., 2026) and TPFreeDAS. In these examples, TPFreeDAS concentrates responses on anomaly regions while suppressing normal-structure and background activations.

5 Conclusion

TPFreeDAS treats synthetic-anomaly quality as a compatibility problem involving spatial support, image structure, and local organization. It reuses frozen DINO features already available to DINO-based detectors to regulate these conditions through FCAS, SASR, and RAQG during training, without a synthesis-specific backbone, learnable synthesis parameters, auxiliary losses, or inference-time computation. Results on VisA, MVTec AD, and PCBMarket, including synthesis replacement in three detector frameworks, suggest that this compatibility view can improve few-shot detection and localization, with the largest gains on PCBMarket.

AI use statement

We used generative AI tools for language editing, manuscript consistency checks, and assistance in interpreting reported results. The authors take responsibility for the final content.

References

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*, pp. 9592–9600, 2019.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9650–9660, 2021.

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

Zheng Fang, Xiaoyang Wang, Haocheng Li, Jiejie Liu, Qiugui Hu, and Jimin Xiao. Fastrecon: Few-shot industrial anomaly detection via fast feature reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 17481–17490, 2023.

Bin-Bin Gao. Learning to detect multi-class anomalies with just one normal image prompt. In *18th European Conference on Computer Vision (ECCV 2024)*, pp. –, 2024.

Weibo Huang, Peng Wei, Manhua Zhang, and Hong Liu. Hripcb: a challenging dataset for pcb defects detection and classification. *The Journal of Engineering*, 2020(13):303–309, 2020.

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

Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9664–9674, 2021.

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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16838–16848, 2024.

Wenxi Lv, Qinliang Su, and Wenchao Xu. One-for-all few-shot anomaly detection via instance-induced prompt learning. In *The Thirteenth International Conference on Learning Representations*, 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, 2021.

Blaž Rolih, Matic Fučka, Skočaj, et al. Supersimplenet: Unifying unsupervised and supervised learning for fast and reliable surface defect detection. In *International Conference on Pattern Recognition*, pp. 47–65. Springer, 2024.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10684–10695, 2022.

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*, pp. 14318–14328, 2022.

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.

Zhenyu Yan, Qingqing Fang, Wenxi Lv, and Qinliang Su. Anomalysd: One-for-all few-shot anomaly detection via pre-trained diffusion models. *Computer Vision and Image Understanding*, 265: 104668, 2026.

Guangyao Zhai, Yue Zhou, Xinyan Deng, Lars Heckler, Nassir Navab, and Benjamin Busam. Foundation visual encoders are secretly few-shot anomaly detectors. In *International Conference on Learning Representations*, 2026.

Jinjin Zhang, Guodong Wang, Yizhou Jin, and Di Huang. Towards training-free anomaly detection with vision and language foundation models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15204–15213, 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*, 2024.

Jiawen Zhu and Guansong Pang. Toward generalist anomaly detection via in-context residual learning with few-shot sample prompts. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 17826–17836, 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.