BEYOND ONE-STEP RESIDUALS: TRAJECTORY-AWARE MANIFOLD PROJECTION FOR FEW-SHOT MEDICAL ANOMALY DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Visual foundation models provide structured normal feature spaces for manifold-based anomaly detection. However, in few-shot medical settings, inter-individual variation induces train-test distribution shifts, leaving limited normal samples insufficient to characterize the normal manifold. Inspired by a dynamical systems perspective, we propose Trajectory-Aware Manifold Projection (TAMP), which exploits trajectory information beyond initial projection residuals. TAMP learns normal-manifold projectors through consistency-guided multi-level learning and repeatedly applies them during inference to construct recovery trajectories. We observe that initial residuals capture deviations, while subsequent projection steps reveal different responses to normal variations and lesions. Using consecutive projection displacements as additional anomaly cues, we introduce the post-projection trajectory ratio and terminal convergence residual to measure relative continued changes after the initial projection and the remaining final-step update, respectively. These complement the initial residual for anomaly scoring. Experiments on multiple medical benchmarks demonstrate improved detection and localization over existing few-shot methods, with competitive results against full-shot unsupervised methods using only a few normal references.

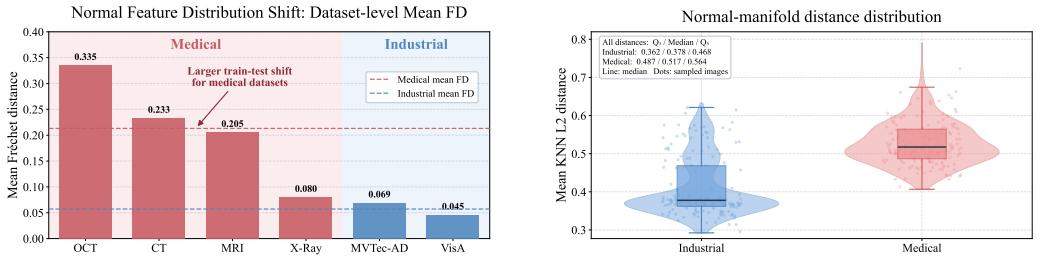
1 INTRODUCTION

Deep learning has advanced medical image analysis tasks such as disease recognition and lesion segmentation (Litjens et al., 2017). However, large annotated datasets are often unavailable for rare diseases, subtle lesions, and heterogeneous pathologies, motivating medical anomaly detection from normal data (Chalapathy & Chawla, 2019; Xiang et al., 2023; Zhou et al., 2022) or only a few normal references (Dong et al., 2026). Visual foundation encoders (Oquab et al., 2023; Siméoni et al., 2025; Liu et al., 2024) provide structured feature spaces for modeling normality. Recent manifold-based methods learn normal feature projectors or estimate normal subspaces from these representations (Zhai et al., 2026; Lendering et al., 2026). Their projection residuals measure the discrepancy between input features and projected representations, providing direct anomaly cues.

However, train-test distribution shifts in medical images can leave the learned normal manifold inadequately covered by a few normal training samples. Figure 1 shows that the evaluated medical datasets exhibit larger normal feature shifts and greater distances to few-shot normal manifold than the industrial benchmarks. Consequently, uncovered normal variations may also produce large one-step residuals, making residual magnitude alone ambiguous for anomaly discrimination. We therefore ask: **How can manifold projection provide anomaly evidence beyond residual magnitude under few-shot learning?**

Recurrent models can be formulated as discrete-time nonlinear dynamical systems, whose states evolve through repeated application of a learned mapping (Miller & Hardt, 2019). This formulation offers a dynamical perspective on iterative feature transformations. Furthermore, stable manifold theory characterizes the local convergence behavior of trajectories under repeated mappings near stable invariant structures (Wiggins, 2003). Inspired by this perspective, we repeatedly apply the learned projector to the input feature, and examine the resulting recovery trajectories for anomaly cues beyond the one-step residual.

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.



(a) Train-test normal feature distribution shift

(b) Distance to few-shot normal manifold

Figure 1: Normal-reference coverage in the evaluated few-shot settings. (a) Dataset-level Fréchet distance between limited training normal samples and test normal features. (b) Distribution of distances between normal test features and few-shot normal reference structures. Both measurements indicate larger coverage gaps in the evaluated medical datasets.

We visualize feature displacements between consecutive projection states along the trajectory (Fig. 2). Both the brain MRI and OCT examples exhibit ambiguous initial displacement responses. This is consistent with limited normal-manifold coverage under few-shot learning. Along the projection trajectory, ambiguous responses in normal regions progressively fade, whereas lesion regions exhibit persistent feature updates. These observations suggest that stepwise changes can complement the initial residual when normal variation and pathology produce ambiguous responses.

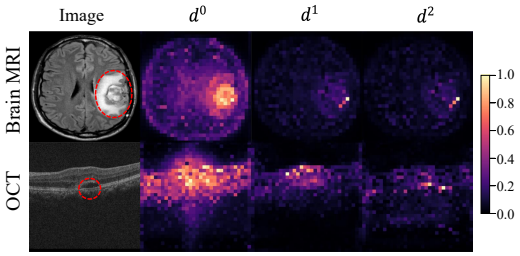


Figure 2: Feature displacements between consecutive projection states for brain MRI and OCT. Dashed circles mark lesions.

Based on this insight, we propose **Trajectory-Aware Manifold Projection (TAMP)** for few-shot medical anomaly detection. During training, consistency-guided multi-level learning obtains normal feature projectors from a frozen visual encoder, building on the projection learning paradigm of FoundAD (Zhai et al., 2026). During inference, TAMP repeatedly applies the learned projectors to construct recovery trajectories. It complements the initial residual with two trajectory measurements: the *post-projection trajectory ratio*, which measures accumulated subsequent displacement relative to the initial residual, and the *terminal convergence residual*, which measures the remaining update at the final step. Together, these cues enable trajectory-aware anomaly detection and localization with only a few normal references.

Our contributions are summarized as follows:

- We investigate the ambiguity of one-step projection residuals under limited normal coverage in few-shot medical anomaly detection and explore recovery trajectories as complementary anomaly evidence.
- We propose TAMP, which learns normal-manifold projectors through consistency-guided multi-level learning and reuses them iteratively at inference. We introduce the post-projection trajectory ratio and terminal convergence residual to capture feature evolution during multi-step projection.
- Experiments on four medical benchmarks demonstrate improved few-shot detection and localization, with competitive performance against full-shot unsupervised methods. Ablations support the contribution of trajectory information beyond the initial residual.

2 Related Work

Medical anomaly detection. Medical anomaly detection has been extensively studied through reconstruction-based normality modeling. Early approaches reconstruct images at the pixel level

using generative adversarial networks (GANs) (Schlegl et al., 2017; 2019) or autoencoders (Baur et al., 2018; 2021; Kascenas et al., 2022; Wyatt et al., 2022; Baur et al., 2020). More recent methods increasingly operate in feature space, including encoder–decoder reconstruction (Zhao et al., 2021; Zhou et al., 2022; Xiang et al., 2023; Tang et al., 2025; Xu et al., 2025; Guo et al., 2025), feature distillation (Deng & Li, 2022; Wang et al., 2021; Zhang et al., 2023), and contrastive reconstruction (Guo et al., 2024; 2023). However, their effectiveness generally relies on the sufficiency of normal training data. In few-shot medical settings, substantial anatomical and acquisition-related variability may remain underrepresented, making discrepancy-based anomaly cues ambiguous and motivating the use of additional evidence beyond a single reconstruction response.

Few-shot anomaly detection. Few-shot methods characterize normality from only a few references. Early approaches use patch memories or feature distributions (Defard et al., 2021; Roth et al., 2022), while recent methods increasingly exploit pretrained CLIP (Radford et al., 2021) and DINO representations (Caron et al., 2021; Oquab et al., 2023; Siméoni et al., 2025). CLIP-based methods align visual features with textual normality or abnormality cues through prompt ensembles or learnable prompts, as in WinCLIP (Jeong et al., 2023), AnomalyCLIP (Zhou et al., 2024), PromptAD (Li et al., 2024), SimCLIP (Deng et al., 2024), and IIPAD (Lv et al., 2025). These methods often either rely on carefully designed text prompts or require considerable additional training overhead. Vision-only methods such as AnomalyDINO (Damm et al., 2025) and SuperAD (Zhang et al., 2025) directly exploit pretrained visual features and normal references. Despite strong low-data performance, these methods rely on normal reference samples and can be sensitive to normal anatomical variations in medical imaging.

Manifold-based anomaly detection. Manifold-based methods explicitly model normal features as a compact geometric structure. FoundAD (Zhai et al., 2026) learns a nonlinear normal-manifold projector, SubspaceAD (Lendering et al., 2026) constructs a low-dimensional normal subspace from pretrained features, and PDD (Lu et al., 2026) incorporates manifold priors through distillation. These studies demonstrate the effectiveness of modeling normality through feature manifolds. However, existing manifold-based approaches typically perform a single projection and use the resulting discrepancy as the primary anomaly cue. Under limited normal coverage, such one-step discrepancies can be ambiguous. In contrast, we formulate projection as an iterative manifold recovery process and exploit multi-step projection trajectories.

3 METHODOLOGY

3.1 PROBLEM DEFINITION AND ONE-STEP PROJECTION

We consider few-shot medical anomaly detection with a small normal training set $\mathcal{D}_{\text{tr}} = \{x_i^n\}_{i=1}^K$, where K is typically set to 1, 2, or 4. Training uses only normal images, without abnormal samples or pixel-level annotations. Given a test image $x \in \mathbb{R}^{H\times W\times C}$, the goal is to predict a pixel-level anomaly map $M(x) \in \mathbb{R}^{H\times W}$ and an image-level anomaly score $s(x) \in \mathbb{R}$, where both outputs are normalized to $[0, 1]$.

Let $\theta(\cdot)$ denote a frozen visual encoder and $f = \theta(x) \in \mathbb{R}^{N\times D}$ the extracted feature, where N denotes the number of spatial tokens and D the feature dimension. Projection-based methods learn a projector $\phi : \mathbb{R}^{N\times D} \to \mathbb{R}^{N\times D}$ to map f toward the normal manifold, yielding $z = \phi(f) \in \mathbb{R}^{N\times D}$, and detect anomalies using the one-step residual $d = \|z - f\|_2^2$. In few-shot medical settings, limited normal references may not fully cover the normal variations encountered at test time, making the initial residual less reliable under limited normal coverage.

3.2 OVERVIEW OF TAMP

Inspired by the dynamical systems perspective on repeated mappings (Miller & Hardt, 2019) and the local convergence behavior characterized by stable manifold theory (Wiggins, 2003), we extend the conventional one-step manifold projection into a multi-step process by repeatedly applying the learned projector and propose **Trajectory-Aware Manifold Projection (TAMP)**, as illustrated in Fig. 3. Rather than relying solely on the initial projection residual, TAMP further exploits the feature

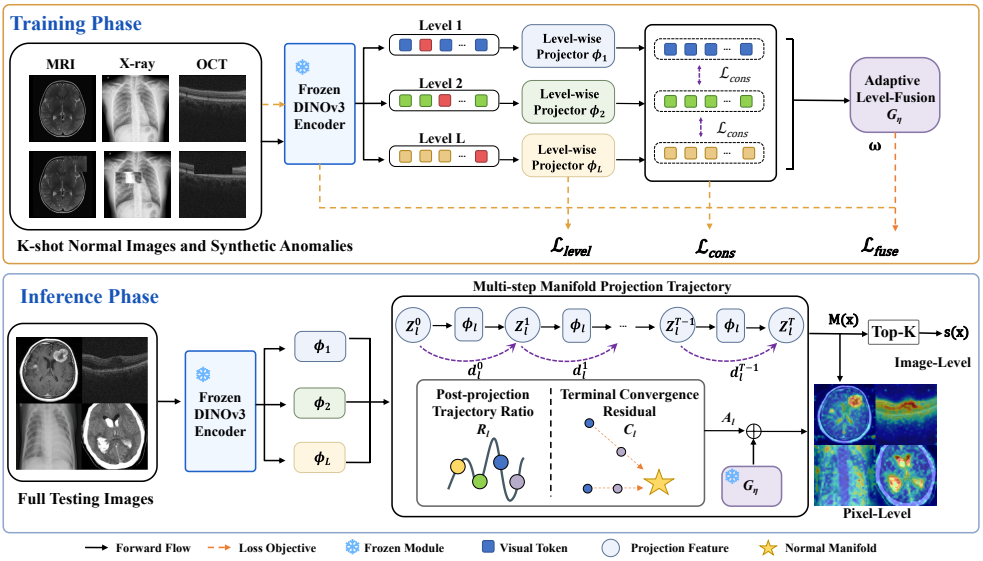


Figure 3: Overall framework of TAMP. Multi-level normal-manifold projectors are learned from frozen foundation features using reconstruction and consistency objectives. At inference, the learned projectors are repeatedly applied as manifold projection operators to construct multi-step trajectories. The initial residual, post-projection trajectory ratio R_l, and terminal convergence residual C_l are jointly used for trajectory-aware anomaly scoring.

changes revealed along the resulting projection trajectory. The following sections describe how the projectors are learned and how the multi-step trajectory is used for anomaly scoring.

3.3 Consistency-Guided Multi-level Manifold Projection Learning

Following FoundAD (Zhai et al., 2026), we train each projector to restore synthetically perturbed features toward their corresponding normal features. Since medical abnormalities and normal variations can appear at different semantic levels, we extend this projection learning paradigm to multi-level feature spaces and encourage consistent normal mappings across levels. Let $\theta_l(\cdot)$ denote the l-th selected layer of the visual encoder. For a normal image x_i^n and its synthetically perturbed counterpart x_i^s, we extract $f_{i,l}^n = \theta_l(x_i^n) \in \mathbb{R}^{N\times D}$ and $f_{i,l}^s = \theta_l(x_i^s) \in \mathbb{R}^{N\times D}$. Each level employs an independent projector $z_{i,l} = \phi_l(f_{i,l}^s) \in \mathbb{R}^{N\times D}$. The learned projectors approximate level-specific mappings toward the normal manifold and are subsequently iteratively applied to construct multi-step projection trajectories during inference.

3.3.1 Level-wise Normal-Manifold Reconstruction

To enable each level-wise projector to independently capture the normal manifold structure of its corresponding feature space, we introduce a level-wise reconstruction objective. Specifically,

$$\mathcal{L}_{\text{level}} = \frac{1}{KLN} \sum_{i=1}^{K} \sum_{l=1}^{L} \sum_{j=1}^{N} \left\| z_{i,l,j} - f_{i,l,j}^n \right\|_2^2, \tag{1}$$

where j denotes the patch index. This objective enables each projector to learn a level-specific feature recovery mapping, providing the foundation for subsequent multi-step manifold projection.

3.3.2 Adaptive Fused Reconstruction Loss

Different feature levels capture complementary characteristics of medical images, and their reliability may vary across modalities, anatomical regions, and abnormality types. Therefore, we introduce

an adaptive fusion module to learn sample-specific level weights. The fusion module G_η predicts $\mathbf{w}_i = (w_{i,1}, \ldots, w_{i,L})$, where $\sum_{l=1}^{L} w_{i,l} = 1$, and the fused reconstruction objective is defined as

$$\mathcal{L}_{\text{fuse}} = \frac{1}{KN} \sum_{i=1}^{K} \sum_{j=1}^{N} \sum_{l=1}^{L} w_{i,l} \left\| z_{i,l,j} - f_{i,l,j}^{n} \right\|_2^2. \tag{2}$$

Details of the adaptive fusion module and weight estimation procedure are provided in Appendix A.

3.3.3 Cross-level Projection Consistency

Although different feature levels capture complementary semantic information, they correspond to the same underlying anatomical structures. Independent manifold projection at each level may therefore result in inconsistent normal recovery across feature spaces. To encourage coherent manifold recovery among different levels, we introduce a cross-level consistency loss that aligns adjacent projected features using cosine similarity:

$$\mathcal{L}_{\text{cons}} = \frac{1}{K(L-1)N} \sum_{i=1}^{K} \sum_{l=1}^{L-1} \sum_{j=1}^{N} \left[1 - \cos\left(z_{i,l,j}, z_{i,l+1,j} \right) \right]. \tag{3}$$

The overall training objective is

$$\mathcal{L}_{\text{train}} = \lambda_{\text{level}} \mathcal{L}_{\text{level}} + \lambda_{\text{fuse}} \mathcal{L}_{\text{fuse}} + \lambda_{\text{cons}} \mathcal{L}_{\text{cons}}, \tag{4}$$

where λ_{level}, λ_{fuse}, and λ_{cons} balance the three loss terms. These training objectives are applied to single-step projection outputs, without unrolling the projector or imposing constraints on multi-step trajectories. During inference, we treat each learned projector as a normal-manifold projection operator and repeatedly apply it to construct a feature recovery trajectory.

3.4 Multi-step Projection Trajectory Scoring

During inference, we treat each learned level-wise projector as a reusable normal-manifold projection operator. Given a test image x, the frozen encoder extracts the l-th level feature $f_l = \theta_l(x) \in \mathbb{R}^{N \times D}$. We initialize the projection state as $z_l^0 = f_l$, and repeatedly apply the same projector:

$$z_l^{t+1} = \phi_l(z_l^t), \qquad t = 0, \ldots, T-1, \tag{5}$$

where T denotes the number of projection steps. For the j-th spatial token, the displacement between two consecutive projection states is defined as

$$d_{l,j}^t = \left\| z_{l,j}^{t+1} - z_{l,j}^t \right\|_2^2. \tag{6}$$

The resulting sequence $\{d_{l,j}^0, d_{l,j}^1, \ldots, d_{l,j}^{T-1}\}$ describes the step-wise changes along the projection trajectory. In particular, $d_{l,j}^0$ corresponds to the conventional one-step projection residual and provides the most direct evidence of the initial feature deviation. Figure 4 visualizes the multi-step projection states and their step-wise projection displacement maps.

To characterize the feature changes remaining after the initial projection, we define the *post-projection trajectory ratio* as $R_{l,j} = \frac{\sum_{t=1}^{T-1} d_{l,j}^t}{d_{l,j}^0 + \epsilon}$, where ϵ is a small constant for numerical stability. $R_{l,j}$ accumulates the displacements after the initial projection and normalizes them by the initial residual, quantifying the relative extent of continued feature changes. This captures the sustained displacement responses observed in lesion regions, providing extra anomaly evidence. We define the *terminal convergence residual* as $C_{l,j} = d_{l,j}^{T-1}$, which measures the remaining feature update at the end of the trajectory. It complements $R_{l,j}$ by capturing the terminal update activity observed in anomalous regions. The trajectory-aware response is computed as

$$A_{l,j} = R_{l,j} + \gamma C_{l,j}, \tag{7}$$

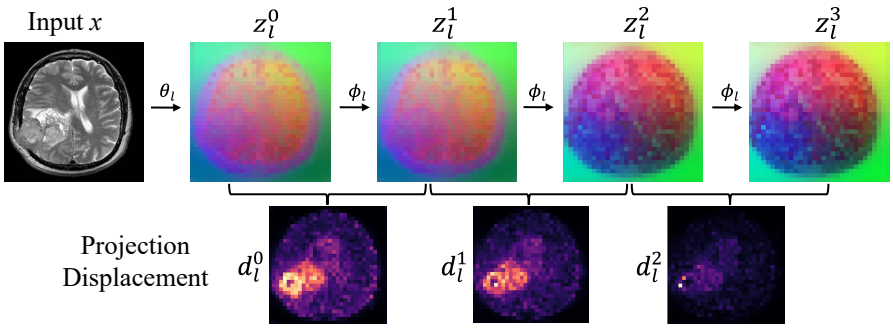


Figure 4: Illustration of multi-step projection and step-wise projection displacement at layer l. The top row shows the input image and projection states from z_l^0 to z_l^3. The bottom row shows the consecutive displacement maps. Lesion regions exhibit persistent displacement responses.

where γ controls the contribution of the terminal response. The final patch-level anomaly response combines the initial projection discrepancy and multi-step trajectory information:

$$M_j(x) = \sum_{l=1}^{L} w_l d_{l,j}^0 + \beta \frac{1}{\rho} \sum_{l \in \Omega_\rho(j)} A_{l,j}. \tag{8}$$

Here, w_l denotes the adaptive level weight predicted by the fusion module, β balances the one-step and multi-step responses, and $\Omega_\rho(j)$ selects the top-ρ levels with the strongest trajectory responses. The patch scores are rearranged into the spatial token grid and upsampled to the input resolution. The image-level anomaly score is computed from the top-κ patch responses:

$$s(x) = \frac{1}{\kappa} \sum_{j \in \Omega_\kappa(x)} M_j(x), \tag{9}$$

where $\Omega_\kappa(x)$ contains the indices of the κ highest-scoring patches.

4 Experiments

4.1 Experimental Setup

Dataset. We evaluate our method on four medical anomaly detection benchmarks: HeadCT (Kitamura, 2018), BrainMRI (Chakrabarty, 2019), and ZhangLab Chest X-ray (Kermany et al., 2018) for image-level anomaly detection, as well as RESC (Hu et al., 2019) for pixel-level anomaly segmentation. These benchmarks cover diverse medical imaging modalities, including CT, MRI, chest X-ray, and retinal OCT. For all datasets, we conduct experiments under 1-shot, 2-shot, and 4-shot settings. Only normal images are used for training. During testing, the full test set is used for evaluation.

Implementation details. We adopt the pretrained DINOv3 ViT-L/16 model (Siméoni et al., 2025) as the visual encoder and keep it frozen during both training and inference. All input images are resized to 512×512. We extract three levels of dense token features from the 20th, 21st, and 22nd Transformer blocks. Synthetic anomalies are generated using CutPaste (Li et al., 2021). The level-wise projectors follow the ViT architecture used in FoundAD (Zhai et al., 2026), and the main experiments use three projectors. All experiments are conducted on a single NVIDIA RTX 4090 GPU. Additional training configurations and hyperparameter settings are provided in Appendix B.

Baselines. We compare TAMP with representative few-shot methods, including PatchCore (Roth et al., 2022), WinCLIP (Jeong et al., 2023), IIPAD (Lv et al., 2025), AnomalyDINO (Damm et al., 2025), SubspaceAD (Lendering et al., 2026), and FoundAD (Zhai et al., 2026), as well as full-shot unsupervised baselines, including SALAD (Zhao et al., 2021), SQUID (Xiang et al., 2023), RD4AD (Deng & Li, 2022), EDC (Guo et al., 2024) and E2AD (Tang et al., 2025). All comparison methods

are implemented based on their publicly released code. Both AnomalyDINO and FoundAD use DINOv3 ViT-L/16 as the backbone for a fair comparison.

Evaluation metrics. For image-level anomaly detection, we report the image-level Area Under the Receiver Operating Characteristic Curve (I-AUROC) and Area Under the Precision-Recall Curve (I-AUPR). For pixel-level anomaly localization, we use pixel-level AUROC (P-AUROC) and Per-Region Overlap (PRO) as evaluation metrics. We report the mean performance over five runs using different random seeds.

Table 1: Comparison of anomaly detection performance on HeadCT, BrainMRI, and ZhangLab Chest X-ray. The best result is marked in bold and the second-best result is marked underlined.

Shot	Method	HeadCT		BrainMRI		ZhangLab		Average	
		I-AUROC	I-AUPR	I-AUROC	I-AUPR	I-AUROC	I-AUPR	I-AUROC	I-AUPR
1	PatchCore	53.7	69.4	64.5	85.2	58.0	68.3	58.7	74.3
	WinCLIP	88.8	**94.6**	85.0	95.5	69.2	79.3	81.0	89.8
	IIPAD	82.6	92.8	**94.1**	97.3	79.4	88.9	85.4	93.0
	AnomalyDINO	89.5	82.9	74.8	78.6	76.9	84.1	80.4	81.9
	SubspaceAD	62.9	78.2	79.4	92.3	80.8	86.9	74.4	85.8
	FoundAD	67.6	81.7	89.4	96.9	82.0	88.0	79.7	88.8
	TAMP (Ours)	**91.8**	93.3	93.0	**98.1**	**87.1**	**91.9**	**90.6**	**94.4**
2	PatchCore	55.3	70.0	75.5	88.7	64.2	76.3	65.0	78.3
	WinCLIP	90.1	**94.9**	92.0	93.5	70.9	79.6	84.3	89.3
	IIPAD	81.1	90.4	93.1	97.5	80.0	86.9	84.7	91.6
	AnomalyDINO	90.6	86.8	78.8	82.3	84.4	90.6	84.6	86.5
	SubspaceAD	61.6	75.3	84.7	94.5	83.8	88.6	76.7	86.1
	FoundAD	74.2	86.6	91.1	97.3	83.7	89.1	83.0	91.0
	TAMP (Ours)	**93.0**	93.5	**93.2**	**98.1**	**91.4**	**94.4**	**92.5**	**95.3**
4	PatchCore	60.4	74.3	73.8	89.1	55.9	65.7	63.4	76.4
	WinCLIP	90.6	**95.5**	92.6	97.9	82.5	89.2	88.6	94.2
	IIPAD	83.0	91.2	93.8	97.4	85.0	90.9	87.3	93.2
	AnomalyDINO	92.6	88.3	88.4	87.3	88.7	93.1	89.9	89.6
	SubspaceAD	62.7	76.5	84.9	94.5	84.8	89.8	77.5	86.9
	FoundAD	88.4	86.4	92.1	91.8	87.9	91.7	89.5	89.9
	TAMP (Ours)	**94.6**	89.6	**96.1**	**98.7**	**93.8**	**96.1**	**94.8**	**94.8**

4.2 Experimental Results

Overall Performance. As shown in Table 1, TAMP achieves the best average I-AUROC/I-AUPR under all shot settings, reaching 90.6%/94.4%, 92.5%/95.3%, and 94.8%/94.8% for 1-, 2-, and 4-shot detection, respectively. Compared with FoundAD (Zhai et al., 2026), TAMP improves the average I-AUROC by **10.9%**, **9.5%**, and **5.3%** under the three shot settings. More importantly, the advantage is also evident when compared with the strongest competing baseline, such as IIPAD (Lv et al., 2025) and AnomalyDINO (Damm et al., 2025). TAMP improves the average I-AUROC by **5.2%**, **7.8%**, and **4.9%** for 1-, 2-, and 4-shot detection, respectively. Table 2 further evaluates pixel-level anomaly localization on RESC. TAMP achieves the best P-AUROC/PRO under all shot settings, with 97.0%/88.4%, 96.6%/87.1%, and 97.1%/87.9% from 1-shot to 4-shot, respectively. Even against the strongest baseline in each setting, TAMP maintains P-AUROC gains of **1.2%**, **1.0%**, and **0.8%** and substantially larger PRO gains of **7.4%**, **5.8%**, and **3.6%**. These results demonstrate that TAMP is consistently effective across diverse anatomical structures and imaging modalities, with multi-step trajectory information providing complementary evidence beyond the initial projection residual.

Table 2: Comparison of anomaly localization performance on RESC. The best result is marked in bold and the second-best result is marked underlined.

Shot	Method	P-AUROC	PRO
1	IIPAD	94.9	81.0
	AnomalyDINO	92.2	76.1
	SubspaceAD	95.8	80.0
	FoundAD	93.2	76.4
	TAMP (Ours)	**97.0**	**88.4**
2	IIPAD	95.6	81.3
	AnomalyDINO	94.8	80.2
	SubspaceAD	95.2	74.7
	FoundAD	93.9	76.8
	TAMP (Ours)	**96.6**	**87.1**
4	IIPAD	96.0	81.9
	AnomalyDINO	96.3	84.3
	SubspaceAD	95.6	80.8
	FoundAD	94.6	79.9
	TAMP (Ours)	**97.1**	**87.9**

Comparison with Full-Shot Methods. Table 3 shows that TAMP remains competitive with full-shot unsupervised methods despite using only four normal samples. It achieves the best average I-AUROC/I-AUPR across HeadCT, BrainMRI, and ZhangLab. Compared with the second-best

results, TAMP improves the average I-AUROC/I-AUPR by **6.2%/6.3%**. More comparisons and further discussions of model performance are provided in Appendix D.

Table 3: Comparison of anomaly detection performance between full-shot unsupervised methods and 4-shot TAMP on HeadCT, BrainMRI, and ZhangLab Chest X-ray. The best result is marked in bold and the second-best result is marked underlined.

Method	Training	HeadCT		BrainMRI		ZhangLab		Average	
		I-AUROC	I-AUPR	I-AUROC	I-AUPR	I-AUROC	I-AUPR	I-AUROC	I-AUPR
SALAD	full-shot	81.1	80.5	67.4	77.4	78.0	84.7	75.5	80.8
RD4AD		62.0	75.5	86.7	<u>95.1</u>	88.1	92.3	78.9	87.6
SQUID		75.4	<u>87.6</u>	74.7	84.8	<u>90.0</u>	<u>93.3</u>	80.0	<u>88.5</u>
EDC		<u>89.9</u>	82.2	<u>92.8</u>	93.2	83.2	87.3	<u>88.6</u>	87.5
E2AD		75.8	78.9	89.4	95.0	86.4	89.6	83.9	87.8
TAMP (Ours)	**4-shot**	**94.6**	**89.6**	**96.1**	**98.7**	**93.8**	**96.1**	**94.8**	**94.8**

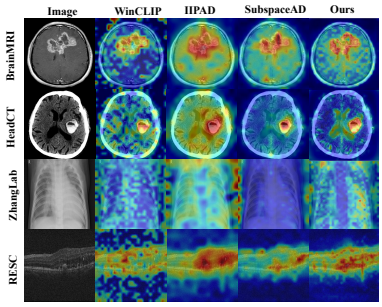


Figure 5: Qualitative anomaly heatmaps on representative abnormal samples.

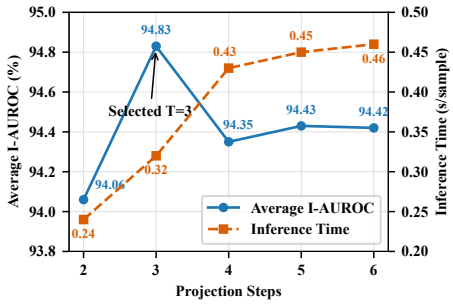


Figure 6: Effect of the number of projection steps on average I-AUROC and inference time under the 4-shot setting.

Qualitative Results. As shown in Fig. 5, existing methods often produce diffuse responses or activate high-contrast normal structures. In contrast, TAMP yields more concentrated and anatomically plausible anomaly maps by incorporating complementary multi-step trajectory information beyond the initial projection residual. More detailed visualizations and analyses of multi-level projection trajectories are provided in Appendix D.7.

4.3 Ablation Studies

Training Objectives. We ablate the level-wise reconstruction loss $\mathcal{L}_{\text{level}}$ and cross-level consistency loss $\mathcal{L}_{\text{cons}}$ while retaining $\mathcal{L}_{\text{fuse}}$. As shown in Table 4, removing $\mathcal{L}_{\text{level}}$ decreases the average I-AUROC/I-AUPR by **2.5%/3.0%**, while removing $\mathcal{L}_{\text{cons}}$ leads to decreases of **2.4%/2.9%**, respectively. The comparable degradations reflect their complementary roles: $\mathcal{L}_{\text{level}}$ provides direct supervision for learning level-specific normal recovery, whereas $\mathcal{L}_{\text{cons}}$ encourages coherent projections across feature levels. This confirms that level-specific manifold reconstruction and cross-level projection consistency are both important for learning multi-level normal-manifold projectors.

Table 4: Ablation study of training objectives on HeadCT, BrainMRI, ZhangLab, and their average performance under the 4-shot setting. The best result is marked in bold and the second-best result is marked underlined.

Training Objective			HeadCT		BrainMRI		ZhangLab		Average	
$\mathcal{L}_{\text{fuse}}$	$\mathcal{L}_{\text{level}}$	$\mathcal{L}_{\text{cons}}$	I-AUROC	I-AUPR	I-AUROC	I-AUPR	I-AUROC	I-AUPR	I-AUROC	I-AUPR
✓	✓	✓	**94.6**	**89.6**	**96.1**	**98.7**	**93.8**	**96.1**	**94.8**	**94.8**
✓		✓	<u>93.0</u>	<u>87.8</u>	<u>90.7</u>	91.6	93.3	<u>95.9</u>	92.3	91.8
✓	✓		92.9	87.7	<u>90.7</u>	<u>91.9</u>	<u>93.7</u>	**96.1**	<u>92.4</u>	<u>91.9</u>

Effect of Projection Steps. As shown in Fig. 6, the average I-AUROC peaks at 94.83% with $T = 3$, while inference time increases with additional steps. Performance slightly degrades for larger T, suggesting that a small number of projections is sufficient to capture useful trajectory information beyond the initial residual, whereas excessive iterations may attenuate discriminative feature variations. We therefore set $T = 3$ in all experiments. We provide additional analysis of the multi-step projection process in Appendix C.

Multi-level Fusion and Multi-step Projection. Table 5 evaluates I-AUROC for multi-level fusion and multi-step projection based on the FoundAD baseline (Zhai et al., 2026). MA, MF, and MP denote Multi-level Average, Multi-level Fusion, and Multi-step Projection, respectively. The results show that naive multi-level averaging is ineffective, while adaptive fusion and multi-step projection consistently improve performance. Moreover, removing R_l, C_l, or the complete trajectory-aware score A_l degrades performance, demonstrating that both recovery evolution and terminal convergence are important for trajectory-aware anomaly scoring.

Table 5: Ablation of multi-level fusion and multi-step projection on anomaly detection datasets. The best result is marked in bold and the second-best result is marked underlined.

Variant	HeadCT	BrainMRI	ZhangLab
Baseline	88.4	92.1	87.9
Baseline + MA	89.6	86.6	89.3
Baseline + MF	90.0	90.8	89.9
Baseline + MP	91.2	90.5	<u>92.9</u>
TAMP w/ d^*	<u>92.9</u>	90.6	92.8
TAMP w/o C_l	92.8	<u>93.2</u>	**93.8**
TAMP w/o R_l	92.8	90.3	92.1
TAMP w/o A_l	90.0	90.8	89.9
TAMP w/ A_l	**94.6**	**96.1**	**93.8**

We further compare TAMP with an endpoint displacement variant, replacing A_l with $d^* = \left\| z^T - z^0 \right\|_2^2$, which measures only the overall feature movement between the initial and final projection states. As shown in Table 5, d^* performs worse than the full TAMP, suggesting that a few normal references provide incomplete coverage of the normal manifold, and repeated projection may also change legitimate normal features. Further analysis of this behavior is provided in Appendix C.2.

Projector Architecture and Depth. We compare ViT and MLP projectors with different depths. As shown in Table 6, ViT yields more consistent performance across datasets, suggesting that modeling interactions among spatial tokens is beneficial for normal-manifold projection. The 4-layer ViT achieves the best overall results, while deeper ViT variants and MLP projectors are less consistent, so we use the 4-layer ViT by default. Appendix D further reports hyperparameter sensitivity and model-efficiency analyses.

Table 6: Ablation of projector architecture and depth on HeadCT, BrainMRI, and ZhangLab Chest X-ray. The best result is marked in bold and the second-best result is marked underlined.

Projector ϕ	Depth	HeadCT		BrainMRI		ZhangLab	
		I-AUROC	I-AUPR	I-AUROC	I-AUPR	I-AUROC	I-AUPR
ViT	4	**94.6**	<u>89.6</u>	**96.1**	**98.7**	**93.8**	**96.1**
	6	93.3	87.6	91.2	96.4	<u>91.9</u>	<u>95.1</u>
	8	<u>94.0</u>	86.7	<u>91.6</u>	92.9	91.5	94.8
MLP	4	91.8	**93.2**	90.1	<u>97.2</u>	91.6	95.0
	6	92.9	87.8	90.1	97.1	91.4	95.0
	8	91.8	**93.2**	90.4	<u>97.2</u>	91.7	<u>95.1</u>

5 Conclusion

In this paper, we proposed Trajectory-Aware Manifold Projection (TAMP) for few-shot medical anomaly detection. Rather than attempting to reconstruct a complete normal manifold from only a few references, TAMP treats incomplete normal-manifold coverage as an inherent challenge of the few-shot setting. Inspired by the perspective of dynamical systems, we move beyond conventional one-step projection paradigms and explore multi-step manifold projection trajectories to expose additional anomaly cues beyond the initial residual. Together with consistency-guided multi-level projection learning, these trajectory cues provide complementary evidence for distinguishing pathological changes from normal variations. Future work will explore adaptive projection strategies and more effective use of limited normal references under heterogeneous imaging conditions.

ETHICS STATEMENT

This work uses only publicly available benchmark datasets. We comply with the licenses, terms of use, and applicable data-governance requirements specified by the original data providers.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, Section 3 specifies the proposed TAMP framework, including the consistency-guided projector learning objectives, multi-step projection procedure, and trajectory-aware anomaly scoring strategy. Appendix A details the adaptive fusion module. Appendix B reports the datasets, preprocessing, backbone, projector architecture, training configuration, inference settings, and evaluation protocol. The implementation, configuration files, and evaluation scripts will be publicly released upon paper acceptance.

AI USE STATEMENT

We used ChatGPT to polish the writing of this manuscript and to assist with code implementation and debugging. All research ideas, methodological decisions, experimental design and execution, and paper illustrations were completed by the authors. We reviewed and validated all AI-assisted edits and code changes, and take full responsibility for the final content of this work.

REFERENCES

Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In *BrainLes*, 2018.

Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Scale-space autoencoders for unsupervised anomaly segmentation in brain MRI. In *MICCAI*, 2020.

Christoph Baur, Stefan Denner, Benedikt Wiestler, Nassir Navab, and Shadi Albarqouni. Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study. *Medical Image Analysis*, 2021.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, 2021.

Navoneel Chakrabarty. Brain MRI images for brain tumor detection. Kaggle, 2019. URL `https://www.kaggle.com/navoneel/brain-mri-images-for-brain-tumor-detection`.

Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.

Norman Cliff. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin*, 1993.

Simon Damm, Mike Laszkiewicz, Johannes Lederer, and Asja Fischer. AnomalyDINO: Boosting patch-based few-shot anomaly detection with DINOv2. In *WACV*, 2025.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: A patch distribution modeling framework for anomaly detection and localization. In *ICPR Workshops*, 2021.

Chenghao Deng, Haote Xu, Xiaolu Chen, Haodi Xu, Xiaotong Tu, Xinghao Ding, and Yue Huang. SimCLIP: Refining image-text alignment with simple prompts for zero-/few-shot anomaly detection. In *ACM MM*, 2024.

Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, 2022.

Le Dong, Qinzhong Tan, Chunlei Li, Jingliang Hu, Yilei Shi, Weisheng Dong, Xiaoxiang Zhu, and Lichao Mou. Dual distillation for few-shot anomaly detection. In *ICLR*, 2026.

Jia Guo, Shuai Lu, Lize Jia, Weihang Zhang, and Huiqi Li. ReContrast: Domain-specific anomaly detection via contrastive reconstruction. In *NeurIPS*, 2023.

Jia Guo, Shuai Lu, Lize Jia, Weihang Zhang, and Huiqi Li. Encoder-decoder contrast for unsupervised anomaly detection in medical images. *IEEE Transactions on Medical Imaging*, 2024.

Jia Guo, Shuai Lu, Weihang Zhang, Fang Chen, Huiqi Li, and Hongen Liao. Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In *CVPR*, 2025.

Junjie Hu, Yuanyuan Chen, and Zhang Yi. Automated segmentation of macular edema in OCT using deep neural networks. *Medical Image Analysis*, 2019.

Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. WinCLIP: Zero-/few-shot anomaly classification and segmentation. In *CVPR*, 2023.

Antanas Kascenas, Nicolas Pugeault, and Alison Q. O'Neil. Denoising autoencoders for unsupervised anomaly detection in brain MRI. In *MIDL*, 2022.

Daniel S. Kermany, Michael Goldbaum, Wenjia Cai, Carolina C.S. Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, Justin Dong, Made K. Prasadha, Jacqueline Pei, Magdalene Y.L. Ting, Jie Zhu, Christina Li, Sierra Hewett, Jason Dong, Ian Ziyar, Alexander Shi, Runze Zhang, Lianghong Zheng, Rui Hou, William Shi, Xin Fu, Yaou Duan, Viet A.N. Huu, Cindy Wen, Edward D. Zhang, Charlotte L. Zhang, Oulan Li, Xiaobo Wang, Michael A. Singer, Xiaodong Sun, Jie Xu, Ali Tafreshi, M. Anthony Lewis, Huimin Xia, and Kang Zhang. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 2018.

Felipe Kitamura. Head CT–hemorrhage. Kaggle, 2018. URL `https://www.kaggle.com/felipekitamura/head-ct-hemorrhage`.

Camile Lendering, Erkut Akdag, and Egor Bondarau. SubspaceAD: Training-free few-shot anomaly detection via subspace modeling. In *CVPR*, 2026.

Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. CutPaste: Self-supervised learning for anomaly detection and localization. In *CVPR*, 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 *CVPR*, 2024.

Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 2017.

Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. Vmamba: Visual state space model. In *NeurIPS*, 2024.

Xijun Lu, Hongying Liu, Fanhua Shang, Yanming Hui, and Liang Wan. PDD: Manifold-prior diverse distillation for medical anomaly detection. In *CVPR*, 2026.

Wenxi Lv, Qinliang Su, and Wenchao Xu. One-for-all few-shot anomaly detection via instance-induced prompt learning. In *ICLR*, 2025.

Henry B. Mann and Donald R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 1947.

John Miller and Moritz Hardt. Stable recurrent models. In *ICLR*, 2019.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russ Howes, Po-Yao (Bernie) Huang, Shang-Wen Li, Ishan Misra, Michael G. Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. 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, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *ICML*, 2021.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, 2022.

Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *IPMI*, 2017.

Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 2019.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3. *arXiv preprint arXiv:2508.10104*, 2025.

Peng Tang, Xiaoxiao Yan, Xiaobin Hu, Kai Wu, Tobias Lasser, and Kuangyu Shi. Anomaly detection in medical images using encoder-attention-2decoders reconstruction. *IEEE Transactions on Medical Imaging*, 2025.

Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for anomaly detection. In *BMVC*, 2021.

Stephen Wiggins. Invariant manifolds: Linear and nonlinear systems. In *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. 2003.

Julian Wyatt, Adam Leach, Sebastian M. Schmon, and Chris G. Willcocks. AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *CVPR Workshops*, 2022.

Tiange Xiang, Yixiao Zhang, Yongyi Lu, Alan L. Yuille, Chaoyi Zhang, Weidong Cai, and Zongwei Zhou. SQUID: Deep feature in-painting for unsupervised anomaly detection. In *CVPR*, 2023.

Shicheng Xu, Wei Li, Zuoyong Li, Tiesong Zhao, and Bob Zhang. Facing differences of similarity: Intra- and inter-correlation unsupervised learning for chest X-ray anomaly detection. *IEEE Transactions on Medical Imaging*, 2025.

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

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.

Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. DeSTSeg: Segmentation-guided denoising student-teacher for anomaly detection. In *CVPR*, 2023.

He Zhao, Yuexiang Li, Nanjun He, Kai Ma, Leyuan Fang, Huiqi Li, and Yefeng Zheng. Anomaly detection for medical images using self-supervised and translation-consistent features. *IEEE Transactions on Medical Imaging*, 2021.

Kang Zhou, Jing Li, Weixin Luo, Zhengxin Li, Jianlong Yang, Huazhu Fu, Jun Cheng, Jiang Liu, and Shenghua Gao. Proxy-bridged image reconstruction network for anomaly detection in medical images. *IEEE Transactions on Medical Imaging*, 2022.

Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. AnomalyCLIP: Object-agnostic prompt learning for zero-shot anomaly detection. In *ICLR*, 2024.

Jiawen Zhu and Guansong Pang. Toward generalist anomaly detection via in-context residual learning with few-shot sample prompts. In *CVPR*, 2024.

APPENDIX

A ADAPTIVE FUSION DETAILS

Different representation levels may respond differently to anatomical structures and pathological patterns. We therefore use a lightweight sample-adaptive fusion module to estimate the contribution of each feature level from its reconstruction statistics.

For the i-th sample, the token-wise reconstruction residual at level l is defined as

$$r_{i,l,j} = \left\| z_{i,l,j} - f^n_{i,l,j} \right\|_2^2, \tag{A1}$$

where j indexes spatial tokens and D denotes the feature dimension. For each level, we summarize the spatial distribution of reconstruction responses using its mean and standard deviation:

$$\mu_{i,l} = \frac{1}{N} \sum_{j=1}^{N} r_{i,l,j}, \tag{A2}$$

and

$$\sigma_{i,l} = \sqrt{\frac{1}{N-1} \sum_{j=1}^{N} \left(r_{i,l,j} - \mu_{i,l} \right)^2}. \tag{A3}$$

The statistics from all L levels are concatenated as

$$g_i = [\mu_{i,1}, \ldots, \mu_{i,L}, \sigma_{i,1}, \ldots, \sigma_{i,L}] \in \mathbb{R}^{2L}. \tag{A4}$$

A lightweight two-layer fusion network then maps these statistics to sample-adaptive level weights. Specifically,

$$h_i = \mathrm{Dropout}_{0.1} \left(\mathrm{ReLU} \left(W_1 g_i + b_1 \right) \right), \qquad h_i \in \mathbb{R}^{64}, \tag{A5}$$

followed by

$$a_i = W_2 h_i + b_2, \qquad a_i \in \mathbb{R}^L, \tag{A6}$$

where $W_1 \in \mathbb{R}^{64 \times 2L}$ and $W_2 \in \mathbb{R}^{L \times 64}$ are learnable parameters. The level weights are obtained through softmax normalization:

$$w_{i,l} = \frac{\exp(a_{i,l})}{\sum_{q=1}^{L} \exp(a_{i,q})}, \qquad \sum_{l=1}^{L} w_{i,l} = 1. \tag{A7}$$

The resulting weights are shared across spatial tokens within each sample and used to aggregate the level-wise reconstruction responses:

$$\bar{r}_{i,j} = \sum_{l=1}^{L} w_{i,l} r_{i,l,j}. \tag{A8}$$

The adaptive fused reconstruction objective is then

$$\mathcal{L}_{\mathrm{fuse}} = \frac{1}{KN} \sum_{i=1}^{K} \sum_{j=1}^{N} \bar{r}_{i,j}. \tag{A9}$$

Unlike fixed averaging, this formulation conditions the fusion weights on the reconstruction-response distribution of each sample, allowing feature levels with more informative reconstruction behavior to contribute more strongly to normal-manifold learning.

B EXPERIMENTAL DETAILS

Implementation Details. We use the pretrained DINOv3 ViT-L/16 (Siméoni et al., 2025) as the default foundation encoder and keep it frozen throughout both training and inference. All input images are resized to 512×512. The projectors follow the ViT-based architecture adopted in baseline FoundAD (Zhai et al., 2026), with a predictor depth of 4 and an embedding dimension of 384. Synthetic anomalies used for projector training are generated with CutPaste (Li et al., 2021).

The overall training objective is optimized with $\lambda_{\text{level}} = 1.0$, $\lambda_{\text{fuse}} = 0.5$, and $\lambda_{\text{cons}} = 0.1$. We use AdamW with a learning rate of 1×10^{-3} and a weight decay of 1×10^{-4}. During inference, the number of projection steps is fixed to $T = 3$, as selected by the ablation study in Sec. 4.3, and the number of selected feature levels is set to $\rho = 2$. The weights β and γ, which balance the trajectory-aware response and terminal convergence residual, are selected from $\{0.1, 0.2\}$ according to the dataset. For image-level anomaly scoring, the top-κ parameter is selected separately for each dataset through repeated empirical experiments to obtain the most suitable setting. All experiments are conducted on a single NVIDIA RTX 4090 GPU.

Dataset Details. We evaluate TAMP on four medical anomaly detection benchmarks covering CT, MRI, chest X-ray, and retinal OCT modalities. For all datasets, we conduct experiments under the 1-shot, 2-shot, and 4-shot settings. The normal training images are randomly sampled from each dataset, and the reported results are averaged over multiple random selections. During training, no abnormal images, category labels, or pixel-level annotations are used. During testing, the full test set is used for evaluation. The dataset details are summarized below:

HeadCT. The HeadCT dataset (Kitamura, 2018) is used for intracranial hemorrhage detection and contains 200 CT images. The training set consists of 60 normal images, while the test set contains 140 images, including 40 normal and 100 hemorrhage samples.

BrainMRI. The BrainMRI dataset (Chakrabarty, 2019) is used for brain tumor detection and contains 252 MRI images. The training set includes 58 normal images, and the test set contains 194 images, including 40 normal and 154 tumor samples.

ZhangLab Chest X-ray. The ZhangLab Chest X-ray dataset (Kermany et al., 2018) contains normal and pneumonia images. It includes 1,249 normal images in the training set, while the test set contains 234 normal and 390 pneumonia images. Pneumonia images are regarded as abnormal samples.

RESC. The RESC dataset (Hu et al., 2019) is used to evaluate anomaly detection and localization on retinal OCT images. The training set contains 4,297 normal images, while the test set includes 1,041 normal and 764 abnormal images. Pixel-level annotations are used only for localization evaluation.

C MULTI-STEP PROJECTION ANALYSIS

C.1 EVOLUTION UNDER REPEATED MANIFOLD PROJECTION

We measure the distance of intermediate states z^t to an evaluation normal manifold across projection steps (Fig. 7). Both normal and abnormal features change substantially after the first projection and continue to evolve over the next few steps, rather than simply reproducing the one-step output. The trajectories then saturate, with the normal curve slightly increasing after an early minimum. These observations support short trajectories for capturing information beyond the initial residual.

C.2 WHY NOT ENDPOINT DISPLACEMENT?

A natural alternative to trajectory-aware scoring is to directly measure the displacement between the initial and terminal states, $d^* = \|z^T - z^0\|_2^2$. However, this quantity collapses all intermediate projection changes into a single accumulated displacement. Under few-shot normal coverage, repeated projection may also modify legitimate anatomical or acquisition-related variations that are underrepresented by the limited normal references. Consequently, a large endpoint displacement does not necessarily indicate pathology, since substantial feature changes may also occur for normal test samples.

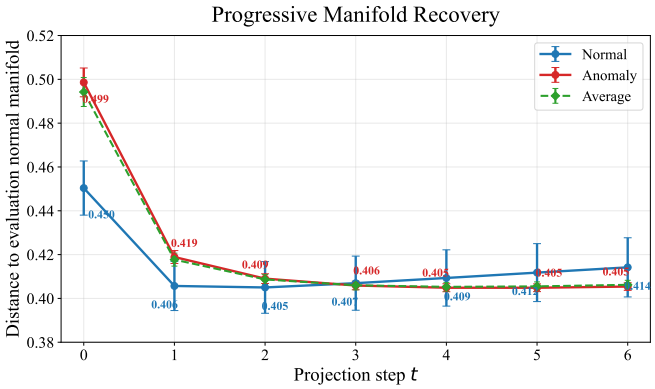


Figure 7: Progressive manifold recovery under repeated projection. The distance to the evaluation normal manifold decreases sharply during the early projection steps and then largely stabilizes, with a slight rebound for normal features after reaching an early minimum.

To examine this behavior, we analyze the evolution of normal representations under repeated projection from three complementary perspectives. Given N normal representations $\{z_i^t\}_{i=1}^N$ at projection step t, we first measure global feature diversity using the mean pairwise Euclidean distance:

$$V_t = \frac{2}{N(N-1)} \sum_{i<j} \left\| z_i^t - z_j^t \right\|_2 . \tag{A10}$$

A decrease in V_t indicates that the normal feature distribution becomes more compact as projection proceeds.

We further evaluate the preservation of local feature geometry using the mean k-nearest-neighbor Jaccard similarity:

$$J_t = \frac{1}{N} \sum_{i=1}^{N} \frac{\left| \mathcal{N}_k^0(i) \cap \mathcal{N}_k^t(i) \right|}{\left| \mathcal{N}_k^0(i) \cup \mathcal{N}_k^t(i) \right|}, \tag{A11}$$

where $\mathcal{N}_k^t(i)$ denotes the k nearest neighbors of the i-th normal representation at step t. A lower J_t indicates increasing changes in the local neighborhood structure.

Finally, we quantify the accumulated feature movement from the initial state:

$$D_t = \frac{1}{N} \sum_{i=1}^{N} \left\| z_i^t - z_i^0 \right\|_2^2 . \tag{A12}$$

Figure 8 shows that V_t initially decreases and then stabilizes, while J_t declines and D_t increases. Thus, repeated projection compresses normal feature diversity, alters local neighborhoods, and accumulates feature movement even without pathology. This explains the ambiguity of d^*, which discards when changes occur. TAMP instead retains the initial residual and complements it with R, the accumulated subsequent displacement relative to d^0, and C, the remaining final-step update.

D Additional Experimental Results

D.1 Comparison with Full-Shot Unsupervised Methods on RESC

Table 7 compares 4-shot TAMP with full-shot unsupervised methods on RESC. Despite using only four normal images, TAMP achieves a P-AUROC of 97.1%, only 0.1% below the best full-shot result, while obtaining the best PRO of 87.9%, outperforming the second-best method by **4.1%**. These results show that TAMP remains highly competitive with full-shot methods for pixel-level anomaly localization, demonstrating the effectiveness of multi-step manifold projection trajectories under extremely limited normal training data.

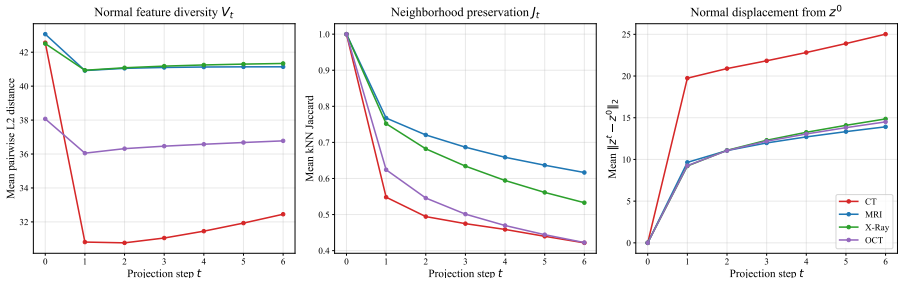


Figure 8: Evolution of normal representations under repeated projection. Feature diversity and neighborhood preservation initially decrease and then stabilize or partially recover, while the accumulated displacement from the initial representation increases.

Table 7: Comparison of anomaly detection performance on RESC between full-shot unsupervised methods and 4-shot TAMP. The best result is marked in bold and the second-best result is marked underlined.

Method	Training	I-AUROC	I-AUPR	P-AUROC	PRO
SALAD	full-shot	43.2	43.2	65.1	30.7
RD4AD	full-shot	**92.6**	**90.3**	**97.2**	<u>83.8</u>
SQUID	full-shot	56.1	43.1	75.2	50.9
EDC	full-shot	90.0	88.0	91.8	69.5
E2AD	full-shot	<u>90.2</u>	<u>88.1</u>	82.2	46.3
TAMP (Ours)	**4-shot**	88.8	85.3	<u>97.1</u>	**87.9**

D.2 Hyperparameter Sensitivity

Table 8: Sensitivity to β and γ under the 4-shot setting. The best result is marked in bold.

β	γ	HeadCT		BrainMRI		ZhangLab	
		I-AUROC	I-AUPR	I-AUROC	I-AUPR	I-AUROC	I-AUPR
0.1	0.1	92.9	87.7	**96.1**	**98.7**	89.5	93.9
0.1	0.5	92.9	87.7	96.1	98.7	89.6	93.9
0.2	0.1	92.8	87.7	95.9	98.7	89.9	94.1
0.2	0.2	**94.6**	**89.6**	86.9	85.8	**93.8**	**96.1**
0.5	0.1	92.8	87.7	93.2	97.5	90.6	94.4
0.8	0.5	92.6	88.1	84.0	83.5	91.1	94.6

Sensitivity to β and γ. We study the sensitivity to the projection trajectory-aware response weight β and terminal-residual weight γ in Table 8. The preferred setting is dataset-dependent: $(\beta, \gamma) = (0.2, 0.2)$ performs best on HeadCT and ZhangLab, while $(0.1, 0.1)$ yields the best performance on BrainMRI. Overall, relatively small values of β and γ provide more robust performance across datasets, whereas larger weights may overemphasize the trajectory-aware terms. We therefore use $\beta, \gamma \in \{0.1, 0.2\}$ in the main experiments.

D.3 Model Efficiency Analysis

Figure 9 compares average detection performance with total model parameters. TAMP achieves strong performance with a moderate model size. Its total parameter count is substantially lower than that of IIPAD (Lv et al., 2025), which combines CLIP ViT-L/14 with BLIP-Diffusion, and SubspaceAD (Lendering et al., 2026), which adopts DINOv2 ViT-G/14. Compared with FoundAD (Zhai et al., 2026), TAMP increases the total parameter count by only 7.3%. Moreover, unlike full-shot baselines (Guo et al., 2024; Tang et al., 2025) that optimize the entire network, TAMP keeps the DINOv3 ViT-L/16 encoder frozen and trains only 34.31M parameters. These results indicate that TAMP achieves a favorable balance between detection performance and training efficiency.

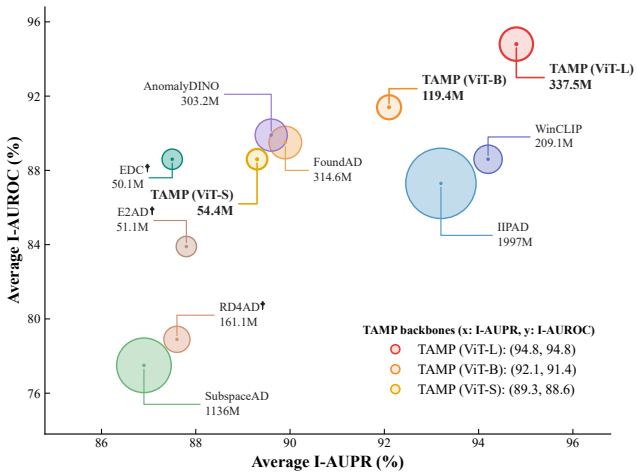


Figure 9: Average detection performance versus total model parameters. Bubble area denotes the total parameter count; † denotes full-shot methods.

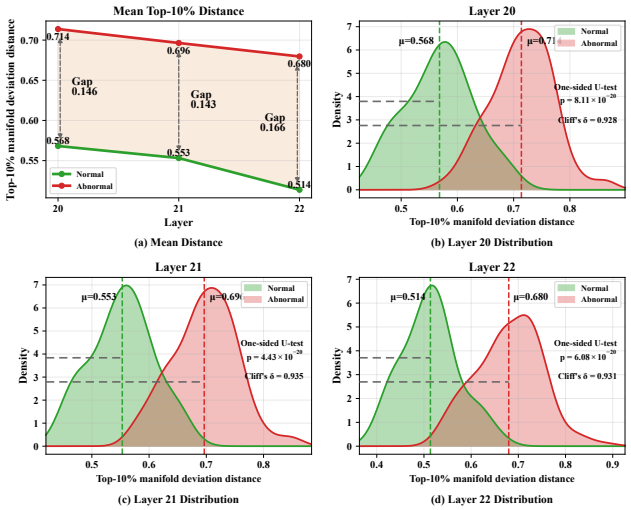


Figure 10: Empirical validation of the normal feature manifold. The image-level deviation score averages the largest 10% patch-level deviation distances. (a) Mean scores for normal and abnormal test images across feature levels. (b)–(d) Score distributions at Layers 20–22.

D.4 Empirical Analysis of Normal-Manifold Deviation

To examine whether normal-manifold modeling provides a meaningful feature-space prior for medical anomaly detection, we construct a point-cloud approximation of the normal feature structure using projected patch features from normal training images. For each test image, we compute the minimum L_2 distance of each patch to the corresponding normal reference structure and aggregate the largest 10% patch-level distances to obtain an image-level manifold-deviation score.

As shown in Fig. 10, abnormal samples tend to exhibit larger manifold deviations than normal samples across all three feature levels. We further conduct one-sided Mann–Whitney U tests (Mann & Whitney, 1947) with the alternative hypothesis abnormal $>$ normal. The resulting $p < 10^{-19}$ at every level indicates a statistically significant shift toward larger deviations for abnormal samples. Cliff's delta (Cliff, 1993) is additionally used to quantify the effect size, with larger values indicating a stronger tendency for abnormal samples to have higher deviation scores. All values exceed 0.92, showing a strong group-level separation between normal and abnormal samples.

The density distributions in Fig. 10(b)–(d) further demonstrate this systematic shift toward larger deviations for abnormal samples. These results support normal-manifold modeling as a useful feature-space prior.

D.5 Supplementary Evaluation on Industrial Anomaly Detection

Table 9: Comparison of anomaly detection performance on MVTec and VisA across different few-shot settings. The best result is marked in bold and the second-best result is marked underlined.

Shot	Method	MVTec				VisA			
		I-AUROC	I-AUPR	P-AUROC	PRO	I-AUROC	I-AUPR	P-AUROC	PRO
1	PatchCore	83.4	92.2	92.0	79.7	79.9	82.8	95.4	80.5
	WinCLIP	93.1	96.5	95.2	87.1	83.8	85.1	96.4	85.1
	InCTRL	91.3	95.2	94.6	87.8	83.2	84.1	89.0	66.7
	AnomalyCLIP	95.2	97.2	94.6	87.6	87.7	87.7	83.2	90.1
	PromptAD	94.6	97.1	95.9	87.9	86.9	88.4	96.7	85.1
	IIPAD	94.2	97.2	96.4	89.8	85.4	87.5	96.9	87.3
	SubspaceAD	**97.1**	**98.6**	97.5	93.7	**93.4**	**93.1**	98.2	93.5
	FoundAD	96.1	97.9	96.8	92.8	92.6	92.0	**99.7**	**98.0**
	TAMP (Ours)	**97.1**	98.5	**97.7**	**94.6**	90.8	90.5	98.3	93.6
2	PatchCore	86.3	93.8	93.3	82.3	81.6	84.8	96.1	82.6
	WinCLIP	94.4	97.0	96.0	88.4	84.6	85.8	96.8	86.2
	InCTRL	91.8	95.5	95.2	88.3	86.3	86.8	89.8	68.1
	AnomalyCLIP	95.4	97.3	94.9	87.8	87.8	89.1	84.5	90.8
	PromptAD	95.7	97.9	96.2	88.5	88.3	90.0	97.1	85.8
	IIPAD	95.7	97.9	96.7	90.3	86.7	88.6	97.2	87.9
	SubspaceAD	**97.6**	98.8	97.7	94.1	93.7	**93.3**	98.3	93.6
	FoundAD	96.9	98.3	97.0	93.2	**93.8**	**93.3**	**99.7**	**98.2**
	TAMP (Ours)	**97.6**	**98.9**	**97.9**	**95.0**	92.1	91.6	98.6	94.8
4	PatchCore	88.8	94.5	94.3	84.3	62.6	69.9	85.4	70.6
	WinCLIP	95.2	97.3	96.2	89.0	84.1	86.1	95.2	82.1
	InCTRL	93.1	96.3	95.8	89.5	87.8	88.0	90.2	68.8
	AnomalyCLIP	96.1	97.8	95.5	88.2	88.9	90.9	97.5	94.3
	PromptAD	96.6	98.5	96.5	90.5	89.1	90.8	97.4	86.2
	IIPAD	96.1	98.1	97.0	91.2	88.3	89.6	97.4	88.3
	SubspaceAD	**97.9**	**99.0**	97.8	94.3	**94.7**	**94.3**	98.4	93.8
	FoundAD	97.1	98.6	97.2	93.6	94.4	94.0	**99.7**	**98.4**
	TAMP (Ours)	97.4	98.8	**97.9**	**95.2**	93.6	92.9	98.8	95.2

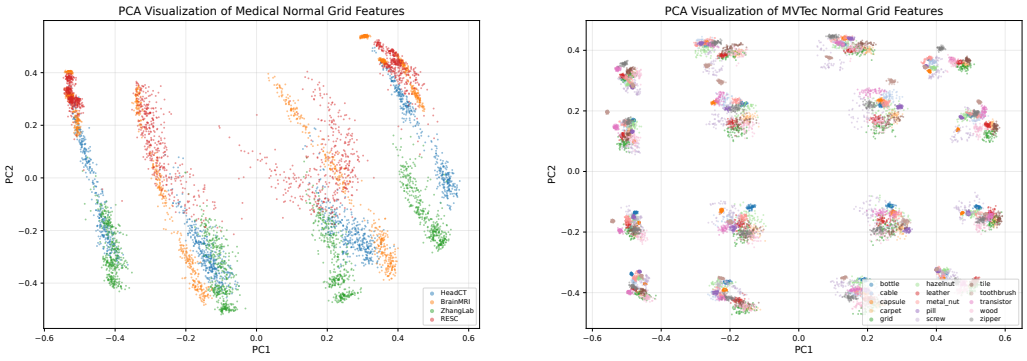
Although TAMP is developed for few-shot medical anomaly detection, we further evaluate its transferability on the industrial MVTec and VisA benchmarks under the 1-, 2-, and 4-shot settings. Table 9 reports image-level and pixel-level results with representative industrial anomaly detection baselines, including PatchCore (Roth et al., 2022), WinCLIP (Jeong et al., 2023), InCTRL (Zhu & Pang, 2024), IIPAD (Lv et al., 2025), AnomalyCLIP (Zhou et al., 2024), PromptAD (Li et al., 2024), SubspaceAD (Lendering et al., 2026) and FoundAD (Zhai et al., 2026). TAMP consistently achieves strong pixel-level localization across different shot settings while maintaining competitive image-level performance. These results demonstrate that multi-step manifold projection trajectories remain effective beyond the medical domain.

Industrial images have more regular structures and compact normal feature distributions, making one-step residuals more discriminative. In medical images, anatomical and acquisition-related variability leaves greater ambiguity under few-shot coverage, motivating complementary trajectory cues.

D.6 Normal Feature Variability Across Domains

Figure 11 visualizes PCA embeddings of spatially pooled normal features from medical datasets and MVTec AD. Each image is partitioned into a 4×4 grid, and one pooled DINOv3 feature is extracted from each spatial cell. Medical normal features exhibit more dispersed and elongated distributions, whereas MVTec features form substantially more compact position-dependent clusters. This suggests greater intra-normal variability in medical images, consistent with anatomical and acquisition-related variations.

This variability makes few-shot normal coverage more difficult: uncovered normal features can produce large initial residuals. The contrast with compact industrial distributions therefore supports using trajectory information to complement one-step anomaly scoring in medical images.



(a) Medical normal features (b) MVTec normal features

Figure 11: Normal feature variability across medical datasets and MVTec AD. PCA embeddings are computed from spatially pooled normal DINOv3 features. Medical features show dispersed and elongated distributions, whereas MVTec features form compact position-dependent clusters.

D.7 Visualization of Multi-Level Projection Trajectories

Figures 12–15 visualize normal and abnormal samples, segmentation heatmaps, and the first three displacement maps at each feature level. Consistent with our motivation, normal samples can produce ambiguous initial residual responses resembling anomalies. Across multiple projections, consecutive-state displacements reveal complementary anomaly cues: lesion regions can retain persistent feature updates, with responses becoming more concentrated around lesions as spurious initial responses are suppressed. This pattern is particularly visible in the BrainMRI and RESC examples, illustrating the discriminative information available beyond the initial residual. The levels also exhibit distinct, complementary spatial patterns that vary across samples, suggesting that a single level may not generalize consistently. These observations support fusing multiple levels to obtain more reliable anomaly localization.

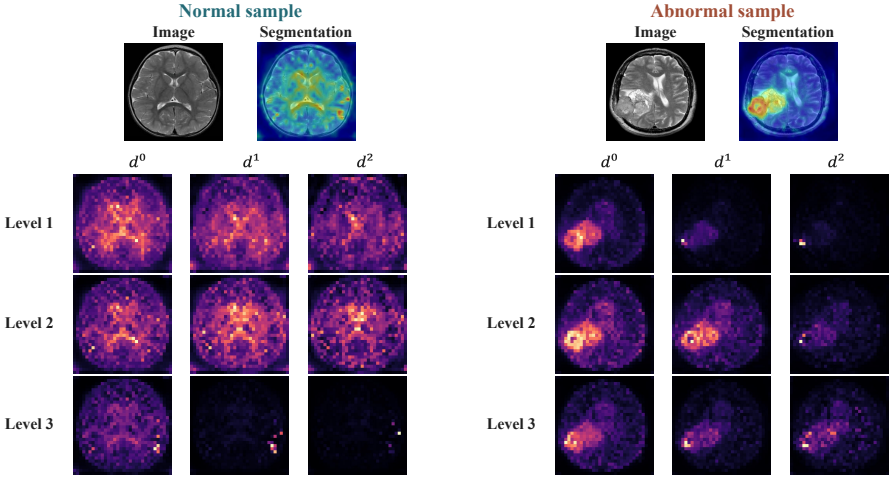


Figure 12: Multi-level displacement visualizations and segmentation heatmaps for normal and abnormal samples from BrainMRI.

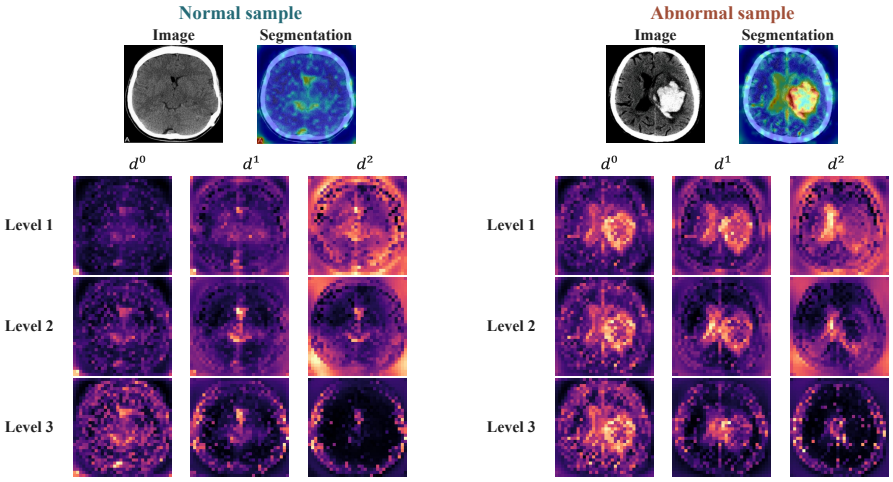


Figure 13: Multi-level displacement visualizations and segmentation heatmaps for normal and abnormal samples from HeadCT.

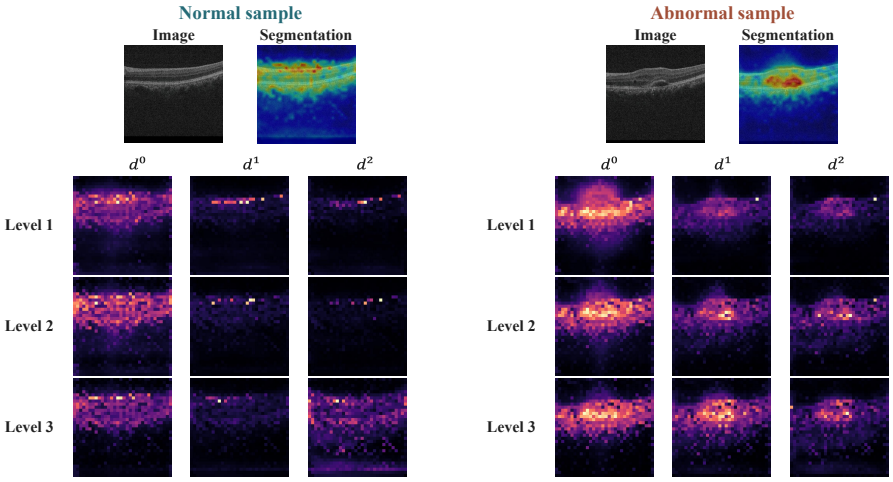


Figure 14: Multi-level displacement visualizations and segmentation heatmaps for normal and abnormal samples from RESC.

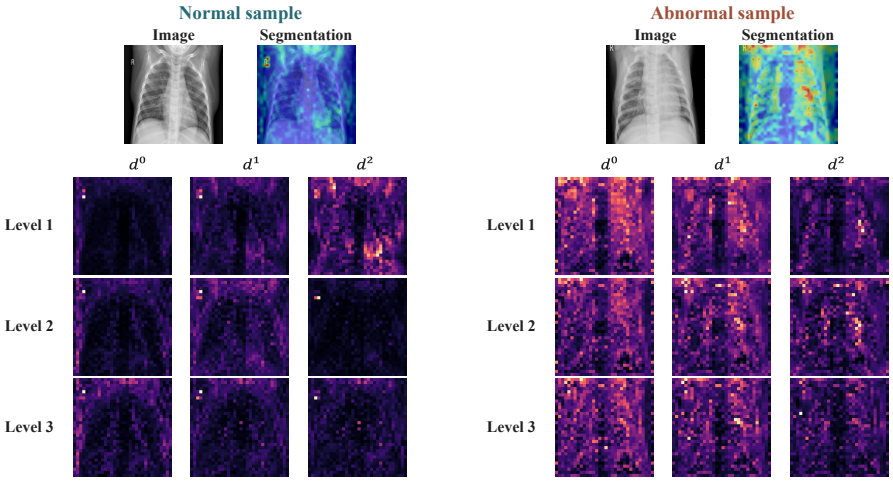


Figure 15: Multi-level displacement visualizations and segmentation heatmaps for normal and abnormal samples from ZhangLab.