HiNorm: Hierarchical Normality Learning for Whole-Slide Image Anomaly Detection

Anonymous authors
Paper under double-blind review

Abstract

Unsupervised anomaly detection (UAD) in whole-slide images (WSIs) provides a scalable approach to detecting pathological abnormalities by learning solely from normal WSIs, without requiring abnormal samples for training or lesion-level annotations. Existing methods, however, typically learn normality from isolated patches and derive slide-level predictions by aggregating patch anomaly scores. This patch-centric paradigm overlooks the spatial organization of tissue, even though pathological assessment depends on both local morphology and global architecture, and may therefore yield unreliable WSI-level decisions. We propose HiNorm, a hierarchical normality learning framework trained exclusively on normal WSIs. HiNorm employs dual-branch reverse distillation, with frozen patch-level and WSI-level pathology foundation models providing complementary references for local tissue morphology and whole-slide organization. Its key component, Asymmetric Cross-Level Reconstruction, uses global slide context to guide patch-level reconstruction while allowing fine-grained patch evidence to inform WSI-level reconstruction. This asymmetric interaction integrates complementary information without conflating the distinct pathological roles of the two representation levels. At inference, patch-level reconstruction discrepancies are mapped spatially for anomaly localization, whereas the WSI-level discrepancy directly measures slide-level abnormality, eliminating heuristic aggregation of patch scores. Across three public WSI benchmarks, HiNorm achieves state-of-the-art performance in both anomaly localization and WSI-level detection, while substantially reducing the WSI–patch performance gap.

1 Introduction

In clinical pathology, whole-slide images (WSIs) provide comprehensive histopathological evidence for disease screening, subtype identification, and prognostic assessment (Pantanowitz et al., 2011; He et al., 2012; Zhou et al., 2020). Unlike natural images, gigapixel WSIs often contain lesions occupying small tissue fractions and appearing as subtle deviations from normal morphology. Accurate pixel-level annotations require exhaustive review by experienced pathologists, making large-scale supervision costly and difficult to scale. Unsupervised anomaly detection (UAD) offers a promising alternative (Fernando et al., 2021; Bao et al., 2024; Cui et al., 2025). By learning normal tissue patterns without lesion annotations, UAD can identify deviations from normality and potentially detect known and previously unseen pathological abnormalities.

Existing WSI UAD methods predominantly follow reconstruction-based paradigms. Earlier approaches (Shvetsova et al., 2021; Linmans et al., 2024; Schlegl et al., 2019) learn normal tissue patterns from conventional or task-specific patch representations, whereas recent methods (Song et al., 2025; Xiao et al., 2026; Wang et al., 2025) use pathology foundation models pretrained on large-scale histopathology data as feature extractors (Lu et al., 2024). Despite stronger representations, both groups retain the patch-centered pipeline in Fig. 1(a). Each gigapixel WSI is partitioned into patches treated as independent samples for normality learning. At inference, each patch receives an anomaly score, which is spatially mapped for localization and aggregated, typically via Top-K pooling, for WSI-level prediction. Thus, normality is modeled explicitly only at the patch level, while slide-level assessment relies indirectly on pooled patch-wise evidence.

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.

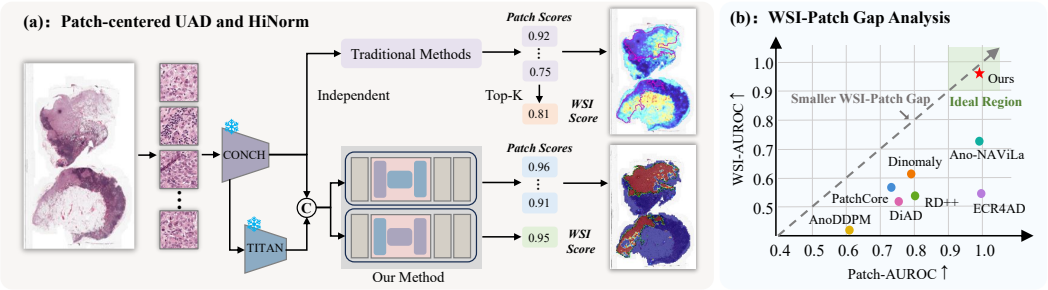

Figure 1: (a) Prevailing patch-centered WSI UAD methods aggregate patch anomaly scores for WSI-level prediction, whereas HiNorm models patch- and WSI-level normality via asymmetric cross-level reconstruction for direct WSI-level assessment and patch-level localization. (b) Patch-level versus WSI-level AUROC of representative WSI UAD methods. The gray dashed diagonal denotes equal performance; proximity to the upper-right indicates stronger overall performance, and proximity to the diagonal a smaller WSI–Patch gap. HiNorm achieves high AUROC at both levels with a substantially reduced gap.

This patch-centered formulation assumes that learning patch-level normality and aggregating anomaly evidence suffice to characterize whole-slide normality and support reliable WSI-level assessment. However, as shown in Fig. 1(b), many methods achieve strong patch-level anomaly discrimination but substantially weaker WSI-level performance, producing a pronounced WSI–Patch gap. This persistent gap despite stronger patch representations suggests that WSI-level limitations arise not only from representation capacity but also from the pathological levels at which normality is modeled. WSI normality reflects both fine-grained local morphology and slide-wide structural relationships, spatial distributions, co-occurrence patterns, and pathological composition. Patch-level normality identifies subtle local morphological deviations and localizes abnormalities, whereas WSI-level normality provides a global reference for assessing their pathological significance within the slide. The two therefore provide related, complementary, but non-interchangeable references. Existing WSI UAD methods explicitly model only patch-level normality, deriving WSI predictions by aggregating patch anomaly evidence. However, such aggregation is not equivalent to explicitly learning whole-slide tissue organization. Thus, the key limitation is not patch decomposition or patch-level normality learning, but restricting explicit normality modeling to patches rather than jointly establishing patch- and WSI-level normality references.

To address this limitation, we propose **HiNorm**, a hierarchical normality learning framework that explicitly models patch- and WSI-level normality. Trained only on normal WSIs, HiNorm uses dual-branch reverse distillation with frozen patch- and WSI-level pathology foundation models as complementary teachers: the former encodes local tissue morphology, while the latter captures whole-slide tissue organization and composition. A shared bottleneck transforms their representations for reconstruction by dedicated student branches, establishing level-specific normality references within a unified framework. A key challenge is enabling patch- and WSI-level normality to interact without collapsing into one representation. Direct fusion or symmetric feature exchange may obscure their distinct pathological roles. We therefore introduce *Asymmetric Cross-Level Reconstruction*: whole-slide context guides patch reconstruction to incorporate broader tissue context, while fine-grained patch evidence informs WSI reconstruction to ground global normality in local tissue patterns. This role-specific interaction preserves complementary responsibilities: local morphology supports precise anomaly localization, while global tissue organization supports direct slide-level assessment.

During inference, HiNorm spatially maps patch reconstruction discrepancies for anomaly localization and averages WSI-level discrepancies to obtain the slide-level anomaly score without aggregating patch predictions. Aligning normality learning with anomaly assessment across levels enables fine-grained localization and direct WSI-level detection. Experiments on three public WSI benchmarks, Camelyon16 (Ehteshami Bejnordi et al., 2017), Camelyon17 (Bandi et al., 2018), and TCGA-NSCLC (Weinstein et al., 2013), show that HiNorm consistently outperforms state-of-the-art WSI UAD methods at both levels while substantially narrowing the WSI–Patch gap.

Our main contributions are as follows:

- We expose a key limitation of existing WSI UAD pipelines: slide-level decisions are typically inferred from pooled patch scores without explicitly modeling whole-slide normality. We therefore establish a hierarchical formulation learning complementary normality references at patch and WSI levels.
- We develop **HiNorm**, a dual-branch reverse distillation framework leveraging frozen pathology foundation models to transfer level-specific knowledge of local morphology and global tissue organization.
- We design *Asymmetric Cross-Level Reconstruction* to couple the two normality levels through role-specific information exchange. Global context guides patch reconstruction, while local evidence grounds WSI reconstruction, enabling cross-level cooperation without collapsing their distinct functions.
- We validate HiNorm on three public WSI benchmarks, achieving state-of-the-art anomaly localization and WSI-level detection while consistently narrowing the WSI–Patch gap.

2 Related Work

2.1 Unsupervised Anomaly Detection in WSIs

To reduce reliance on exhaustive lesion-level annotations, unsupervised anomaly detection (UAD) has gained attention in computational pathology (Bao et al., 2024; Cui et al., 2025). Existing UAD methods mainly follow reconstruction-based or embedding-based paradigms. Reconstruction-based methods, including autoencoder (Baur et al., 2021; Shvetsova et al., 2021), generative adversarial network (Schlegl et al., 2019; Akcay et al., 2018), and diffusion approaches (Linmans et al., 2024; Wyatt et al., 2022), reconstruct normal samples and detect anomalies from reconstruction discrepancies. Embedding-based methods, such as teacher–student frameworks (Deng & Li, 2022; Tien et al., 2023) and memory banks (Roth et al., 2022), model feature-space normality and quantify anomalies through feature discrepancies or distances to normal representations.

Although effective in natural images, industrial inspection, and general medical imaging, these methods remain challenging to apply directly to whole-slide images (WSIs). Gigapixel resolution precludes end-to-end processing, while tissue heterogeneity complicates normality modeling. Existing approaches therefore partition WSIs into patches and learn normal tissue patterns from conventional or task-specific patch representations (Hirra et al., 2021; Shvetsova et al., 2021; Linmans et al., 2024; Schlegl et al., 2019). This makes processing tractable and enables fine-grained localization by spatially mapping patch anomaly scores. However, treating patches independently restricts explicit normality modeling to local patterns, while WSI-level predictions rely on aggregating patch scores, typically via max pooling or Top-K averaging (Cui et al., 2025; Linmans et al., 2024). Such aggregation summarizes local anomaly evidence without explicitly modeling structural relationships or whole-slide tissue organization.

2.2 Foundation Models for WSI UAD

Recent pathology foundation models pretrained on large-scale histopathology data have advanced computational pathology representation learning. Patch-level vision and vision–language models (Lu et al., 2024; Filiot et al., 2023; Nechaev et al., 2024; Chen et al., 2024; Vorontsov et al., 2023) learn transferable representations of cellular morphology, tissue texture, and local pathological semantics. WSI-level foundation models (Ding et al., 2025; Xu et al., 2024; Wang et al., 2024; Shaikovski et al., 2024) instead aggregate tissue regions and spatial relationships to capture global histological context, tissue organization, and slide composition. Thus, patch-level models emphasize fine-grained local morphology, whereas WSI-level models capture broader tissue organization.

Building on these models, recent WSI UAD methods (Song et al., 2025; Xiao et al., 2026; Wang et al., 2025) use CONCH (Lu et al., 2024) as a patch-level feature extractor and model anomalies from its representations. Some reconstruct normal patch features or measure feature discrepancies, while others incorporate vision–language cues or predefined pathological priors. Despite these differences, normality modeling remains patch-centered: normality is learned from patch representations, while WSI-level predictions are obtained by aggregating patch anomaly scores. Stronger foundation representations therefore improve local anomaly characterization but do not establish

an explicit normality reference for whole-slide tissue organization. In contrast, HiNorm combines complementary patch-level and WSI-level foundation models to explicitly learn normality at both levels through asymmetric cross-level reconstruction.

3 METHODOLOGY

3.1 OVERVIEW OF HINORM

HiNorm is a hierarchical framework that learns complementary patch- and WSI-level normality references. As shown in Fig. 2, HiNorm uses dual-branch reverse distillation with frozen patch- and WSI-level pathology foundation models, a shared MLP bottleneck, and reconstruction branches. Given a WSI, CONCH (Lu et al., 2024) extracts all valid patch representations, while TITAN (Ding et al., 2025) aggregates them with their spatial coordinates into a WSI representation. For reconstruction, patch and WSI representations are paired and passed through the shared bottleneck. HiNorm introduces *Asymmetric Cross-Level Reconstruction*, where whole-slide context guides patch reconstruction and fine-grained patch evidence informs WSI reconstruction, enabling cross-level interaction while preserving distinct pathological roles. HiNorm is trained only on normal WSIs. During training, a patch subset is reconstructed, while the WSI representation uses all valid patches. During inference, all valid patches are reconstructed. Patch reconstruction discrepancies yield anomaly scores and localization maps, whereas WSI reconstruction discrepancies are averaged into the slide-level anomaly score without pooling patch anomaly scores.

3.2 DUAL-LEVEL PATHOLOGY REPRESENTATION

Given a WSI, we partition its tissue regions into N_{all} valid patches, denoted by $\{P_i\}_{i=1}^{N_{\text{all}}}$. A frozen CONCH encoder (Lu et al., 2024) extracts each patch representation as

$$F_{P,i} = \text{CONCH}(P_i) \in \mathbb{R}^{L\times d}, \quad i = 1, \ldots, N_{\text{all}}, \tag{1}$$

where L is the number of tokens per patch representation and d is the feature dimension. These representations capture fine-grained local tissue morphology. For a complementary WSI-level representation, frozen TITAN (Ding et al., 2025) aggregates all valid patch representations and their spatial coordinates:

$$F_W = \text{TITAN}\left(\{(F_{P,i}, \mathbf{c}_i)\}_{i=1}^{N_{\text{all}}}\right) \in \mathbb{R}^{1\times d}, \tag{2}$$

where $\mathbf{c}_i$ denotes the spatial coordinate of patch P_i. Thus, $F_{P,i}$ and F_W represent local tissue morphology and whole-slide tissue organization, respectively. The WSI-level representation uses all valid patches during training and inference.

During training, we sample N of the N_{all} valid patches for reconstruction. Their representations $\{F_{P,n}\}_{n=1}^{N}$ are stacked as $F_P \in \mathbb{R}^{N\times L\times d}$. Each sampled patch representation is paired with the same slide's WSI-level representation by concatenation along the token dimension:

$$F_n = [F_{P,n}; F_W] \in \mathbb{R}^{(L+1)\times d}, \quad n = 1, \ldots, N. \tag{3}$$

The paired representations are stacked as $F \in \mathbb{R}^{N\times(L+1)\times d}$, and passed to the shared bottleneck for cross-level reconstruction. During inference, patch sampling is omitted, and all valid patch representations are paired with F_W and processed identically.

3.3 ASYMMETRIC CROSS-LEVEL RECONSTRUCTION

Given the paired representation F_n of the n-th sampled patch, a shared MLP bottleneck transforms the features:

$$H_n = B(F_n) \in \mathbb{R}^{(L+1)\times d}, \tag{4}$$

where $B(\cdot)$ denotes the shared bottleneck. Dropout is applied during training to discourage trivial identity reconstruction. Following the token organization of F_n, the output is separated into patch-level and WSI-level components: $H_n = [H_{P,n}; H_{W,n}]$, where $H_{P,n} \in \mathbb{R}^{L\times d}$ and $H_{W,n} \in \mathbb{R}^{1\times d}$.

We introduce *Asymmetric Cross-Level Reconstruction* for role-specific interaction between the two pathological levels. The patch and WSI branches each contain eight stacked Transformer decoder

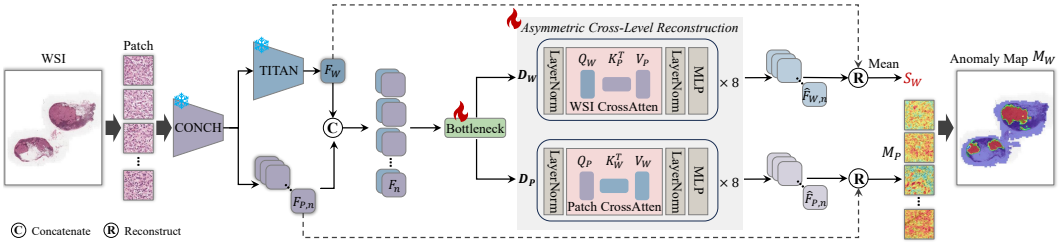

Figure 2: Overview of HiNorm. HiNorm learns complementary patch- and WSI-level normality via dual-branch reconstruction with asymmetric cross-level interaction. WSI context guides patch-level reconstruction, while fine-grained patch evidence informs WSI-level reconstruction, enabling anomaly detection and localization at both pathological levels.

blocks, incorporating cross-level information through oppositely directed cross-attention. The patch branch uses WSI-level context to condition patch reconstruction, whereas the WSI branch uses patch-level evidence to inform WSI reconstruction. We denote cross-attention as

$$\mathcal{A}(Q, K, V) = \text{softmax}\left(\frac{\theta(Q)\phi(K)^T}{\sqrt{d_h}}\right) g(V), \tag{5}$$

where $\theta(\cdot)$, $\phi(\cdot)$, and $g(\cdot)$ denote learnable linear projections for the query, key, and value, respectively, and d_h is the feature dimension of each attention head.

In the patch branch, $H_{P,n}$ serves as the query, while $H_{W,n}$ provides the key and value:

$$Z_{P,n} = \mathcal{A}\left(H_{P,n}, H_{W,n}, H_{W,n}\right). \tag{6}$$

As $H_{W,n}$ is a single WSI-level token, this interaction provides shared whole-slide conditioning to the patch tokens, allowing patch reconstruction to incorporate whole-slide pathological context.

Conversely, the WSI branch uses $H_{W,n}$ as the query and $H_{P,n}$ as the key and value:

$$Z_{W,n} = \mathcal{A}\left(H_{W,n}, H_{P,n}, H_{P,n}\right). \tag{7}$$

The WSI token attends to the L patch tokens, allowing fine-grained evidence from the n-th patch to inform WSI reconstruction. Thus, the two directions play distinct but complementary roles: whole-slide context conditions patch reconstruction, while local pathological evidence informs WSI reconstruction. After eight decoder blocks, the two branches produce $\hat{F}_{P,n} \in \mathbb{R}^{L\times d}$ and $\hat{F}_{W,n} \in \mathbb{R}^{1\times d}$, where $\hat{F}_{P,n}$ denotes the reconstructed representation of the n-th patch and $\hat{F}_{W,n}$ the WSI-level representation reconstructed under that patch's conditioning.

3.4 Reconstruction Objective

HiNorm minimizes reconstruction discrepancies between frozen teacher representations and their student reconstructions. Following Dinomaly (Guo et al., 2025), we use a hard-mining cosine reconstruction objective $D_{\cos}(\cdot,\cdot)$. For the n-th sampled patch, patch- and WSI-level reconstruction losses are

$$\mathcal{L}_{P,n} = D_{\cos}\left(F_{P,n}, \hat{F}_{P,n}\right), \qquad \mathcal{L}_{W,n} = D_{\cos}\left(F_W, \hat{F}_{W,n}\right). \tag{8}$$

Since each $\hat{F}_{W,n}$ is conditioned on the n-th patch, all patch-conditioned WSI reconstructions use the same teacher target F_W, with losses averaged over the N sampled patches:

$$\mathcal{L}_P = \frac{1}{N}\sum_{n=1}^{N}\mathcal{L}_{P,n}, \qquad \mathcal{L}_W = \frac{1}{N}\sum_{n=1}^{N}\mathcal{L}_{W,n}. \tag{9}$$

The overall training objective is $\mathcal{L} = \mathcal{L}_P + \mathcal{L}_W$.

3.5 Anomaly Scoring and Inference

During inference, patch sampling is omitted and all N_{all} valid patches are processed. For the n-th patch, the l-th token reconstruction discrepancy is the cosine distance between teacher and reconstructed representations:

$$e_{P,n}^{(l)} = 1 - \frac{\left(F_{P,n}^{(l)}\right)^T \hat{F}_{P,n}^{(l)}}{\left\|F_{P,n}^{(l)}\right\|_2 \left\|\hat{F}_{P,n}^{(l)}\right\|_2}, \quad l = 1, \ldots, L. \tag{10}$$

The patch anomaly score averages the largest 1% of token-wise reconstruction errors:

$$S_{P,n} = \text{Mean}\left(\text{Top}_{1\%}\left(\{e_{P,n}^{(l)}\}_{l=1}^{L}\right)\right). \tag{11}$$

Token-wise errors are spatially arranged within each patch and upsampled to its original resolution to generate the corresponding patch anomaly heatmap $M_{P,n}$. The whole-slide heatmap M_W is then obtained by mapping patch anomaly scores $\{S_{P,n}\}_{n=1}^{N_{\text{all}}}$ to their corresponding WSI locations and spatially upsampling them over the tissue region.

For WSI-level detection, each valid patch yields a patch-conditioned WSI reconstruction $\hat{F}_{W,n}$. Its discrepancy from the WSI teacher representation is

$$e_{W,n} = 1 - \frac{F_W^T \hat{F}_{W,n}}{\|F_W\|_2 \left\|\hat{F}_{W,n}\right\|_2}. \tag{12}$$

The WSI anomaly score averages reconstruction discrepancies over all valid patches:

$$S_W = \frac{1}{N_{\text{all}}} \sum_{n=1}^{N_{\text{all}}} e_{W,n}. \tag{13}$$

Unlike conventional WSI predictions obtained by pooling patch anomaly scores, S_W is derived directly from the WSI reconstruction branch.

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate HiNorm on three public WSI datasets: Camelyon16 (C16) (Ehteshami Bejnordi et al., 2017), Camelyon17 (C17) (Bandi et al., 2018), and TCGA-NSCLC (Weinstein et al., 2013). C16 and C17 are breast cancer metastasis datasets, with C17 used for cross-center generalization evaluation, while TCGA-NSCLC evaluates anomaly detection across lung cancer subtypes. Detailed dataset statistics and splits are provided in the Appendix A.1.

Evaluation Metrics. We evaluate HiNorm at both WSI and patch levels using image-level detection metrics (I-AUROC, I-AP, and I-F1) and pixel-level localization metrics (P-AUROC and P-AUPRO). Detailed metric definitions and evaluation protocols are provided in Appendix A.2.

Compared Methods. We compare HiNorm with representative unsupervised anomaly detection methods in two categories: natural-image UAD (PatchCore (Roth et al., 2022), DiAD (He et al., 2024), RD++ (Tien et al., 2023), and Dinomaly (Guo et al., 2025)) and pathology-specific UAD (DPA (Shvetsova et al., 2021), AnoDDPM (Linmans et al., 2024), ECR4AD (Xiao et al., 2026), and Ano-NAViLa (Song et al., 2025)). For reference, we report weakly supervised MIL methods: DSMIL (Li et al., 2021), TransMIL (Shao et al., 2021), CaMIL (Fourkioti et al., 2024), and HD-MIL (Dong et al., 2025), trained with WSI-level supervision.

Implementation Details. Each WSI is partitioned into non-overlapping 512×512 patches at $20\times$ magnification. Pretrained CONCH (Lu et al., 2024) and TITAN (Ding et al., 2025) extract patch- and WSI-level representations, respectively, and remain frozen throughout training and inference. We sample $N = 50$ patches per WSI for reconstruction training. All experiments are conducted on a single NVIDIA RTX 4090 GPU. Additional implementation details are provided in Appendix A.3.

Table 1: Quantitative comparison of WSI-level and patch-level anomaly detection and localization performance on Camelyon16. Among the unsupervised methods, the best and second-best results are highlighted in **bold** and underlined, respectively. An en dash indicates results that are not reported for the corresponding setting.

Paradigm	Method	WSI-level					Patch-level				
		I-AUROC	I-AP	I-F1	P-AUROC	P-AUPRO	I-AUROC	I-AP	I-F1	P-AUROC	P-AUPRO
Weakly Supervised	DSMIL (Li et al., 2021)	0.9370	0.9439	0.9032	0.9995	0.9085	0.9992	0.9985	0.9316	–	–
	TransMIL (Shao et al., 2021)	0.9674	0.9736	0.9661	0.9977	0.8862	0.9334	0.8507	0.8030	–	–
	CaMIL (Fourkioti et al., 2024)	0.9500	0.9578	0.8913	0.9962	0.7932	0.9977	0.9964	0.9836	–	–
	HDMIL (Dong et al., 2025)	0.9586	0.9631	0.9321	0.9887	0.8124	0.9967	0.9955	0.8728	–	–
Unsupervised	Dinomaly (Guo et al., 2025)	0.6125	0.4974	0.5750	0.8250	0.4550	0.7886	0.4773	0.6137	0.7972	0.4505
	PatchCore (Roth et al., 2022)	0.5768	0.4671	0.5614	0.8369	0.5354	0.7490	0.4460	0.5760	0.7550	0.3960
	DiAD (He et al., 2024)	0.5031	0.4189	0.5563	0.8735	0.6086	0.7580	0.4700	0.5760	0.7710	0.4330
	RD++ (Tien et al., 2023)	0.5255	0.4233	0.5517	0.8494	0.5524	0.7898	0.5043	0.6080	0.7931	0.4828
	AnoDDPM (Linmans et al., 2024)	0.4018	0.3451	0.5465	0.9250	0.7428	0.6205	0.3818	0.4600	0.4356	0.0936
	DPA (Shvetsova et al., 2021)	0.3909	0.3539	0.5455	0.8384	0.4791	0.5914	0.3502	0.4455	0.5309	0.1642
	ECR4AD (Xiao et al., 2026)	0.5365	0.3940	0.5629	0.9484	0.5628	0.9944	0.9934	0.9456	0.9696	0.8942
	Ano-NAViLa (Song et al., 2025)	0.7302	0.6516	0.6613	0.9899	0.8308	0.9934	0.9812	0.9509	–	–
	HiNorm	**0.9581**	**0.9531**	**0.9293**	**0.9980**	**0.9662**	**0.9978**	**0.9959**	**0.9729**	**0.9700**	**0.9028**

Table 2: Quantitative comparison of HiNorm with representative methods on Camelyon17 and TCGA-NSCLC. Among the unsupervised methods, the best and second-best results are highlighted in **bold** and underlined, respectively.

Paradigm	Method	Camelyon17					TCGA-NSCLC		
		I-AUROC	I-AP	I-F1	P-AUROC	P-AUPRO	I-AUROC	I-AP	I-F1
Weakly Supervised	DSMIL (Li et al., 2021)	0.6142	0.7475	0.7525	0.9992	0.8680	0.9679	0.9613	0.9180
	TransMIL (Shao et al., 2021)	0.6287	0.7760	0.6073	0.9979	0.7882	0.9749	0.9767	0.9212
	CaMIL (Fourkioti et al., 2024)	0.6200	0.7388	0.7573	0.9989	0.7992	0.9722	0.9645	0.9293
	HDMIL (Dong et al., 2025)	0.6596	0.7732	0.7600	0.9937	0.7607	0.9781	0.9810	0.9273
Unsupervised	Dinomaly (Guo et al., 2025)	0.5541	0.7208	0.8167	0.8037	0.4057	0.4810	0.4762	0.6443
	PatchCore (Roth et al., 2022)	0.4925	0.6866	0.8065	0.8087	0.4259	0.5467	0.5591	0.6486
	DiAD (He et al., 2024)	0.5017	0.6663	0.8065	0.8569	0.6179	0.4071	0.4280	0.6421
	RD++ (Tien et al., 2023)	0.5100	0.6798	0.8130	0.9156	0.7321	0.5631	0.5229	0.6508
	AnoDDPM (Linmans et al., 2024)	0.5358	0.7169	0.8197	0.9056	0.7481	0.6029	0.5648	0.6523
	DPA (Shvetsova et al., 2021)	0.5342	0.7206	0.8167	0.8197	0.5422	0.5326	0.5336	0.6505
	ECR4AD (Xiao et al., 2026)	0.5583	0.7063	0.8142	0.9758	0.5034	0.5624	0.5663	0.6528
	Ano-NAViLa (Song et al., 2025)	0.4917	0.7158	0.8067	0.9885	0.5361	0.6887	0.6005	0.7179
	HiNorm	**0.5642**	**0.7256**	**0.8205**	**0.9983**	**0.9164**	**0.8371**	**0.7642**	**0.8054**

4.2 Quantitative Comparison

Table 1 reports results on Camelyon16. HiNorm ranks first among unsupervised methods at the WSI level, exceeding the second-best I-AUROC, I-AP, and I-F1 by 22.79, 30.15, and 26.80 percentage points, respectively. It also improves P-AUPRO by 13.54 points and achieves the best patch-level results across all metrics, showing that stronger WSI-level assessment does not compromise local discrimination or localization. Its WSI–Patch I-AUROC gap is only 3.97 percentage points, versus 26.32 for Ano-NAViLa, supporting complementary patch- and WSI-level normality modeling. Despite training only on normal WSIs, HiNorm remains competitive with weakly supervised MIL methods: its WSI-level P-AUPRO exceeds the best baseline by 5.77 percentage points, while image-level detection remains comparable. These results show that explicit WSI-level normality modeling enables reliable whole-slide assessment without WSI-level anomaly labels.

Table 2 reports results on Camelyon17 and TCGA-NSCLC. On Camelyon17, HiNorm leads all unsupervised methods across five metrics under cross-center evaluation. Its largest gain is in localization, exceeding the second-best P-AUPRO by 16.83 percentage points while retaining the best image-level detection performance among unsupervised methods. Despite training without WSI-level anomaly labels, HiNorm also surpasses all weakly supervised MIL methods in I-F1 and P-AUPRO, demonstrating strong cross-center generalization and reliable localization across medical centers. On TCGA-NSCLC, HiNorm leads all unsupervised methods across three image-level metrics, improving I-AUROC, I-AP, and I-F1 over the second-best Ano-NAViLa by 14.84, 16.37, and 8.75 percentage points, respectively. These gains in histological subtype discrimination demonstrate generalization beyond breast cancer metastasis detection to a distinct pathological setting.

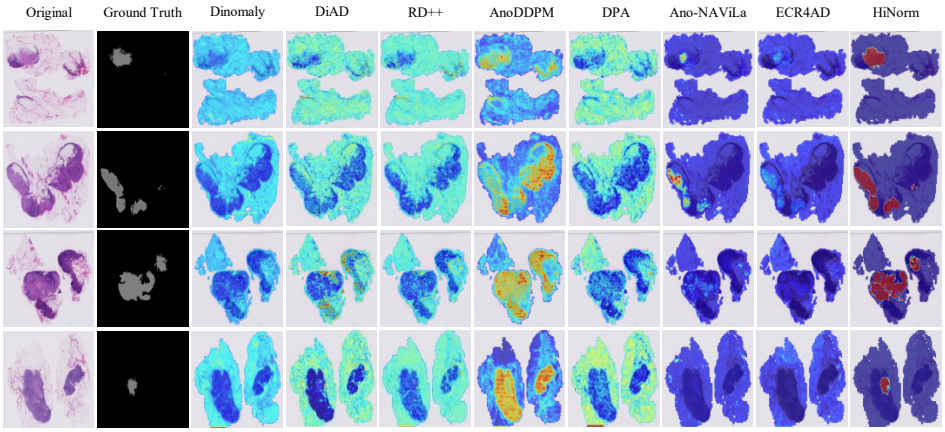

Figure 3: Qualitative comparison of WSI-level anomaly localization. HiNorm achieves more complete lesion coverage with fewer false-positive responses in normal tissue.

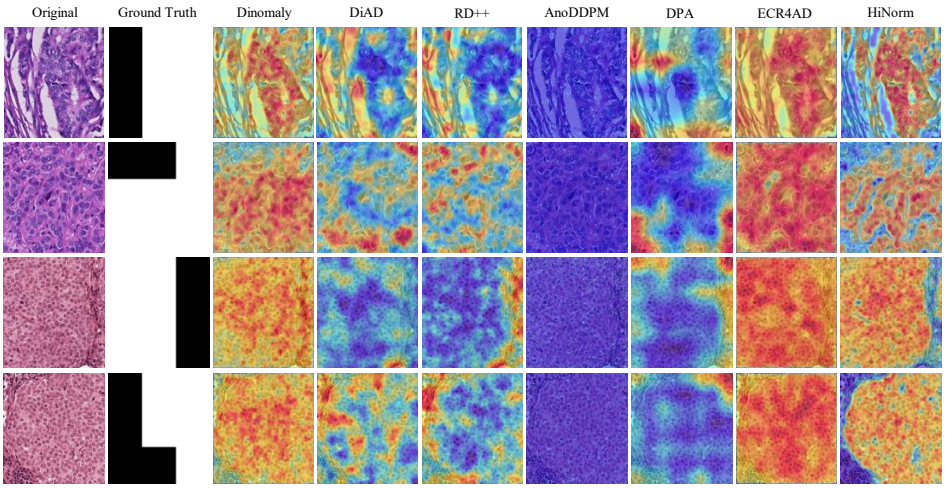

Figure 4: Qualitative comparison of patch-level anomaly localization. HiNorm achieves more complete fine-grained lesion localization with fewer false-positive responses in normal tissue.

4.3 Qualitative Comparison

We further conduct qualitative comparisons of anomaly localization at both the WSI and patch levels. As shown in Fig. 3, HiNorm consistently provides more complete coverage of abnormal regions at the WSI level while effectively suppressing spurious responses in normal tissue. In contrast, competing unsupervised methods often capture only partial lesion regions, leading to missed abnormalities, or produce widespread responses outside the lesions, resulting in false positives. Fig. 4 presents representative results at the patch level. HiNorm similarly produces more complete and concentrated responses within abnormal regions, whereas the competing methods frequently exhibit incomplete lesion coverage or scattered activations in normal tissue. These consistent observations across representative samples demonstrate that HiNorm provides reliable anomaly localization at both levels, capturing abnormal regions comprehensively across whole slides while preserving fine-grained localization within individual patches.

4.4 Ablation Study

We conduct ablation studies on Camelyon16 to evaluate HiNorm's core architecture, patch-level feature encoder, and training patch quantity. Results are summarized in Table 3, with additional WSI-level scoring ablations in Appendix B.1.

Table 3: Comprehensive ablation studies on Camelyon16, covering the core architecture (WB: WSI Branch; PB: Patch Branch; ACLR: Asymmetric Cross-Level Reconstruction), patch-level feature encoder, and number of patches sampled during training. Best results in each ablation group are highlighted in **bold**. An en dash indicates results not reported for the corresponding setting.

Ablation Factor	Setting	WSI-level Evaluation					Patch-level Evaluation				
		I-AUROC	I-AP	I-F1	P-AUROC	P-AUPRO	I-AUROC	I-AP	I-F1	P-AUROC	P-AUPRO
Architecture	WB Only	**0.9641**	**0.9579**	0.9000	**0.9980**	0.9647	–	–	–	–	–
	PB Only	0.8443	0.7730	0.7069	0.9973	0.9568	0.9973	0.9945	0.9702	0.9671	0.8832
	WB + PB	0.9276	0.9107	0.8654	0.9973	0.9568	0.9975	0.9949	0.9711	0.9673	0.8864
	WB + PB + ACLR	0.9581	0.9531	**0.9293**	**0.9980**	**0.9662**	**0.9978**	**0.9959**	**0.9729**	**0.9700**	**0.9028**
Encoder	DINOv2 (Oquab et al., 2023)	0.9555	0.9494	0.9200	0.8815	0.5436	0.7979	0.5086	0.6201	0.7821	0.3422
	Phikon (Filiot et al., 2023)	**0.9625**	**0.9558**	0.9109	0.9976	0.9183	0.9973	0.9951	**0.9738**	0.9633	0.8824
	Hibou-B (Nechaev et al., 2024)	0.9573	0.9510	**0.9293**	0.9949	0.8956	0.9937	0.9873	0.9490	0.9482	0.8400
	CONCH (Lu et al., 2024)	0.9581	0.9531	**0.9293**	**0.9980**	**0.9662**	**0.9978**	**0.9959**	0.9729	**0.9700**	**0.9028**
Sampled Patches	50 Patches	0.9581	0.9531	**0.9293**	0.9980	0.9662	**0.9978**	0.9959	0.9729	0.9700	**0.9028**
	100 Patches	**0.9615**	**0.9573**	0.9200	0.9980	0.9647	0.9975	0.9954	0.9722	0.9700	0.8840
	150 Patches	0.9581	0.9540	0.9184	**0.9981**	**0.9670**	**0.9978**	**0.9960**	**0.9760**	0.9707	0.9021
	All Patches	0.9609	0.9558	0.9091	0.9980	0.9642	0.9977	0.9958	0.9748	**0.9710**	0.8932

Effect of Core Architecture. We evaluate the patch and WSI reconstruction branches and Asymmetric Cross-Level Reconstruction (ACLR). As shown in Table 3, WB achieves strong WSI-level performance but provides no patch-level predictions, whereas PB maintains strong patch-level performance but performs substantially worse at the WSI level. Combining both branches without ACLR improves WSI-level performance while preserving patch-level performance. ACLR further improves all metrics, with gains of 6.39 percentage points in WSI-level I-F1 and 1.64 points in patch-level P-AUPRO. Although WB yields slightly higher WSI-level I-AUROC and I-AP, the full model achieves the best overall performance across both levels, supporting complementary patch- and WSI-level normality modeling and effective cross-level cooperation through ACLR.

Impact of Patch Feature Encoders. We compare CONCH (Lu et al., 2024) with the general-purpose DINOv2 (Oquab et al., 2023) and pathology-specific Phikon (Filiot et al., 2023) and Hibou-B (Nechaev et al., 2024). As shown in Table 3, encoder choice has limited impact on WSI-level detection but substantially affects localization and patch-level performance. DINOv2 performs markedly worse than pathology-specific encoders, highlighting the importance of pathology-aware representations for fine-grained tissue abnormalities. Among them, CONCH performs best on all four localization metrics, exceeding the second-best P-AUPRO by 4.79 and 2.04 percentage points at the WSI and patch levels, respectively, while remaining competitive in image-level detection. We therefore use CONCH as the default patch encoder.

Influence of Patch Sampling Quantity. We vary the number of patches sampled per WSI during training, with $N \in \{50, 100, 150, \text{all}\}$. As shown in Table 3, performance remains stable across sampling quantities, with no consistent gains as N increases. Although different metrics favor different settings, $N = 50$ achieves the best WSI-level I-F1 and patch-level P-AUPRO while remaining competitive on the others, indicating that a small subset suffices for effective reconstruction learning. We therefore use $N = 50$ by default, maintaining strong overall performance with substantially fewer patch-wise reconstructions.

5 Conclusion

We propose **HiNorm**, a hierarchical framework for unsupervised WSI anomaly detection that learns complementary patch- and WSI-level normality references. Unlike patch-centered formulations, HiNorm uses frozen patch- and WSI-level pathology foundation models as complementary teachers for fine-grained local morphology and whole-slide tissue organization. We introduce *Asymmetric Cross-Level Reconstruction*, where whole-slide context guides patch reconstruction and fine-grained patch evidence informs WSI reconstruction, exploiting their complementarity while preserving distinct pathological roles. During inference, patch reconstruction discrepancies enable fine-grained localization, while averaged WSI reconstruction discrepancies yield the slide-level anomaly score without post-hoc patch aggregation. Experiments on three public WSI benchmarks show that HiNorm consistently outperforms existing WSI UAD methods in localization and WSI-level detection while substantially narrowing the WSI–Patch gap. Together, these designs enable joint normality learning across the two levels. These results highlight the importance of aligning normality learning with anomaly assessment levels for reliable WSI anomaly detection.

AI USE STATEMENT

We used AI tools to assist with English-language polishing, including improvements to clarity and grammar. The authors reviewed and revised the AI-assisted edits to ensure that they preserved the intended scientific meaning. The authors take full responsibility for the final content of the manuscript, including all AI-assisted content.

REPRODUCIBILITY STATEMENT

We provide detailed descriptions of the model architecture, training objective, anomaly scoring, and inference procedure in the main text. Dataset statistics and splits, evaluation protocols, and implementation details are provided in Appendix A. WSI-level scoring ablations are reported in Appendix B. These details are provided to facilitate reproduction of the reported experiments.

REFERENCES

Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*, pp. 622–637. Springer, 2018.

Peter Bandi, Oscar Geessink, Quirine Manson, Marcory Van Dijk, Maschenka Balkenhol, Meyke Hermsen, Babak Ehteshami Bejnordi, Byungjae Lee, Kyunghyun Paeng, Aoxiao Zhong, et al. From detection of individual metastases to classification of lymph node status at the patient level: the camelyon17 challenge. *IEEE transactions on medical imaging*, 38(2):550–560, 2018.

Jinan Bao, Hanshi Sun, Hanqiu Deng, Yinsheng He, Zhaoxiang Zhang, and Xingyu Li. Bmad: Benchmarks for medical anomaly detection. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 4042–4053. IEEE, 2024.

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*, 69:101952, 2021.

Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature medicine*, 30(3):850–862, 2024.

Can Cui, Xindong Zheng, Ruining Deng, Quan Liu, Tianyuan Yao, Keith T Wilson, Lori A Coburn, Bennett A Landman, Haichun Yang, Yaohong Wang, et al. Quantitative benchmarking of anomaly detection methods in digital pathology. *arXiv preprint arXiv:2506.19234*, 2025.

Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9737–9746, 2022.

Tong Ding, Sophia J Wagner, Andrew H Song, Richard J Chen, Ming Y Lu, Andrew Zhang, Anurag J Vaidya, Guillaume Jaume, Muhammad Shaban, Ahrong Kim, et al. A multimodal whole-slide foundation model for pathology. *Nature medicine*, pp. 1–13, 2025.

Jiuyang Dong, Junjun Jiang, Kui Jiang, Jiahan Li, and Yongbing Zhang. Fast and accurate gigapixel pathological image classification with hierarchical distillation multi-instance learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30818–30828, 2025.

Babak Ehteshami Bejnordi, Mitko Veta, Paul Johannes van Diest, Bram Van Ginneken, Nico Karssemeijer, Geert Litjens, Jeroen AWM Van Der Laak, CAMELYON16 consortium, Meyke Hermsen, Quirine F Manson, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22):2199–2210, 2017.

Tharindu Fernando, Harshala Gammulle, Simon Denman, Sridha Sridharan, and Clinton Fookes. Deep learning for medical anomaly detection–a survey. *ACM computing surveys (CSUR)*, 54(7): 1–37, 2021.

Alexandre Filiot, Ridouane Ghermi, Antoine Olivier, Paul Jacob, Lucas Fidon, Axel Camara, Alice Mac Kain, Charlie Saillard, and Jean-Baptiste Schiratti. Scaling self-supervised learning for histopathology with masked image modeling. *MedRxiv*, pp. 2023–07, 2023.

Olga Fourkioti, Matt De Vries, and Chris Bakal. CAMIL: Context-aware multiple instance learning for cancer detection and subtyping in whole slide images. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=rzBskAEmoc.

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 *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 20405–20415, 2025.

Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pp. 8472–8480, 2024.

Lei He, L Rodney Long, Sameer Antani, and George R Thoma. Histology image analysis for carcinoma detection and grading. *Computer methods and programs in biomedicine*, 107(3):538–556, 2012.

Irum Hirra, Mubashir Ahmad, Ayaz Hussain, M Usman Ashraf, Iftikhar Ahmed Saeed, Syed Furqan Qadri, Ahmed M Alghamdi, and Ahmed S Alfakeeh. Breast cancer classification from histopathological images using patch-based deep learning modeling. *IEEE Access*, 9:24273–24287, 2021.

Bin Li, Yin Li, and Kevin W Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14318–14328, 2021.

Jasper Linmans, Gabriel Raya, Jeroen van der Laak, and Geert Litjens. Diffusion models for out-of-distribution detection in digital pathology. *Medical Image Analysis*, 93:103088, 2024.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

Ming Y Lu, Bowen Chen, Drew FK Williamson, Richard J Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, et al. A visual-language foundation model for computational pathology. *Nature medicine*, 30(3):863–874, 2024.

Dmitry Nechaev, Alexey Pchelnikov, and Ekaterina Ivanova. Hibou: A family of foundational vision transformers for pathology. *arXiv preprint arXiv:2406.05074*, 2024.

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.

Liron Pantanowitz, Paul N Valenstein, Andrew J Evans, Keith J Kaplan, John D Pfeifer, David C Wilbur, Laura C Collins, and Terence J Colgan. Review of the current state of whole slide imaging in pathology. *Journal of pathology informatics*, 2(1):36, 2011.

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.

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*, 54:30–44, 2019.

George Shaikovski, Adam Casson, Kristen Severson, Eric Zimmermann, Yi Kan Wang, Jeremy D Kunz, Juan A Retamero, Gerard Oakley, David Klimstra, Christopher Kanan, et al. Prism: A multi-modal generative foundation model for slide-level histopathology. *arXiv preprint arXiv:2405.10254*, 2024.

Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, et al. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. *Advances in neural information processing systems*, 34:2136–2147, 2021.

Nina Shvetsova, Bart Bakker, Irina Fedulova, Heinrich Schulz, and Dmitry V Dylov. Anomaly detection in medical imaging with deep perceptual autoencoders. *IEEE Access*, 9:118571–118583, 2021.

Jinsol Song, Jiamu Wang, Anh Tien Nguyen, Keunho Byeon, Sangjeong Ahn, Sung Hak Lee, and Jin Tae Kwak. Normal and abnormal pathology knowledge-augmented vision-language model for anomaly detection in pathology images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 22066–22076, 2025.

Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan Duong, Chanh D Tr Nguyen, and Steven QH Truong. Revisiting reverse distillation for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 24511–24520, 2023.

Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Siqi Liu, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, et al. Virchow: A million-slide digital pathology foundation model. *arXiv preprint arXiv:2309.07778*, 2023.

Jiamu Wang, Keunho Byeon, Jinsol Song, Anh Nguyen, Sangjeong Ahn, Sung Hak Lee, and Jin Tae Kwak. Pathology-informed latent diffusion model for anomaly detection in lymph node metastasis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 453–462. Springer, 2025.

Xiyue Wang, Junhan Zhao, Eliana Marostica, Wei Yuan, Jietian Jin, Jiayu Zhang, Ruijiang Li, Hongping Tang, Kanran Wang, Yu Li, et al. A pathology foundation model for cancer diagnosis and prognosis prediction. *Nature*, 634(8035):970–978, 2024.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Mitchell Wortsman, Tim Dettmers, Luke Zettlemoyer, Ari Morcos, Ali Farhadi, and Ludwig Schmidt. Stable and low-precision training for large-scale vision-language models. *Advances in Neural Information Processing Systems*, 36:10271–10298, 2023.

Julian Wyatt, Adam Leach, Sebastian M Schmon, and Chris G Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 649–655. IEEE, 2022.

Bin Xiao, Collins Wangulu, Theodorus van der Kwast, George M Yousef, and Fatemeh Zabihollahy. Revisiting reconstruction-based anomaly detection for whole slide image. *IEEE Transactions on Medical Imaging*, 2026.

Hanwen Xu, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, et al. A whole-slide foundation model for digital pathology from real-world data. *Nature*, 630(8015):181–188, 2024.

Xiaomin Zhou, Chen Li, Md Mamunur Rahaman, Yudong Yao, Shiliang Ai, Changhao Sun, Qian Wang, Yong Zhang, Mo Li, Xiaoyan Li, et al. A comprehensive review for breast histopathology image analysis using classical and deep neural networks. *IEEE Access*, 8:90931–90956, 2020.

A Additional Experimental Details

A.1 Dataset Details

Camelyon16 (C16) (Ehteshami Bejnordi et al., 2017) is a breast cancer metastasis dataset of lymph-node WSIs. The training set contains 160 normal WSIs, while the test set contains 129 WSIs, including 80 normal and 49 abnormal slides. For patch-level evaluation, the test set further contains 4,000 normal patches and 1,546 abnormal patches. **Camelyon17 (C17)** (Bandi et al., 2018) is another breast cancer metastasis dataset collected from multiple medical centers. To evaluate cross-center generalization, the model trained on C16 is directly evaluated on the C17 test set without retraining or fine-tuning. The test set contains 74 WSIs, including 50 abnormal slides with pixel-level annotations and 24 normal slides. **TCGA-NSCLC** (Weinstein et al., 2013) comprises two non-small cell lung cancer subtypes: lung adenocarcinoma (TCGA-LUAD) and lung squamous cell carcinoma (TCGA-LUSC). For UAD, LUAD is treated as normal and LUSC as anomalous. The training set contains 423 LUAD WSIs, while the test set contains 203 WSIs, including 107 LUAD and 96 LUSC slides.

A.2 Evaluation Metrics

We evaluate HiNorm at both the WSI and patch levels following the standard image-level detection and pixel-level localization protocol used in anomaly detection. At the WSI level, each WSI is treated as one evaluation unit, with S_W used for image-level anomaly detection and M_W for pixel-level localization over the whole slide. At the patch level, each patch is treated as one evaluation unit, with $S_{P,n}$ and $M_{P,n}$ used for image-level detection and pixel-level localization, respectively. We report AUROC, average precision (AP), and F1-score for image-level evaluation, denoted as I-AUROC, I-AP, and I-F1, and AUROC and the area under the per-region overlap curve (AUPRO) for pixel-level evaluation, denoted as P-AUROC and P-AUPRO.

A.3 Implementation Details

Each WSI is partitioned into non-overlapping 512×512 patches at $20\times$ magnification. Pretrained CONCH (Lu et al., 2024) extracts patch-level features, while pretrained TITAN (Ding et al., 2025) aggregates all valid patch features and their spatial coordinates into a WSI-level representation. During training, $N = 50$ patches are sampled from each WSI for reconstruction learning. Both pathology foundation models remain frozen throughout training and inference. The network is optimized with StableAdamW (Wortsman et al., 2023; Loshchilov & Hutter, 2017) using $\beta_1 = 0.9$, $\beta_2 = 0.999$, a weight decay of 10^{-4}, and an initial learning rate of 2×10^{-3}. All experiments are implemented in PyTorch and conducted on a single NVIDIA RTX 4090 GPU.

B Additional Experimental Results

B.1 WSI-level Anomaly Scoring Strategy

We evaluate different WSI-level anomaly scoring strategies. HiNorm directly derives the WSI-level anomaly score from reconstruction discrepancies in the WSI representation space. To assess this design, we keep all other components unchanged and replace only the scoring strategy with mean, max, or Top-1% aggregation of patch-level anomaly scores. As shown in Table 4, WSI reconstruction scoring consistently outperforms all aggregation strategies, exceeding the best aggregation result for each metric by 10.24, 14.50, and 16.29 percentage points in I-AUROC, I-AP, and I-F1, respectively. These results support directly assessing anomalies in the WSI representation space rather than deriving WSI-level predictions through post-hoc aggregation of patch-level anomaly scores.

Table 4: Ablation study of the WSI-level anomaly scoring strategy on Camelyon16. The best results are highlighted in **bold**.

Ablation Factor	Setting	WSI-level Evaluation		
		I-AUROC	I-AP	I-F1
Scoring Strategy	Mean Aggregation	0.7594	0.6735	0.6609
	Max Aggregation	0.8445	0.7176	0.7664
	Top-1% Aggregation	0.8557	0.8081	0.7253
	WSI Reconstruction Scoring	**0.9581**	**0.9531**	**0.9293**