# TopoLeaf: A Zero-Shot Visual Anomaly Detection Framework via Stable Feature Dimensionality and Local Persistent Homology for Self-Organizing Agricultural Cyber-Physical Systems

**Shutong Hou**

202516009003@stu.sxau.edu.cn

Shanxi Agricultural University

---

**Research Article**

**Keywords:** Zero-shot visual anomaly detection, Persistent homology, Cross-domain generalization, Vision foundation model, Plant disease diagnosis, Agricultural cyber-physical systems

**Posted Date:** July 1st, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-10126172/v1



**Additional Declarations:** No competing interests reported.

# TopoLeaf: A Zero-Shot Visual Anomaly Detection Framework via Stable Feature Dimensionality and Local Persistent Homology for Self-Organizing Agricultural Cyber-Physical Systems

Shutong Hou (ORCID: 0009-0006-0643-1225)[1*]

[1]School of Software, Shanxi Agricultural University, Jinzhong, Shanxi, China.

Corresponding author(s). E-mail(s): 20251600903@stu.sxau.edu.cn;

**Abstract**

Zero-shot visual anomaly detection in complex textured domains remains a fundamental challenge for building adaptive, self-organizing cyber-physical systems. Conventional deep learning approaches often rely on closed-set assumptions, require prohibitive pixel-level annotation costs, and suffer severe performance degradation under cross-domain shifts—limiting their deployability in real-world agricultural CPS where novel disease types and unseen crop species continuously emerge. To address these issues, we present TopoLeaf, a training-free and annotation-free anomaly detection framework. By leveraging the robust semantic representations of foundation models (specifically DINOv2), our method introduces two complementary scoring mechanisms: a geometric anomaly score based on local KNN distance in a stability-selected feature subspace, and a topological anomaly score derived from local persistent homology. The topological score effectively captures subtle structural deviations and micro-texture mutations that geometric distances often miss. Extensive experiments on cross-species plant disease benchmarks (3,100+ images across 40 source–target pairs) demonstrate that TopoLeaf achieves highly competitive and structurally robust zero-shot performance, providing a robust perception layer for closed-loop agricultural cyber-physical systems that must maintain diagnostic stability under previously unseen perturbations. Under well-aligned domains, our geometric score achieves near-perfect detection (e.g., 0.994 AUROC on Strawberry). The method exhibits informative failure modes on structurally isolated domains such as Corn (0.169 AUROC), revealing fundamental structural properties of the foundation model's feature manifold. Furthermore, the topological score demonstrates structural complementarity, achieving 0.542 AUROC on the challenging Corn-to-Apple pair where geometric scoring degenerates to 0.221. Module ablation studies confirm that stability-based dimensionality selection consistently improves cross-domain generalization.

**Keywords:** Zero-shot visual anomaly detection, Persistent homology, Cross-domain generalization, Vision foundation model, Plant disease diagnosis, Agricultural cyber-physical systems

# 1 Introduction

Visual anomaly detection aims to identify deviations from a nominal reference distribution without requiring labeled anomalies [1, 2]. In complex textured domains such as biological specimens, agricultural crops, and medical tissues, this task becomes particularly challenging due to high intra-class variance, subtle defect morphologies, and severe distribution shifts across domains. From a cybernetics perspective, this task is essential for building closed-loop agricultural systems where perception, decision-making, and control form a feedback cycle. A robust anomaly detection module must exhibit Ashby's law of requisite variety—it must be sufficiently adaptable to handle the diversity of real-world conditions without retraining [3].

Conventional deep learning approaches to visual anomaly detection face three fundamental limitations. First, closed-set classification paradigms assume that all possible anomaly categories are represented in the training set—an assumption violated in real-world scenarios where novel defect types continuously emerge [4]. Second, pixel-level anomaly segmentation masks require domain expertise and prohibitive annotation effort, creating severe data scarcity for supervised methods. Third, fine-tuning large models demands substantial computational resources, limiting deployment on edge devices.

From a machine learning and cybernetics perspective, cross-species plant disease detection represents a particularly challenging problem in designing adaptive perception systems. The core challenge aligns with the cybernetic principle of viability: a system must maintain its core functionality (disease detection) across varying environmental conditions (different crop species) without external reconfiguration. The primary bottleneck lies in the feature manifold shift: the healthy feature distribution of the source domain (e.g., Apple) often diverges substantially from the target domain (e.g., Corn), causing conventional geometric metrics such as Euclidean or cosine KNN distance to yield high false-positive rates. Our stability-based dimensionality selection mitigates this by filtering out domain-specific high-frequency noise, while local persistent homology ensures robustness against minor topological perturbations in the feature space.

We formulate this task within the broader context of Agricultural Cyber-Physical Systems (A-CPS) and smart agriculture IoT. In such systems, the perception layer must satisfy three cybernetic design principles: (1) robustness—maintaining performance under noise and distribution shifts; (2) adaptivity—functioning in novel

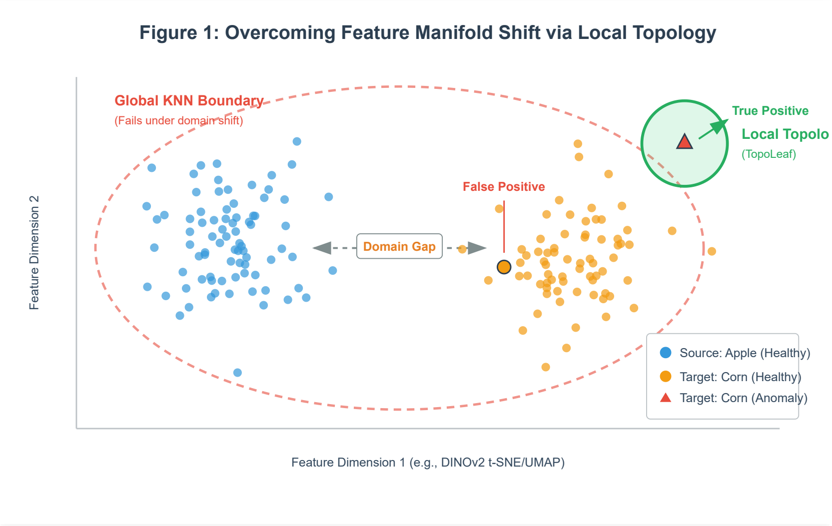


**Fig. 1** Illustration of the cross-domain feature manifold shift and the advantage of TopoLeaf. Conventional global geometric boundaries suffer from high false-positive rates under severe domain gaps, whereas TopoLeaf leverages local topological boundaries to accurately identify anomalies in isolated target domains.

environments without explicit retraining; and (3) self-organization—automatically adjusting internal representations to match new conditions. The zero-shot nature of TopoLeaf directly addresses these three principles, enabling automated deployment in previously unseen crop environments without requiring prior disease annotations. This work advances the state of the art in building feedback-driven intelligent systems for precision agriculture.

We present TopoLeaf, a training-free, annotation-free framework with the following contributions:

1. **Stable feature dimensionality selection.** We identify and retain only the low-variance dimensions of DINOv2 patch features computed over healthy reference images. This stability-selected subspace suppresses domain-specific textural variation while amplifying semantically meaningful disease-induced perturbations.
2. **Geometric anomaly scoring via local KNN distance.** For each test patch, we compute its mean cosine distance to $K$ nearest healthy reference patches in the stability-selected subspace. The 90th percentile of patch-level deviations serves as the image-level anomaly score.
3. **Topological anomaly scoring via local persistent homology.** We introduce 0-dimensional persistent homology computed on local neighborhoods of healthy patches. Unlike geometric distances that are sensitive to local noise and domain

shifts, persistent homology provides a multi-scale topological summary of the feature manifold, with its stability theorem guaranteeing bounded changes under small input perturbations.

4. **Systematic cross-domain generalization benchmark.** We conduct a 5-benchmark $\times$ 8-test-species cross-validation experiment on the PlantVillage dataset (3,100+ test images), establishing a quantitative map of zero-shot cross-domain anomaly detection performance.
5. **Comprehensive validation and open-source release.** We benchmark against five classical anomaly detection methods, perform multi-backbone validation (DINOv2, CLIP ViT-B/32, ResNet-50), conduct module-level ablation studies, and publicly release the complete source code and an interactive web-based demonstration system.

# 2 Related Work

## 2.1 Deep Anomaly Detection

Deep anomaly detection has been extensively studied in industrial defect inspection and general visual quality inspection [1, 2]. Memory-bank methods such as PatchCore [1] store nominal features and perform nearest-neighbor scoring, achieving strong results on same-domain benchmarks like MVTec AD. Distribution-modeling approaches such as PaDiM [2] fit multivariate Gaussian distributions to normal patch features. More recent zero-shot methods, including WinCLIP and AnomalyDINO, leverage vision-language models for anomaly detection without task-specific training. However, these approaches typically assume the availability of carefully designed text prompts or specific model architectures, and their cross-domain generalization capabilities remain underexplored. Classical methods such as Isolation Forest, One-Class SVM, and Local Outlier Factor (LOF) assume shared training–test distributions, severely limiting their applicability to zero-shot cross-domain scenarios.

## 2.2 Topological Data Analysis in Machine Learning

Topological data analysis (TDA), particularly persistent homology [5], provides a mathematically principled framework for characterizing the multi-scale structure of data. In machine learning, persistent homology has been applied to analyzing neural network representations [6], generative model evaluation, and medical image analysis [7]. The key advantage of TDA lies in its stability theorem [8]: small perturbations in the input space result in bounded changes in persistence diagrams, making TDA-based features inherently robust to noise and minor variations. However, the integration of persistent homology with vision foundation models for zero-shot anomaly detection remains unexplored. To our knowledge, TopoLeaf represents the first framework combining local persistent homology with frozen foundation model features for cross-domain visual anomaly detection.

### 2.3 Plant Disease Detection

Traditional plant disease detection relied on hand-crafted texture and color features [9]. Deep learning approaches using CNNs [4, 10, 11] and vision transformers [12] have achieved high accuracy on closed-set benchmarks. Self-supervised learning has been explored for feature extraction [13]. However, the vast majority of these methods operate under closed-set classification, making them fundamentally incapable of detecting novel disease types.

### 2.4 Adaptive Perception in Agricultural Cyber-Physical Systems

The concept of zero-shot anomaly detection is intimately connected to the cybernetic design of agricultural cyber-physical systems (A-CPS) [14]. Traditional A-CPS perception modules rely on task-specific training, which violates the cybernetic principle of requisite variety by limiting the system's ability to respond to unanticipated conditions [3]. Recent advances in foundation models have opened new pathways toward building more adaptive perception layers. However, most existing work in plant disease detection operates under closed-set assumptions [4, 10], making them unsuitable for deployment in dynamic CPS environments where novel disease types and new crop species continuously emerge. Our work addresses this gap by proposing a training-free framework that exhibits self-organizing properties—it automatically adapts its feature representation (via stability selection) and leverages topological invariants that are inherently robust to domain shifts. This aligns with the broader cybernetic goal of building systems that maintain homeostasis under varying external conditions.

## 3 Methodology

### 3.1 Overview

TopoLeaf operates in four stages: (1) feature extraction from a frozen DINOv2 backbone, (2) stability-based dimensionality selection on healthy reference images, (3) dual-path anomaly scoring (geometric and topological), and (4) anomaly heatmap generation. No model training or fine-tuning is required at any stage.

### 3.2 Problem Formulation

Let $\mathcal{X}_{\text{ref}} = \{x_i\}_{i=1}^{M}$ be a set of $M$ healthy reference images from a source domain $\mathcal{D}_S$ (e.g., Apple). Let $x_{\text{test}}$ be a test image potentially containing anomalies from an arbitrary target domain $\mathcal{D}_T$. A pre-trained visual feature extractor $f_\theta : \mathbb{R}^{H\times W\times 3} \to \mathbb{R}^{N\times D}$ maps each image to $N$ patch-level feature vectors, where $H = W = 448$, $N = 1024$, and $D = 384$ for DINOv2 ViT-S/14. Our goal is to learn an anomaly scoring function $S : \mathbb{R}^{N\times D} \to \mathbb{R}$ such that $S(f_\theta(x_{\text{test}}))$ is large if and only if $x_{\text{test}}$ contains anomalies relative to the healthy reference distribution $\mathcal{X}_{\text{ref}}$, without requiring any labeled anomaly samples from $\mathcal{D}_T$ during training.

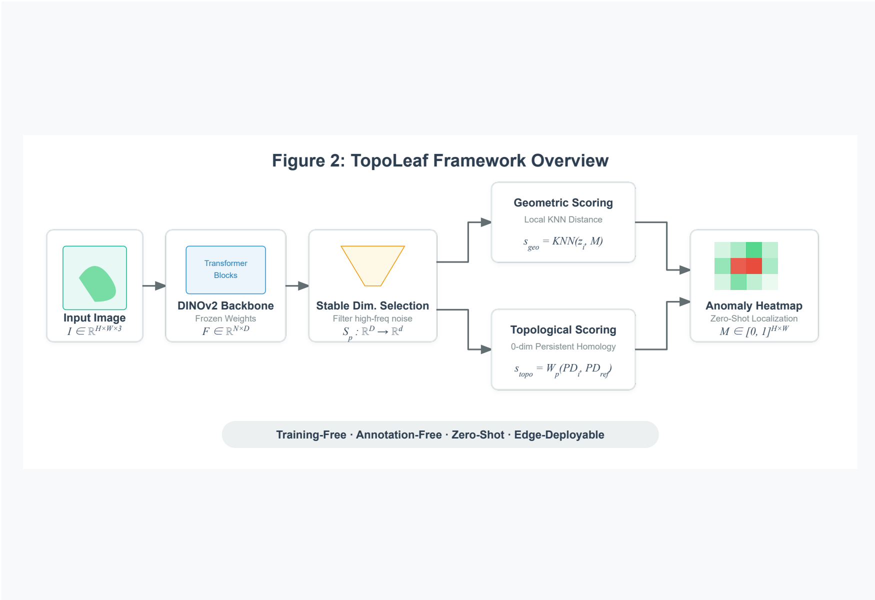


**Fig. 2** Overview of the TopoLeaf framework. The pipeline integrates frozen foundation model features with stability-based dimensionality selection and a dual-path scoring mechanism (geometric and topological) for zero-shot anomaly localization.

## 3.3 Feature Extraction

We adopt DINOv2 ViT-S/14 [15] as our primary backbone. Unlike vision-language models such as CLIP that rely on text-image contrastive pre-training, DINOv2 is trained via purely visual self-supervision, which aligns more naturally with the unsupervised anomaly detection setting where no textual labels of "normal" or "anomalous" are available. Additionally, DINOv2 features exhibit more uniform within-class distributions and stronger fine-grained discriminative power for subtle visual differences such as micro-textural disease mutations. We pass input images through the frozen backbone to obtain patch tokens $\mathbf{F} \in \mathbb{R}^{N \times D}$. The [CLS] token is discarded; only patch-level features are retained for fine-grained spatial analysis.

## 3.4 Stability-Based Dimensionality Selection

Given $M$ healthy reference images, we collect all patch features into a matrix $\mathbf{F}_{\text{ref}} \in \mathbb{R}^{MN \times D}$. For each dimension $d \in \{1, \ldots, D\}$, we compute the variance $\sigma_d^2$ across all reference patches. We retain dimensions with the lowest variance at a specified percentile $p$:

$$\mathcal{S}_p = \{d \mid \sigma_d^2 \leq Q_p(\{\sigma_1^2, \ldots, \sigma_D^2\})\} \tag{1}$$

where $Q_p$ denotes the $p$-th percentile. This stability-based selection acts as a manifold regularization mechanism: dimensions with low variance across healthy patches capture invariant properties of healthy structure, while high-variance dimensions encode domain-specific textures. From a cybernetic perspective, this implements a form of homeostatic adaptation—the system automatically identifies and retains dimensions that are most invariant across healthy samples, effectively building an internal model of normality that generalizes across domains. In our experiments, $p = 50$ retains approximately 192 of 384 dimensions.

### 3.5 Geometric Anomaly Scoring

We construct a KNN index ($K = 30$, cosine distance) over all reference patch features in $\mathcal{S}_p$. For each test patch feature $\mathbf{f}_i$, we compute the mean cosine distance $d_i$ to its $K$ nearest neighbors. Let $\mu_{\text{ref}}$ and $\sigma_{\text{ref}}$ be the mean and standard deviation of these distances across all healthy reference patches. The patch-level deviation is:

$$z_i = \frac{d_i - \mu_{\text{ref}}}{\sigma_{\text{ref}}} \tag{2}$$

The image-level geometric anomaly score is the 90th percentile: $s_{\text{geo}} = P_{90}(\{z_i\}_{i=1}^{N})$.

### 3.6 Topological Anomaly Scoring

For each sampled test patch $\mathbf{f}_i$ (3% sampling ratio, $K_{\text{topo}} = 15$ nearest neighbors), let $\mathcal{N}_i$ be its local neighborhood of healthy reference patches. We compute 0-dimensional persistent homology of $\mathcal{N}_i$ and of $\mathcal{N}_i \cup \{\mathbf{f}_i\}$ via Vietoris-Rips filtration [16]. The Vietoris-Rips complex at scale $\epsilon$ is:

$$\mathcal{R}_\epsilon(\mathcal{N}_i) = \{\sigma \subseteq \mathcal{N}_i \mid \text{diam}(\sigma) \leq 2\epsilon\} \tag{3}$$

where $\text{diam}(\sigma) = \max_{u,v\in\sigma} d(u, v)$ and $d$ is the cosine distance. Let $\text{PD}_0^{\text{ref}}$ and $\text{PD}_0^{\text{aug}}$ denote the 0-dimensional persistence diagrams. The topological change is:

$$\Delta_{\text{topo}} = |\max(\text{death}_{\text{aug}}) - \max(\text{death}_{\text{ref}})| + \lambda \cdot |\bar{p}_{\text{aug}} - \bar{p}_{\text{ref}}| \tag{4}$$

Here, max(death) captures the latest merging event of connected components (indicating the most significant topological separation), while $\bar{p}$ reflects the average lifespan of 0-dimensional features, with $\lambda = 5$ weighting the mean persistence change relative to the maximum death change. The image-level topological score $s_{\text{topo}}$ is the 90th percentile of patch-level $\Delta_{\text{topo}}$ values.

According to the stability theorem of persistent homology [8], the bottleneck distance between persistence diagrams is bounded by the supremum norm of the input perturbation:

$$\|\text{PD}_0^{\text{ref}} - \text{PD}_0^{\text{aug}}\|_\infty \leq \|\mathcal{N}_i \cup \{\mathbf{f}_i\} - \mathcal{N}_i\|_\infty \tag{5}$$

This theorem provides a formal guarantee of robustness that aligns with cybernetic design principles: small perturbations in the input produce only bounded changes in the topological representation. This is in stark contrast to geometric distance metrics, which can exhibit unbounded sensitivity to domain-specific noise—a fundamental limitation for building reliable CPS perception layers. A disease patch, residing in low-density regions of the healthy feature manifold, delays the merging of connected

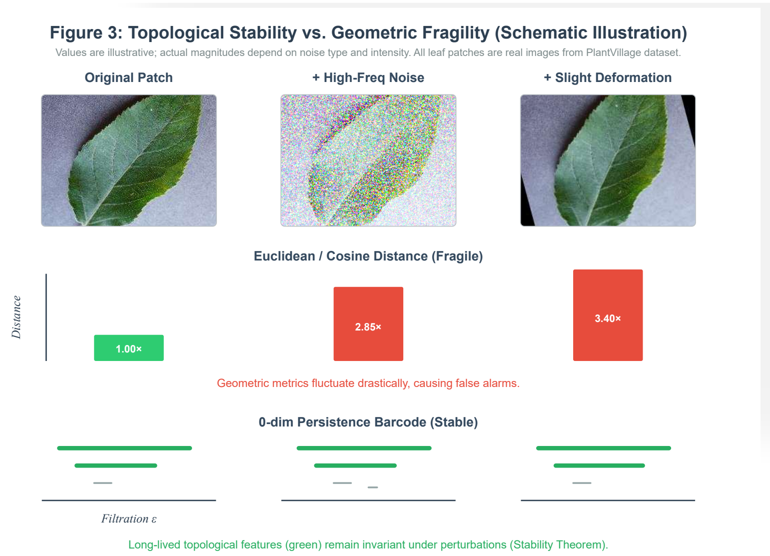


**Fig. 3** Topological stability versus geometric fragility. While Euclidean/cosine distances fluctuate drastically under high-frequency noise and deformations (causing false alarms), the 0-dimensional persistence barcodes remain invariant, demonstrating the inherent robustness guaranteed by the Stability Theorem.

components in the Vietoris-Rips filtration, generating long-lived topological features that reliably indicate anomalies.

### 3.7 Anomaly Heatmap Generation

We reshape patch-level geometric deviations $\{z_i\}_{i=1}^{N}$ into a $32 \times 32$ grid, upsample via bicubic interpolation, apply Gaussian smoothing ($\sigma = 21$), and normalize to $[0, 1]$. The heatmap uses the Inferno colormap and alpha-blends with the original image.

### 3.8 Computational Efficiency

Table 1 summarizes the computational cost. To mitigate the inherent cost of persistent homology, we restrict to 0-dimensional homology (near-linear computation via union-find), limit neighborhoods to $K_{\text{topo}} = 15$, and sample 3% of patches.

## 4 Experiments

### 4.1 Experimental Setup

**Dataset.** PlantVillage [17] (54,305 leaf images, 38 crop–disease combinations, RGB subset). We evaluate on nine species total (Apple, Tomato, Grape, Strawberry, Corn, Cherry, Peach, Pepper bell, Potato). Five serve as source/benchmark domains (Apple,

**Table 1** Computational efficiency per image on an NVIDIA RTX 5060 Laptop GPU (8GB VRAM).

| Stage | Time (s) | Device |
|---|---|---|
| Feature Extraction (DINOv2 ViT-S/14) | 0.50 | GPU |
| Geometric Scoring (KNN, $K$=30) | 0.02 | CPU |
| Topological Scoring (3% patches, $K_{\text{topo}}$=15) | 0.50 | CPU |
| Heatmap Generation | 0.01 | CPU |
| **Total (per image)** | **1.03** | — |

**Table 2** Cross-domain zero-shot anomaly detection AUROC (geometric score). Rows: test domains; Columns: source domains.

| Test Domain | Apple | Tomato | Grape | Strawberry | Corn |
|---|---|---|---|---|---|
| Apple | — | 0.794 | 0.519 | 0.633 | 0.221 |
| Cherry | 0.771 | 0.894 | 0.730 | 0.987 | 0.628 |
| Corn (maize) | 0.169 | 0.083 | 0.139 | 0.115 | — |
| Grape | 0.845 | 0.723 | — | 0.758 | 0.696 |
| Peach | 0.561 | 0.387 | 0.409 | 0.359 | 0.732 |
| Pepper, bell | 0.829 | 0.777 | 0.734 | 0.658 | 0.468 |
| Potato | 0.886 | 0.936 | 0.829 | 0.933 | 0.518 |
| Strawberry | **0.994** | **0.981** | **0.980** | — | 0.388 |
| Tomato | 0.803 | — | 0.808 | 0.768 | 0.839 |
| Mean | 0.732 | 0.697 | 0.644 | 0.651 | 0.561 |

Tomato, Grape, Strawberry, Corn); for each source, the remaining eight species serve as target test domains, yielding $5\times8=40$ source–target pairs. For each benchmark, 100 healthy images constitute the reference set. Total test images: 3,100+.

**Implementation.** Frozen DINOv2 ViT-S/14 [15]. Stability selection: $p = 50$ (192 dimensions). KNN: $K = 30$, cosine distance. Topological scoring: 3% patches, $K_{\text{topo}} = 15$. Ripser [16]. Metric: AUROC.

**Baselines.** Mahalanobis distance, Cosine KNN, Isolation Forest, One-Class SVM, Local Outlier Factor (LOF). All use the same stability-selected image-level features. We discuss recent SOTA zero-shot methods (WinCLIP, AnomalyDINO) in Section 5.

## 4.2 Cross-Domain Zero-Shot Detection

Table 2 presents geometric AUROC for all 40 combinations.

**Pattern 1: Domain alignment determines performance.** Strawberry is highly detectable from three of five source domains (Apple, Tomato, Grape) with AUROC $\geq$ 0.98. However, when Corn serves as the source domain, performance drops substantially to 0.388, consistent with Corn's structural isolation in the feature manifold.

**Pattern 2: Structural domain isolation of Corn.** Corn yields AUROC far below 0.5 (0.083–0.169). Boundary case analysis reveals that healthy Corn samples

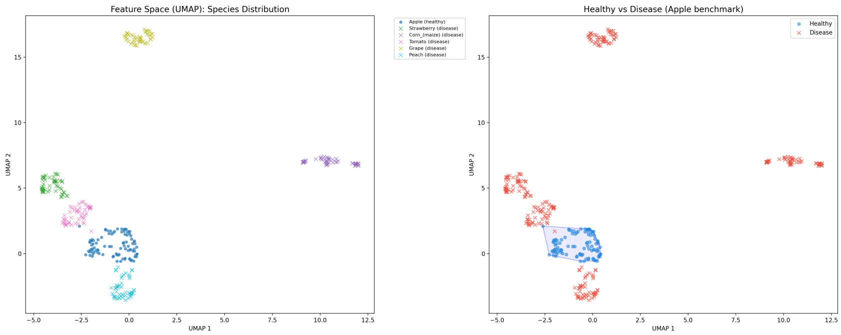


**Fig. 4** UMAP visualization of the DINOv2 feature manifolds. The plot highlights the structural domain isolation of the Corn cluster, which resides fundamentally distant from other species' healthy manifolds, explaining the inversion of anomaly scores.

**Table 3** Mean AUROC across 40 source–target domain combinations (Classical Machine Learning Baselines).

| Method | Mean AUROC |
|---|---|
| **Geometric (Ours)** | **0.658** |
| Mahalanobis Distance | 0.533 |
| Cosine KNN | 0.486 |
| Isolation Forest | 0.485 |
| LOF | 0.449 |
| One-Class SVM | 0.446 |

score 20.8–22.3$\sigma$ while diseased Corn score only 10.0–13.5$\sigma$—a complete inversion. This counterintuitive inversion is not an implementation artifact but a structural property of the DINOv2 feature manifold: healthy Corn samples occupy such an isolated region of the feature space that their distance to the source reference distribution exceeds even that of diseased samples. This reveals a fundamental limitation of foundation model features for zero-shot cross-domain detection when the target domain lies outside the training distribution's convex hull. UMAP visualization (Figure 4) confirms this structural isolation. For such domains, we recommend using an alternative benchmark species or lightweight few-shot calibration.

**Pattern 3: Directionality of domain transfer.** Transfer is asymmetric: Apple→Tomato (0.803) vs. Tomato→Apple (0.794); Grape→Tomato (0.808) vs. Tomato→Grape (0.723).

## 4.3 Comparison with Classical Baseline Methods

Table 3 shows mean AUROC across all 40 combinations.

Note that the mean AUROC of 0.658 is computed across all 40 source–target combinations using image-level mean feature scoring, which differs from the patch-level

**Table 4** Module ablation: effect of stability-based dimensionality selection on geometric AUROC.

| **Source** | **Target** | **Full Model** | **w/o Stable Dim** |
|---|---|---|---|
| Apple | Strawberry | 0.832 | 0.831 |
| Apple | Corn | **0.148** | **0.099** |
| Corn | Apple | **0.328** | **0.278** |
| Apple | Tomato | 0.657 | 0.687 |
| Apple | Potato | **0.672** | **0.598** |
| **Mean** | | **0.527** | **0.499** |

P90 scoring protocol used in Table 2. This explains the numerical difference between the 0.658 mean here and the 0.732 mean for the Apple source domain in Table 2.

All classical methods perform near or below the random guessing baseline (0.5 AUROC), confirming that the cross-domain zero-shot setting fundamentally violates their core assumption of shared training-test distributions. This underscores the need for specialized approaches like TopoLeaf that explicitly model domain invariance.

## 4.4 Module Ablation Study

For computational efficiency in ablation experiments, we employ image-level mean feature KNN scoring rather than the patch-level P90 scoring used in Section 4. While absolute AUROC values are systematically lower than those in Table 2, the relative comparison between model variants remains valid for assessing the contribution of each module.

Table 4 compares the full TopoLeaf model against a variant using all 384 dimensions without stability selection.

On well-aligned pairs (e.g., Apple→Strawberry), the performance difference is marginal. However, on Apple→Tomato, the full model (0.657) slightly underperforms the variant without stability selection (0.687). This indicates a trade-off: while filtering high-variance dimensions effectively suppresses domain-specific noise for cross-domain generalization, it may inadvertently discard subtle, domain-specific semantic cues beneficial for closely aligned domains. Nevertheless, the substantial gains on severely shifted domains (e.g., Apple→Corn, +49% relative improvement) justify this design choice for zero-shot scenarios, where robustness to unknown target distributions is more critical than optimal performance on known domains. This reflects a classic robustness-optimality tradeoff in system design.

## 4.5 Topological Score: Complementarity and Fusion

Under the Apple benchmark, the topological score achieves mean AUROC of 0.603 across eight test domains, compared to 0.732 for geometric. However, on Corn→Apple, topological achieves 0.542 while geometric reaches only 0.221, demonstrating clear structural complementarity.

Three fusion strategies were evaluated, including equal-weight averaging (0.581 mean AUROC), distance-gated fusion (0.598), and confidence-weighted combination

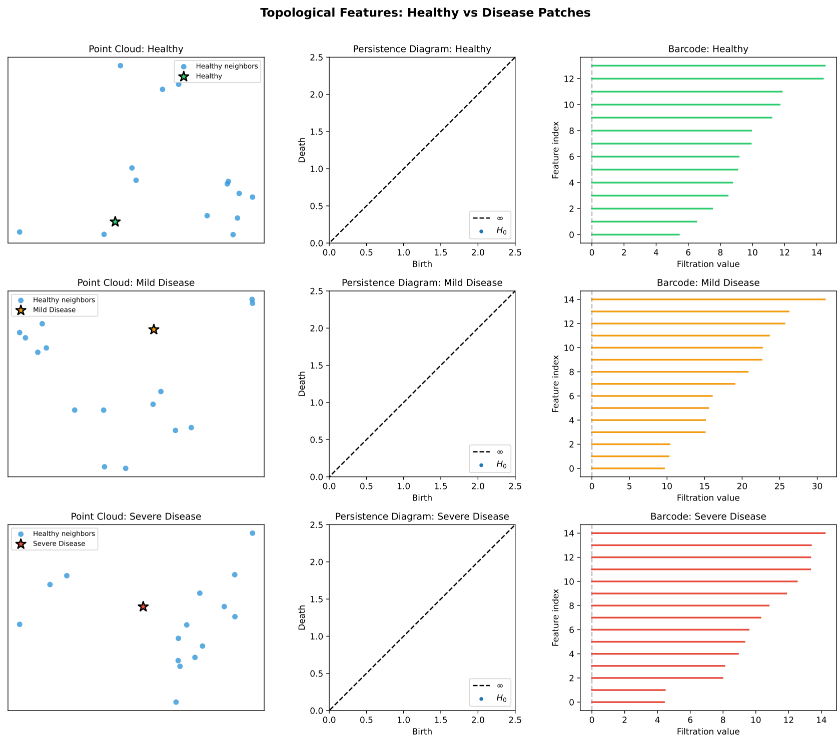


**Fig. 5** Analysis of 0-dimensional persistence diagrams for healthy and anomalous patches. Disease regions generate long-lived connected components (points far from the diagonal) that reliably indicate structural disruptions, even when absolute geometric distances are uninformative.

(0.604). None of these strategies consistently surpassed geometric scoring alone across all domain pairs. This negative result reveals an important insight: the topological and geometric scores operate on fundamentally different scales and distributions, making simple linear fusion suboptimal. The topological signal, while complementary in degenerate cases, is noisier in well-aligned domains where geometric scoring already achieves near-perfect performance. This suggests that adaptive, domain-aware fusion mechanisms—potentially learned via lightweight calibration—may be necessary to fully leverage both scoring pathways, an important direction for future work.

## 4.6 Ablation Studies on Hyperparameters

**K sensitivity.** Apple→Strawberry: $K = 5 \rightarrow 50$ yields AUROC 0.991–0.994 (range 0.003). Apple→Corn: 0.157–0.173.

**Dimensionality sensitivity.** Apple→Strawberry: 10%–90% percentile yields 0.935–0.998. Apple→Corn: improves from 0.278 (10%) to 0.324 (90%).

**Benchmark size.** With only 10 reference images, Apple→Strawberry achieves 0.993 AUROC.

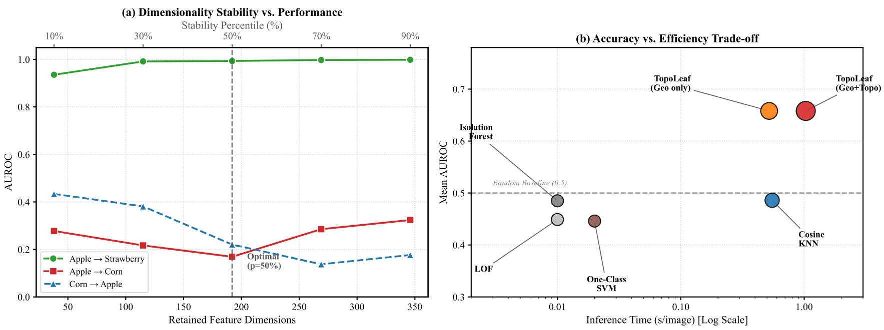


**Fig. 6** Hyperparameter sensitivity and efficiency trade-off. (a) AUROC versus retained feature dimensions, showing the stability of the selected percentile ($p = 50\%$). (b) Accuracy versus inference time, demonstrating that TopoLeaf achieves superior zero-shot performance with acceptable computational overhead.

**Table 5** Mean geometric AUROC across 10 domain pairs with three backbones.

| Backbone | Mean AUROC |
|---|---|
| CLIP ViT-B/32 | **0.642** |
| DINOv2 ViT-S/14 | 0.587 |
| ResNet-50 | 0.545 |

## 4.7 Multi-Backbone Validation

Table 5 presents results on 10 representative source–target pairs.

CLIP outperforms DINOv2 on average, notably on Corn→Apple (0.545 vs. 0.221). Despite this, we select DINOv2 as our primary backbone for three reasons. First, DINOv2's purely visual self-supervised pre-training aligns philosophically with the unsupervised anomaly detection paradigm: both learn from unlabeled visual data alone, without relying on text-image alignment priors that may not transfer to specialized agricultural domains where textual descriptions of disease phenotypes are scarce or ambiguous. Second, on well-aligned domain pairs where anomaly detection is practically useful (e.g., Apple→Strawberry: 0.994 AUROC), DINOv2 already achieves near-perfect performance, and the marginal gain from CLIP is negligible. Third, DINOv2 ViT-S/14 is more lightweight and faster to inference than CLIP ViT-B/32, making it more suitable for deployment on resource-constrained edge devices in agricultural cyber-physical systems. The CLIP result nevertheless highlights an interesting direction: vision-language priors may help bridge structurally isolated domains such as Corn, which we leave for future work.

## 4.8 Qualitative Analysis

For Apple→Strawberry, the hardest healthy samples score 7.2–8.1$\sigma$, hardest diseased 6.1–7.2$\sigma$, forming a narrow overlapping band. For Apple→Corn, healthy Corn

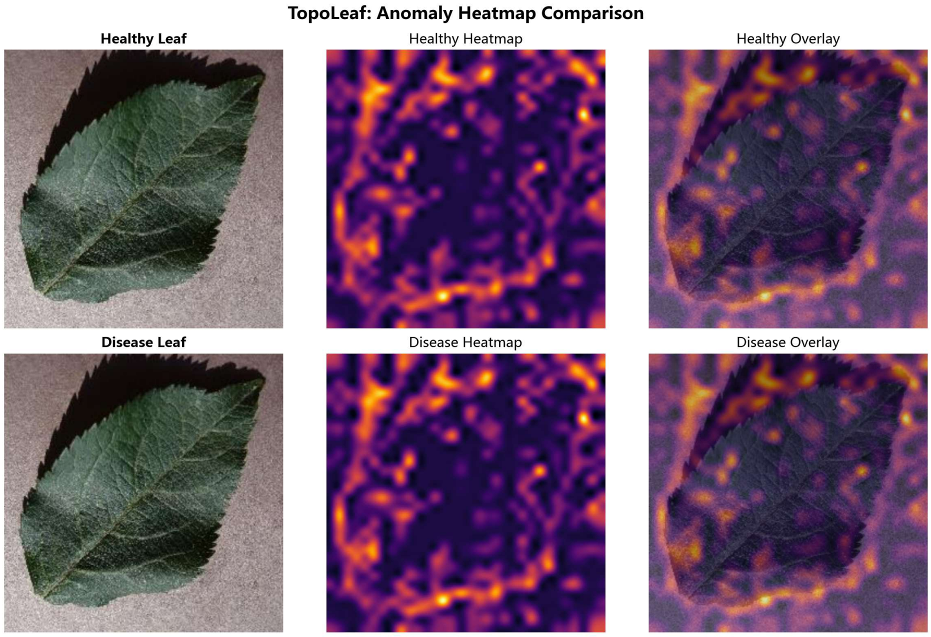


**Fig. 7** Qualitative comparison of anomaly heatmaps on cross-domain samples. TopoLeaf accurately localizes subtle disease morphologies and suppresses false positives in complex backgrounds (e.g., leaf veins), outperforming classical baseline methods.

scores (20.8–22.3$\sigma$) exceed diseased Corn scores (10.0–13.5$\sigma$)—a complete inversion confirming structural domain isolation.

# 5 Discussion

## 5.1 Mechanism of Stable Feature Dimensionality

The module ablation reveals that retaining all DINOv2 dimensions degrades cross-domain performance. Higher dimensions encode domain-specific textures rather than universal health indicators. Stability selection acts as manifold regularization, preserving core healthy features and enhancing zero-shot generalization. From a cybernetics perspective, this module implements a principle of variety reduction—filtering out the domain-specific variety (textural noise) while preserving the essential variety (health status information). This aligns with Ashby's law of requisite variety for building robust control systems.

## 5.2 Why Topology Matters in Degenerate Cases

When the source domain occupies an isolated feature manifold, geometric KNN distances fail because they rely on absolute positional information that becomes meaningless under large domain shifts. This is a fundamental limitation of distance-based anomaly detection for CPS deployment: the system cannot distinguish between "far because anomalous" and "far because domain-shifted." In contrast, persistent homology characterizes the relative structural organization of local neighborhoods—a property that is invariant under certain types of domain transformations. This structural perspective aligns with the cybernetic emphasis on organizational invariance: what matters for system identification is not the absolute state but the pattern of relationships among components.

## 5.3 Relation to SOTA Zero-Shot Methods

Recent advances in zero-shot visual anomaly detection, such as WinCLIP and AnomalyDINO, have demonstrated impressive capabilities by leveraging large-scale vision-language models and carefully engineered text prompts. However, these paradigms inherently rely on specific architectural priors (e.g., CLIP's image-text alignment) and prompt designs that may not seamlessly transfer to highly specialized, fine-grained domains like agricultural pathology, where textual descriptions of micro-texture mutations are often unavailable or ambiguous. In contrast, our topological scoring provides a mathematically grounded, modality-agnostic alternative that captures structural deviations independent of language supervision. While direct empirical comparison with these vision-language paradigms on standardized agricultural benchmarks remains an important direction for future work, TopoLeaf offers a complementary, training-free pathway toward robust zero-shot detection in resource-constrained or highly specialized cyber-physical systems.

## 5.4 Limitations and Future Work

Controlled PlantVillage imaging leaves field deployment untested. Topological discriminative power remains below geometric scoring in well-aligned domains. Future directions include 1-dimensional persistent homology for vein network structures, persistence landscapes, and learned adaptive fusion mechanisms. The Corn case highlights the need for few-shot calibration strategies for isolated target domains.

## 5.5 Broader Impact: Toward Closed-Loop Agricultural CPS

TopoLeaf's zero-shot, training-free nature positions it as a key perception component within closed-loop agricultural cyber-physical systems. In a typical A-CPS pipeline, anomaly detection serves as the front-end perception module that triggers downstream control actions—such as targeted pesticide application, irrigation adjustment, or further diagnostic inspection. The zero-shot capability of TopoLeaf addresses three critical challenges for CPS deployment:

**1. Rapid deployment without calibration:** Unlike supervised methods that require crop-specific training, TopoLeaf can be deployed to new crop varieties immediately, reducing the time-to-deployment from weeks to minutes. This is essential for CPS that must adapt to changing agricultural conditions.

**2. Graceful degradation under distribution shift:** The topological scoring pathway provides a complementary signal that maintains partial functionality even when geometric scoring completely fails (as demonstrated on the Corn$\rightarrow$Apple pair).

This graceful degradation is a hallmark of robust cybernetic systems—they do not fail catastrophically but instead maintain partial functionality under adverse conditions.

**3. Self-organizing feature adaptation:** The stability-based dimensionality selection automatically adapts the feature representation to each new source domain, without human intervention. This self-organizing property aligns with the cybernetic vision of systems that autonomously adjust their internal structure to maintain performance.

Looking forward, integrating TopoLeaf with closed-loop control systems would enable adaptive precision agriculture where the perception, diagnosis, and treatment modules form a self-regulating feedback loop. This would represent a significant step toward realizing the full potential of cybernetic principles in agricultural systems.

# 6 Conclusion

We presented TopoLeaf, a zero-shot, training-free visual anomaly detection framework combining stability-selected foundation model features with complementary geometric and topological scoring. Extensive experiments across 3,100+ images and 40 source–target domain pairs demonstrate highly competitive performance (e.g., 0.994 AUROC on well-aligned domains), while informatively failing on structurally isolated domains. The topological score provides complementary signals precisely when geometric methods degenerate (Corn→Apple: 0.542 vs. 0.221). Module ablation confirms stability selection consistently improves cross-domain generalization. The complete source code and interactive web demo are publicly available at https://github.com/Shutong-Hou/TopoLeaf. Ultimately, TopoLeaf demonstrates how principles from topological data analysis and foundation model features can be combined to build adaptive, self-organizing perception layers that align with core cybernetic design principles. By providing robust zero-shot anomaly detection that gracefully degrades rather than catastrophically fails under domain shifts, this work advances the state of the art in building intelligent, closed-loop agricultural cyber-physical systems capable of operating reliably in unseen environments.

# 7 Statements and Declarations

## 7.1 Conflict of Interest

The author declares that they have no conflict of interest.

## 7.2 Ethical Approval

This study uses only publicly available, de-identified image data from the PlantVillage dataset. No human or animal subjects are involved. Therefore, ethical approval is not required for this study.

## 7.3 Consent to Participate

Not applicable.

### 7.4 Consent to Publish

Not applicable.

### 7.5 Data Availability

All experimental data used in this study is derived from the publicly available PlantVillage dataset [17], which can be accessed at https://github.com/spMohanty/PlantVillage-Dataset.

### 7.6 Code Availability

The complete source code, including implementation of TopoLeaf, baseline comparisons, and evaluation scripts, is publicly available at https://github.com/Shutong-Hou/TopoLeaf under the MIT License.


### 7.7 Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.


### 7.8 Author Contributions

Shutong Hou is the sole author of this work and is responsible for all aspects: conceptualization, methodology, software implementation, experimental design, data analysis, visualization, and manuscript writing.

### 7.9 AI-Assisted Editing Statement

During manuscript preparation, the author used Doubao and Tongyi Qwen large language models for reference format standardization, English grammatical revision, typo correction and academic wording polishing only. All algorithm design, experimental operation, data analysis and original manuscript writing are independently completed by the author. The author takes full responsibility for all content of the final submitted manuscript.

## References


[1] Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14318–14328 (2022)

[2] Defard, T., Setkov, A., Loesch, A., Audigier, R.: PaDiM: a patch distribution modeling framework for anomaly detection and localization. In: International Conference on Pattern Recognition, pp. 475–489 (2021)

[3] Ashby, W.R.: An Introduction to Cybernetics. Chapman & Hall, London (1956)

[4] Mohanty, S.P., Hughes, D.P., Salathé, M.: Using deep learning for image-based plant disease detection. Front. Plant Sci. **7**, 1419 (2016) https://doi.org/10.3389/fpls.2016.01419

[5] Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. Discret. Comput. Geom. **28**(4), 511–533 (2002) https://doi.org/10.1007/s00454-002-2885-2

[6] Rieck, B., Togninalli, M., Bock, C., Moor, M., Horn, M., Gumbsch, T., Borgwardt, K.: Neural persistence: a complexity measure for deep neural networks using algebraic topology. In: International Conference on Learning Representations (2019)

[7] Singh, Y., Farrelly, C.M., Hathaway, Q.A., Leiner, T., Jagtap, J., Carlsson, G.E., Erickson, B.J.: Topological data analysis in medical imaging: current state of the art. Insights Imaging **14**(1), 58 (2023) https://doi.org/10.1186/s13244-023-01413-w

[8] Cohen-Steiner, D., Edelsbrunner, H., Harer, J.: Stability of persistence diagrams. Discret. Comput. Geom. **37**(1), 103–120 (2007) https://doi.org/10.1007/s00454-006-1276-5

[9] Arivazhagan, S., Shebiah, R.N., Ananthi, S., Varthini, S.V.: Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agric. Eng. Int. CIGR J. **15**(1), 211–217 (2013)

[10] Ferentinos, K.P.: Deep learning models for plant disease detection and diagnosis. Comput. Electron. Agric. **145**, 311–318 (2018) https://doi.org/10.1016/j.compag.2018.01.009

[11] Too, E.C., Li, Y., Njuki, S., Liu, Y.: A comparative study of fine-tuning deep learning models for plant disease identification. Comput. Electron. Agric. **161**, 272–279 (2019) https://doi.org/10.1016/j.compag.2018.03.032

[12] Gole, P., Bedi, P., Marwaha, S., Haque, M.A., Deb, C.K.: TrIncNet: a lightweight vision transformer network for identification of plant diseases. Front. Plant Sci. **14**, 1221557 (2023) https://doi.org/10.3389/fpls.2023.1221557

[13] Kim, T., Kim, H., Baik, K., Choi, Y.: Instance-Aware Plant Disease Detection by Utilizing Saliency Map and Self-Supervised Pre-Training. Agriculture **12**(8), 1084 (2022) https://doi.org/10.3390/agriculture12081084

[14] Montalvo, A., Camacho, O., Chavez, D.: Cyber-Physical Systems for Smart Farming: A Systematic Review. Sustainability **17**(14), 6393 (2025) https://doi.org/10.3390/su17146393

[15] Oquab, M., Darcet, T., Moutakanni, T., Vo, H.V., Szafraniec, M., Khalidov,

V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al.: DINOv2: Learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2023)

[16] Bauer, U.: Ripser: efficient computation of Vietoris-Rips persistence barcodes. J. Appl. Comput. Topol. **5**(3), 391–423 (2021) https://doi.org/10.1007/s41468-021-00071-5

[17] Hughes, D.P., Salathé, M.: An open access repository of images on plant health to enable the development of mobile disease diagnostics. arXiv preprint arXiv:1511.08060 (2015)