How Much Source Geometry to Trust? Adaptive Spiked Supports for Open-Set Domain Generalization

Anonymous authors
Paper under double-blind review

Abstract

Open-set domain generalization (OSDG) requires acceptance regions to be loose enough to accommodate domain-shifted known samples yet compact enough to reject unknowns. Relying more on source-domain class geometry makes these regions more compact but less tolerant of domain shift, and since the reliability of that geometry varies across classes and source-domain configurations, no fixed degree of reliance suits all cases. This raises a central question: how much source geometry should we trust? We propose Adaptive Spiked Support (AdaSpike), which adapts this trust through cross-domain geometric consistency. Our key insight, which our robust OSDG analysis formalizes, is that greater source consistency suggests more stable target geometry and warrants greater trust, while greater discrepancy suggests more uncertainty and warrants less. Specifically, we show that the optimal acceptance region under a robust risk bound has a closed-form spiked structure, which preserves anisotropic scales along the k leading source directions and assigns a shared isotropic scale to the rest, interpolating between isotropic and fully anisotropic supports. AdaSpike determines k from its cross-domain discrepancy, having no need of target data or manual tuning. The supports are learned with objectives formulated from an OSDG risk bound, and a single support score serves both classification and rejection. Across three OSDG benchmarks and two backbones, AdaSpike outperforms existing methods in all six settings, improving average OSCR by 2.16 points over state-of-the-art methods.

1 Introduction

Open-set domain generalization (OSDG) trains models on multiple source domains to recognize known classes and reject unknowns in an unseen target domain. This requires acceptance regions to accommodate domain-shifted known samples while remaining compact enough to reject unknowns. Source class geometry helps constrain these regions, but following it too closely may reject shifted known samples, while relaxing it too broadly may admit unknowns. Its reliability also varies across source configurations, raising a central question in OSDG: *How much source geometry should we trust to adaptively infer generalizable and compact acceptance regions?*

Geometry-based open-set recognition (OSR) methods exploit directional variation within known classes to construct anisotropic acceptance regions (Zou et al., 2026), but their direction-dependent constraints, fitted to a single domain only, are unreliable under domain shift. Domain generalization (DG) methods use variation across source domains as evidence of uncertainty beyond the observed domains (Wang et al., 2024), but only apply it to closed-set robustness. Current OSDG methods exploit multiple source domains to improve representations and decision boundaries (Wang et al., 2026b), but do not explicitly account for how uncertainty under domain shift should influence the amount of source geometry retained in acceptance regions.

To answer this question, we propose Adaptive Spiked Support (AdaSpike), a theoretically motivated framework that uses cross-domain geometric discrepancy to determine how much directional source structure to retain under domain shift. Our intuition is that, for each class, greater source discrepancy suggests greater uncertainty about whether its geometry will remain stable in an unseen target domain. We therefore place less trust in source geometry and retain less of its directional structure.

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.

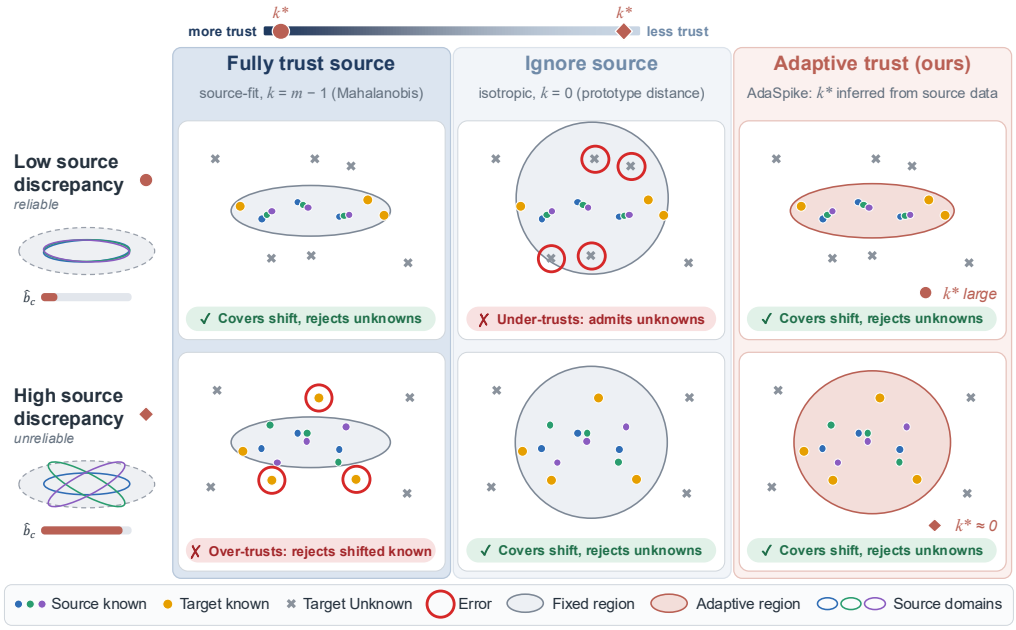


Figure 1: **Why OSDG calls for adaptive reliance on source geometry.** For a fixed class c, different source-domain configurations can exhibit different levels of source geometric discrepancy $\hat{b}_c$. A fixed acceptance region therefore cannot suit all cases: when the source geometry is consistent, under-trusting it yields overly loose acceptance regions that admit unknowns, whereas under large discrepancy, over-trusting it yields overly restrictive acceptance regions that reject shifted known samples. AdaSpike adapts the degree of reliance through the trusted spike dimension k, retaining more source geometry under low discrepancy and less under high discrepancy.

Lower discrepancy gives us greater confidence that the target geometry will resemble the source geometry, warranting greater trust and preservation of more source structure (Fig 1).

For each known class, we model a geometric support describing where its samples are expected to lie, which determines its acceptance region. We show that minimizing a robust upper bound on OSDG risk yields an optimal support with a closed-form spiked structure. It retains distinct scales along the k leading source directions and assigns a shared isotropic scale to the rest. The trusted spike dimension k thus measures how much directional source structure is preserved, with isotropic and fully anisotropic supports as limiting cases. Motivated by the spike structure, we introduce a spectral criterion for selecting k based on class-wise geometric discrepancy across source domains. This enables us to adaptively determine k directly from source data. This matters, both theoretically and practically, because no fixed k suits all discrepancy regimes (Fig. 1), the best k varies across unseen target domains, and our k requires neither target information nor manual tuning. We then formulate three surrogate learning objectives motivated by an OSDG risk bound, each targeting one of its error terms, so that training jointly addresses known-class classification, false rejection of domain-shifted known samples, and false acceptance of unknowns. A single acceptance region score serves both classification and rejection.

Our main contributions are summarized as follows.

- We reveal that robust optimal class support for OSDG has a closed-form spiked structure, which selectively preserves the leading directions of the source geometry and shares a common isotropic scale across the rest. This structure recovers isotropic and fully anisotropic supports as limiting cases, and its trusted spike dimension k provides an explicit measure of how much source geometry should be trusted.

- We propose a source-only spectral rule that selects k from source discrepancy before training, requiring neither target data nor manual tuning. The support thus adapts to how much the source geometry varies across domains, rather than being fixed by design.

- We develop a practical framework AdaSpike that jointly learns representations and support scales with objectives motivated by an OSDG risk bound, using a single support score for both classification and rejection. AdaSpike achieves the best closed-set accuracy, H-score, and OSCR in all six dataset–backbone settings, improving average OSCR by 2.94 points with ResNet-18 and 1.37 points with CLIP.

2 Theoretical Analysis

Problem Setup. In open-set domain generalization (OSDG), we are given S labeled source domains $\mathcal{S} = \{D_1, \dots, D_S\}$, where $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$. Let $\mathcal{Y}^*$ be the universal label space. All source domains share a common known label space $\mathcal{K} \subseteq \mathcal{Y}^*$, and the remaining labels $\mathcal{U} = \mathcal{Y}^* \setminus \mathcal{K}$ are unknown classes, which are unobserved during training. The target domain D_T is unavailable during training and may contain samples from both $\mathcal{K}$ and $\mathcal{U}$. The goal is to learn a predictor $h : \mathcal{X} \to \mathcal{K} \cup \{y_{\text{unk}}\}$ that generalizes to unseen domains, where y_{unk} denotes a single rejection label for samples predicted as belonging to any class in $\mathcal{U}$. Let $\phi : \mathcal{X} \to \mathbb{S}^{m-1}$ be an ℓ_2-normalized representation map. Following classical work on classification with a reject option (Chow, 1970; Scheirer et al., 2013), we build the open-set predictor from a closed-set classifier $f : \mathcal{X} \to \mathcal{K}$ and an acceptance region $A \subseteq \mathbb{S}^{m-1}$:

$$h_{A,f}(x) = \begin{cases} f(x), & \phi(x) \in A, \\ y_{\text{unk}}, & \phi(x) \notin A, \end{cases}$$

where f handles known-class classification, while A determines whether a sample is accepted as known or rejected as unknown.

Following the generative view of domain generalization (Blanchard et al., 2021), each domain is associated with a joint distribution $P \in \mathfrak{P}_{\mathcal{X}\times\mathcal{Y}^*}$ drawn from a meta-distribution μ over $\mathfrak{P}_{\mathcal{X}\times\mathcal{Y}^*}$. The source-domain distributions $P_1, \dots, P_S \overset{\text{iid}}{\sim} \mu$, and, conditional on P_s, the observed source samples are drawn independently from $P_s(\cdot \mid Y^* \in \mathcal{K})$. We assume $P(Y^* \in \mathcal{K}) > 0$ for μ-almost every P. We write $P \sim \mu$ for a generic domain distribution and $Z = \phi(X)$ for the feature representation.

Definition 1 (OSDG Risk)**.** The risk of $h_{A,f}$ on a randomly drawn unseen domain $P \sim \mu$ is

$$\mathcal{E}(h_{A,f}) = \mathbb{E}_{P\sim\mu}\mathbb{E}_{(X,Y^*)\sim P}\left[\ell_{\text{os}}\big(h_{A,f}(X), Y^*\big)\right], \tag{1}$$

where

$$\ell_{\text{os}}(\hat{y}, y) = \mathbf{1}\{y \in \mathcal{K}, \hat{y} \neq y\} + \mathbf{1}\{y \in \mathcal{U}, \hat{y} \neq y_{\text{unk}}\}.$$

Let $\pi_{\mathcal{K}}(P) = P(Y^* \in \mathcal{K})$ and $\pi_{\mathcal{U}}(P) = P(Y^* \in \mathcal{U})$, and let $P^P_{Z_{\mathcal{K}}}$ and $P^P_{Z_{\mathcal{U}}}$ denote the feature distributions conditioned on known and unknown labels, respectively. Then the OSDG risk decomposes as

$$\mathcal{E}(h_{A,f}) = \underbrace{\mathbb{E}_{P\sim\mu}\left[\pi_{\mathcal{K}}(P) R^P_{\text{mf}}(A, f)\right]}_{B_{\text{cls}}} + \underbrace{\mathbb{E}_{P\sim\mu}\left[\pi_{\mathcal{K}}(P) P^P_{Z_{\mathcal{K}}}(A^c)\right]}_{R_{\text{FR}}(A)} + \underbrace{\mathbb{E}_{P\sim\mu}\left[\pi_{\mathcal{U}}(P) P^P_{Z_{\mathcal{U}}}(A)\right]}_{R_{\text{FA}}(A)}, \tag{2}$$

where $R^P_{\text{mf}}(A, f) = P(Z \in A, f(X) \neq Y^* \mid Y^* \in \mathcal{K})$. The three terms correspond to misclassification of accepted known samples, false rejection of known samples, and false acceptance of unknown samples. Hence, enlarging A reduces R_{FR} but increases R_{FA}.

Since unknown classes are unobserved during training, their contribution $R_{\text{FA}}(A)$ cannot be estimated directly. We therefore impose a regularity condition on the aggregate unknown measure. Define $\bar{P}_{\mathcal{U}}(B) := \mathbb{E}_{P\sim\mu}\left[\pi_{\mathcal{U}}(P)\, P^P_{Z_{\mathcal{U}}}(B)\right], B \subseteq \mathbb{S}^{m-1}$, so that $R_{\text{FA}}(A) = \bar{P}_{\mathcal{U}}(A)$.

Assumption 1 (Bounded Aggregate Unknown Density)**.** Let ν denote the spherical surface measure on $\mathbb{S}^{m-1}$. Assume that $\bar{P}_{\mathcal{U}} \ll \nu$ and that there exists a constant $C_{\mathcal{U}} < \infty$ such that

$$\frac{d\bar{P}_{\mathcal{U}}}{d\nu}(z) \leq C_{\mathcal{U}}, \qquad \text{for } \nu\text{-almost every } z \in \mathbb{S}^{m-1}.$$

Definition 2 (Spherical Anisotropic Class Support)**.** For each known class $c \in \mathcal{K}$, let $p_c \in \mathbb{S}^{m-1}$ be its prototype and define the tangent displacement

$$\xi_c(z) = \text{Log}_{p_c}(z) = \frac{\theta_c}{\sin\theta_c}\left(z - \cos\theta_c\, p_c\right), \qquad \theta_c = \arccos(p_c^\top z).$$

Let $M_c \succ 0$ be a symmetric positive-definite support matrix. We define

$$S_c(z) = \xi_c(z)^\top M_c^{-1} \xi_c(z), \qquad A_c = \{z \in \mathbb{S}^{m-1} : S_c(z) \le \rho\}, \qquad A = \bigcup_{c \in \mathcal{K}} A_c,$$

where $\rho > 0$ is the acceptance threshold.

The eigenvectors and eigenvalues of M_c determine the principal directions and squared extents of the class support, respectively. For the theoretical analysis, let $\bar{P} = \mathbb{E}_{P \sim \mu}[P]$ and define the class-c tangent-space second moment

$$T_c = \mathbb{E}_{(Z, Y^*) \sim \bar{P}} \left[\xi_c(Z) \xi_c(Z)^\top \mid Y^* = c \right] \succ 0.$$

Thus, T_c characterizes the population class geometry, whereas M_c parameterizes the corresponding acceptance support.

Domain generalization typically requires restricting unseen-domain shift, often through bounded discrepancies or uncertainty sets around source domains (Albuquerque et al., 2019; Zhou et al., 2023; Dai et al., 2023). We adopt this principle for class-support geometry by assuming bounded cross-domain variation in each class's tangent-space second moment.

Assumption 2 (Bounded Cross-domain Second-moment Variation)**.** For a domain distribution P, let $T_c^P = \mathbb{E}_{Z \sim P(\cdot \mid Y^* = c)} \left[\xi_c(Z) \xi_c(Z)^\top \right]$ denote the class-c tangent-space second moment under P. For each known class $c \in \mathcal{K}$, there exists a constant $b_c < \infty$ such that

$$\| T_c^P - T_c^{P'} \|_* \le b_c$$

for any two plausible domain distributions P and P', where $\| \cdot \|_*$ denotes the nuclear norm.

Proposition 1 (High-probability Second-moment Uncertainty Set)**.** *Let $\widehat{T}_c$ be any convex combination of the empirical source-domain second moments $\{\widehat{T}_c^{(s)}\}_{s=1}^S$.*

Under Assumption 2, for any $c \in \mathcal{K}$ and $\delta \in (0, 1)$, there exists $\varepsilon_c > 0$ such that, with probability at least $1 - \delta$,

$$T_c \in \mathcal{T}_c(\varepsilon_c) := \left\{ T \succeq 0 : \| T - \widehat{T}_c \|_* \le \varepsilon_c \right\}. \tag{3}$$

Details are provided in Appendix A.

The proposition formalizes the uncertainty of transferring source geometry to an unseen domain: the true class geometry T_c need not equal its source estimate $\widehat{T}_c$, but is constrained to lie within an uncertainty set around it. This uncertainty set is the basis of our robust support design in Theorem 1, which determines how much of the source geometry should be retained when constructing the acceptance region.

Theorem 1 (Robust Optimal Support)**.** *Under Assumption 1,*

$$R_{\mathrm{FR}}(A) + R_{\mathrm{FA}}(A) \le \sum_{c \in \mathcal{K}} \left[\alpha_c \operatorname{tr}(T_c M_c^{-1}) + \beta_c \det(M_c)^{1/2} \right]. \tag{4}$$

Combining this bound with the uncertainty set $\mathcal{T}_c(\varepsilon_c)$, the robust optimal support has the form

$$M_c^\star = t_c^\star \tau_c I + t_c^\star \sum_{j : \widehat{\gamma}_{c,j} > \tau_c} (\widehat{\gamma}_{c,j} - \tau_c) q_{c,j} q_{c,j}^\top = \sigma_c^{2\star} I + U_c^\star \Lambda_c^\star U_c^{\star\top}, \tag{5}$$

where $\sum_{j=1}^{m-1} (\tau_c - \widehat{\gamma}_{c,j})_+ = \varepsilon_c$. Thus the anisotropic dimension is $k_c^\star = \#\{j : \widehat{\gamma}_{c,j} > \tau_c\}$. The closed-form scale parameters are given in Appendix A.

The robust optimum balances covering plausible domain shifted known samples against limiting the support volume exposed to unknowns. Its spectral form shows that source directions with $\widehat{\gamma}_{c,j} > \tau_c$ retain their direction-dependent scales, whereas the remaining directions are collapsed to the common isotropic scale $\sigma_c^{2\star}$. Thus, $k_c^\star$ is the number of source directions whose anisotropic structure is retained under the estimated uncertainty. As ε_c increases, the floor τ_c rises and fewer directions remain above it, reducing $k_c^\star$ and making the support more isotropic. Conversely, lower uncertainty preserves more source geometry through a larger $k_c^\star$.

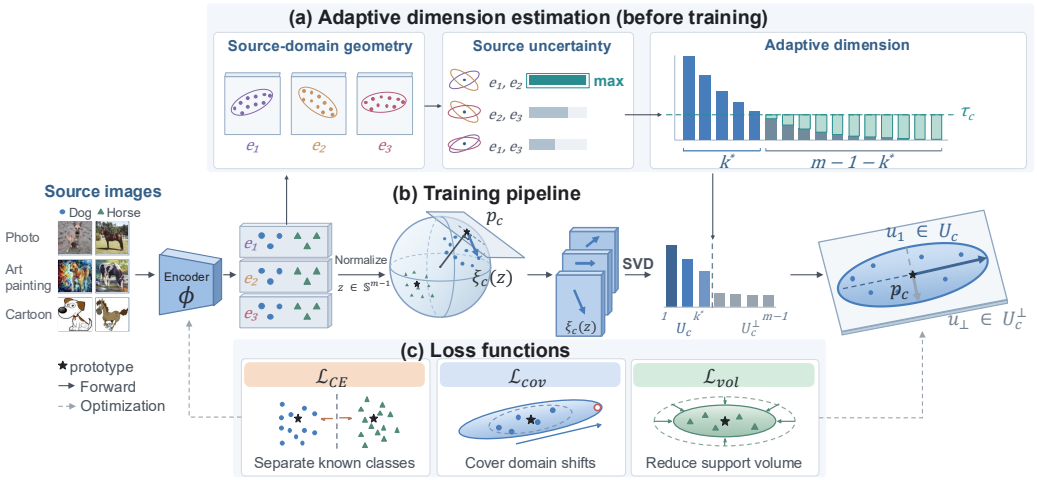


Figure 2: **Overview of the proposed method.**(a) Source discrepancies are used to adaptively select the dimension k^* with threshold τ_c. (b) Source features are mapped to the tangent space via $\xi_c(z) = \mathrm{Log}_{p_c}(z)$, and SVD yields the retained subspace U_c and its complement $U_c^\perp$ the class-wise spike acceptance region. (c) Three objectives jointly shape the class acceptance regions. Here, e_1, e_2, e_3 denote source domains.

Theorem 2 (Finite-Sample OSDG Risk Bound). *Let $N_c = \sum_{s=1}^{S}\sum_{i=1}^{N_s} \mathbf{1}\{Y_i^s = c\}$ denote the number of source samples from class c. Suppose Assumptions 1 and 2 hold, and let $A^\star$ be the robust acceptance region induced by Theorem 1. Then, with probability at least $1-\delta$, for every $f \in \mathcal{F}$,*

$$\mathcal{E}(h_{A^\star,f}) \le \underbrace{\frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s}\mathbf{1}\{f(X_i^s)\neq Y_i^s\}}_{\widehat{B}_{\mathrm{cls}}} + \underbrace{\sum_{c\in\mathcal{K}}\frac{\alpha_c}{N_c}\sum_{i:y_i=c} S_c^\star(z_i)}_{\widehat{B}_{\mathrm{cov}}} + \underbrace{\sum_{c\in\mathcal{K}}\beta_c(\sigma_c^{2\star})^{\frac{m-1-k_c}{2}}\prod_{j=1}^{k_c}(\sigma_c^{2\star}+\lambda_{c,j}^\star)^{1/2}}_{\widehat{B}_{\mathrm{vol}}} + B_{\mathrm{aux}}, \tag{6}$$

where $S_c^\star(z) = \xi_c(z)^\top (M_c^\star)^{-1}\xi_c(z)$, and B_{aux} collects the remaining finite-sample robustness and generalization terms.

The first three terms directly motivate our three learning objectives. The remaining uncertainty-dependent term instead guides the shape of the acceptance region by determining how much anisotropic source geometry should be retained. All explicit forms are provided in Appendix A.

3 Method

Guided by Theorem 1, we estimate the anisotropic dimension k from source domain before training. We then learn spiked class supports using objectives motivated by Theorem 2 (Fig. 2). At inference, the learned support scores serve both known-class prediction and unknown rejection.

3.1 Adaptive Anisotropic Dimension Estimation

Before training, we estimate the support structure from source data only. We ℓ_2-normalize source features, construct each class prototype p_c from the source-domain class means, and map features to its tangent space using the logarithmic map $\xi_c(z)$ defined in Sec. 2. For each source domain s, we estimate the class geometry by the tangent space second moment $\widehat{T}_c^{(s)}$, and quantify the source uncertainty of class c by the largest pairwise discrepancy between these domain-specific geometries:

$$\widehat{T}_c^{(s)} = \frac{1}{N_c^{(s)}}\sum_{i:y_i=c,\, e_i=s}\xi_c(z_i)\xi_c(z_i)^\top, \qquad \widehat{b}_c = \max_{s\neq s'}\left\|\widehat{T}_c^{(s)} - \widehat{T}_c^{(s')}\right\|_*.$$

Let $\widehat{T}_c = \frac{1}{|\mathcal{E}_s|}\sum_s \widehat{T}_c^{(s)}$ have eigenvalues $\gamma_{c,1} \geq \cdots \geq \gamma_{c,d}$. Motivated by Theorem 1, we set $\epsilon_c = \widehat{b}_c$ and determine τ_c from $\sum_j(\tau_c - \gamma_{c,j})_+ = \epsilon_c$. The numbers of eigenvalues above these floors determine the trusted spike dimension k used throughout training and inference.

3.2 Learning Spiked Class Supports

Spiked Support Construction. Given the adaptively estimated dimension k from source domain, we construct a spiked support for each known class. As shown in Fig. 2, we maintain a per-class memory bank of tangent displacements $\xi_c(z)$ and recompute the leading k singular directions $U_c = [u_{c,1}, \ldots, u_{c,k}]$ by SVD every 10 steps. The class prototype p_c is updated by exponential moving average (EMA). Both p_c and U_c are treated as stop-gradient quantities and are not optimized by backpropagation. In contrast, the residual scale σ_c^2 and directional scales $\{\lambda_{c,j}\}_{j=1}^k$ are learnable parameters optimized by gradient descent. Thus, U_c determines the empirical orientation of the anisotropic component, while its extent along each direction is learned during training.

The support matrix takes the form

$$M_c = \sigma_c^2 I + U_c \Lambda_c U_c^\top, \qquad \Lambda_c = \mathrm{diag}(\lambda_{c,1}, \ldots, \lambda_{c,k}), \quad \lambda_{c,j} \geq 0,\ j = 1, \ldots, k,$$

where the adaptive dimension k determines how much directional structure is retained. The support score is

$$S_c(z) = \sum_{j=1}^{k} \frac{\langle \xi_c(z), u_{c,j}\rangle^2}{\sigma_c^2 + \lambda_{c,j}} + \frac{\|P_{\perp,c}\xi_c(z)\|_2^2}{\sigma_c^2}, \qquad P_{\perp,c} = I - U_c U_c^\top,$$

where $\sigma_c^2 > 0$ is the common scale of the residual isotropic component, while $\sigma_c^2 + \lambda_{c,j}$ determines the support extent along the j-th retained anisotropic direction. Thus, deviations along directions supported by the source geometry can be assigned distinct scales, whereas all remaining directions share the same residual scale.

Learning Objectives. The first term in Theorem 2 requires accurate discrimination among known classes. We therefore use a prototype cross-entropy loss to encourage the representations of different known classes to be well separated:

$$\mathcal{L}_{\mathrm{CE}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{\exp(-S_{y_i}(z_i))}{\sum_{c\in\mathcal{K}} \exp(-S_c(z_i))},$$

where $-S_c(z)$ is used as the classification logit. This objective updates only the representation, with gradients to the support parameters stopped. Otherwise, since $-S_c(z)$ serves as the classification logit, increasing the support scale of the ground-truth class can reduce $\mathcal{L}_{\mathrm{CE}}$ even without changing the feature representation. The support geometry is therefore optimized only by $\mathcal{L}_{\mathrm{cov}}$ and $\mathcal{L}_{\mathrm{vol}}$.

Motivated by the coverage term in Theorem 2, we use leave-one-domain-out (LODO) estimation to expose the support to domain shift during training. For each source domain $e \in \mathcal{E}_s$, we construct the class supports from $\mathcal{E}_s \setminus \{e\}$, yielding $S_c^{(-e)}$, and evaluate them on the held-out samples $\mathcal{I}_e$. We optimize

$$\mathcal{L}_{\mathrm{cov}} = \frac{1}{N_v}\sum_{e\in\mathcal{E}_s}\sum_{i\in\mathcal{I}_e}\left[S_{y_i}^{(-e)}(z_i) - \rho\right]_+^2.$$

Since rejection depends on whether a known sample exceeds the support boundary, the squared hinge acts only on held-out samples that violate the boundary, avoiding unnecessary expansion for samples that are already accepted.

The volume term B_{vol} in Theorem 2 motivates penalizing the extent of the learned class supports. Replacing the optimal support parameters in the bound with the current learnable parameters, we use the class-wise log-volume surrogate

$$\mathcal{L}_{\mathrm{vol}} = \frac{1}{|\mathcal{K}|}\sum_{c\in\mathcal{K}}\left[\frac{m-1-k}{2}\log\sigma_c^2 + \frac{1}{2}\sum_{j=1}^{k}\log(\sigma_c^2 + \lambda_{c,j})\right].$$

As the support volume is multiplicative across directions, we optimize its logarithm rather than using an additive penalty on the scales. The volume term penalizes unnecessary support extent, while the coverage objective counteracts this compression along directions required to accommodate shifted known samples. Their interaction therefore yields anisotropic support contraction: directions needed for cross-domain coverage remain extended, whereas unsupported extent is compressed to limit unknown-class acceptance.

Overall Training Objective. The overall training objective is

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \lambda_{\text{cov}}\mathcal{L}_{\text{cov}} + \lambda_{\text{vol}}\mathcal{L}_{\text{vol}}, \tag{7}$$

where λ_{cov} and λ_{vol} control the relative contributions of the two objectives. For CLIP, we additionally introduce an exclusion loss on pseudo-unknown samples, with the corresponding loss formulation provided in Appendix B.

3.3 Open-Set Inference

At inference, a target sample is scored against every known class by $S_c(z)$ and assigned to the class attaining the minimum if that score does not exceed ρ:

$$\hat{y}(x) = \begin{cases} \arg\min_{c\in\mathcal{K}} S_c(z), & \min_{c\in\mathcal{K}} S_c(z) \le \rho, \\ \text{unknown}, & \min_{c\in\mathcal{K}} S_c(z) > \rho, \end{cases} \qquad z = \phi(x).$$

Classification and unknown rejection are therefore governed by the same support score without a separate rejection head or unknown-class logit.

4 Experiments

4.1 Setup

Datasets. We evaluate on PACS (Li et al., 2017), OfficeHome (Venkateswara et al., 2017), and VLCS (Fang et al., 2013) under leave-one-domain-out evaluation, where each domain is held out in turn as the unseen target. Details of the dataset protocols and splits are provided in Appendix B.

Backbones. We evaluate our method with two backbone configurations. For the CNN-based setting, we use an ImageNet-pretrained ResNet-18 (He et al., 2016; Deng et al., 2009), fine-tuning the entire network. For the CLIP-based setting, we use OpenAI CLIP ViT-B/32 (Radford et al., 2021), keeping all pretrained CLIP parameters frozen and training only lightweight residual adapters.

Baselines. For ResNet-18, we compare against MEDIC (Wang et al., 2023), EBiL-HaDS (Peng et al., 2024), and MEDIC++ (Wang et al., 2026a), reporting both the `cls` and `bcls` variants where applicable. For CLIP, we compare against the OSDG methods ODG-CLIP (Singha et al., 2024), SeeCLIP (Wang et al., 2026b), and OSLoPrompt (C et al., 2025), together with SPG (Bai et al., 2024) and DPStyler (Tang et al., 2025).

Evaluation Metrics. Following MEDIC (Wang et al., 2023), we report Closed-set Accuracy, H-score, and Open-Set Classification Rate (OSCR). Since H-score depends on the choice of threshold, we use H-score as a complementary metric (Wang et al., 2026a). Detailed metric definitions are provided in Appendix B.

4.2 Results

We report results averaged across leave-one-domain-out (LODO) target domains. Full per-domain results are provided in Appendix B. Tables 1 and 2 report the average results with ResNet-18 and CLIP, respectively. Our method achieves the best Closed-set Accuracy, H-score, and OSCR in all six dataset–backbone combinations. Compared with the state-of-the-art methods, OSCR improves by 2.70, 0.89, and 2.35 points with ResNet-18 and by 1.12, 1.57, and 1.26 points with CLIP on PACS, OfficeHome, and VLCS, respectively. Closed-set Accuracy also improves in every setting, indicating that better unknown rejection does not sacrifice known-class recognition.

Table 1: Average (%) results (%) on PACS, OfficeHome, and VLCS with ResNet-18.

	PACS			OfficeHome			VLCS			Avg.		
Method	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR
MEDIC-cls	86.20	71.47	75.69	73.42	62.22	61.40	88.97	65.05	69.09	82.86	66.25	68.73
MEDIC-bcls	86.20	75.98	77.89	73.42	60.84	61.61	88.97	65.43	67.57	82.86	67.42	69.02
EBiL-HaDS-cls	87.31	76.71	78.72	74.69	65.15	63.45	89.17	67.75	69.66	83.72	69.87	70.61
EBiL-HaDS-bcls	87.31	78.66	80.34	74.69	63.02	64.47	89.17	65.61	68.05	83.72	69.10	70.95
MEDIC++-cls	86.91	76.64	77.54	74.37	65.31	62.58	90.25	68.95	70.92	83.84	70.30	70.35
MEDIC++-bcls	86.91	77.02	78.24	74.37	65.32	62.68	90.25	67.75	69.79	83.84	70.03	70.24
Ours	**88.03**	**80.69**	**83.04**	**76.00**	**67.93**	**65.36**	**91.32**	**69.81**	**73.27**	**85.12**	**72.81**	**73.89**

Table 2: Average results (%) on PACS, OfficeHome, and VLCS using CLIP.

	PACS			OfficeHome			VLCS			Avg.		
Method	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR
SPG	94.31	85.02	93.52	71.93	65.21	61.40	90.67	69.10	71.83	85.64	73.11	75.58
DPStyler	94.64	93.56	94.26	84.08	68.90	74.59	92.76	80.10	83.28	90.49	80.85	84.04
OSLoPrompt	94.83	94.41	94.21	84.34	70.20	72.72	94.06	78.79	82.50	91.08	81.13	83.14
ODG-CLIP	92.76	90.66	90.71	85.41	73.19	74.28	92.00	74.98	78.20	90.06	79.61	81.06
SeeCLIP	94.60	94.83	94.41	84.72	72.72	73.12	95.18	76.68	80.53	91.50	81.41	82.69
Ours	**95.84**	**95.76**	**95.53**	**86.92**	**74.19**	**76.16**	**95.37**	**80.88**	**84.54**	**92.71**	**83.61**	**85.41**

4.3 Analysis and Ablation Studies

Evidence for Source-to-Target Generalization Since our support geometry is estimated solely from source domains, we examine whether the source-estimated uncertainty and anisotropic structure remain informative for unseen target domains. As shown in Fig. 3, the source-only uncertainty estimate remains close to its source-plus-target counterpart, with a median ratio of $1.09\times$, while the source-estimated anisotropic subspace captures 74% of the unseen domain shift on average. These results support the transferability of the source-estimated geometry to unseen domains.

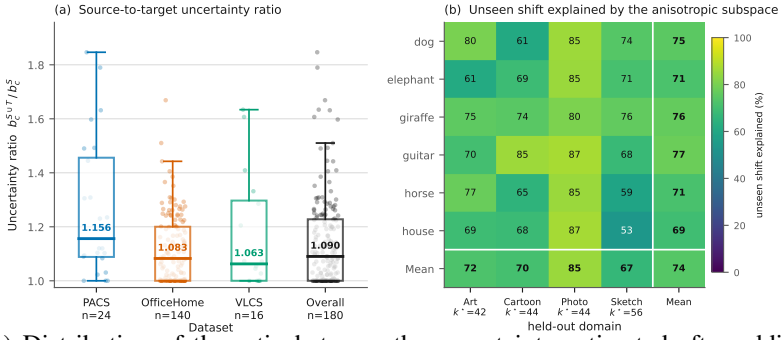


Figure 3: (a) Distribution of the ratio between the uncertainty estimated after adding the target domain and that estimated from source domains only. (b) Percentage of unseen domain shift energy captured by the source-estimated anisotropic subspace U_c.

Adaptive vs. Fixed Anisotropic Dimension. We first examine whether the theoretically derived anisotropic dimension $k^\star$ offers an advantage over using a fixed k. As shown in Fig. 4, the best fixed choice of k varies across held-out domains, whereas the source-adaptively estimated $k^\star$ consistently matches or approaches the best-performing region of the fixed-k sweeps in both Closed-set Accuracy and OSCR. This shows that the theoretically derived $k^\star$ also provides an effective choice in practice. Moreover, unlike fixed k, which requires manual tuning, $k^\star$ is determined directly from source-domain geometry and uncertainty, removing the need to treat the anisotropic dimension as an additional hyperparameter.

Effect of Each Objective. $\mathcal{L}_{\text{cov}}$ and $\mathcal{L}_{\text{vol}}$ need to act jointly: the former improves coverage of domain-shifted known samples but may over-expand the support, whereas the latter can over-contract it and increase false rejection. Their combination improves H-score and OSCR by 1.49 and 1.25 points over $\mathcal{L}_{\text{CE}}$ alone; detailed ablation results are provided in Appendix B.

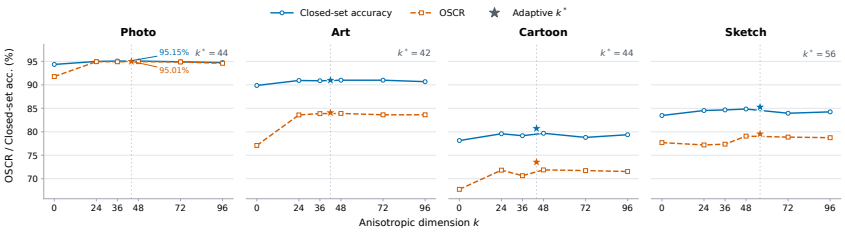


Figure 4: Comparison of the source-adaptive anisotropic dimension $k^\star$ with fixed choices of k on PACS. Curves report Closed-set Accuracy and OSCR for models retrained with fixed $k \in \{0, 24, 36, 48, 72, 96\}$, while stars denote the performance obtained using the source-adaptively estimated $k^\star$ for each held-out domain.

Support Parameterization. Table 3 compares the theoretically derived spiked structure with isotropic, diagonal, and full parameterizations. The spiked support achieves the best H-score and OSCR (80.69% and 83.04%), outperforming the full parameterization by 2.38 OSCR points while using only $k + 1$ learnable scale parameters per class.

Table 3: Ablation results for different support matrix parameterizations on PACS with ResNet-18. "Params" denotes the number of learnable support scale parameters per class, where m is the feature dimension and k is the anisotropic dimension.

Support Structure	Params	Closed Acc. ↑	H-score ↑	OSCR ↑
Isotropic	1	87.68	78.07	79.55
Diagonal	m	87.70	77.52	79.65
full(Mahalanobis)	$m(m+1)/2$	87.88	79.80	80.66
Spiked (Ours)	$k+1$	**88.03**	**80.69**	**83.04**

5 Related Work

OSR. Open-set recognition (OSR) requires recognizing known classes while rejecting unseen ones. Distance-based methods characterize known classes using structures such as Mahalanobis distances, class-conditional Gaussian models, or ellipsoidal supports (Geng et al., 2021; Liu et al., 2020; Sun et al., 2020; Zou et al., 2026). While such geometric modeling can constrain open space, the source geometry may become unreliable under domain shift.

OSDG. Open-set domain generalization (OSDG) further requires generalization to an unseen target domain (Shu et al., 2021). Existing methods mainly improve robustness through meta-learning (Wang et al., 2023; Peng et al., 2024; Wang et al., 2026a) or vision-language priors and pseudo-unknown generation (Singha et al., 2024; C et al., 2025; Wang et al., 2026b). However, they do not explicitly determine how strongly the acceptance region should rely on source geometry as source discrepancy varies. Our method instead uses source discrepancy to adaptively determine how much source geometry is retained in the support.

Uncertainty and Robustness in DG. DG methods have used source domain variation to model uncertainty or robustness, including risk variation (Krueger et al., 2021), feature statistic uncertainty (Li et al., 2022), and distributional uncertainty sets (Wang et al., 2024). In contrast, we translate class-conditional source discrepancy directly into the trusted spike dimension, thereby controlling the geometry of the acceptance region.

6 Conclusion

We propose AdaSpike, a theoretically motivated OSDG framework that establishes a direct relationship between source discrepancy and the trusted spike dimension k. This enables the acceptance region to adapt its shape to different levels of source discrepancy, adjusting how much source geometry is preserved. Extensive experiments validate this adaptive design and show strong overall improvements over state-of-the-art methods. We further discuss the limitations of the current framework in Appendix C.

AI USE STATEMENT

Generative AI tools were used to assist with language editing, literature search, refining the presentation of scientific figures, and implementing a small portion of the code. The research methodology, experimental design, analysis, and scientific conclusions were developed and validated by the authors. All AI-assisted content was carefully reviewed and revised by the authors, who take full responsibility for the final content of this work.

ETHICS STATEMENT

This work adheres to the ICLR Code of Ethics. In this study, no animal experimentation was involved. All datasets used were sourced in compliance with relevant usage guidelines, ensuring no violation of privacy. We have taken care to avoid any biases or discriminatory outcomes in our research process. No personally identifiable information was used, and no experiments were conducted that could raise privacy or security concerns. We are committed to maintaining transparency and integrity throughout the research process.

REPRODUCIBILITY STATEMENT

We have made every effort to ensure that the results presented in this paper are reproducible. All code will be made publicly available to facilitate replication and verification after the paper gets accepted. The experimental setup, including model configurations, resources, and hardware details, is described in detail in Appendix B. We also provide a complete proof of the core claims in the Appendix A. Additionally, the datasets used in this paper are publicly available, ensuring consistent and reproducible evaluation results. We believe these measures will enable other researchers to reproduce our work and further advance the field.

REFERENCES

Isabela Albuquerque, João Monteiro, Mohammad Darvishi, Tiago H. Falk, and Ioannis Mitliagkas. Generalizing to unseen domains via distribution matching. *arXiv preprint arXiv:1911.00804*, 2019.

Shuanghao Bai, Yuedi Zhang, Wanqi Zhou, Zhirong Luan, and Badong Chen. Soft prompt generation for domain generalization. In *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part III*, pp. 434–450. Springer Nature Switzerland, 2024.

Gilles Blanchard, Aniket Anand Deshmukh, Ürün Dogan, Gyemin Lee, and Clayton Scott. Domain generalization by marginal transfer learning. *Journal of Machine Learning Research*, 22(2):1–55, 2021.

Mohamad Hassan N C, Divyam Gupta, Mainak Singha, Sai Bhargav Rongali, Ankit Jha, Muhammad Haris Khan, and Biplab Banerjee. Osloprompt: Bridging low-supervision challenges and open-set domain generalization in CLIP. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 10110–10120, 2025.

C. Chow. On optimum recognition error and reject tradeoff. *IEEE Transactions on Information Theory*, 16(1):41–46, 1970.

Rui Dai, Yonggang Zhang, Zhen Fang, Bo Han, and Xinmei Tian. Moderately distributional exploration for domain generalization. In *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202, pp. 6786–6817. PMLR, 2023.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA*, pp. 248–255, 2009.

Chen Fang, Ye Xu, and Daniel N. Rockmore. Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1657–1664, 2013.

Chuanxing Geng, Sheng-Jun Huang, and Songcan Chen. Recent advances in open set recognition: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3614–3631, 2021.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pp. 770–778, 2016.

David Krueger, Ethan Caballero, Jörn-Henrik Jacobsen, Amy Zhang, Jonathan Binas, Dinghuai Zhang, Rémi Le Priol, and Aaron C. Courville. Out-of-distribution generalization via risk extrapolation (REx). In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139, pp. 5815–5826. PMLR, 2021.

Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. Deeper, broader and artier domain generalization. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pp. 5542–5550, 2017.

Xiaotong Li, Yongxing Dai, Yixiao Ge, Jun Liu, Ying Shan, and Lingyu Duan. Uncertainty modeling for out-of-distribution generalization. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*, 2022.

Bo Liu, Hao Kang, Haoxiang Li, Gang Hua, and Nuno Vasconcelos. Few-shot open-set recognition using meta-learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pp. 8795–8804, 2020.

Kunyu Peng, Di Wen, Kailun Yang, Ao Luo, Yufan Chen, Jia Fu, M. Saquib Sarfraz, Alina Roitberg, and Rainer Stiefelhagen. Advancing open-set domain generalization using evidential bi-level hardest domain scheduler. In *Advances in Neural Information Processing Systems 37: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024*, 2024.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, pp. 8748–8763. PMLR, 2021.

Walter J. Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E. Boult. Toward open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(7): 1757–1772, 2013.

Yang Shu, Zhangjie Cao, Chenyu Wang, Jianmin Wang, and Mingsheng Long. Open domain generalization with domain-augmented meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9619–9628, 2021.

Mainak Singha, Ankit Jha, Shirsha Bose, Ashwin R. Nair, Moloud Abdar, and Biplab Banerjee. Unknown prompt, the only lacuna: Unveiling clip's potential for open domain generalization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pp. 13309–13319, 2024.

Xin Sun, Zhenning Yang, Chi Zhang, Keck Voon Ling, and Guohao Peng. Conditional gaussian distribution learning for open set recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pp. 13477–13486, 2020.

Yunlong Tang, Yuxuan Wan, Lei Qi, and Xin Geng. Dpstyler: Dynamic promptstyler for source-free domain generalization. *IEEE Transactions on Multimedia*, 27:120–132, 2025.

Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pp. 5018–5027, 2017.

Jingge Wang, Liyan Xie, Yao Xie, Shao-Lun Huang, and Yang Li. Generalizing to unseen domains with Wasserstein distributional robustness under limited source knowledge. *IEEE Journal on Selected Topics in Signal Processing*, 19(1):103–114, 2024.

Xiran Wang, Jian Zhang, Lei Qi, and Yinghuan Shi. Generalizable decision boundaries: Dualistic meta-learning for open set domain generalization. In *IEEE/CVF International Conference on Computer Vision, ICCV 2023, Paris, France, October 1-6, 2023*, pp. 11530–11539, 2023.

Xiran Wang, Jian Zhang, Lei Qi, Yang Gao, and Yinghuan Shi. Exploring dualistic meta-learning to enhance domain generalization in open set scenarios. *IEEE Transactions on Knowledge and Data Engineering*, 2026a.

Yunyun Wang, Zheng Duan, Xinyue Liao, Ke-Jia Chen, and Songcan Chen. The finer the better: Towards granular-aware open-set domain generalization. In *Fortieth AAAI Conference on Artificial Intelligence, Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence, Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2026, Singapore, January 20-27, 2026*, volume 40, pp. 26732–26740, 2026b.

Fan Zhou, Yuyi Chen, Shichun Yang, Boyu Wang, and Brahim Chaib-draa. On the value of label and semantic information in domain generalization. *Neural Networks*, 163:244–255, 2023.

Yuetian Zou, Hanlei Zhang, Hua Xu, Songze Li, and Long Xiao. Ellipsoid-based decision boundaries for open intent classification. In *Fortieth AAAI Conference on Artificial Intelligence, Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence, Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2026, Singapore, January 20-27, 2026*, volume 40, pp. 35293–35301, 2026.

A Proof of Theoretical Analysis

A.1 Notations

Table 4: **Summary of notation used in the theoretical analysis.**

Notation	Meaning
$\mathcal{X}$	Input space.
$\mathcal{Y}^*$	Universal label space.
$\mathcal{K}$, $\mathcal{U}$	Known and unknown label sets, with $\mathcal{U} = \mathcal{Y}^* \setminus \mathcal{K}$.
P, μ	Domain distribution over $\mathcal{X} \times \mathcal{Y}^*$ and meta-distribution over domains, with $P \sim \mu$.
$\bar{P}$	Mixture distribution induced by the domain meta-distribution, $\bar{P} = \mathbb{E}_{P\sim\mu}[P]$.
$\pi_c(P)$	Probability mass of class c under domain P, $\pi_c(P) = P(Y^* = c)$.
$\pi_{\mathcal{K}}(P)$, $\pi_{\mathcal{U}}(P)$	Probability masses of known and unknown classes under P.
$\bar{P}_{\mathcal{U}}$	Aggregate unknown measure induced by the domain meta-distribution.
ϕ	ℓ_2-normalized representation map $\phi : \mathcal{X} \to \mathbb{S}^{m-1}$.
$Z = \phi(X)$	Feature representation on the unit sphere $\mathbb{S}^{m-1}$.
m	Ambient feature dimension; $\mathbb{S}^{m-1}$ has intrinsic dimension $m-1$.
p_c	Prototype of known class c, with $p_c \in \mathbb{S}^{m-1}$.
$\xi_c(z)$	Tangent displacement $\xi_c(z) = \mathrm{Log}_{p_c}(z) \in T_{p_c}\mathbb{S}^{m-1}$.
T_c^P	Class-c tangent second moment under domain P.
T_c	Population class-conditional tangent second moment under $\bar{P}$.
$\widehat{T}_c$	Empirical class-conditional tangent second moment estimated from source domains.
$\widehat{\gamma}_{c,j}$, $q_{c,j}$	Eigenvalue and eigenvector of $\widehat{T}_c$, ordered such that $\widehat{\gamma}_{c,1} \geq \cdots \geq \widehat{\gamma}_{c,m-1}$.
$\mathcal{T}_c(\varepsilon_c)$	Nuclear-norm uncertainty set around $\widehat{T}_c$.
b_c, $r_{N,c}(\delta)$	Cross-domain variation radius and finite-sample estimation radius.
ε_c	Total uncertainty radius, $\varepsilon_c = b_c + r_{N,c}(\delta)$.
M_c	Positive-definite anisotropic support matrix for class c.
$S_c(z)$	Anisotropic support score, $S_c(z) = \xi_c(z)^\top M_c^{-1} \xi_c(z)$.
ρ	Acceptance threshold.
A_c, A	Class-wise and overall known acceptance regions, $A = \bigcup_{c\in\mathcal{K}} A_c$.
τ_c	Robust spectral floor determined by $\sum_j (\tau_c - \widehat{\gamma}_{c,j})_+ = \varepsilon_c$.
$t_c^\star$	Optimal global scale of the robust support spectrum.
$\sigma_c^{2\star}$	Optimal residual-direction support scale, $\sigma_c^{2\star} = t_c^\star \tau_c$.
$\lambda_{c,j}^\star$	Additional optimal support scale along active direction $q_{c,j}$.
k_c	Number of active directions satisfying $\widehat{\gamma}_{c,j} > \tau_c$.
$M_c^\star$	Robust optimal anisotropic support matrix.
f	Closed-set classifier $f : \mathcal{X} \to \mathcal{K}$.
$h_{A,f}$	Open-set predictor determined by classifier f and acceptance region A.
$\mathcal{E}(h_{A,f})$	OSDG risk over unseen domains.
$R_{\mathrm{mf}}^P(A, f)$	Misclassification risk of accepted known samples under domain P.
$R_{\mathrm{FR}}(A)$, $R_{\mathrm{FA}}(A)$	False-rejection risk of known samples and false-acceptance risk of unknown samples.
ν	Spherical surface measure on $\mathbb{S}^{m-1}$.
$C_{\mathcal{U}}$	Constant satisfying $\bar{P}_{\mathcal{U}}(B) \leq C_{\mathcal{U}} \nu(B)$.
ω_{m-1}	Volume of the Euclidean unit ball in $\mathbb{R}^{m-1}$.
α_c, β_c	Class-wise coefficients controlling the coverage and volume terms in the open-set risk bound.
$\mathcal{F}$	Hypothesis class of closed-set classifiers.
Γ_{cls}	Uniform finite-sample generalization term for classification.
N_c	Number of source training samples belonging to class c.

A.2 Details of Spherical Class Supports

For completeness, we provide the explicit construction of the spherical class support used in Definition 2. For $z, p_c \in \mathbb{S}^{m-1}$, let

$$\theta_c = \arccos(p_c^\top z).$$

The spherical logarithmic map at p_c is

$$\xi_c(z) = \mathrm{Log}_{p_c}(z) = \frac{\theta_c}{\sin\theta_c}\,(z - \cos\theta_c\, p_c) \in T_{p_c}\mathbb{S}^{m-1},$$

with

$$\|\xi_c(z)\|_2 = \theta_c,$$

so the tangent-space norm equals the geodesic distance from z to p_c.

For a symmetric positive-definite matrix M_c, the class score is

$$S_c(z) = \xi_c(z)^\top M_c^{-1}\xi_c(z).$$

Hence, within the injectivity radius, the spherical acceptance region A_c is induced by the tangent-space ellipsoid

$$E_c = \left\{v \in T_{p_c}\mathbb{S}^{m-1} : v^\top M_c^{-1} v \le \rho\right\}.$$

The eigenvectors of M_c determine the principal directions of the support, while its eigenvalues determine their squared extents.

In contrast to M_c, which parameterizes the acceptance support,

$$T_c = \mathbb{E}\left[\xi_c(Z)\xi_c(Z)^\top \mid Y^* = c\right]$$

describes the population second-moment geometry of class c. The robust optimization in Theorem 1 determines $M_c^\star$ from uncertainty about T_c.

A.3 Proof of Proposition 1

Proof. Fix a class $c \in \mathcal{K}$. Write

$$\pi_c(P) := P(Y^* = c), \qquad \bar{\pi}_c := \mathbb{E}_{P\sim\mu}[\pi_c(P)].$$

Since $\bar{\pi}_c > 0$, define the class-conditional measure on domain distributions by

$$d\mu_c(P) = \frac{\pi_c(P)}{\bar{\pi}_c}\, d\mu(P).$$

By the definition of T_c under the mixture distribution $\bar{P}$,

$$T_c = \int T_c^P \, d\mu_c(P).$$

Let

$$T_{S,c} = \sum_{s=1}^{S} w_{s,c} T_c^{P_s}, \qquad w_{s,c} \ge 0, \qquad \sum_{s=1}^{S} w_{s,c} = 1,$$

be any convex combination of the source-domain second moments. Then

$$T_c - T_{S,c} = \int \sum_{s=1}^{S} w_{s,c}\left(T_c^P - T_c^{P_s}\right) d\mu_c(P).$$

Using convexity of the nuclear norm and Assumption 2, we obtain

$$\begin{aligned}\|T_c - T_{S,c}\|_* &\le \int \sum_{s=1}^{S} w_{s,c}\|T_c^P - T_c^{P_s}\|_*\, d\mu_c(P) \\ &\le \int \sum_{s=1}^{S} w_{s,c} b_c \, d\mu_c(P) = b_c.\end{aligned}$$

We next make the finite-sample radius explicit. Let $N_{s,c}$ denote the number of class-c samples used from source domain s, and define

$$\widehat{T}_c = \sum_{s=1}^{S} w_{s,c} \widehat{T}_c^{\,s}, \qquad \widehat{T}_c^{\,s} = \frac{1}{N_{s,c}} \sum_{i=1}^{N_{s,c}} \xi_{s,i}^{c} \xi_{s,i}^{c\top}.$$

Suppose $\|\xi_c(Z)\| \le R_c$ almost surely, and let $d = m - 1$. Conditional on the sampled source domains and the counts $N_{s,c}$, independence of the within-domain samples gives

$$\mathbb{E}\left[\|\widehat{T}_c - T_{S,c}\|_F^2\right] \le R_c^4 \sum_{s=1}^{S} \frac{w_{s,c}^2}{N_{s,c}}.$$

Moreover, by bounded-difference concentration, with probability at least $1 - \delta$,

$$\|\widehat{T}_c - T_{S,c}\|_F \le R_c^2 \sqrt{\sum_{s=1}^{S} \frac{w_{s,c}^2}{N_{s,c}}} \left(1 + \sqrt{2 \log \frac{1}{\delta}}\right).$$

Since $\|A\|_* \le \sqrt{d}\, \|A\|_F$, one valid estimation radius is

$$r_{N,c}(\delta) = R_c^2 \sqrt{d \sum_{s=1}^{S} \frac{w_{s,c}^2}{N_{s,c}}} \left(1 + \sqrt{2 \log \frac{1}{\delta}}\right).$$

Equivalently, defining $N_{\mathrm{eff},c} = \left(\sum_{s=1}^{S} w_{s,c}^2 / N_{s,c}\right)^{-1}$,

$$r_{N,c}(\delta) = R_c^2 \sqrt{\frac{d}{N_{\mathrm{eff},c}}} \left(1 + \sqrt{2 \log \frac{1}{\delta}}\right).$$

Hence the event

$$\mathcal{E}_c(\delta) = \left\{ \|\widehat{T}_c - T_{S,c}\|_* \le r_{N,c}(\delta) \right\}$$

has probability at least $1 - \delta$. On this event, the triangle inequality gives

$$\begin{aligned} \|T_c - \widehat{T}_c\|_* &\le \|T_c - T_{S,c}\|_* + \|T_{S,c} - \widehat{T}_c\|_* \\ &\le b_c + r_{N,c}(\delta). \end{aligned}$$

Therefore, with probability at least $1 - \delta$,

$$T_c \in \mathcal{T}_c(\varepsilon_c), \qquad \varepsilon_c = b_c + r_{N,c}(\delta).$$

□

A.4 Robust Optimal Support 1

Theorem 3 (Robust Optimal Support)**.** *Under Assumption 1, the open-set terms in Definition 2 satisfy*

$$R_{\mathrm{FR}}(A) + R_{\mathrm{FA}}(A) \le \sum_{c \in \mathcal{K}} \left[\alpha_c \operatorname{tr}(T_c M_c^{-1}) + \beta_c \det(M_c)^{1/2}\right].$$

Combining this bound with the uncertainty set $\mathcal{T}_c(\varepsilon_c)$ in Proposition 1, consider

$$M_c^\star \in \arg\min_{M \succ 0} \sup_{T \in \mathcal{T}_c(\varepsilon_c)} \left[\alpha_c \operatorname{tr}(TM^{-1}) + \beta_c \det(M)^{1/2}\right].$$

Suppose $\alpha_c > 0$, $\beta_c > 0$, and $\varepsilon_c > 0$. Let $\widehat{T}_c = \sum_{j=1}^{m-1} \widehat{\gamma}_{c,j} q_{c,j} q_{c,j}^\top$, with $\widehat{\gamma}_{c,1} \ge \cdots \ge \widehat{\gamma}_{c,m-1} \ge 0$. Then the robust optimum has the spiked form

$$M_c^\star = t_c^\star \tau_c I + t_c^\star \sum_{j:\widehat{\gamma}_{c,j} > \tau_c} (\widehat{\gamma}_{c,j} - \tau_c) q_{c,j} q_{c,j}^\top = \sigma_c^{2\star} I + U_c^\star \Lambda_c^\star U_c^{\star\top}.$$

The floor τ_c is uniquely determined by $\sum_{j=1}^{m-1} (\tau_c - \widehat{\gamma}_{c,j})_+ = \varepsilon_c$. Let $a_{c,j} = \max\{\widehat{\gamma}_{c,j}, \tau_c\}$ and $G_c = \prod_{j=1}^{m-1} a_{c,j}^{1/2}$. Then $t_c^\star = (2\alpha_c / (\beta_c G_c))^{2/(m+1)}$, with $\sigma_c^{2\star} = t_c^\star \tau_c$ and $\lambda_{c,j}^\star = t_c^\star (\widehat{\gamma}_{c,j} - \tau_c)$ for $\widehat{\gamma}_{c,j} > \tau_c$. Here $\varepsilon_c = b_c + r_{N,c}(\delta)$, $\alpha_c = \rho^{-1} \mathbb{E}_{P \sim \mu}[\pi_c(P)]$, and $\beta_c = C_{\mathcal{U}} \omega_{m-1} \rho^{(m-1)/2}$.

A.5 Proof of Theorem 1

We first establish a geometric lemma for the spherical support regions.

Lemma 1 (Spherical volume of an anisotropic support)**.** Let $p \in \mathbb{S}^{m-1}$ and let M be positive definite on $T_p\mathbb{S}^{m-1}$. Define

$$A(M) = \left\{ z \in \mathbb{S}^{m-1} : \operatorname{Log}_p(z)^\top M^{-1} \operatorname{Log}_p(z) \le \rho \right\}.$$

Then

$$\nu(A(M)) \le \omega_{m-1} \rho^{(m-1)/2} \det(M)^{1/2},$$

where ω_{m-1} is the Euclidean volume of the unit ball in $\mathbb{R}^{m-1}$.

Proof. Set $d = m - 1$ and define the tangent-space ellipsoid

$$E(M) = \left\{ v \in T_p\mathbb{S}^{m-1} : v^\top M^{-1} v \le \rho \right\}.$$

Away from the cut locus of p, the support satisfies

$$A(M) = \operatorname{Exp}_p \left(E(M) \cap B(0, \pi) \right).$$

The cut locus has zero spherical surface measure and does not affect the volume.

In geodesic normal coordinates, the Jacobian determinant of the exponential map is

$$J(v) = \left(\frac{\sin \|v\|}{\|v\|} \right)^{d-1} \le 1 \qquad \text{for } 0 \le \|v\| < \pi.$$

Therefore,

$$\begin{aligned} \nu(A(M)) &= \int_{E(M) \cap B(0,\pi)} J(v)\, dv \\ &\le \operatorname{Vol}_d(E(M)). \end{aligned}$$

The change of variables

$$v = \rho^{1/2} M^{1/2} u$$

maps the Euclidean unit ball to $E(M)$. Hence

$$\operatorname{Vol}_d(E(M)) = \omega_d \rho^{d/2} \det(M)^{1/2}.$$

Substituting $d = m - 1$ proves the result. □

Proof of Theorem 1. Fix a collection of positive-definite matrices $\{M_c\}_{c \in \mathcal{K}}$ and let

$$A = \bigcup_{c \in \mathcal{K}} A_c.$$

(A) Bounding the open-set risk. Write

$$\pi_c(P) := P(Y^* = c).$$

Since $A_c \subseteq A$, we have $A^c \subseteq A_c^c$. Therefore,

$$\begin{aligned} R_{\mathrm{FR}}(A) &= \mathbb{E}_{P \sim \mu} \left[\sum_{c \in \mathcal{K}} \pi_c(P) P_Z^P(A^c \mid Y^* = c) \right] \\ &\le \sum_{c \in \mathcal{K}} \mathbb{E}_{P \sim \mu} \left[\pi_c(P) P_Z^P(A_c^c \mid Y^* = c) \right]. \end{aligned}$$

By the definition of A_c,

$$P_Z^P(A_c^c \mid Y^* = c) = P_Z^P\left(S_c(Z) > \rho \mid Y^* = c \right).$$

Since $S_c(Z) \ge 0$, Markov's inequality gives

$$P_Z^P\left(S_c(Z) > \rho \mid Y^* = c \right) \le \frac{1}{\rho} \mathbb{E}_P \left[S_c(Z) \mid Y^* = c \right].$$

Moreover,

$$\mathbb{E}_P\left[S_c(Z) \mid Y^* = c\right] = \mathbb{E}_P\left[\xi_c(Z)^\top M_c^{-1}\xi_c(Z) \mid Y^* = c\right] = \operatorname{tr}\left(T_c^P M_c^{-1}\right).$$

It follows that

$$R_{\mathrm{FR}}(A) \le \frac{1}{\rho}\sum_{c\in\mathcal{K}} \operatorname{tr}\left(\mathbb{E}_{P\sim\mu}\left[\pi_c(P)T_c^P\right] M_c^{-1}\right).$$

By the definition of the class-conditional second moment under $\bar{P}$,

$$\mathbb{E}_{P\sim\mu}\left[\pi_c(P)T_c^P\right] = \mathbb{E}_{P\sim\mu}[\pi_c(P)]T_c.$$

Therefore,

$$R_{\mathrm{FR}}(A) \le \sum_{c\in\mathcal{K}} \alpha_c \operatorname{tr}\left(T_c M_c^{-1}\right),$$

where

$$\alpha_c = \frac{1}{\rho}\mathbb{E}_{P\sim\mu}[\pi_c(P)].$$

By Assumption 1,

$$R_{\mathrm{FA}}(A) = \bar{P}_{\mathcal{U}}(A) \le C_{\mathcal{U}}\nu(A).$$

Since $A = \bigcup_{c\in\mathcal{K}} A_c$, subadditivity of ν gives

$$\nu(A) \le \sum_{c\in\mathcal{K}} \nu(A_c).$$

Applying Lemma 1 to each class,

$$\nu(A_c) \le \omega_{m-1}\rho^{(m-1)/2}\det(M_c)^{1/2}.$$

Consequently,

$$R_{\mathrm{FA}}(A) \le \sum_{c\in\mathcal{K}} \beta_c \det(M_c)^{1/2},$$

where

$$\beta_c = C_{\mathcal{U}}\omega_{m-1}\rho^{(m-1)/2}.$$

Combining the false-rejection and false-acceptance bounds yields

$$R_{\mathrm{FR}}(A) + R_{\mathrm{FA}}(A) \le \sum_{c\in\mathcal{K}} \left[\alpha_c \operatorname{tr}(T_c M_c^{-1}) + \beta_c \det(M_c)^{1/2}\right].$$

(B) Robust spectral reduction. Fix a class c and abbreviate

$$\widehat{T} = \widehat{T}_c, \qquad \varepsilon = \varepsilon_c, \qquad \alpha = \alpha_c, \qquad \beta = \beta_c.$$

For any $M \succ 0$ and any $T \in \mathcal{T}_c(\varepsilon)$, trace duality gives

$$\begin{aligned}\operatorname{tr}(TM^{-1}) &= \operatorname{tr}(\widehat{T}M^{-1}) + \operatorname{tr}\left((T-\widehat{T})M^{-1}\right)\\ &\le \operatorname{tr}(\widehat{T}M^{-1}) + \|T-\widehat{T}\|_*\|M^{-1}\|_{\mathrm{op}}\\ &\le \operatorname{tr}(\widehat{T}M^{-1}) + \frac{\varepsilon}{\lambda_{\min}(M)}.\end{aligned}$$

This upper bound is attained. Let u be a unit eigenvector of M associated with $\lambda_{\min}(M)$ and set

$$T^\dagger = \widehat{T} + \varepsilon uu^\top.$$

Then $T^\dagger \succeq 0$,

$$\|T^\dagger - \widehat{T}\|_* = \varepsilon,$$

and

$$\operatorname{tr}\left((T^\dagger - \widehat{T})M^{-1}\right) = \frac{\varepsilon}{\lambda_{\min}(M)}.$$

Thus,

$$\sup_{T\in\mathcal{T}_c(\varepsilon)} \operatorname{tr}(TM^{-1}) = \operatorname{tr}(\widehat{T}M^{-1}) + \frac{\varepsilon}{\lambda_{\min}(M)}.$$

The robust problem is therefore equivalent to

$$\min_{M\succ 0} \left[\alpha \operatorname{tr}(\widehat{T}M^{-1}) + \frac{\alpha\varepsilon}{\lambda_{\min}(M)} + \beta \det(M)^{1/2}\right].$$

Let $d = m - 1$ and write

$$\widehat{T} = \sum_{j=1}^{d} \widehat{\gamma}_j q_j q_j^\top, \qquad \widehat{\gamma}_1 \geq \cdots \geq \widehat{\gamma}_d \geq 0.$$

For fixed eigenvalues of M, both $\lambda_{\min}(M)$ and $\det(M)$ are independent of its eigenvectors. By the von Neumann trace inequality, the trace term is minimized when M and $\widehat{T}$ share eigenvectors, with their eigenvalues aligned in the same order. Hence an optimizer can be written as

$$M = \sum_{j=1}^{d} m_j q_j q_j^\top, \qquad m_j > 0.$$

The robust objective becomes

$$\alpha \sum_{j=1}^{d} \frac{\widehat{\gamma}_j}{m_j} + \frac{\alpha\varepsilon}{\min_j m_j} + \beta \left(\prod_{j=1}^{d} m_j\right)^{1/2}.$$

(C) KKT characterization of the support spectrum. Let

$$v := \min_j m_j,$$

so that $m_j \geq v$ for all j. Introducing multipliers $\eta_j \geq 0$ for the constraints $m_j \geq v$, the Lagrangian is

$$\mathcal{L} = \alpha \sum_{j=1}^{d} \frac{\widehat{\gamma}_j}{m_j} + \frac{\alpha\varepsilon}{v} + \beta \left(\prod_{\ell=1}^{d} m_\ell\right)^{1/2} + \sum_{j=1}^{d} \eta_j (v - m_j).$$

The stationarity condition with respect to m_j is

$$-\frac{\alpha\widehat{\gamma}_j}{m_j^2} + \frac{\beta}{2m_j}\left(\prod_{\ell=1}^{d} m_\ell\right)^{1/2} - \eta_j = 0.$$

For an eigenvalue strictly above the floor, $m_j > v$, complementary slackness gives $\eta_j = 0$, and hence

$$m_j = \frac{2\alpha}{\beta\left(\prod_{\ell=1}^{d} m_\ell\right)^{1/2}} \widehat{\gamma}_j.$$

Thus all eigenvalues above the floor are proportional to the corresponding $\widehat{\gamma}_j$. Defining the common scale

$$t := \frac{2\alpha}{\beta\left(\prod_{\ell=1}^{d} m_\ell\right)^{1/2}} > 0$$

and $\tau := v/t$, the unconstrained stationary value $t\widehat{\gamma}_j$ lies above the floor exactly when $\widehat{\gamma}_j > \tau$. Otherwise the constraint is active and $m_j = v = t\tau$. Therefore,

$$m_j = t \max\{\widehat{\gamma}_j, \tau\}.$$

Equivalently,

$$M = t\tau I + t \sum_{j:\widehat{\gamma}_j > \tau} (\widehat{\gamma}_j - \tau) q_j q_j^\top.$$

(D) Optimal floor and scale. Define the active and floor index sets

$$\mathcal{I}(\tau) = \{j : \widehat{\gamma}_j > \tau\}, \qquad \mathcal{I}(\tau)^c = \{j : \widehat{\gamma}_j \le \tau\}.$$

Thus, $m_j = t\widehat{\gamma}_j > v$ for $j \in \mathcal{I}(\tau)$, while $m_j = v = t\tau$ for $j \in \mathcal{I}(\tau)^c$.

We first determine the optimal floor. Recall the Lagrangian is

$$\mathcal{L} = \alpha \sum_{j=1}^{d} \frac{\widehat{\gamma}_j}{m_j} + \frac{\alpha\varepsilon}{v} + \beta \left(\prod_{\ell=1}^{d} m_\ell \right)^{1/2} + \sum_{j=1}^{d} \eta_j (v - m_j),$$

with multipliers $\eta_j \ge 0$. Stationarity with respect to v gives

$$-\frac{\alpha\varepsilon}{v^2} + \sum_{j=1}^{d} \eta_j = 0.$$

For $j \in \mathcal{I}(\tau)$, complementary slackness gives $\eta_j = 0$. For $j \in \mathcal{I}(\tau)^c$, we have $m_j = v$, and the stationarity condition with respect to m_j gives

$$\eta_j = -\frac{\alpha\widehat{\gamma}_j}{v^2} + \frac{\beta}{2v} \left(\prod_{\ell=1}^{d} m_\ell \right)^{1/2}.$$

Using the common-scale relation from Step 4,

$$t = \frac{2\alpha}{\beta \left(\prod_{\ell=1}^{d} m_\ell \right)^{1/2}},$$

together with $v = t\tau$, this becomes

$$\eta_j = \frac{\alpha}{v^2}(\tau - \widehat{\gamma}_j), \qquad j \in \mathcal{I}(\tau)^c.$$

Substituting into the stationarity condition for v yields

$$\varepsilon = \sum_{j \in \mathcal{I}(\tau)^c} (\tau - \widehat{\gamma}_j),$$

or equivalently,

$$\sum_{j=1}^{d} (\tau - \widehat{\gamma}_j)_+ = \varepsilon.$$

It remains to show that this equation uniquely determines the floor. Define

$$H(\tau) := \sum_{j=1}^{d} (\tau - \widehat{\gamma}_j)_+.$$

The function H is continuous and nondecreasing, satisfies $H(\tau) = 0$ for $\tau \le \widehat{\gamma}_d$, is strictly increasing for $\tau > \widehat{\gamma}_d$, and satisfies $H(\tau) \to \infty$ as $\tau \to \infty$. Hence, since $\varepsilon > 0$, the equation $H(\tau) = \varepsilon$ admits a unique solution $\tau > 0$.

We next determine the global scale. Define

$$a_j := \max\{\widehat{\gamma}_j, \tau\}, \qquad G := \prod_{j=1}^{d} a_j^{1/2}.$$

Then $m_j = t a_j$. Moreover, using the floor equation,

$$\begin{aligned} \sum_{j=1}^{d} \frac{\widehat{\gamma}_j}{a_j} + \frac{\varepsilon}{\tau} &= |\mathcal{I}(\tau)| + \sum_{j \in \mathcal{I}(\tau)^c} \frac{\widehat{\gamma}_j}{\tau} + \frac{1}{\tau} \sum_{j \in \mathcal{I}(\tau)^c} (\tau - \widehat{\gamma}_j) \\ &= |\mathcal{I}(\tau)| + |\mathcal{I}(\tau)^c| = d. \end{aligned}$$

Substituting $m_j = ta_j$ and $v = t\tau$ into the robust objective therefore gives the one-dimensional problem

$$\Psi(t) = \frac{\alpha d}{t} + \beta G t^{d/2}, \qquad t > 0.$$

Since $\Psi(t) \to \infty$ as $t \downarrow 0$ and as $t \to \infty$, its unique minimizer is characterized by

$$\Psi'(t) = -\frac{\alpha d}{t^2} + \frac{\beta d}{2} G t^{d/2-1} = 0.$$

Solving for t gives

$$t^\star = \left(\frac{2\alpha}{\beta G}\right)^{\frac{2}{d+2}} = \left(\frac{2\alpha}{\beta G}\right)^{\frac{2}{m+1}}.$$

Restoring the class index,

$$M_c^\star = t_c^\star \tau_c I + t_c^\star \sum_{j:\widehat{\gamma}_{c,j} > \tau_c} (\widehat{\gamma}_{c,j} - \tau_c) q_{c,j} q_{c,j}^\top,$$

where

$$\sum_{j=1}^{m-1} (\tau_c - \widehat{\gamma}_{c,j})_+ = \varepsilon_c,$$

and

$$t_c^\star = \left(\frac{2\alpha_c}{\beta_c G_c}\right)^{\frac{2}{m+1}}, \qquad G_c = \prod_{j=1}^{m-1} \max\{\widehat{\gamma}_{c,j}, \tau_c\}^{1/2}.$$

Finally, setting

$$\sigma_c^{2\star} = t_c^\star \tau_c$$

and

$$\lambda_{c,j}^\star = t_c^\star (\widehat{\gamma}_{c,j} - \tau_c), \qquad \widehat{\gamma}_{c,j} > \tau_c,$$

gives

$$M_c^\star = \sigma_c^{2\star} I + U_c^\star \Lambda_c^\star U_c^{\star\top}.$$

□

A.6 Finite-Sample OSDG Risk Bound

Theorem 4 (Finite-Sample OSDG Risk Bound). *Let $N_c = \sum_{s=1}^{S} \sum_{i=1}^{N_s} \mathbf{1}\{Y_i^s = c\}$ denote the number of source samples from class c. Suppose Assumptions 1 and 2 hold.*

Let $\mathcal{F}$ be a fixed hypothesis class and $\Gamma_{\mathrm{cls}}(\mathcal{F}, S, \{N_s\}, \delta/2)$ be a valid uniform finite-sample generalization term. For each $c \in \mathcal{K}$, set $\varepsilon_c = b_c + r_{N,c}(\delta/(2|\mathcal{K}|))$ as in Proposition 1.

Let $A^\star = \bigcup_{c\in\mathcal{K}} A_c(M_c^\star)$. Then, with probability at least $1-\delta$ over the sampled source domains and training samples, for every $f \in \mathcal{F}$,

$$\mathcal{E}(h_{A^\star,f}) \le \underbrace{\frac{1}{S}\sum_{s=1}^{S} \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{1}\{f(X_i^s) \neq Y_i^s\}}_{\widehat{B}_{\mathrm{cls}}} + \underbrace{\sum_{c\in\mathcal{K}} \frac{\alpha_c}{N_c} \sum_{i:y_i=c} S_c^\star(z_i)}_{\widehat{B}_{\mathrm{cov}}}$$
$$+ \underbrace{\sum_{c\in\mathcal{K}} \beta_c (\sigma_c^{2\star})^{\frac{m-1-k_c}{2}} \prod_{j=1}^{k_c} (\sigma_c^{2\star} + \lambda_{c,j}^\star)^{1/2}}_{\widehat{B}_{\mathrm{vol}}} + \underbrace{\sum_{c\in\mathcal{K}} \frac{\alpha_c \varepsilon_c}{\sigma_c^{2\star}}}_{B_{\mathrm{rob}}} + \Gamma_{\mathrm{cls}}.$$

Here $S_c^\star(z) = \xi_c(z)^\top (M_c^\star)^{-1} \xi_c(z)$, $k_c = |\{j : \widehat{\gamma}_{c,j} > \tau_c\}|$, and α_c and β_c are defined in Theorem 1.

A.7 Proof of Theorem 2

Proof. For each domain distribution P, define the known-class classification risk

$$r_P(f) = P\left(f(X) \neq Y^* \mid Y^* \in \mathcal{K}\right).$$

(A) Bounding classification risk from samples. For every acceptance region A,

$$R^P_{\mathrm{mf}}(A, f) \leq r_P(f).$$

Since $\pi_{\mathcal{K}}(P) \leq 1$, it follows that

$$\begin{aligned} B_{\mathrm{cls}} &= \mathbb{E}_{P\sim\mu}\left[\pi_{\mathcal{K}}(P) R^P_{\mathrm{mf}}(A, f)\right] \\ &\leq \mathbb{E}_{P\sim\mu}\left[r_P(f)\right]. \end{aligned}$$

Let $\mathcal{E}_{\mathrm{cls}}$ denote the uniform classification generalization event. By the definition of Γ_{cls}, this event has probability at least $1 - \delta/2$, and on this event every $f \in \mathcal{F}$ satisfies

$$\mathbb{E}_{P\sim\mu}[r_P(f)] \leq \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s}\mathbf{1}\{f(X_i^s) \neq Y_i^s\} + \Gamma_{\mathrm{cls}}.$$

Consequently, on $\mathcal{E}_{\mathrm{cls}}$,

$$B_{\mathrm{cls}} \leq \widehat{B}_{\mathrm{cls}} + \Gamma_{\mathrm{cls}}$$

uniformly over $f \in \mathcal{F}$, where

$$\widehat{B}_{\mathrm{cls}} = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N_s}\sum_{i=1}^{N_s}\mathbf{1}\{f(X_i^s) \neq Y_i^s\}.$$

(B) Controlling second-moment uncertainty. For each $c \in \mathcal{K}$, apply Proposition 1 with confidence level

$$\delta_c = \frac{\delta}{2|\mathcal{K}|}.$$

Define

$$\varepsilon_c = b_c + r_{N,c}\left(\frac{\delta}{2|\mathcal{K}|}\right).$$

Then

$$P\left(\|T_c - \widehat{T}_c\|_* > \varepsilon_c\right) \leq \frac{\delta}{2|\mathcal{K}|}.$$

A union bound over all $c \in \mathcal{K}$ gives an event $\mathcal{E}_{\mathrm{T}}$ such that

$$P(\mathcal{E}_{\mathrm{T}}) \geq 1 - \frac{\delta}{2}$$

and

$$\|T_c - \widehat{T}_c\|_* \leq \varepsilon_c \qquad \text{for every } c \in \mathcal{K}$$

on this event.

For any $M_c \succ 0$, trace duality gives

$$\begin{aligned} \mathrm{tr}(T_c M_c^{-1}) &= \mathrm{tr}(\widehat{T}_c M_c^{-1}) + \mathrm{tr}\left((T_c - \widehat{T}_c)M_c^{-1}\right) \\ &\leq \mathrm{tr}(\widehat{T}_c M_c^{-1}) + \|T_c - \widehat{T}_c\|_* \|M_c^{-1}\|_{\mathrm{op}} \\ &\leq \mathrm{tr}(\widehat{T}_c M_c^{-1}) + \frac{\varepsilon_c}{\lambda_{\min}(M_c)}. \end{aligned}$$

For $M_c = M_c^\star$, the spiked representation implies

$$\lambda_{\min}(M_c^\star) = \sigma_c^{2\star}.$$

Therefore,

$$\mathrm{tr}\left(T_c (M_c^\star)^{-1}\right) \leq \mathrm{tr}\left(\widehat{T}_c (M_c^\star)^{-1}\right) + \frac{\varepsilon_c}{\sigma_c^{2\star}}.$$

(C) Assembling the finite-sample bound. Let

$$A^\star = \bigcup_{c\in\mathcal{K}} A_c(M_c^\star).$$

By Definition 2,

$$\mathcal{E}(h_{A^\star,f}) = B_{\mathrm{cls}} + R_{\mathrm{FR}}(A^\star) + R_{\mathrm{FA}}(A^\star).$$

Theorem 1 gives

$$R_{\mathrm{FR}}(A^\star) + R_{\mathrm{FA}}(A^\star) \le \sum_{c\in\mathcal{K}} \left[\alpha_c \operatorname{tr}\left(T_c (M_c^\star)^{-1}\right) + \beta_c \det(M_c^\star)^{1/2}\right].$$

On $\mathcal{E}_{\mathrm{T}}$, this is at most

$$\sum_{c\in\mathcal{K}} \Bigg[\alpha_c \operatorname{tr}\left(\widehat{T}_c (M_c^\star)^{-1}\right) + \frac{\alpha_c \varepsilon_c}{\sigma_c^{2\star}} + \beta_c \det(M_c^\star)^{1/2}\Bigg].$$

Using the pooled class-wise empirical second moment

$$\widehat{T}_c = \frac{1}{N_c} \sum_{s=1}^{S} \sum_{i=1}^{N_s} \mathbf{1}\{Y_i^s = c\} \xi_c(z_i^s) \xi_c(z_i^s)^\top,$$

we have

$$\begin{aligned}\operatorname{tr}\left(\widehat{T}_c (M_c^\star)^{-1}\right) &= \frac{1}{N_c} \sum_{s=1}^{S} \sum_{i=1}^{N_s} \mathbf{1}\{Y_i^s = c\} \\ &\qquad \cdot \xi_c(z_i^s)^\top (M_c^\star)^{-1} \xi_c(z_i^s) \\ &= \frac{1}{N_c} \sum_{i:y_i=c} S_c^\star(z_i).\end{aligned}$$

Hence

$$\sum_{c\in\mathcal{K}} \alpha_c \operatorname{tr}\left(\widehat{T}_c (M_c^\star)^{-1}\right) = \sum_{c\in\mathcal{K}} \frac{\alpha_c}{N_c} \sum_{i:y_i=c} S_c^\star(z_i).$$

Let

$$k_c = \left|\{j : \widehat{\gamma}_{c,j} > \tau_c\}\right|.$$

The eigenvalues of $M_c^\star$ are

$$\sigma_c^{2\star} + \lambda_{c,j}^\star, \qquad j = 1, \dots, k_c,$$

along the active directions, and

$$\sigma_c^{2\star}$$

along the remaining $m - 1 - k_c$ directions. Therefore,

$$\det(M_c^\star)^{1/2} = (\sigma_c^{2\star})^{(m-1-k_c)/2} \prod_{j=1}^{k_c} (\sigma_c^{2\star} + \lambda_{c,j}^\star)^{1/2}.$$

Combining these identities with the classification bound gives

$$\begin{aligned}\mathcal{E}(h_{A^\star,f}) \le{}& \frac{1}{S} \sum_{s=1}^{S} \frac{1}{N_s} \sum_{i=1}^{N_s} \mathbf{1}\{f(X_i^s) \neq Y_i^s\} \\ &+ \sum_{c\in\mathcal{K}} \frac{\alpha_c}{N_c} \sum_{i:y_i=c} S_c^\star(z_i) \\ &+ \sum_{c\in\mathcal{K}} \beta_c (\sigma_c^{2\star})^{(m-1-k_c)/2} \prod_{j=1}^{k_c} (\sigma_c^{2\star} + \lambda_{c,j}^\star)^{1/2} \\ &+ \sum_{c\in\mathcal{K}} \frac{\alpha_c \varepsilon_c}{\sigma_c^{2\star}} + \Gamma_{\mathrm{cls}}.\end{aligned}$$

The classification event satisfies

$$P(\mathcal{E}_{\text{cls}}) \geq 1 - \frac{\delta}{2},$$

and the simultaneous moment event satisfies

$$P(\mathcal{E}_{\text{T}}) \geq 1 - \frac{\delta}{2}.$$

A final union bound gives

$$P\left(\mathcal{E}_{\text{cls}} \cap \mathcal{E}_{\text{T}}\right) \geq 1 - \delta.$$

Since the classification event is uniform over $f \in \mathcal{F}$, the stated risk bound holds simultaneously for every $f \in \mathcal{F}$ with probability at least $1 - \delta$. □

B Experimental Details

B.1 Dataset and Protocol Details

PACS (Li et al., 2017) contains 9,991 images from seven classes across four domains: *photo*, *art painting*, *cartoon*, and *sketch*. OfficeHome (Venkateswara et al., 2017) contains 15,588 images from 65 classes across *Art*, *Clipart*, *Product*, and *Real World*. VLCS (Fang et al., 2013) contains 10,729 images from five classes across *Pascal VOC*, *LabelMe*, *Caltech*, and *SUN*.

We evaluate both ResNet-18 and CLIP on all three datasets under a leave-one-domain-out protocol. For each dataset, one domain is held out in turn as the unseen target, while the remaining domains are used as sources. The target domain is never accessed during training or model selection.

For the known/unknown class partition, we follow the protocols used in prior OSDG work. For PACS and OfficeHome, we use the partitions adopted by MEDIC (Wang et al., 2023; Peng et al., 2024), where classes are ordered alphabetically, yielding known-to-unknown ratios of 6:1 and 35:30, respectively. For VLCS, we follow ODG-CLIP (Singha et al., 2024), using the first four alphabetically ordered classes as known and the remaining class as unknown, resulting in a 4:1 ratio. For the source training/validation split, PACS and OfficeHome use the official lists provided by MEDIC (Wang et al., 2023). Since no corresponding split is provided for VLCS, we randomly split each known class within each source domain into training and validation subsets at a 9:1 ratio. In all datasets, the entire held-out target domain is used for testing. All known classes are present in every source domain, ensuring that the class-conditional second moments can be estimated consistently across domains. This makes the cross-domain uncertainty $\hat{b}_c$, and consequently the adaptive dimension $k_c^\star$, well defined for every known class.

B.2 Metrics

For a target sample x, let

$$\hat{c}(x) = \arg\min_{c \in \mathcal{K}} S_c(\phi(x)), \qquad r(x) = \min_{c \in \mathcal{K}} S_c(\phi(x)),$$

where a smaller $r(x)$ indicates stronger evidence that the sample belongs to a known class. Let $\mathcal{D}_K$ and $\mathcal{D}_U$ denote the known- and unknown-class subsets of the target domain, with sizes N_K and N_U, respectively.

Closed-set Accuracy. Closed-set Accuracy evaluates classification performance on target-domain known samples without applying the rejection threshold:

$$\text{Acc}_{\text{closed}} = \frac{1}{N_K} \sum_{(x_i, y_i) \in \mathcal{D}_K} \mathbf{1}\{\hat{c}(x_i) = y_i\}.$$

H-score. For a rejection threshold ρ, the known-class accuracy and unknown-class rejection accuracy are

$$\text{Acc}_K(\rho) = \frac{1}{N_K} \sum_{(x_i, y_i) \in \mathcal{D}_K} \mathbf{1}\{\hat{c}(x_i) = y_i,\ r(x_i) \leq \rho\},$$

and

$$\mathrm{Acc}_U(\rho) = \frac{1}{N_U} \sum_{x_i \in \mathcal{D}_U} \mathbf{1}\{r(x_i) > \rho\}\,.$$

The H-score is their harmonic mean:

$$H(\rho) = \frac{2\,\mathrm{Acc}_K(\rho)\,\mathrm{Acc}_U(\rho)}{\mathrm{Acc}_K(\rho) + \mathrm{Acc}_U(\rho)}.$$

Since the H-score depends on a manually specified rejection threshold ρ, we report it as a complementary open-set metric rather than our primary metric.

OSCR. The Open-Set Classification Rate (OSCR) evaluates the trade-off between correct classification of known samples and false acceptance of unknown samples over all rejection thresholds. For threshold t, define

$$\mathrm{CCR}(t) = \frac{1}{N_K} \sum_{(x_i,y_i) \in \mathcal{D}_K} \mathbf{1}\{\hat{c}(x_i) = y_i,\ r(x_i) \le t\}\,,$$

and

$$\mathrm{FPR}(t) = \frac{1}{N_U} \sum_{x_i \in \mathcal{D}_U} \mathbf{1}\{r(x_i) \le t\}\,.$$

OSCR is the area under the resulting CCR–FPR curve:

$$\mathrm{OSCR} = \int \mathrm{CCR}(t)\, d\,\mathrm{FPR}(t).$$

Because OSCR evaluates performance across all rejection thresholds and does not depend on a single manually chosen threshold, we use it as our primary open-set evaluation metric.

B.3 Extension with Pseudo Unknowns.

The volume term controls unknown exposure without requiring unknown samples. When pseudo-unknowns are available, we use them as additional negative evidence. Let $z_i^{\mathcal{U}}$ denote a pseudo-unknown feature. Since $z \in \mathcal{A}$ iff $\min_{c\in\mathcal{K}} S_c(z) \le \rho$, we encourage pseudo-unknowns to lie outside all known-class supports using

$$\mathcal{L}_{\mathrm{excl}} = \frac{1}{n_u} \sum_{i=1}^{n_u} \left[\tau_u - \min_{c\in\mathcal{K}} S_c(z_i^{\mathcal{U}})\right]_+^2, \qquad \tau_u \ge \rho. \tag{8}$$

For CLIP, we follow OSLoPrompt (C et al., 2025) to generate pseudo-unknowns from source-domain styles and optimize

$$\mathcal{L}_{\mathrm{total}}^{\mathrm{CLIP}} = \mathcal{L}_{\mathrm{total}} + \lambda_{\mathrm{excl}} \mathcal{L}_{\mathrm{excl}}.$$

B.4 Implementation Details

Table 5 summarizes the optimization, sampling, and training hyperparameters used in all experiments. SGD is used to optimize the trainable network parameters, while Adam is used for the support parameters. For ResNet-18, the full backbone is fine-tuned with a StepLR scheduler; for CLIP, the pretrained image and text encoders remain frozen and SGD updates only the residual adapter. All experiments use domain- and class-balanced sampling, with an equal number of samples drawn from each source domain- class pair at every optimization step. All ResNet-18 experiments were conducted on a single NVIDIA RTX 6000 Ada GPU with 48 GB of GPU memory, while all CLIP ViT-B/32 experiments were conducted on a single NVIDIA RTX PRO 6000 Blackwell Max-Q GPU with 96 GB of GPU memory.

Table 5: Implementation details. **(a) Optimization.** n is the number of samples per source domain–class pair per step (batch size $= 3 \times |\mathcal{K}| \times n$). Network parameters are optimized with SGD; support parameters with Adam (LR 5×10^{-3} for ResNet-18 and 1×10^{-5} for CLIP). $k^\star$ is estimated from source data before training (Sec. 3.1) and listed per held-out domain in the order PACS: A/C/P/S; OfficeHome: A/C/P/R; VLCS: C/L/P/S. **(b) Objectives.** The hinge thresholds are fixed to $\tau_{\text{cov}} = \tau_u = 1$ in all experiments. All hyperparameters are selected on source-domain validation splits; the target domain is never used.

(a) Optimization

Backbone	Dataset	n	Batch	LR	WD	Support LR	$k^\star$
ResNet-18	PACS	4	72	2×10^{-3}	5×10^{-4}	5×10^{-3}	42/44/44/56
	OfficeHome	2	210	3×10^{-3}	2×10^{-3}	5×10^{-3}	48/50/50/48
	VLCS	4	48	2×10^{-3}	5×10^{-4}	5×10^{-3}	48/48/54/48
CLIP ViT-B/32	PACS	20	360	2×10^{-4}	5×10^{-4}	1×10^{-5}	37/35/34/47
	OfficeHome	3	315	2×10^{-4}	2×10^{-2}	1×10^{-5}	45/47/49/46
	VLCS	14	168	3×10^{-3}	2×10^{-3}	1×10^{-5}	55/50/52/52

(b) Objectives

Backbone	Dataset	λ_{cov}	λ_{vol}	λ_{excl}
ResNet-18	PACS	1×10^{-3}	4×10^{-4}	–
	OfficeHome	1×10^{-3}	4×10^{-4}	–
	VLCS	1×10^{-3}	4×10^{-4}	–
CLIP ViT-B/32	PACS	2×10^{-4}	0.01	0.05
	OfficeHome	2×10^{-4}	0.05	0.05
	VLCS	2×10^{-4}	0.01	0.1

B.5 Ablation Results

Effect of Each Objective. $\mathcal{L}_{\text{cov}}$ and $\mathcal{L}_{\text{vol}}$ need to act jointly: the former improves coverage of domain-shifted known samples but may over-expand the support, whereas the latter can over-contract it and increase false rejection. Their combination improves H-score and OSCR by 1.49 and 1.25 points over $\mathcal{L}_{\text{CE}}$ alone. Ablation results on PACS with ResNet-18 are shown in Table 6.

Adaptive vs. Fixed Anisotropic Dimension. Comparisons of the source-adaptive anisotropic dimension $k^\star$ with fixed choices of k on Officehome and VLCS are shown in Fig. 5 and Fig. 6 respectively.

Table 6: Ablation results on PACS with ResNet-18.

$\mathcal{L}_{\text{CE}}$	$\mathcal{L}_{\text{cov}}$	$\mathcal{L}_{\text{vol}}$	**Closed Acc.** ↑	**H-score** ↑	**OSCR** ↑
✓			87.27	79.20	81.79
✓		✓	84.20	61.16	59.25
✓	✓		86.06	76.77	78.33
✓	✓	✓	**88.03**	**80.69**	**83.04**

B.6 Computational Overhead

SVD cost is amortized over the refresh interval (every 10 iterations); values are averaged over the four target domains. Training-time overhead of subspace estimation (ResNet-18, single RTX 6000 Ada) is shown in Table 7. Training-time overhead of subspace estimation (CLIP ViT-B/32, single RTX PRO 6000 Blackwell Max-Q)is shown in Table 8.

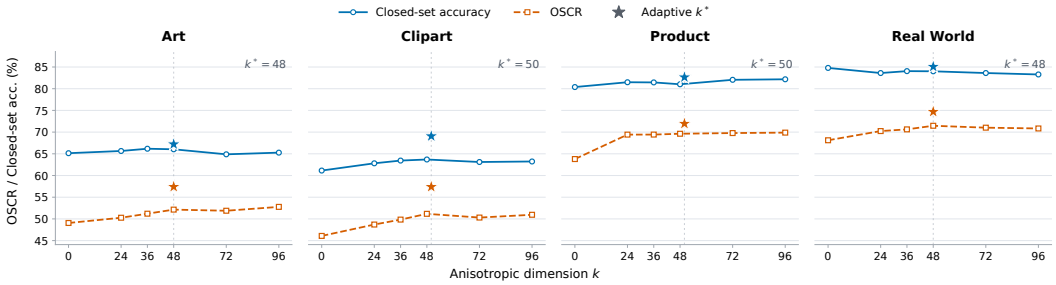


Figure 5: Comparison of the source-adaptive anisotropic dimension $k^\star$ with fixed choices of k on Officehome. Curves report Closed-set Accuracy and OSCR for models retrained with fixed $k \in \{0, 24, 36, 48, 72, 96\}$, while stars denote the performance obtained using the source-adaptively estimated $k^\star$ for each held-out domain.

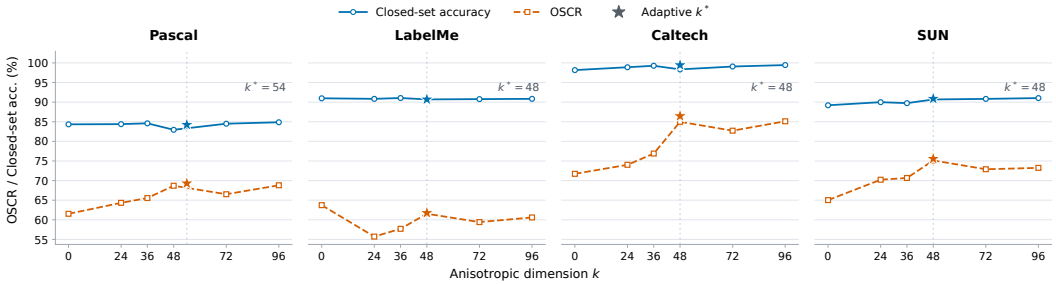


Figure 6: Comparison of the source-adaptive anisotropic dimension $k^\star$ with fixed choices of k on VLCS. Curves report Closed-set Accuracy and OSCR for models retrained with fixed $k \in \{0, 24, 36, 48, 72, 96\}$, while stars denote the performance obtained using the source-adaptively estimated $k^\star$ for each held-out domain.

Table 7: Training-time overhead of subspace estimation (ResNet-18, single RTX 6000 Ada).

Dataset	SVD / iter	Training step	Ratio
PACS	39.5 ms	199 ms	19.8%
OfficeHome	230.2 ms	1112 ms	20.7%
VLCS	26.6 ms	185 ms	14.4%

Table 8: Training-time overhead of subspace estimation (CLIP ViT-B/32, single RTX PRO 6000 Blackwell Max-Q).

Dataset	SVD / iter	Training step	Ratio
PACS	59.1 ms	1454 ms	4.1%
OfficeHome	192.4 ms	4025 ms	4.8%
VLCS	90.1 ms	1570 ms	5.7%

B.7 Full Results

Full per-domain results with ResNet-18 are provided in Tables 9 10 11 respectively.

Full per-domain results using CLIP are provided in Tables 12 13 14 respectively.

Table 9: Results (%) on PACS with ResNet-18. The open-set ratio is 6:1.

	Photo			Art			Cartoon			Sketch			Avg.		
Method	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR
MEDIC-cls	94.83	83.68	90.30	86.20	69.35	74.16	81.94	63.26	67.43	81.84	69.60	70.85	86.20	71.47	75.69
MEDIC-bcls	94.83	89.49	92.40	86.20	73.82	75.58	81.94	66.26	69.04	81.84	74.37	74.52	86.20	75.98	77.89
EBiL-HaDS-cls	95.80	91.54	94.62	87.24	71.87	74.15	**82.98**	68.55	71.62	83.21	74.89	74.50	87.31	76.71	78.72
EBiL-HaDS-bcls	95.80	**93.10**	94.42	87.24	75.66	77.19	**82.98**	67.57	72.22	83.21	**78.29**	77.52	87.31	78.66	80.34
MEDIC++-cls	**95.96**	90.16	92.94	87.80	75.37	77.16	82.88	68.75	69.32	80.98	72.28	70.74	86.91	76.64	77.54
MEDIC++-bcls	**95.96**	90.81	93.50	87.80	75.54	77.73	82.88	68.85	69.52	80.98	72.87	72.21	86.91	77.02	78.24
Ours	95.15	90.38	**95.01**	**90.99**	**81.62**	**84.07**	80.71	**72.49**	**73.54**	**85.27**	78.26	**79.55**	**88.03**	**80.69**	**83.04**

Table 10: Results (%) on OfficeHome with ResNet-18. The open-set ratio is 35:30

	Art			Clipart			Product			Real World			Avg.		
Method	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR
MEDIC-cls	66.81	55.78	55.85	61.14	54.21	48.51	85.03	71.16	73.15	80.69	67.72	68.09	73.42	62.22	61.40
MEDIC-bcls	66.81	51.76	56.21	61.14	53.28	48.97	85.03	70.61	74.08	80.69	67.70	67.17	73.42	60.84	61.61
EBiL-HaDS-cls	68.18	59.66	56.83	63.48	57.01	52.26	**85.48**	72.88	74.45	81.61	71.03	70.25	74.69	65.15	63.45
EBiL-HaDS-bcls	68.18	53.57	**57.49**	63.48	52.12	53.14	**85.48**	**74.20**	**75.64**	81.61	72.20	71.62	74.69	63.02	64.47
MEDIC++-cls	**69.34**	61.65	56.66	61.27	55.49	49.30	81.77	70.47	70.26	**85.11**	73.61	74.11	74.37	65.31	62.58
MEDIC++-bcls	**69.34**	61.36	56.48	61.27	55.52	49.10	81.77	70.45	70.36	**85.11**	73.94	**74.77**	74.37	65.32	62.68
Ours	67.21	**62.59**	57.40	**69.07**	**62.56**	**57.39**	82.64	72.35	71.94	85.07	**74.23**	74.69	**76.00**	**67.93**	**65.36**

Table 11: Results (%) on VLCS with ResNet-18. The open-set ratio is 4:1.

	Caltech			LabelMe			Pascal			Sun			Avg.		
Method	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR
MEDIC-cls	91.19	75.77	80.33	**91.40**	56.82	64.54	83.16	61.06	61.51	90.14	66.55	69.96	88.97	65.05	69.09
MEDIC-bcls	91.19	72.05	75.01	**91.40**	60.10	62.76	83.16	63.46	63.70	90.14	66.10	68.80	88.97	65.43	67.57
EBiL-HaDS-cls	96.51	**84.50**	**89.61**	91.19	58.70	59.08	78.21	62.33	60.83	90.78	65.47	69.12	89.17	67.75	69.66
EBiL-HaDS-bcls	96.51	77.08	83.80	91.19	57.90	59.73	78.21	61.87	60.33	90.78	65.59	68.34	89.17	65.61	68.05
MEDIC++-cls	97.43	79.38	85.59	91.19	64.88	64.52	80.50	61.48	60.27	**91.87**	70.06	73.28	90.25	68.95	70.92
MEDIC++-bcls	97.43	74.97	80.58	91.19	**65.01**	**64.60**	80.50	61.09	60.23	**91.87**	69.92	73.75	90.25	67.75	69.79
Ours	**99.45**	78.40	86.46	90.70	60.84	61.76	**84.23**	**67.83**	**69.30**	90.88	**72.18**	**75.56**	**91.32**	**69.81**	**73.27**

Table 12: Results (%) on PACS using CLIP. The open-set ratio is 6:1.

	Photo			Art			Cartoon			Sketch			Avg.		
Method	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR
SPG	**99.68**	99.33	99.61	93.43	91.46	92.03	97.11	88.61	96.84	87.02	60.68	85.60	94.31	85.02	93.52
DPStyler	**99.68**	94.26	**99.68**	96.37	**95.61**	95.61	**98.66**	97.64	98.50	83.84	86.72	83.24	94.64	93.56	94.26
OSLoPrompt	99.35	99.03	99.28	**97.00**	94.93	**96.31**	98.09	95.10	97.32	84.88	88.57	83.92	94.83	94.41	94.21
ODG-CLIP	91.24	87.43	88.00	91.90	86.20	87.92	98.63	**98.50**	**98.62**	89.25	90.50	88.29	92.76	90.66	90.71
SeeCLIP	99.43	99.43	99.43	95.75	95.30	95.47	97.42	94.03	97.13	85.81	90.57	85.62	94.60	94.83	94.41
Ours	**99.68**	**99.80**	**99.68**	96.75	94.52	96.11	97.58	95.60	97.11	**89.36**	**93.10**	**89.23**	**95.84**	**95.76**	**95.53**

Table 13: Results (%) on OfficeHome using CLIP. The open-set ratio is 35:30.

	Art			Clipart			Product			Real World			Avg.		
Method	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR
SPG	68.69	62.08	57.86	61.69	53.78	46.96	75.50	70.63	67.18	81.85	74.34	73.59	71.93	65.21	61.40
DPStyler	**84.54**	70.52	**73.18**	71.79	51.67	61.61	89.99	75.21	80.71	90.01	78.21	**82.87**	84.08	68.90	74.59
OSLoPrompt	81.57	67.63	67.25	75.30	64.63	61.64	91.27	74.97	83.96	89.23	73.56	78.03	84.34	70.20	72.72
ODG-CLIP	83.31	69.49	69.21	77.13	66.35	64.65	89.99	77.49	80.96	**91.19**	**79.42**	82.31	85.41	73.19	74.28
SeeCLIP	83.76	**71.81**	70.50	75.08	63.47	60.88	90.43	78.14	80.92	89.60	77.45	80.18	84.72	72.72	73.12
Ours	81.82	70.08	69.89	**81.76**	**67.83**	**68.01**	**93.29**	**79.67**	**84.18**	90.82	79.19	82.54	**86.92**	**74.19**	**76.16**

Table 14: Results (%) on VLCS using CLIP. The open-set ratio is 4:1.

	Caltech			LabelMe			Pascal			SUN			Avg.		
Method	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR	Closed	H-score	OSCR
SPG	97.25	83.68	90.53	89.36	56.44	55.80	92.17	71.05	74.45	83.89	65.22	66.53	90.67	69.10	71.83
DPStyler	99.82	99.52	99.81	90.77	**70.97**	**74.42**	94.14	**81.98**	**87.43**	86.32	67.92	71.46	92.76	80.10	83.28
OSLoPrompt	**100.00**	**99.81**	**100.00**	91.31	64.44	69.35	**94.49**	80.06	85.38	90.43	70.85	75.27	94.06	78.79	82.50
ODG-CLIP	99.27	90.94	95.56	90.42	67.00	69.20	85.88	69.90	71.51	92.42	72.08	76.53	92.00	74.98	78.20
SeeCLIP	99.82	91.45	93.79	92.88	68.33	70.08	94.19	72.89	78.51	93.81	74.06	79.74	95.18	76.68	80.53
Ours	99.63	98.46	99.45	**93.09**	67.11	70.19	93.23	79.08	83.90	**95.54**	**78.86**	**84.63**	**95.37**	**80.88**	**84.54**

C LIMITATIONS

First, although our theory includes a finite-sample estimation radius $r_{N,c}$, existing OSDG benchmarks typically contain only a small number of source domains. Under this limited-source setting, the resulting bound on $r_{N,c}$ can be relatively loose. We therefore omit this term in practice and rely on the empirical source discrepancy, leaving tighter estimation of $r_{N,c}$ for future work. Second, our current formulation assumes that each known class is observed in every source domain, which simplifies the estimation of cross-domain class-conditional discrepancy. Extending AdaSpike to source domains with missing or imbalanced class coverage is an important direction for future work.