HyCal: A Training-Free Prototype Calibration Method for Cross-Discipline Few-Shot Class-Incremental Learning

Supplementary Material

7. Supplementary on the complementary roles and mutual information between cosine and Mahalanobis measures

This supplementary section provides additional analysis supporting the complementary relationship between cosine similarity and Mahalanobis distance, as formalized in Theorem 1 and Theorem 2. While both measures are computed from the same embedding pairs, they capture fundamentally different geometric and statistical aspects—angular alignment versus covariance-adjusted magnitude—which together yield richer and more discriminative information than either measure alone.

7.1. Complementary roles of cosine and Mahalanobis measures

7.1.1. Preliminaries and statistical independence

Let $X, Y \in \mathbb{R}^d$ be random vectors and define their difference:

$$\Delta = X - Y.$$

We consider the following similarity measures:

$$C = \frac{X^\top Y}{\|X\|\|Y\|},$$
$$M = \sqrt{(X-Y)^\top \Sigma^{-1}(X-Y)}.$$

Cosine similarity C captures angular alignment between X and Y, while Mahalanobis distance M reflects the covariance-adjusted magnitude of their difference. Although both are derived from the same pair (X, Y), they emphasize distinct geometric aspects.

To obtain a tractable theoretical perspective, we adopt an idealized isotropic surrogate model. Assume that

$$\Delta \sim \mathcal{N}(0, \sigma^2 I_d),$$

and that Mahalanobis distance is computed with the true covariance $\Sigma = \sigma^2 I_d$. In this case,

$$M = \sqrt{\Delta^\top (\sigma^2 I_d)^{-1} \Delta} = \frac{\|\Delta\|}{\sigma}.$$

Moreover, we can represent Δ in polar form as

$$\Delta = RU,$$

where

$$R = \|\Delta\| \sim \sigma \cdot \chi_d, \qquad U = \frac{\Delta}{\|\Delta\|} \in \mathbb{S}^{d-1}.$$

For an isotropic Gaussian, the magnitude R and the direction U are statistically independent:

$$R \perp U.$$

In our surrogate model, we further idealize cosine similarity and Mahalanobis distance as being driven by these independent radial and angular factors: Mahalanobis distance depends only on R, while cosine similarity depends only on an angular component (which we denote generically by U). Formally, we model

$$C = f(U), \qquad M = g(R),$$

for some measurable functions f and g. This does not claim that empirical cosine similarity is exactly a function of $\Delta/\|\Delta\|$, but instead captures the intuition that cosine similarity is dominantly governed by angular variation. In contrast, the Mahalanobis distance is primarily determined by radial variation.

Theorem 1 (Independence of Cosine Similarity and Mahalanobis Distance under an Isotropic Model)**.** *Under the isotropic surrogate model described above, cosine similarity depends only on the angular factor U and Mahalanobis distance depends only on the radial factor R:*

$$C = f(U), \qquad M = g(R).$$

Since $R \perp U$, it follows that $C \perp M$, and thus their joint entropy decomposes as

$$H(C, M) = H(C) + H(M).$$

Proof. For an isotropic Gaussian $\Delta \sim \mathcal{N}(0, \sigma^2 I_d)$, the density of $\Delta \in \mathbb{R}^d$ is rotationally invariant and depends only on $\|\Delta\|$. In hyperspherical coordinates $\Delta = ru$ with $r = \|\Delta\|$ and $u \in \mathbb{S}^{d-1}$, the joint density factorizes as

$$p(r, u) = c\, r^{d-1} \exp\left(-\frac{r^2}{2\sigma^2}\right) p(u),$$

where c is a normalization constant and $p(u)$ is the uniform density on the unit sphere. Hence, $R \perp U$.

By our surrogate model assumption, $C = f(U)$ and $M = g(R)$ for measurable functions f and g. Since measurable functions of independent random variables remain

independent, we obtain $C \perp M$. Therefore, their joint entropy factorizes:

$$\begin{aligned} H(C, M) &= -\int p(c, m) \log p(c, m)\, dc\, dm \\ &= -\int p(c)p(m) \log\big(p(c)p(m)\big)\, dc\, dm \\ &= -\int p(c) \log p(c)\, dc - \int p(m) \log p(m)\, dm \\ &= H(C) + H(M). \end{aligned}$$

□

This result formalizes the idea that cosine similarity and Mahalanobis distance capture statistically independent aspects of the pair (X, Y) under an isotropic surrogate model: one is driven by angular variation, and the other by radial variation. In particular, the pair (C, M) encodes the sum of their individual entropies, reflecting complementary information about the underlying embedding geometry.

7.2. Mutual information interpretation

Let $L \in \{0, 1, \ldots, n\}$ denote a class label, and suppose that L depends on the relative relationship between C and M. We are interested in the mutual information quantities

$$I(L; C), \quad I(L; M), \quad I(L; C, M).$$

Theorem 2 (Information Gain from Combination)**.** *The mutual information between the label L and the combined similarity vector (C, M) satisfies*

$$\begin{aligned} I(L; C, M) &= I(L; C) + I(L; M \mid C) \\ &\geq \max\{I(L; C), I(L; M)\}. \end{aligned}$$

Proof. By the chain rule for mutual information, we have

$$I(L; C, M) = I(L; C) + I(L; M \mid C).$$

By definition of conditional mutual information,

$$I(L; M \mid C) = \mathbb{E}_{c\sim p(C)}\Big[D_{\mathrm{KL}}\big(p(L, M \mid C = c) \,\|\, p(L \mid C = c)\, p(M \mid C = c)\big)\Big].$$

and since the Kullback–Leibler divergence satisfies $D_{\mathrm{KL}}(\cdot \,\|\, \cdot) \geq 0$, it follows that

$$I(L; M \mid C) \geq 0.$$

Therefore,

$$\begin{aligned} I(L; C, M) &= I(L; C) + I(L; M \mid C) \\ &\geq I(L; C). \end{aligned}$$

By symmetry, $I(L; C, M) \geq I(L; M)$ as well, which yields

$$I(L; C, M) \geq \max\{I(L; C), I(L; M)\}.$$

□

Together with the independence result above, this theorem shows that cosine similarity and Mahalanobis distance are information-theoretically complementary: under the isotropic surrogate model, they encode independent angular and radial factors, and when labels depend on both, their combination (C, M) provides at least as much mutual information with L as either measure alone. This complementarity enables richer information extraction from frozen embeddings while keeping the backbone entirely fixed.

7.3. Real VLM embeddings: Why complementarity becomes stronger

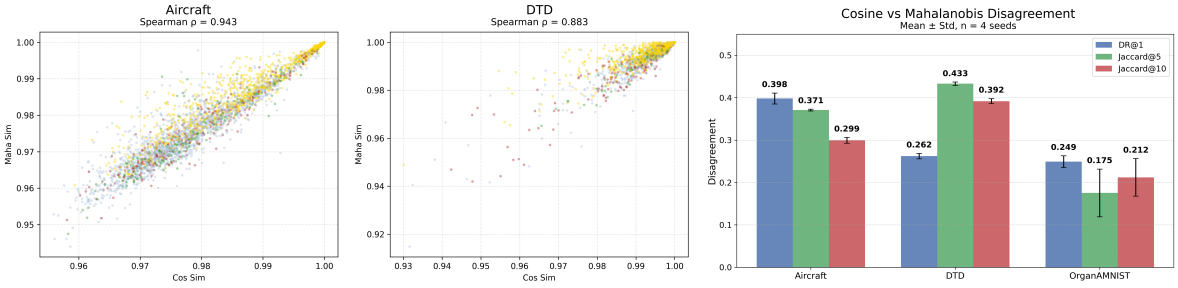
The independence result above is derived under an idealized isotropic Gaussian surrogate, in which the angular and radial components factorize. In real pretrained VLMs, however, the embedding geometry is inherently anisotropic: feature norms exhibit heavy-tailed distributions, covariance structures vary significantly across domains, and domain shifts induce direction–magnitude distortions in heterogeneous ways [25, 27].

These anisotropic effects break the exact statistical independence between cosine similarity and Mahalanobis distance. Yet, paradoxically, this makes the two measures *more* complementary in practice. Under anisotropy, directional cues captured by cosine similarity and covariance-aware magnitude cues captured by Mahalanobis distance tend to decouple, encoding increasingly orthogonal information regarding the data distribution, especially under domain shifts [14, 26].

Therefore, the theoretical independence shown in the isotropic case should be viewed as a conceptual lower bound of complementarity. In real VLM embeddings, the geometric anisotropy amplifies the distinct information encoded by each measure, making their combination strictly more discriminative than either measure alone.

7.4. Theoretical–empirical connection in HyCal

Secs. 7.1 to 7.3 establish that cosine similarity and Mahalanobis distance provide different but complementary views of prototype matching: cosine similarity captures directional alignment, whereas Mahalanobis distance measures deviation relative to class-specific covariance. HyCal translates this observation into a practical inference rule by combining both cues when evaluating each seen class. This is particularly useful in XD-VSCIL, where heterogeneous domains arrive sequentially and often with severe imbalance. Under such conditions, domain-level asymmetry can



(a) Relationship between cosine similarity and Mahalanobis similarity

(b) Decision-level disagreement at Top-1, Top-5, and Top-10

Figure 6. Cosine–Mahalanobis: (a) Relationship and (b) Ranking where cosine-correct (green) and Mahalanobis-correct (red) samples only partially overlap (yellow).

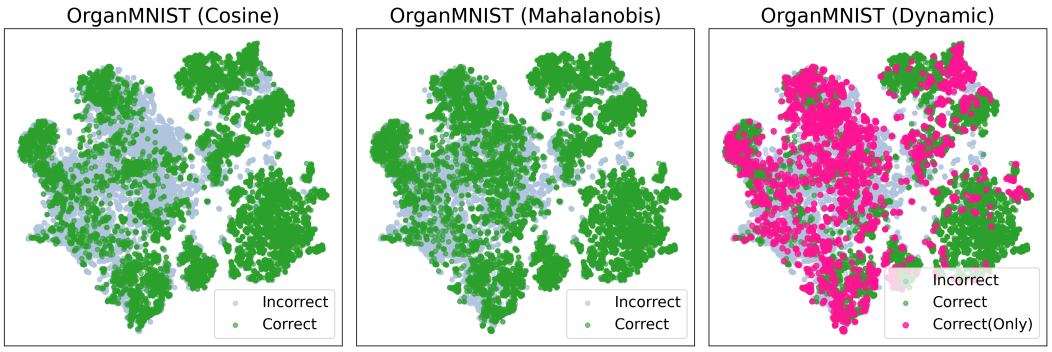


Figure 7. Qualitative comparison of distance metrics. Pink points show samples only correctly classified by dynamic summation.

distort the representation space and make a single metric overly biased toward only one geometric aspect. By reusing frozen pretrained representations and calibrating decisions at the prototype level, HyCal reduces sensitivity to such domain-level dominance without requiring backbone adaptation or task-specific retraining.

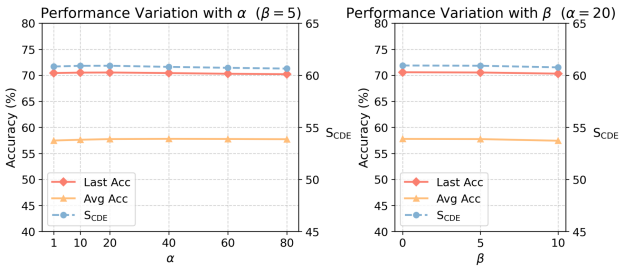
The empirical results are consistent with this interpretation, while also clarifying the limits of the theory. Real CLIP embeddings are anisotropic, so our analysis is not intended as an exact statistical description of the feature space. Nevertheless, the diagnostics in Figs. 6 and 7 show that cosine similarity and Mahalanobis distance make complementary decisions: they disagree on nontrivial subsets of samples and recover different correct predictions, especially in sparse domains such as Aircraft, DTD, and OrganMNIST. This complementarity translates into strong behavior under challenging imbalance settings, where HyCal remains robust on both overall metrics and low-data domains. Tab. 6 further shows that the same qualitative tendency extends across pretrained models. With `OpenCLIP ViT-B/16` pretrained on `DataComp-1B` [12], HyCal achieves the best performance across all three metrics, and under the X-TAIL [35] setting it attains the best Last Acc. while remaining competitive on the other metrics. Taken together, these results support the practical value of the theoretical complementarity developed in this section, while making clear that the theory serves primarily as intuition for real VLM embeddings rather than a strict independence claim.

Table 6. Performance under High-scale domain imbalance.

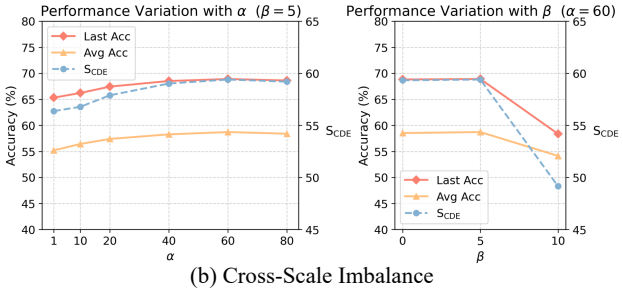
	PTM: `DataComp-1B` [12]			Setting: X-TAIL [35]		
Method	Last Acc.	Avg Acc.	S_{CDE}	Last Acc.	Avg Acc.	S_{CDE}
Primal-RAIL [35]	65.89	58.10	56.40	73.82	66.79	69.48
RanPAC [22]	66.20	55.78	55.64	69.74	67.60	68.20
KLDA [23]	63.95	41.24	51.50	76.71	**71.38**	**72.42**
HyCal (Ours)	**68.88**	**59.89**	**57.94**	**77.44**	64.41	70.66

8. Additional ablation study

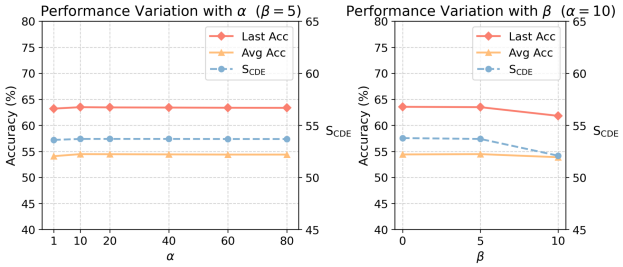
To further evaluate the stability and design choices of HyCal, we conduct ablation studies on three components: robustness to domain order, sensitivity to hyperparameters, and the effect of different image–text embedding fusion strategies. These analyses assess whether the method maintains consistent performance under different task permutations, whether its calibration behavior is sensitive to the selection of (α, β), and whether our sum-based image–text fusion scheme provides advantages over the commonly used concatenation strategy.



(a) Balanced-in-Class Domain



(b) Cross-Scale Imbalance



(c) High-Scale Domain Imbalance

Figure 8. Hyperparameter sensitivity analysis for α and β. Each curve shows the performance variation when sweeping one parameter while keeping the other fixed.

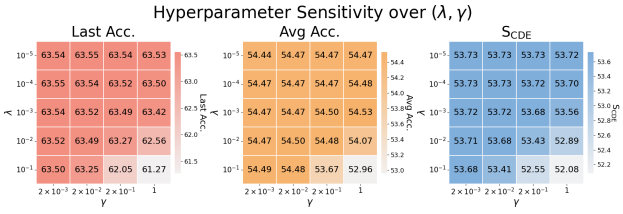


Figure 9. Hyperparameter sensitivity analysis for λ and γ under the High-scale domain imbalance setting. Heatmaps show Last Acc., Avg Acc., and S_{CDE} over different (λ, γ) configurations.

8.1. Robustness to domain order

To examine the robustness of each method under variations in domain sequences, we evaluate all approaches across four random domain orders, derived from fixed random seeds $\{0, 1, 42, 1993\}$. Each seed generates a unique permutation of the eight domains—Aircraft, ArtBench, DTD, EuroSAT, Galaxy, MNIST, Organ, and Flower—which are presented sequentially as incremental tasks. These four permutations are referred to as Random order I–IV.

- *Random order I (seed = 0)*
  Galaxy → ArtBench → MNIST → DTD → Aircraft → EuroSAT → Flower → Organ
- *Random order II (seed = 1)*
  EuroSAT → Organ → ArtBench → MNIST → Flower → Aircraft → Galaxy → DTD
- *Random order III (seed = 42)*
  EuroSAT → Galaxy → Organ → Flower → DTD → MNIST → Aircraft → ArtBench
- *Random order IV (seed = 1993)*
  ArtBench → MNIST → Organ → Aircraft → EuroSAT → DTD → Galaxy → Flower

The overall performance of all compared methods under Random order I–IV is summarized in Tab. 8, covering three XD-VSCIL settings.

Across all domain orders and settings, HyCal consistently achieves the highest performance, indicating that its calibration mechanism remains stable and effective regardless of task permutation. Furthermore, to provide a more fine-grained perspective, Tab. 9 reports the class-wise performance under the High-scale domain imbalanced setting for each of the four random domain orders.

Table 7. Comparison of fusion strategies under three domain conditions.

Fusion	Last Acc.	Avg Acc.	S_{CDE}
Balanced-in-Class Domain			
Concat	$70.25_{\pm 0.20}$	$57.33_{\pm 0.32}$	$60.70_{\pm 0.15}$
Sum (Ours)	$\mathbf{70.54}_{\pm \mathbf{0.23}}$	$\mathbf{57.75}_{\pm \mathbf{0.45}}$	$\mathbf{60.92}_{\pm \mathbf{0.17}}$
Cross-Scale Imbalance			
Concat	$68.30_{\pm 1.02}$	$57.96_{\pm 1.16}$	$58.95_{\pm 0.79}$
Sum (Ours)	$\mathbf{68.69}_{\pm \mathbf{0.44}}$	$\mathbf{58.43}_{\pm \mathbf{0.51}}$	$\mathbf{59.24}_{\pm \mathbf{0.41}}$
High-Scale Domain Imbalance			
Concat	$63.06_{\pm 0.63}$	$53.94_{\pm 0.76}$	$53.39_{\pm 0.71}$
Sum (Ours)	$\mathbf{63.50}_{\pm \mathbf{0.73}}$	$\mathbf{54.48}_{\pm \mathbf{0.78}}$	$\mathbf{53.70}_{\pm \mathbf{0.74}}$

8.2. Hyperparameter sensitivity

8.2.1. Sensitivity to fusion weights α and β

We examine the influence of the hyperparameters α and β by evaluating performance across the search ranges $\alpha \in \{1, 10, 20, 40, 60, 80\}$ and $\beta \in \{0, 5, 10\}$. Overall, HyCal shows stable performance across this hyperparameter space, with calibration strength varying smoothly as the activation point and smoothness parameters change. The effect of varying α and β is visualized in Fig. 8.

Across the examined range, performance is generally robust with respect to α, indicating that the calibration activation point does not strongly influence the overall behav-

Table 8. Performance comparison across random-order domain setting. Results are averaged over 4 seeds with 95% confidence intervals in subscripts.

	Random Order I			*Random Order II*			*Random Order III*			*Random Order IV*		
	Last Acc.	Avg Acc.	S_{CDE}	Last Acc.	Avg Acc.	S_{CDE}	Last Acc.	Avg Acc.	S_{CDE}	Last Acc.	Avg Acc.	S_{CDE}
Balanced-in-Class Domain												
Primal-RAIL [35]	$67.90_{\pm 0.59}$	$63.53_{\pm 0.85}$	$59.77_{\pm 0.40}$	$67.67_{\pm 0.38}$	$76.41_{\pm 0.31}$	$59.45_{\pm 0.26}$	$67.63_{\pm 0.87}$	$74.26_{\pm 1.00}$	$59.42_{\pm 0.51}$	$67.58_{\pm 0.56}$	$68.18_{\pm 0.38}$	$59.78_{\pm 0.40}$
FeCAM [6]	$15.26_{\pm 2.99}$	$18.17_{\pm 2.30}$	$10.16_{\pm 1.99}$	$12.63_{\pm 2.07}$	$21.95_{\pm 6.73}$	$10.24_{\pm 1.20}$	$13.97_{\pm 4.85}$	$20.81_{\pm 5.63}$	$9.30_{\pm 3.23}$	$11.70_{\pm 5.90}$	$27.09_{\pm 7.01}$	$12.90_{\pm 4.08}$
RanPAC [22]	$67.60_{\pm 0.25}$	$63.43_{\pm 0.13}$	$59.13_{\pm 0.17}$	$67.21_{\pm 1.00}$	$75.84_{\pm 1.27}$	$59.05_{\pm 0.63}$	$66.03_{\pm 3.00}$	$74.16_{\pm 0.18}$	$58.71_{\pm 1.16}$	$67.30_{\pm 0.42}$	$66.81_{\pm 0.60}$	$59.04_{\pm 0.20}$
KLDA [23]	$\mathbf{70.80_{\pm 0.30}}$	$51.39_{\pm 0.36}$	$56.41_{\pm 0.06}$	$70.14_{\pm 0.21}$	$58.11_{\pm 0.47}$	$53.47_{\pm 0.28}$	$70.56_{\pm 0.59}$	$57.61_{\pm 0.66}$	$55.04_{\pm 0.42}$	$70.50_{\pm 0.59}$	$53.95_{\pm 0.27}$	$54.42_{\pm 0.37}$
HyCal (Ours)	$70.58_{\pm 0.32}$	$\mathbf{64.40_{\pm 0.59}}$	$\mathbf{60.91_{\pm 0.23}}$	$\mathbf{70.53_{\pm 0.24}}$	$\mathbf{78.53_{\pm 0.33}}$	$\mathbf{60.86_{\pm 0.16}}$	$\mathbf{70.69_{\pm 0.38}}$	$\mathbf{76.18_{\pm 0.24}}$	$\mathbf{61.07_{\pm 0.27}}$	$\mathbf{70.65_{\pm 0.30}}$	$\mathbf{70.24_{\pm 0.35}}$	$\mathbf{60.95_{\pm 0.23}}$
Cross-Scale Imbalance												
Primal-RAIL [35]	$58.40_{\pm 1.54}$	$53.57_{\pm 3.12}$	$52.76_{\pm 1.65}$	$58.31_{\pm 1.18}$	$67.40_{\pm 1.81}$	$52.74_{\pm 1.25}$	$58.09_{\pm 1.16}$	$64.70_{\pm 1.94}$	$52.05_{\pm 1.49}$	$58.14_{\pm 0.68}$	$59.18_{\pm 0.56}$	$52.94_{\pm 0.92}$
FeCAM [6]	$16.65_{\pm 3.51}$	$19.75_{\pm 5.18}$	$13.41_{\pm 3.88}$	$17.56_{\pm 6.27}$	$22.46_{\pm 10.36}$	$13.57_{\pm 4.80}$	$16.98_{\pm 3.92}$	$20.12_{\pm 7.07}$	$13.03_{\pm 3.41}$	$17.31_{\pm 2.68}$	$21.23_{\pm 6.35}$	$13.85_{\pm 2.13}$
RanPAC [22]	$61.77_{\pm 6.12}$	$56.97_{\pm 4.13}$	$57.49_{\pm 4.03}$	$59.27_{\pm 7.85}$	$69.77_{\pm 3.21}$	$55.11_{\pm 4.44}$	$59.39_{\pm 7.56}$	$68.39_{\pm 3.33}$	$56.46_{\pm 4.52}$	$61.50_{\pm 5.64}$	$61.49_{\pm 3.46}$	$54.98_{\pm 2.97}$
KLDA [23]	$\mathbf{69.26_{\pm 0.44}}$	$43.53_{\pm 0.74}$	$48.14_{\pm 0.90}$	$69.34_{\pm 0.78}$	$46.79_{\pm 0.61}$	$47.68_{\pm 0.36}$	$\mathbf{69.31_{\pm 0.16}}$	$47.67_{\pm 0.38}$	$48.29_{\pm 0.39}$	$\mathbf{69.12_{\pm 1.04}}$	$44.27_{\pm 0.55}$	$46.76_{\pm 1.18}$
HyCal (Ours)	$68.69_{\pm 0.96}$	$\mathbf{61.27_{\pm 1.76}}$	$\mathbf{58.99_{\pm 0.71}}$	$\mathbf{68.60_{\pm 0.58}}$	$\mathbf{73.70_{\pm 0.35}}$	$\mathbf{58.99_{\pm 0.46}}$	$68.64_{\pm 0.85}$	$\mathbf{71.94_{\pm 0.81}}$	$\mathbf{59.06_{\pm 0.76}}$	$68.64_{\pm 1.39}$	$\mathbf{66.52_{\pm 1.06}}$	$\mathbf{58.97_{\pm 1.08}}$
High-Scale Domain Imbalance												
Primal-RAIL [35]	$59.80_{\pm 1.18}$	$52.60_{\pm 3.58}$	$52.18_{\pm 1.23}$	$59.28_{\pm 1.57}$	$65.78_{\pm 1.53}$	$51.51_{\pm 1.38}$	$59.62_{\pm 0.55}$	$61.48_{\pm 1.27}$	$52.10_{\pm 0.45}$	$60.28_{\pm 0.61}$	$61.58_{\pm 0.48}$	$52.74_{\pm 0.60}$
FeCAM [6]	$5.68_{\pm 0.73}$	$8.93_{\pm 0.74}$	$2.53_{\pm 0.21}$	$5.53_{\pm 1.31}$	$7.90_{\pm 0.25}$	$2.70_{\pm 2.70}$	$5.66_{\pm 0.40}$	$8.64_{\pm 3.29}$	$2.35_{\pm 0.56}$	$6.00_{\pm 1.11}$	$10.15_{\pm 1.25}$	$2.94_{\pm 0.57}$
RanPAC [22]	$59.80_{\pm 1.68}$	$51.61_{\pm 1.85}$	$\mathbf{54.40_{\pm 1.62}}$	$59.79_{\pm 1.38}$	$65.28_{\pm 1.53}$	$52.81_{\pm 0.65}$	$59.77_{\pm 1.51}$	$63.91_{\pm 1.21}$	$\mathbf{54.37_{\pm 1.12}}$	$60.31_{\pm 0.63}$	$56.08_{\pm 2.20}$	$51.24_{\pm 0.91}$
KLDA [23]	$61.68_{\pm 1.24}$	$31.03_{\pm 0.81}$	$41.85_{\pm 0.94}$	$60.67_{\pm 1.13}$	$32.89_{\pm 0.74}$	$34.25_{\pm 1.08}$	$61.13_{\pm 0.52}$	$39.60_{\pm 1.12}$	$40.81_{\pm 0.28}$	$60.99_{\pm 0.68}$	$31.44_{\pm 1.53}$	$39.28_{\pm 0.51}$
HyCal (Ours)	$\mathbf{63.39_{\pm 0.48}}$	$\mathbf{54.96_{\pm 0.81}}$	$53.47_{\pm 0.37}$	$\mathbf{63.01_{\pm 1.08}}$	$\mathbf{65.71_{\pm 1.17}}$	$\mathbf{53.13_{\pm 0.94}}$	$\mathbf{63.15_{\pm 0.74}}$	$\mathbf{64.73_{\pm 1.70}}$	$53.15_{\pm 0.73}$	$\mathbf{63.03_{\pm 1.11}}$	$\mathbf{59.69_{\pm 1.70}}$	$\mathbf{53.06_{\pm 1.22}}$

ior of the method. In contrast, β exhibits a clearer trend: larger values relax the threshold that determines when Mahalanobis distance becomes active, effectively reducing its contribution and allowing cosine similarity to dominate. As a result, performance consistently drops when moving from small values such as $\beta = 0, 5$ to higher values like $\beta = 10$, indicating that weakening this threshold diminishes the benefit obtained from the Mahalanobis component.

8.2.2. Sensitivity to covariance regularization (λ and γ)

We also examine the influence of the covariance-regularization parameters λ and γ in Eq. (10). For this analysis, we fix $\alpha = 10$ and $\beta = 5$, and evaluate the method under the High-scale domain imbalance setting. We use $\lambda = 10^{-4}$ and $\gamma = 1$ by default, and vary λ from 10^{-1} to 10^{-5} and γ from 1 to 2×10^{-3}, as shown in Fig. 9. Overall, performance remains stable across a wide range of moderate settings, indicating that HyCal does not require delicate tuning of these parameters.

8.3. Fusion strategies

We compare two image–text fusion strategies in Tab. 7: summation, which preserves the 512-dimensional CLIP embedding, and concatenation, which corresponds to the 1024-dimensional representation. Across all settings, the sum-based fusion consistently achieves better performance than the concatenation fusion.

9. Efficiency analysis

Because HyCal keeps the pretrained backbone frozen and updates only class prototypes and regularized precision matrices for newly introduced classes, its computational cost scales with the number of new classes rather than with full model retraining. Unlike prior approaches that recompute statistics over all classes or domains in XD-VSCIL, HyCal computes and stores covariance-related statistics only for newly added classes while reusing previously constructed prototypes. This design yields linear growth in both memory and computation as the class set expands, making the method well suited to incremental deployment.

This efficiency is also reflected in practice. When 10 new classes are added, HyCal requires only 5.4 GFLOPs, whereas full retraining in Primal-RAIL requires 12 GFLOPs. The required storage remains moderate, amounting to 301 MB when maintaining statistics for $C = 300$ classes. In addition, prototype construction and covariance estimation are independent across classes, enabling straightforward parallelization across newly introduced categories. Taken together, these properties make HyCal an efficient and scalable alternative for XD-VSCIL, especially in settings where new classes or domains must be incorporated quickly without repeated parameter optimization.

10. Detailed experimental setting

10.1. Implementation details

We use the Vision Transformer (ViT-B/16) model with a frozen CLIP text encoder for all experiments. The model weights are loaded from the `openai/clip-vit-base-patch16` checkpoint. The image encoder's parameters are kept frozen throughout all experiments. All experiments were conducted using Py-

Torch version 2.1.1. All training and evaluation runs were performed on a single NVIDIA GeForce RTX 3090 GPU. All images are resized to 224×224 and normalized using standard CLIP statistics.

HyCal is implemented on frozen `CLIP ViT-B/16` [28] features using the original 512-dimensional embedding, whereas Primal-RAIL and other training-free baselines operate on 1024-dimensional embeddings. This setup ensures a fair comparison by contrasting HyCal's training-free, low-dimensional configuration against both training-free and fully trainable alternatives.

To ensure the robustness of our results, all experiments were run four times with different random seeds: {0, 1, 42, 1993}. These seeds control the initialization of any trainable parameters, the selection of few-shot examples, and the order of data shuffling during training. The results reported in the main paper are the mean and standard deviation over these four runs.

HyCal uses two fusion weights hyperparameters, α and β. We perform a lightweight grid search over $\alpha \in \{1, 10, 20, 40, 60, 80\}$ and $\beta \in \{0, 5, 10\}$ and select the best configuration for each setting. For XD-VSCIL, we set $(\alpha, \beta) = (20, 5)$ for the Balanced-in-class domain setting, $(60, 5)$ for the Cross-scale imbalance setting, and $(10, 5)$ for the High-scale domain imbalance setting. For FSCIL, we use $(1, 0)$ in the 5-shot case to allow immediate activation under extremely sparse data, and $(10, 5)$ for the 10-, 15-, and 20-shot settings. We use $\lambda = 10^{-4}$ and $\gamma = 1$ in all reported experiments.

10.2. Evaluation metrics

We report three metrics: Last accuracy, Average accuracy, and the proposed S_{CDE}. As S_{CDE} is formally defined and analyzed in the main text, we provide concise descriptions of the remaining two metrics. Last accuracy denotes the Top-1 accuracy on the final task after all incremental stages and reflects the model's ability to retain previously learned knowledge. Average accuracy is computed as the mean Top-1 accuracy across all tasks,

$$\text{Avg Acc.} = \frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} \text{Acc}(t),$$

where $\mathcal{T}$ is the total number of steps, and provides a summary of overall performance throughout the entire learning trajectory.

10.3. XD-VSCIL settings

We evaluate our training-free hybrid calibration method HyCal by comparing it against both training-free and trainable baselines. As training-free approaches, we include FeCAM [6], RanPAC [22], and KLDA [23], which represent recent state-of-the-art methods relying solely on frozen visual encoders. For the trainable baseline, we adopt Primal-RAIL [35] and report its performance as a reference point for methods that fine-tune task-specific modules.

The training-sample distributions for three XD-VSCIL settings are illustrated in Fig. 3(a)–(c), which highlight the distinct balance/imbalance patterns observed in each setting. In the Cross-scale imbalance setting, a variable number of training samples per class (K_c) is used, and the exact values depend on the random seed. For reference, Tab. 10 and Tab. 11 report the specific K_c values employed when the random seed is set to 42.

10.3.1. FSCIL setting.

We follow the standard few-shot evaluation protocol. For each dataset, we conduct experiments in 5, 10, 15, 20-shot settings. In a K-shot setting, we randomly sample K images per class from the original training set.

11. Numerical results of Balanced-in-class domain and Cross-scale imbalance settings

For completeness, we provide the numerical values for the Balanced-in-class domain and Cross-scale imbalance settings. The Balanced-in-Class Domain setting results are provided in Tab. 12, and the Cross-scale imbalance setting results are summarized in Tab. 13.

12. Limitations

HyCal is designed for settings in which frozen pretrained representations are already sufficiently informative and where prototype-level calibration is preferable to backbone adaptation. Accordingly, its effectiveness depends on the quality of the underlying representation space. When newly arriving tasks come from domains that lie far outside the support of the pretrained model, or when image and text embeddings are strongly misaligned, prototype calibration alone may be insufficient to recover fully discriminative structure.

In addition, although the cosine similarity and regularized Mahalanobis term improve stability in few-shot regimes, covariance estimation can still become noisy when only extremely small numbers of samples are available or when class distributions are highly multimodal. Finally, HyCal deliberately avoids parameter updates in order to preserve efficiency and retention, which is advantageous in data-scarce continual settings, but may limit the attainable gains in scenarios where abundant in-domain data and larger adaptation budgets make trainable methods viable.

Table 9. Performance across random domain orders (I–IV) under the High-scale domain imbalanced setting. Results are averaged over 4 seeds with 95% confidence intervals in subscripts. Domain names are abbreviated.

Random Order I

	Galaxy	ArtBench	MNIST	DTD	Aircraft	EuroSAT	Flower	Organ	**Average**	σ
Zero-shot	9.80	50.88	44.01	41.90	23.91	37.58	67.40	17.97	56.41	18.76
Average Acc.										
Primal-RAIL [35]	$25.37_{\pm 10.57}$	$60.23_{\pm 0.72}$	$76.99_{\pm 6.49}$	$68.93_{\pm 1.35}$	$37.45_{\pm 1.18}$	$72.42_{\pm 1.14}$	$90.49_{\pm 0.58}$	$53.83_{\pm 3.88}$	$52.60_{\pm 3.58}$	$21.47_{\pm 2.62}$
FeCAM [6]	$2.89_{\pm 1.46}$	$5.76_{\pm 0.64}$	$0.00_{\pm 0.00}$	$31.77_{\pm 5.68}$	$11.88_{\pm 0.98}$	$0.00_{\pm 0.00}$	$1.76_{\pm 3.04}$	$0.00_{\pm 0.00}$	$8.93_{\pm 0.74}$	$10.92_{\pm 1.86}$
RanPAC [22]	$37.02_{\pm 5.25}$	$40.30_{\pm 5.96}$	$\mathbf{77.63_{\pm 3.51}}$	$66.33_{\pm 0.55}$	$35.95_{\pm 1.11}$	$71.66_{\pm 4.17}$	$92.45_{\pm 0.48}$	$56.07_{\pm 3.90}$	$51.61_{\pm 1.85}$	$20.96_{\pm 0.72}$
KLDA [23]	$19.98_{\pm 0.50}$	$22.89_{\pm 2.32}$	$53.28_{\pm 3.68}$	$54.83_{\pm 1.18}$	$39.53_{\pm 1.95}$	$\mathbf{76.60_{\pm 2.75}}$	$\mathbf{95.74_{\pm 0.50}}$	$\mathbf{58.78_{\pm 2.59}}$	$31.03_{\pm 0.81}$	$25.65_{\pm 0.52}$
HyCal(Ours)	$\mathbf{38.32_{\pm 2.02}}$	$\mathbf{50.27_{\pm 2.52}}$	$77.42_{\pm 5.24}$	$\mathbf{69.59_{\pm 0.23}}$	$\mathbf{41.67_{\pm 0.34}}$	$76.59_{\pm 3.27}$	$95.69_{\pm 0.38}$	$58.42_{\pm 2.72}$	$\mathbf{54.96_{\pm 0.81}}$	$19.87_{\pm 0.65}$
Last Acc.										
Primal-RAIL [35]	$19.90_{\pm 8.04}$	$60.32_{\pm 0.59}$	$76.76_{\pm 6.69}$	$68.03_{\pm 1.51}$	$37.23_{\pm 1.14}$	$71.89_{\pm 1.64}$	$90.46_{\pm 0.54}$	$53.83_{\pm 3.88}$	$59.80_{\pm 1.18}$	$22.69_{\pm 2.04}$
FeCAM [6]	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$31.77_{\pm 5.68}$	$11.88_{\pm 0.98}$	$0.00_{\pm 0.00}$	$1.76_{\pm 3.04}$	$0.00_{\pm 0.00}$	$5.68_{\pm 0.73}$	$11.34_{\pm 1.83}$
RanPAC [22]	$37.42_{\pm 4.29}$	$40.41_{\pm 6.27}$	$76.91_{\pm 3.83}$	$67.02_{\pm 1.02}$	$36.87_{\pm 1.04}$	$71.24_{\pm 4.36}$	$92.44_{\pm 0.49}$	$56.07_{\pm 3.90}$	$59.80_{\pm 1.68}$	$20.64_{\pm 0.71}$
KLDA [23]	$36.13_{\pm 2.28}$	$37.82_{\pm 4.62}$	$\mathbf{78.97_{\pm 5.62}}$	$\mathbf{69.39_{\pm 1.70}}$	$39.86_{\pm 1.98}$	$\mathbf{76.74_{\pm 2.68}}$	$\mathbf{95.77_{\pm 0.41}}$	$\mathbf{58.78_{\pm 2.59}}$	$61.68_{\pm 1.24}$	$22.27_{\pm 0.89}$
HyCal(Ours)	$\mathbf{38.15_{\pm 1.82}}$	$\mathbf{49.66_{\pm 2.42}}$	$77.72_{\pm 5.20}$	$69.27_{\pm 0.43}$	$\mathbf{41.67_{\pm 0.36}}$	$76.55_{\pm 3.27}$	$95.69_{\pm 0.38}$	$58.42_{\pm 2.72}$	$\mathbf{63.39_{\pm 0.48}}$	$19.98_{\pm 0.64}$

Random Order II

	EuroSAT	Organ	ArtBench	MNIST	Flower	Aircraft	Galaxy	DTD	**Average**	σ
Zero-shot	37.58	17.97	50.88	44.01	67.40	23.91	9.80	41.90	56.41	18.76
Average Acc.										
Primal-RAIL [35]	$72.84_{\pm 3.93}$	$54.25_{\pm 3.81}$	$60.03_{\pm 0.66}$	$77.83_{\pm 6.33}$	$91.75_{\pm 0.58}$	$37.86_{\pm 0.67}$	$22.22_{\pm 10.58}$	$67.77_{\pm 1.44}$	$65.78_{\pm 0.87}$	$22.48_{\pm 2.61}$
FeCAM [6]	$3.24_{\pm 1.05}$	$2.67_{\pm 1.31}$	$4.71_{\pm 1.08}$	$0.00_{\pm 0.00}$	$23.22_{\pm 7.27}$	$12.44_{\pm 1.70}$	$0.00_{\pm 0.00}$	$31.24_{\pm 8.07}$	$7.90_{\pm 0.25}$	$11.80_{\pm 2.76}$
RanPAC [22]	$72.90_{\pm 3.17}$	$56.52_{\pm 3.40}$	$45.90_{\pm 3.80}$	$\mathbf{80.32_{\pm 6.70}}$	$91.77_{\pm 1.33}$	$36.54_{\pm 0.14}$	$\mathbf{39.21_{\pm 4.30}}$	$67.32_{\pm 1.28}$	$65.28_{\pm 1.53}$	$20.14_{\pm 2.11}$
KLDA [23]	$39.06_{\pm 1.55}$	$33.26_{\pm 1.15}$	$27.34_{\pm 1.10}$	$64.01_{\pm 2.90}$	$90.94_{\pm 0.28}$	$38.64_{\pm 1.43}$	$33.27_{\pm 7.29}$	$68.63_{\pm 0.92}$	$32.89_{\pm 0.74}$	$22.53_{\pm 1.25}$
HyCal(Ours)	$\mathbf{73.33_{\pm 3.32}}$	$\mathbf{56.56_{\pm 3.97}}$	$\mathbf{51.28_{\pm 2.75}}$	$76.76_{\pm 5.23}$	$\mathbf{95.76_{\pm 0.51}}$	$\mathbf{41.01_{\pm 0.76}}$	$38.67_{\pm 5.34}$	$\mathbf{69.41_{\pm 1.01}}$	$\mathbf{65.71_{\pm 1.17}}$	$19.59_{\pm 1.13}$
Last Acc.										
Primal-RAIL [35]	$68.99_{\pm 5.05}$	$51.98_{\pm 3.55}$	$59.91_{\pm 0.75}$	$77.36_{\pm 6.73}$	$90.42_{\pm 0.72}$	$37.50_{\pm 0.73}$	$20.28_{\pm 9.73}$	$67.77_{\pm 1.44}$	$59.28_{\pm 1.57}$	$22.50_{\pm 2.69}$
FeCAM [6]	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$1.12_{\pm 3.21}$	$11.90_{\pm 1.72}$	$0.00_{\pm 0.00}$	$31.24_{\pm 8.07}$	$5.53_{\pm 1.31}$	$11.21_{\pm 2.58}$
RanPAC [22]	$70.02_{\pm 4.13}$	$54.34_{\pm 3.54}$	$41.03_{\pm 3.95}$	$77.57_{\pm 7.28}$	$91.90_{\pm 1.03}$	$37.08_{\pm 0.81}$	$\mathbf{39.08_{\pm 4.09}}$	$67.32_{\pm 1.28}$	$59.79_{\pm 1.38}$	$20.20_{\pm 1.99}$
KLDA [23]	$\mathbf{75.43_{\pm 3.99}}$	$57.80_{\pm 2.93}$	$37.35_{\pm 2.72}$	$\mathbf{78.22_{\pm 4.61}}$	$95.33_{\pm 0.58}$	$39.15_{\pm 1.15}$	$33.44_{\pm 7.53}$	$68.63_{\pm 0.92}$	$60.67_{\pm 1.13}$	$22.59_{\pm 1.50}$
HyCal(Ours)	$74.01_{\pm 3.21}$	$\mathbf{58.46_{\pm 4.30}}$	$\mathbf{49.65_{\pm 2.44}}$	$77.69_{\pm 5.15}$	$\mathbf{95.71_{\pm 0.53}}$	$\mathbf{40.95_{\pm 0.74}}$	$38.18_{\pm 5.02}$	$\mathbf{69.41_{\pm 1.01}}$	$\mathbf{63.01_{\pm 1.08}}$	$19.89_{\pm 1.11}$

Random Order III

	EuroSAT	Galaxy	Organ	Flower	DTD	MNIST	Aircraft	ArtBench	**Average**	σ
Zero-shot	37.58	9.80	17.97	67.40	41.90	44.01	23.91	50.88	56.41	18.76
Average Acc.										
Primal-RAIL [35]	$72.40_{\pm 4.22}$	$25.96_{\pm 2.91}$	$54.38_{\pm 1.74}$	$91.44_{\pm 0.46}$	$68.09_{\pm 1.01}$	$80.68_{\pm 3.75}$	$37.02_{\pm 0.93}$	$59.24_{\pm 1.67}$	$61.48_{\pm 1.27}$	$21.89_{\pm 0.96}$
FeCAM [6]	$5.44_{\pm 1.07}$	$4.53_{\pm 0.67}$	$3.20_{\pm 2.38}$	$6.24_{\pm 1.93}$	$35.83_{\pm 9.74}$	$0.00_{\pm 0.00}$	$11.60_{\pm 2.93}$	$0.00_{\pm 0.00}$	$10.15_{\pm 1.25}$	$11.76_{\pm 3.11}$
RanPAC [22]	$72.72_{\pm 1.99}$	$\mathbf{39.72_{\pm 3.65}}$	$56.02_{\pm 3.30}$	$91.60_{\pm 0.42}$	$65.82_{\pm 0.36}$	$78.17_{\pm 4.58}$	$35.60_{\pm 0.78}$	$39.72_{\pm 2.28}$	$63.91_{\pm 1.21}$	$20.60_{\pm 0.55}$
KLDA	$46.86_{\pm 1.94}$	$23.35_{\pm 1.93}$	$46.52_{\pm 3.05}$	$86.14_{\pm 2.78}$	$67.35_{\pm 1.40}$	$\mathbf{81.35_{\pm 2.06}}$	$39.46_{\pm 0.66}$	$37.17_{\pm 2.54}$	$39.60_{\pm 1.12}$	$22.38_{\pm 0.93}$
HyCal(Ours)	$\mathbf{73.81_{\pm 3.36}}$	$38.20_{\pm 5.59}$	$\mathbf{58.84_{\pm 1.22}}$	$\mathbf{95.64_{\pm 0.65}}$	$\mathbf{69.54_{\pm 0.60}}$	$79.06_{\pm 3.41}$	$\mathbf{41.43_{\pm 1.16}}$	$\mathbf{48.51_{\pm 1.85}}$	$\mathbf{64.73_{\pm 1.70}}$	$20.00_{\pm 1.69}$
Last Acc.										
Primal-RAIL [35]	$69.10_{\pm 5.00}$	$19.83_{\pm 2.46}$	$53.43_{\pm 1.22}$	$90.40_{\pm 0.49}$	$67.51_{\pm 0.90}$	$80.49_{\pm 3.89}$	$36.97_{\pm 0.94}$	$59.24_{\pm 1.67}$	$59.62_{\pm 0.55}$	$22.94_{\pm 1.03}$
FeCAM [6]	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.57_{\pm 0.51}$	$35.83_{\pm 9.74}$	$0.00_{\pm 0.00}$	$11.60_{\pm 2.93}$	$0.00_{\pm 0.00}$	$6.00_{\pm 1.11}$	$12.75_{\pm 3.18}$
RanPAC [22]	$70.36_{\pm 4.04}$	$\mathbf{39.95_{\pm 4.09}}$	$55.10_{\pm 3.53}$	$91.92_{\pm 0.68}$	$66.92_{\pm 0.68}$	$77.37_{\pm 5.98}$	$36.84_{\pm 1.31}$	$39.72_{\pm 2.28}$	$59.77_{\pm 1.51}$	$20.24_{\pm 1.12}$
KLDA	$\mathbf{75.57_{\pm 4.02}}$	$35.46_{\pm 3.58}$	$57.51_{\pm 3.50}$	$95.10_{\pm 1.51}$	$68.11_{\pm 1.35}$	$\mathbf{80.69_{\pm 2.11}}$	$39.44_{\pm 0.81}$	$37.17_{\pm 2.54}$	$61.13_{\pm 0.52}$	$22.44_{\pm 0.80}$
HyCal(Ours)	$74.01_{\pm 3.22}$	$38.09_{\pm 5.88}$	$\mathbf{58.81_{\pm 1.27}}$	$\mathbf{95.62_{\pm 0.64}}$	$\mathbf{69.65_{\pm 0.65}}$	$79.07_{\pm 3.42}$	$\mathbf{41.44_{\pm 1.16}}$	$\mathbf{48.51_{\pm 1.85}}$	$\mathbf{63.15_{\pm 0.74}}$	$20.04_{\pm 1.76}$

Random order IV

	ArtBench	MNIST	Organ	Aircraft	EuroSAT	DTD	Galaxy	Flower	**Average**	σ
Zero-shot	50.88	44.01	17.97	23.91	37.58	41.90	9.80	67.40	56.41	18.76
Average Acc.										
Primal-RAIL	$60.55_{\pm 0.19}$	$77.25_{\pm 5.23}$	$55.28_{\pm 5.19}$	$37.01_{\pm 0.37}$	$74.29_{\pm 7.47}$	$68.54_{\pm 0.96}$	$24.15_{\pm 8.06}$	$90.49_{\pm 0.32}$	$61.58_{\pm 0.48}$	$21.96_{\pm 1.96}$
FeCAM	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$0.00_{\pm 0.00}$	$11.93_{\pm 0.00}$	$0.00_{\pm 0.00}$	$33.17_{\pm 3.67}$	$0.00_{\pm 0.00}$	$0.14_{\pm 0.20}$	$5.66_{\pm 0.40}$	$11.89_{\pm 1.12}$
RanPAC	$39.93_{\pm 4.15}$	$\mathbf{79.06_{\pm 7.55}}$	$56.45_{\pm 6.11}$	$\mathbf{36.22_{\pm 2.13}}$	$71.24_{\pm 3.93}$	$67.45_{\pm 1.89}$	$\mathbf{39.71_{\pm 0.89}}$	$92.43_{\pm 0.80}$	$56.08_{\pm 2.20}$	$20.77_{\pm 1.02}$
KLDA [23]	$21.56_{\pm 1.54}$	$49.22_{\pm 5.31}$	$40.30_{\pm 2.21}$	$27.68_{\pm 0.45}$	$70.05_{\pm 4.84}$	$68.01_{\pm 1.04}$	$36.23_{\pm 5.53}$	$95.26_{\pm 1.04}$	$31.44_{\pm 1.53}$	$25.01_{\pm 1.09}$
HyCAL (Ours)	$\mathbf{50.94_{\pm 2.66}}$	$76.84_{\pm 5.18}$	$\mathbf{58.46_{\pm 3.45}}$	$\mathbf{41.21_{\pm 0.81}}$	$\mathbf{72.29_{\pm 4.85}}$	$\mathbf{69.56_{\pm 1.24}}$	$39.35_{\pm 4.00}$	$\mathbf{95.62_{\pm 0.41}}$	$\mathbf{59.69_{\pm 1.70}}$	$19.29_{\pm 1.31}$
Last Acc.										
Primal-RAIL	$60.55_{\pm 0.08}$	$76.69_{\pm 5.99}$	$54.19_{\pm 4.98}$	$36.60_{\pm 0.56}$	$73.05_{\pm 7.54}$	$67.75_{\pm 0.94}$	$22.90_{\pm 7.91}$	$90.49_{\pm 0.32}$	$60.28_{\pm 0.61}$	$22.17_{\pm 1.98}$
FeCAM	$8.56_{\pm 4.51}$	$0.00_{\pm 0.00}$	$0.95_{\pm 1.74}$	$12.31_{\pm 2.19}$	$0.00_{\pm 0.00}$	$33.19_{\pm 3.67}$	$0.00_{\pm 0.00}$	$0.14_{\pm 0.20}$	$8.64_{\pm 3.29}$	$11.67_{\pm 1.08}$
RanPAC	$38.00_{\pm 2.78}$	$78.88_{\pm 7.08}$	$57.08_{\pm 3.78}$	$37.43_{\pm 1.28}$	$72.07_{\pm 5.04}$	$67.26_{\pm 1.38}$	$\mathbf{39.35_{\pm 1.31}}$	$92.43_{\pm 0.80}$	$60.31_{\pm 0.63}$	$20.91_{\pm 0.94}$
KLDA [23]	$37.59_{\pm 4.24}$	$\mathbf{79.63_{\pm 4.88}}$	$58.08_{\pm 2.55}$	$39.58_{\pm 0.75}$	$\mathbf{72.63_{\pm 6.17}}$	$68.79_{\pm 0.94}$	$36.33_{\pm 5.01}$	$95.26_{\pm 1.04}$	$60.99_{\pm 0.68}$	$21.94_{\pm 1.24}$
HyCAL (Ours)	$\mathbf{49.42_{\pm 2.29}}$	$77.68_{\pm 5.13}$	$\mathbf{59.06_{\pm 2.58}}$	$\mathbf{41.25_{\pm 0.85}}$	$72.40_{\pm 4.78}$	$\mathbf{69.47_{\pm 1.25}}$	$39.30_{\pm 3.83}$	$\mathbf{95.62_{\pm 0.41}}$	$\mathbf{63.03_{\pm 1.11}}$	$19.50_{\pm 1.27}$

Table 10. Per-class sample counts (K_c) for the XD-VSCIL setting on the Aircraft, ArtBench, DTD, and EuroSAT datasets, with `random_seed=42`. Classes are sorted alphabetically.

Aircraft (aeronautics) [19]

Class Name	K_c	Class Name	K_c	Class Name	K_c	Class Name	K_c
707-320	45	727-200	12	737-200	6	737-300	22
737-400	20	737-500	19	737-600	13	737-700	11
737-800	48	737-900	39	747-100	10	747-200	42
747-300	32	747-400	7	757-200	6	757-300	10
767-200	18	767-300	19	767-400	37	777-200	43
777-300	6	A300B4	40	A310	17	A318	50
A319	46	A320	49	A321	39	A330-200	31
A330-300	19	A340-200	33	A340-300	42	A340-500	22
A340-600	5	A380	15	ATR-42	49	ATR-72	32
An-12	26	BAE 146-200	22	BAE 146-300	14	BAE-125	18
Beechcraft 1900	26	Boeing 717	11	C-130	10	C-47	29
CRJ-200	11	CRJ-700	27	CRJ-900	27	Cessna 172	43
Cessna 208	21	Cessna 525	7	Cessna 560	34	Challenger 600	39
DC-10	12	DC-3	29	DC-6	10	DC-8	40
DC-9-30	23	DH-82	45	DHC-1	44	DHC-6	28
DHC-8-100	41	DHC-8-300	17	DR-400	50	Dornier 328	9
E-170	7	E-190	47	E-195	19	EMB-120	23
ERJ 135	10	ERJ 145	19	Embraer Legacy 600	11	Eurofighter Typhoon	29
F-16A/B	22	F/A-18	34	Falcon 2000	45	Falcon 900	28
Fokker 100	15	Fokker 50	28	Fokker 70	27	Global Express	18
Gulfstream IV	47	Gulfstream V	22	Hawk T1	49	Il-76	48
L-1011	46	MD-11	9	MD-80	43	MD-87	45
MD-90	15	Metroliner	39	Model B200	20	PA-28	15
SR-20	34	Saab 2000	29	Saab 340	22	Spitfire	45
Tornado	49	Tu-134	40	Tu-154	19	Yak-42	48

ArtBench (art) [17]

Class Name	K_c	Class Name	K_c	Class Name	K_c	Class Name	K_c
Art Nouveau	46	Impressionism	13	Renaissance	50	Ukiyo-e	48
Baroque	13	Post-impressionism	20	Romanticism	25		
Expressionism	49	Realism	14	Surrealism	20		

DTD (textures) [3]

Class Name	K_c	Class Name	K_c	Class Name	K_c	Class Name	K_c
banded	30	flecked	10	matted	34	sprinkled	22
blotchy	36	freckled	30	meshed	45	stained	25
braided	14	frilly	37	paisley	47	stratified	23
bubbly	41	gauzy	22	perforated	49	striped	26
bumpy	45	grid	49	pitted	27	studded	40
chequered	22	grooved	28	pleated	9	swirly	41
cobwebbed	45	honeycombed	34	polka-dotted	39	veined	13
cracked	31	interlaced	36	porous	30	waffled	41
crosshatched	29	knitted	25	potholed	18	woven	36
crystalline	33	lacelike	42	scaly	32	wrinkled	26
dotted	9	lined	5	smeared	18	zigzagged	48
fibrous	45	marbled	10	spiralled	36		

EuroSAT (remote sensing) [9]

Class Name	K_c	Class Name	K_c	Class Name	K_c	Class Name	K_c
Annual Crop Land	26	Highway or Road	19	Permanent Crop Land	40	Sea or Lake	35
Forest	28	Industrial Buildings	34	Residential Buildings	44		
Herbaceous Vegetation Land	8	Pasture Land	26	River	43		

Table 11. Per-class sample counts (K_c) for the XD-VSCIL setting on the Galaxy, MNIST, OrganMNIST, and OxfordFlowers, with `random_seed=42`. Classes are sorted alphabetically.

Galaxy (astronomy) [13]

Class Name	K_c	Class Name	K_c	Class Name	K_c	Class Name	K_c
Disturbed Galaxies	7	In-between Round Smooth Galaxies	27	Unbarred Tight Spiral Galaxies	23	Edge-on Galaxies with Bulge	43
Merging Galaxies	49	Cigar Shaped Smooth Galaxies	37	Unbarred Loose Spiral Galaxies	48		
Round Smooth Galaxies	9	Barred Spiral Galaxies	11	Edge-on Galaxies without Bulge	24		

MNIST(fundamental visual perception) [4]

Class Name	K_c	Class Name	K_c	Class Name	K_c	Class Name	K_c
0	9	3	29	6	6	9	37
1	19	4	13	7	17		
2	29	5	11	8	38		

OrganMNIST (medical imaging) [36]

Class Name	K_c	Class Name	K_c	Class Name	K_c	Class Name	K_c
liver	48	pancreas	43	bladder	35	right femur	7
right lung	31	left lung	11	spleen	44	left femur	39
right kidney	14	left kidney	16	heart	43		

OxfordFlowers (fine-grained biology) [24]

Class Name	K_c	Class Name	K_c	Class Name	K_c	Class Name	K_c
passion flower	42	fritillary	16	sunflower	24	pink primrose	12
water lily	11	sweet william	41	magnolia	21	fire lily	27
cyclamen	25	azalea	13	osteospermum	40	red ginger	48
watercress	28	primula	41	garden phlox	41	prince of wales feathers	43
frangipani	44	cape flower	48	sweet pea	8	carnation	17
wallflower	21	purple coneflower	19	daffodil	30	mexican aster	41
rose	16	colt's foot	11	king protea	11	alpine sea holly	41
petunia	40	artichoke	22	great masterwort	49	siam tulip	46
poinsettia	21	wild pansy	20	black-eyed susan	15	spring crocus	46
clematis	20	peruvian lily	40	bearded iris	8	globe thistle	39
hibiscus	22	ruby-lipped cattleya	32	windflower	18	bolero deep blue	16
lotus	35	canna lily	29	ball moss	21	tiger lily	23
anthurium	44	gazania	17	spear thistle	35	moon orchid	27
thorn apple	26	lenten rose	14	silverbush	45	gaura	25
barbeton daisy	28	buttercup	14	balloon flower	11	japanese anemone	48
sword lily	16	pelargonium	35	oxeye daisy	33	foxglove	37
morning glory	36	desert-rose	5	cautleya spicata	26	bougainvillea	36
columbine	32	hippeastrum	16	common dandelion	21	camellia	33
geranium	18	giant white arum lily	31	yellow iris	28	mallow	10
bishop of llandaff	31	marigold	36	monkshood	28	mexican petunia	29
tree mallow	49	orange dahlia	13	love in the mist	40	bromelia	44
pink-yellow dahlia	28	hard-leaved pocket orchid	45	corn poppy	40	blanket flower	23
bee balm	28	english marigold	38	grape hyacinth	21	trumpet creeper	34
snapdragon	30	stemless gentian	39	canterbury bells	27	blackberry lily	45
californian poppy	22	tree poppy	25	globe-flower	6		
bird of paradise	33	pincushion flower	36	toad lily	31		

Table 12. Performance under Balanced-in-class domain setting. Average and Last accuracy (%) across 8 domains. Results are averaged over 4 seeds with 95% confidence intervals in subscripts. Domain names are abbreviated.

Balanced-in-Class Domain										
	Aircraft	ArtBench	DTD	EuroSAT	Galaxy	MNIST	Organ	Flower	**Average**	σ
Zero-shot	23.91	50.88	41.90	37.58	9.80	44.01	17.97	67.40	56.41	18.76
Average Acc.										
Primal-RAIL [35]	$32.64_{\pm 0.88}$	$60.60_{\pm 0.67}$	$64.54_{\pm 0.68}$	$88.16_{\pm 0.73}$	$51.55_{\pm 1.93}$	$91.58_{\pm 0.70}$	$67.94_{\pm 2.90}$	$86.85_{\pm 0.56}$	$56.57_{\pm 0.55}$	$20.38_{\pm 0.46}$
FeCAM [6]	$3.94_{\pm 0.67}$	$40.69_{\pm 2.23}$	$3.60_{\pm 2.67}$	$17.51_{\pm 9.16}$	$17.46_{\pm 7.04}$	$0.47_{\pm 1.14}$	$9.72_{\pm 10.18}$	$0.00_{\pm 0.00}$	$14.88_{\pm 1.39}$	$13.96_{\pm 0.63}$
RanPAC [22]	$30.18_{\pm 1.99}$	$58.88_{\pm 1.22}$	$59.20_{\pm 2.93}$	$87.26_{\pm 1.57}$	$\mathbf{57.60_{\pm 1.11}}$	$93.85_{\pm 0.82}$	$72.28_{\pm 1.90}$	$87.26_{\pm 1.14}$	$55.60_{\pm 1.10}$	$21.15_{\pm 0.81}$
KLDA [23]	$24.63_{\pm 0.50}$	$50.53_{\pm 0.54}$	$60.76_{\pm 1.23}$	$\mathbf{90.44_{\pm 0.53}}$	$55.85_{\pm 2.74}$	$95.04_{\pm 0.66}$	$74.94_{\pm 1.07}$	$92.85_{\pm 0.53}$	$46.57_{\pm 0.62}$	$24.75_{\pm 0.38}$
HyCal(Ours)	$\mathbf{34.74_{\pm 1.69}}$	$\mathbf{60.41_{\pm 0.59}}$	$\mathbf{64.54_{\pm 1.03}}$	$89.25_{\pm 0.24}$	$51.65_{\pm 1.80}$	$\mathbf{95.14_{\pm 0.49}}$	$\mathbf{76.36_{\pm 1.14}}$	$\mathbf{93.07_{\pm 0.75}}$	$\mathbf{57.75_{\pm 0.45}}$	$21.63_{\pm 0.57}$
Last Acc.										
Primal-RAIL [35]	$31.77_{\pm 0.54}$	$60.67_{\pm 0.56}$	$62.56_{\pm 0.97}$	$87.89_{\pm 0.80}$	$50.70_{\pm 2.29}$	$91.32_{\pm 0.83}$	$67.73_{\pm 2.91}$	$86.85_{\pm 0.56}$	$67.44_{\pm 0.68}$	$20.67_{\pm 0.44}$
FeCAM [6]	$3.05_{\pm 0.93}$	$40.69_{\pm 2.23}$	$3.60_{\pm 2.67}$	$17.51_{\pm 9.16}$	$17.46_{\pm 7.04}$	$0.47_{\pm 1.14}$	$9.72_{\pm 10.18}$	$0.00_{\pm 0.00}$	$11.56_{\pm 1.94}$	$14.04_{\pm 0.62}$
RanPAC [22]	$29.03_{\pm 1.88}$	$52.14_{\pm 24.34}$	$58.27_{\pm 2.60}$	$86.05_{\pm 2.59}$	$\mathbf{56.75_{\pm 1.44}}$	$92.83_{\pm 1.67}$	$70.92_{\pm 3.06}$	$87.26_{\pm 1.14}$	$67.64_{\pm 0.62}$	$22.16_{\pm 3.21}$
KLDA [23]	$33.40_{\pm 1.00}$	$\mathbf{61.49_{\pm 0.62}}$	$62.34_{\pm 1.56}$	$\mathbf{90.13_{\pm 0.36}}$	$55.69_{\pm 2.79}$	$94.78_{\pm 0.64}$	$74.66_{\pm 1.17}$	$92.85_{\pm 0.53}$	$\mathbf{70.67_{\pm 0.90}}$	$21.50_{\pm 0.57}$
HyCal(Ours)	$\mathbf{34.77_{\pm 1.72}}$	$60.70_{\pm 0.41}$	$\mathbf{63.67_{\pm 1.21}}$	$89.22_{\pm 0.24}$	$51.34_{\pm 1.59}$	$\mathbf{95.16_{\pm 0.47}}$	$\mathbf{76.41_{\pm 1.11}}$	$\mathbf{93.07_{\pm 0.75}}$	$70.54_{\pm 0.23}$	$21.68_{\pm 0.55}$

Table 13. Performance under Cross-scale imbalance setting. Average and Last accuracy (%) across 8 domains. Results are averaged over 4 seeds with 95% confidence intervals in subscripts. Domain names are abbreviated.

Cross-Scale Imbalance										
	Aircraft	ArtBench	DTD	EuroSAT	Galaxy	MNIST	Organ	Flower	**Average**	σ
Zero-shot	23.91	50.88	41.90	37.58	9.80	44.01	17.97	67.40	56.41	18.76
Average Acc.										
Primal-RAIL	$36.72_{\pm 1.63}$	$59.77_{\pm 1.10}$	$65.60_{\pm 2.81}$	$76.42_{\pm 3.42}$	$36.49_{\pm 8.47}$	$57.61_{\pm 2.17}$	$53.59_{\pm 3.54}$	$87.71_{\pm 3.18}$	$53.12_{\pm 0.71}$	$17.85_{\pm 2.14}$
FeCAM	$7.84_{\pm 5.90}$	$29.20_{\pm 9.53}$	$20.46_{\pm 2.70}$	$35.55_{\pm 9.63}$	$15.42_{\pm 14.00}$	$12.99_{\pm 17.42}$	$18.38_{\pm 25.53}$	$8.13_{\pm 4.92}$	$19.14_{\pm 4.82}$	$11.92_{\pm 2.54}$
RanPAC	$37.75_{\pm 3.73}$	$51.65_{\pm 2.78}$	$63.09_{\pm 5.28}$	$76.03_{\pm 6.98}$	$45.11_{\pm 5.95}$	$78.20_{\pm 12.34}$	$64.02_{\pm 7.98}$	$87.30_{\pm 7.11}$	$53.43_{\pm 3.94}$	$17.40_{\pm 2.76}$
KLDA	$43.11_{\pm 0.85}$	$\mathbf{55.68_{\pm 1.12}}$	$66.66_{\pm 3.95}$	$\mathbf{85.61_{\pm 1.82}}$	$48.36_{\pm 5.27}$	$89.17_{\pm 2.10}$	$70.04_{\pm 1.51}$	$\mathbf{96.11_{\pm 0.65}}$	$58.60_{\pm 0.59}$	$19.68_{\pm 0.69}$
HyCal(Ours)	$\mathbf{44.72_{\pm 1.91}}$	$54.09_{\pm 1.67}$	$\mathbf{66.73_{\pm 2.16}}$	$82.30_{\pm 2.05}$	$\mathbf{48.84_{\pm 2.90}}$	$\mathbf{89.72_{\pm 0.81}}$	$\mathbf{70.83_{\pm 3.35}}$	$95.56_{\pm 0.68}$	$\mathbf{58.71_{\pm 0.99}}$	$19.07_{\pm 0.70}$
Last Acc.										
Primal-RAIL	$35.78_{\pm 1.89}$	$59.68_{\pm 1.08}$	$63.80_{\pm 2.79}$	$75.47_{\pm 3.44}$	$35.47_{\pm 9.42}$	$56.32_{\pm 2.15}$	$52.95_{\pm 3.64}$	$87.71_{\pm 3.18}$	$58.40_{\pm 0.86}$	$18.06_{\pm 2.21}$
FeCAM	$6.32_{\pm 5.01}$	$26.79_{\pm 10.63}$	$19.28_{\pm 2.80}$	$35.37_{\pm 10.16}$	$15.40_{\pm 13.96}$	$12.99_{\pm 17.42}$	$18.38_{\pm 25.53}$	$8.13_{\pm 4.92}$	$17.83_{\pm 5.73}$	$11.79_{\pm 2.92}$
RanPAC	$36.98_{\pm 4.30}$	$51.41_{\pm 3.23}$	$62.19_{\pm 5.63}$	$74.71_{\pm 7.99}$	$44.69_{\pm 6.36}$	$75.78_{\pm 13.80}$	$63.11_{\pm 8.14}$	$87.30_{\pm 7.11}$	$62.02_{\pm 5.68}$	$17.23_{\pm 2.72}$
KLDA	$43.33_{\pm 0.95}$	$\mathbf{56.34_{\pm 0.91}}$	$\mathbf{66.43_{\pm 3.75}}$	$\mathbf{84.92_{\pm 1.94}}$	$\mathbf{47.66_{\pm 5.18}}$	$88.54_{\pm 2.08}$	$69.52_{\pm 1.44}$	$\mathbf{96.11_{\pm 0.65}}$	$\mathbf{69.11_{\pm 0.68}}$	$19.51_{\pm 0.69}$
HyCal(Ours)	$\mathbf{44.79_{\pm 1.73}}$	$54.00_{\pm 1.91}$	$65.91_{\pm 1.99}$	$82.27_{\pm 2.04}$	$\mathbf{47.66_{\pm 2.26}}$	$\mathbf{90.09_{\pm 0.88}}$	$\mathbf{70.97_{\pm 3.49}}$	$95.56_{\pm 0.68}$	$68.91_{\pm 0.78}$	$19.33_{\pm 0.61}$