SAFEROUTE: RETENTION-CONSTRAINED LOCAL EVIDENCE CONTROL FOR MIXED-SHOT RECOGNITION

Anonymous authors
Paper under double-blind review

ABSTRACT

Vision-language adaptation under heterogeneous support can improve average accuracy while overturning correct predictions for classes with no labeled examples. We formulate this mixed-shot setting as constrained inference over a fixed vocabulary and introduce SAFEROUTE, a score-level interface that separates the permission to change a prediction from the strength of adaptation. We prove an exact observable characterization of universal pointwise retention: without the query label, every unsupported frozen winner must be preserved, while supported frozen winners define the maximal open route on which top-1 changes are admissible. Inside this route, local evidence control (LEC) adjusts a support-selected interpolation scalar using endpoint margins, winner disagreement, residual direction, and local support contrast. LEC contains the matched scalar controller as a special case and is conditionally invariant to noncompeting vocabulary additions. Across seven datasets, three vision-language encoders, four support profiles, and 420 TaskRes episodes, SAFEROUTE-LEC improves class-macro accuracy over the budget-matched scalar by 1.278 percentage points under mixed support and by 1.211 to 1.402 points across all profiles. It also improves the same-loss constant by 1.048 points under mixed support, ranks first in every encoder-profile mean and every mixed-support dataset mean, retains all 173,043 protected decisions, and produces no change in 3,780 specified dummy-class tests. These results establish a modular retention interface and show that query-local evidence provides substantial utility beyond strong support-only calibration without modifying parent training.

1 INTRODUCTION

Vision-language models can recognize a class from its name before any labeled image is available (Radford et al., 2021; Jia et al., 2021; Zhai et al., 2022). Real deployments rarely preserve this balanced zero-shot setting. In an evolving catalog, some classes have no verified example, some have one, and others have many, while every class remains a valid prediction. We call this setting *mixed-shot recognition*: a fixed text-defined vocabulary paired with a heterogeneous support-count vector that may contain zeros.

The main difficulty is not only class imbalance. A balanced training objective changes how supported classes update an adapter, but it does not determine how frozen and adapted scores compete at inference. A supported class can overtake an unsupported class even when the frozen prediction was correct. Figure 1 illustrates this failure. Mixed-shot adaptation therefore requires an inference rule that protects correct unsupported predictions while preserving the ability to exploit support for other classes.

We express this requirement as a pointwise retention contract. A query is protected when its true class has zero support and the frozen model predicts that class correctly. The true label is unavailable at inference, but the contract admits an exact observable implementation: preserve the frozen winner whenever that winner has zero support. We prove that this condition is necessary and sufficient for universal retention under unknown query labels. The same result identifies the maximal label-agnostic region in which top-1 predictions may change.

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.

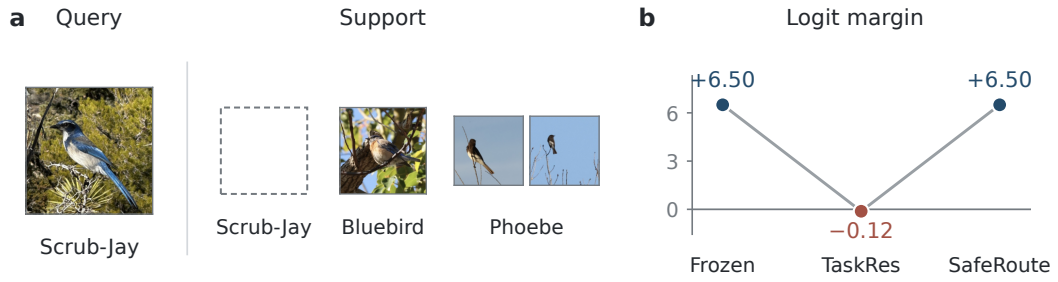

Figure 1: **Mixed-shot adaptation can reverse a correct unsupported prediction.** The query class has no support, while two competitors have labeled examples. The balanced parent reverses the frozen winner. SafeRoute closes the adaptation route and copies the frozen scores for this query.

This characterization separates *permission* from *strength*. The route decides whether adaptation may act. Once the route opens, a controller determines how much of the parent residual to use. Global count schedules couple a query to unrelated classes through episode-wide reference statistics, while a single episode-level scalar cannot respond to query-specific competition. We therefore introduce local evidence control (LEC), a six-parameter correction around a support-selected scalar. LEC uses only endpoint-local margins, winner disagreement, residual contrast, and support contrast among active competitors. It requires no query labels, no query-batch statistics, and no modification to parent training.

The resulting method is a lightweight score interface. A frozen predictor and any aligned parent score vector define two endpoints. The exact route protects unsupported frozen winners, and LEC controls residual interpolation only on the open route. The controller contains the matched scalar exactly, so gains cannot be attributed to a stronger endpoint or a larger information budget. Conditional locality further guarantees that noncompeting vocabulary additions do not alter the original prediction.

Our evaluation closes the loop between theorem, design, and observed behavior. On seven datasets, three encoders, four support profiles, and five support seeds, SAFEROUTE-LEC improves every encoder-profile mean and every mixed-support dataset mean. Under mixed support it gains 1.278 class-macro points over the budget-matched scalar, 1.048 points over the same-loss constant, and 1.444 points over the original global count controller. All corresponding mixed-support intervals remain above zero. The route retains all 173,043 protected decisions, and the locality audit produces zero prediction changes across 3,780 specified additions. A separate natural-history extension across three encoders and three parent families yields a 0.896-point mean gain over matched scalar selection.

Our contributions are:

- an exact observable characterization of universal pointwise retention in mixed-shot recognition, which identifies the maximal admissible route for query-label-agnostic adaptation;
- SAFEROUTE, a score-level route-and-residual interface, and LEC, a local support-only controller with scalar containment, exact endpoint behavior, and conditional locality;
- a matched evaluation that isolates query-local control from endpoint quality, routing, supervision, and optimization objective, while verifying retention and locality over the complete benchmark audit.

2 RELATED WORK

Vision-language pretraining and adaptation. CLIP, ALIGN, LiT, SigLIP, SigLIP2, and MobileCLIP2 establish transferable image-text representations at different data and efficiency scales (Radford et al., 2021; Jia et al., 2021; Zhai et al., 2022; 2023; Tschannen et al., 2025; Faghri et al., 2025). Prompt-based adaptation learns textual or multimodal context, including CoOp, CoCoOp, MaPLe, PromptSRC, and knowledge-guided prompt tuning (Zhou et al., 2022b;a; Khattak et al., 2023a;b; Yao et al., 2023). Feature, cache, residual, and statistical alternatives include CLIP-Adapter,

Tip-Adapter, TaskRes, GDA, and label propagation (Gao et al., 2024; Zhang et al., 2022; Yu et al., 2023; Wang et al., 2024; Li et al., 2025). Recent methods also study multi-step cross-modal transport, manifold-preserving tuning, and long-tailed test-time adaptation (Jiang et al., 2026; Chen et al., 2026; Wang et al., 2026). SAFEROUTE is complementary to these methods. It constrains how frozen and parent scores compete at inference while leaving the parent objective unchanged.

Few-shot learning under heterogeneous supervision. Metric and optimization based few-shot learning include Matching Networks, Prototypical Networks, and MAML (Vinyals et al., 2016; Snell et al., 2017; Finn et al., 2017). Meta-Dataset and class-imbalanced few-shot learning expose variable support size as a first-class evaluation factor (Triantafillou et al., 2019; Ochal et al., 2023). Generalized few-shot, generalized zero-shot, and any-shot learning study competition among differently supervised class groups (Kim & Choi, 2023; Chao et al., 2016; Su et al., 2022; Xian et al., 2019; Xu et al., 2022). DeCoOp explicitly combines base and new classes through out-of-distribution detection (Zhou et al., 2024). Mixed-shot recognition differs by imposing a pointwise retention contract over one text-initialized vocabulary. Its route depends on the observable support count of the frozen winner rather than the unknown support status of the true class.

Retention, interpolation, and matched evaluation. WiSE-FT and task arithmetic show that interpolation in weight or task-vector space can preserve useful pretrained behavior (Wortsman et al., 2022; Ilharco et al., 2022). Learning without Forgetting and positive-congruent training study retention and negative prediction flips through training objectives (Li & Hoiem, 2017; Yan et al., 2021). Test-time prompt tuning uses unlabeled deployment batches and therefore assumes a different information regime (Shu et al., 2022). Calibration work cautions that a score margin is not automatically a correctness probability (Guo et al., 2017). Recent evaluations also show that inductive protocols and strong simple controls can change conclusions about lightweight VLM adaptation (Silva-Rodriguez et al., 2024; Kravets et al., 2025). These observations motivate our nested scalar baseline, same-loss constant, and fixed-parent protocol.

3 METHOD

3.1 MIXED-SHOT PREDICTION AND THE RETENTION CONTRACT

Let $\mathcal{C}$ be a fixed vocabulary of C classes. A frozen vision-language model produces scores $s^0(x) \in \mathbb{R}^C$ and winner

$$\hat{c}^0(x) = \arg\max_{c \in \mathcal{C}} s^0_c(x). \tag{1}$$

A support set $S = \{(x_i, y_i)\}_{i=1}^N$ induces counts n_c, supported classes $\mathcal{C}_+ = \{c : n_c > 0\}$, and unsupported classes $\mathcal{C}_0 = \{c : n_c = 0\}$. A fitted parent adapter produces aligned scores $s^a(x)$ over the same ordered vocabulary. All predictors use the same deterministic tie rule.

A query with label y is protected when $n_y = 0$ and $\hat{c}^0(x) = y$. The pointwise contract is

$$n_y = 0 \ \wedge \ \hat{c}^0(x) = y \quad \Longrightarrow \quad \arg\max_c s_c(x) = y. \tag{2}$$

The contract is pointwise. Correcting one query cannot compensate for losing another protected decision.

3.2 THE EXACT ADMISSIBLE ROUTE

The wrapper observes $(s^0, s^a, \mathbf{n})$ but not y. Universal retention requires Eq. 2 for every label consistent with this observable tuple.

Proposition 1 (Observable retention condition)**.** *A deterministic query-label-agnostic wrapper W satisfies Eq. 2 for all admissible instances if and only if*

$$n_{\hat{c}^0(x)} = 0 \quad \Longrightarrow \quad \arg\max_c W_c(s^0(x), s^a(x), \mathbf{n}) = \hat{c}^0(x). \tag{3}$$

Proof. If a wrapper changes an unsupported frozen winner, choose the admissible unknown label $y = \hat{c}^0(x)$. The original decision was protected and is lost, violating Eq. 2. Conversely, every protected query has $y = \hat{c}^0(x)$ and $n_{\hat{c}^0(x)} = 0$. Preserving that winner preserves the correct label. □

Define the closed route $\Omega_0 = \{x : n_{\hat{c}^0(x)} = 0\}$ and the open route $\Omega_+ = \{x : n_{\hat{c}^0(x)} > 0\}$. The open route is the maximal label-agnostic region in which top-1 changes are admissible. We implement the route with the residual family

$$s^\eta(x) = s^0(x) + \eta(x)\big(s^a(x) - s^0(x)\big), \qquad \eta(x) = 0 \text{ on } \Omega_0. \tag{4}$$

The theorem determines whether adaptation may act. The controller below determines its strength on Ω_+.

3.3 Local Evidence Control

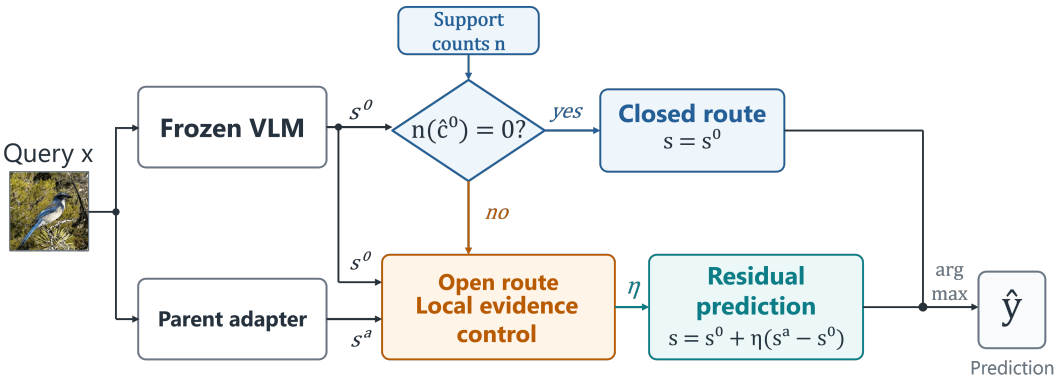

Figure 2: **SafeRoute-LEC inference.** The frozen-winner count selects a closed branch that copies s^0 or an open branch where LEC computes η from s^0, s^a, $\mathbf{n}$, and support-only parameters (ρ, θ). The two score paths merge before the final argmax.

Figure 2 separates route selection from residual control. The straight s^0 path reaches the count test and supplies the frozen endpoint to the open branch; the parallel path supplies s^a. The closed branch returns s^0, while the open branch computes η and blends the same endpoints by Eq. 4. The original count-only controller, denoted SafeRoute-Count, sets open-route strength from the frozen-winner count and a global episode reference. It satisfies Proposition 1, but unrelated high-count classes can perturb every open-route coefficient. LEC instead anchors strength at a support-selected scalar ρ and learns a regularized local correction.

For each query, let $\hat{c}^a = \arg\max_c s^a_c$, let m^0 and m^a be the top-1 minus top-2 margins of the frozen and parent scores, and let $r = s^a - s^0$. Let $U_2(x)$ be the union of the frozen and parent top-2 class sets. With deterministic class-index tie breaking, define

$$z_1 = \tanh(m^0/5), \qquad z_2 = \tanh(m^a/5), \tag{5}$$

$$z_3 = \mathbf{1}[\hat{c}^0 \neq \hat{c}^a], \qquad z_4 = \tanh\big((r_{\hat{c}^a} - r_{\hat{c}^0})/10\big). \tag{6}$$

The count feature compares the frozen winner with only its local competitors:

$$z_5 = \tanh\left(\log(1 + n_{\hat{c}^0}) - \frac{1}{|V|}\sum_{c \in V}\log(1 + n_c)\right), \quad V = U_2 \setminus \{\hat{c}^0\}, \tag{7}$$

with $z_5 = 0$ when V is empty, and $z_6 = 1$. The first four signals describe the local geometry of the endpoint scores, z_5 supplies support evidence only for active competitors, and z_6 permits a regularized episode-level offset. The controller is

$$\eta_\theta(x) = \begin{cases} 0, & x \in \Omega_0, \\ \mathrm{clip}_{[0,1]}\big(\rho + z(x)^\top\theta\big), & x \in \Omega_+. \end{cases} \tag{8}$$

The final predictor is Eq. 4 with η_θ, which we call SafeRoute-LEC.

Residual-path geometry. Equation 4 defines a one-dimensional path between two fixed score vectors rather than an unconstrained classifier. Writing $r = s^a - s^0$, any pair of classes j and k can exchange order only at

$$\eta_{jk} = \frac{s^0_j - s^0_k}{r_k - r_j}, \qquad r_j \neq r_k. \tag{9}$$

The top-1 prediction is constant between consecutive breakpoints. Endpoint margins locate the query relative to active competition, winner disagreement indicates whether the endpoints occupy different ranking regions, and residual contrast identifies the direction of the relevant crossing. LEC therefore uses margins as local coordinates on a finite decision path, not as calibrated correctness probabilities. The exact route removes every closed-route crossing by fixing $\eta = 0$.

The base scalar $\rho \in \{0, .25, .50, .75, 1\}$ is selected with the same legal inner holdout and tie rule as the budget-matched scalar baseline. The six coefficients are fitted on cross-fitted support holdouts by class-macro cross-entropy plus $\lambda\|\theta\|_2^2$, with $\lambda = \max(1, 20/N_{\text{open}})$ and box constraint $\theta_j \in [-.5, .5]$. If fewer than eight open holdout examples or fewer than two represented open classes are available, we set $\theta = 0$. The regularizer and fallback return the matched scalar whenever local evidence is weak.

Table 1: **LEC evidence map.** Every feature is computed from the two endpoint score vectors and support counts inside the local top-2 union.

Feature	Signal	Control role
z_1	Frozen top-1 margin	Frozen endpoint certainty
z_2	Parent top-1 margin	Parent endpoint certainty
z_3	Endpoint winner disagreement	Proposed label change
z_4	Residual contrast of endpoint winners	Direction of parent movement
z_5	Local log-count contrast	Evidence relative to competitors
z_6	Constant bias	Episode-level correction

Implementation and cost. The route needs one frozen argmax and one support-count lookup. LEC reads the top two classes of each endpoint, constructs six scalars, and evaluates one dot product. The score blend remains $O(C)$, controller state is six coefficients plus ρ, and no query label or query-batch statistic is used. Parent training and final-query endpoint scores are identical across all nested controls.

3.4 Guarantees and Containment

Corollary 1 (Retention and scalar containment)**.** SafeRoute-LEC *satisfies Proposition 1. Setting* $\theta = 0$ *returns the same routed scalar* ρ *on every open query.*

Thus the matched scalar is a nested baseline with the same information budget. The parent scores, route, holdout labels, scalar search, and fallback conditions are identical; the only additional degree of freedom is the regularized query-dependent correction. Improvement over this scalar therefore isolates the value of local control rather than a stronger endpoint or extra supervision.

Proposition 2 (Conditional locality)**.** *Fix the parent scores, controller parameters, scalar, and support counts on the original vocabulary. Append classes whose frozen and parent scores are strictly below every original score and remain outside both original top-2 sets. Then* $z(x)$, $\eta_\theta(x)$, *all original output scores, and the final prediction are unchanged.*

The proof and scope are given in Appendix A. The theorem supplies exact retention to every routed controller, while LEC adds query-specific strength only inside the admissible region.

For any fixed unrouted comparator M, let $R(M)$ copy frozen predictions only on Ω_0. With fixed evaluation weights, define B_M as closed-route cases where Frozen is correct and M is wrong, and C_M as closed-route cases where Frozen is wrong and M is correct. Then

$$A(R(M)) - A(M) = B_M - C_M. \tag{10}$$

This identity gives exact benefit-cost accounting for any routed comparator: retained correct decisions contribute B_M, while only corrections actually made by M contribute to the route cost C_M.

Table 2: **Mixed-support TaskRes accuracy.** Equal-dataset class-macro accuracy (%) across seven datasets and five support seeds. Matched is the budget-matched selected scalar. Constant uses the same holdout loss as LEC but no query-dependent features.

Encoder	Frozen	SAFEROUTE-COUNT	Matched	Constant	SAFEROUTE-LEC
CLIP L/14	80.901	83.581	83.643	83.833	**84.893**
SigLIP2 B/16	85.828	86.981	87.432	87.550	**88.576**
MobileCLIP2-S0	81.292	83.352	83.338	83.720	**84.778**
Mean	82.674	84.638	84.805	85.034	**86.082**

Table 3: **Matched mixed-support gains.** LEC minus comparator in class-macro percentage points. Seed CI is conditional on the fixed datasets; Dataset CI resamples datasets and seeds.

Comparator	Δ	Seed CI	Dataset CI
SAFEROUTE-COUNT	+1.444	$[+1.282, +1.601]$	$[+0.983, +1.838]$
Budget-matched scalar	+1.278	$[+1.169, +1.403]$	$[+1.099, +1.561]$
Same-loss constant	+1.048	$[+1.007, +1.094]$	$[+0.979, +1.139]$
Frozen	+3.409	$[+3.265, +3.536]$	$[+1.049, +6.401]$

4 EXPERIMENTS

4.1 PROTOCOL AND COMPARATORS

Data and encoders. The primary benchmark contains CIFAR-10 and CIFAR-100 (Krizhevsky et al., 2009), DTD (Cimpoi et al., 2014), Oxford-IIIT Pets (Parkhi et al., 2012), Flowers-102 (Nilsback & Zisserman, 2008), Stanford Cars (Krause et al., 2013), and Food-101 (Bossard et al., 2014). We evaluate CLIP ViT-L/14, SigLIP2 B/16-256, and MobileCLIP2-S0. TaskRes is the fixed parent for the primary controller study. Each dataset uses five support seeds and four prespecified profiles: mixed $(0, 1, 2, 4, 8)$, positive heterogeneous $(1, 1, 2, 4, 8)$, a near-uniform allocation with the same budget, and uniform eight-shot support. Cyclic class assignment follows a seeded permutation, yielding 420 TaskRes episodes.

Controls. All retaining methods use the same full-support parent and the route in Proposition 1. We compare Frozen, SAFEROUTE-COUNT, the budget-matched selected scalar, a continuous same-loss constant trained with the same holdout objective, and SAFEROUTE-LEC. The matched scalar receives the same inner fits, held-out labels, and fallback conditions as LEC. Query labels are used only for final evaluation.

Metrics and uncertainty. We report class-macro accuracy. Seeds are averaged within each dataset, and datasets are weighted equally. Nominal 95% paired seed intervals use 5,000 support-seed resamples conditional on fixed datasets and queries. Dataset-cluster sensitivity resamples the seven datasets and then seeds. Appendix B specifies preprocessing, parent optimization, holdout construction, and the complete statistical protocol.

4.2 MIXED-SUPPORT PERFORMANCE

Table 2 reports absolute mixed-support accuracy. SAFEROUTE-LEC ranks first for all three encoder means. Averaged over encoders and equal-weight datasets, it reaches 86.082%, compared with 85.034% for the same-loss constant, 84.805% for the budget-matched scalar, 84.638% for SAFEROUTE-COUNT, and 82.674% for Frozen.

Table 3 provides the matched differences. Relative to the budget-matched scalar, LEC gains 1.278 points with paired seed interval $[+1.169, +1.403]$ and dataset-cluster sensitivity $[+1.099, +1.561]$. Matching the continuous holdout objective still leaves a 1.048-point gain over the same-loss constant, with both intervals entirely above zero. The 1.444-point gain over SAFEROUTE-COUNT shows that local evidence also improves substantially over vocabulary-global count normalization.

Table 4: **Mixed-support group analysis.** Equal-dataset, equal-encoder means. A_0 and A_+ are class-macro accuracy (%) for zero-support and supported true classes. $\bar{\eta}$ is the query-average interpolation coefficient.

Method	Overall	A_0	A_+	$\bar{\eta}$
Frozen	82.674	82.561	82.705	0.000
SAFEROUTE-COUNT	84.638	82.989	85.075	0.506
Matched scalar	84.805	82.862	85.315	0.249
Same-loss constant	85.034	82.890	85.597	0.268
SAFEROUTE-LEC	**86.082**	**83.893**	**86.658**	0.273

Table 5: **Support-profile robustness.** Equal-dataset, equal-encoder TaskRes means (%). Deltas compare LEC with the named retaining control.

Profile	LEC	Matched	Constant	SAFEROUTE-COUNT	Δ_{Mat}	Δ_{Count}
Mixed	86.082	84.805	85.034	84.638	+1.278	+1.444
Positive heterogeneous	86.630	85.260	85.546	85.369	+1.370	+1.262
Near-uniform	88.475	87.073	87.403	86.301	+1.402	+2.174
Uniform 8	89.938	88.728	88.895	87.941	+1.211	+1.997

Supported and unsupported groups. Table 4 separates the two true-label groups. Every retaining method has zero protected loss. LEC obtains the highest accuracy for both zero-support and supported true classes while keeping the average interpolation coefficient close to the scalar controls. The gain therefore reflects selective decisions within the open route rather than uniform parent amplification.

4.3 SUPPORT, ENCODER, AND DATASET CONSISTENCY

The four support profiles test whether local control only repairs the original mixed pattern. Table 5 retains all seven datasets, three encoders, and five seeds. LEC exceeds the matched scalar by 1.211 to 1.402 points and SAFEROUTE-COUNT by 1.262 to 2.174 points across the four profiles. The same-loss comparison remains positive in every profile, with gains from 1.043 to 1.084 points.

Table 6 expands the aggregate result into all twelve encoder-profile combinations. Every entry exceeds one point. The improvement is therefore not tied to one representation family or one support allocation.

Table 7 reports mixed-support consistency across datasets. LEC exceeds both the matched scalar and SAFEROUTE-COUNT on all seven datasets after averaging the three encoders. The same-loss comparison is also positive on every dataset in Appendix C, so the aggregate gain is not driven by one dataset family or by the choice of scalar objective.

4.4 RETENTION, LOCALITY, AND TRANSFER

Every retaining method loses zero of 173,043 protected query-episode decisions under mixed support. Across all four profiles, the structural audit appends one, two, or five noncompeting classes with counts in $\{0, 4, 128\}$ while holding endpoints and fitted parameters fixed. Table 8 verifies retention, locality, scalar containment, and coefficient range over the complete benchmark audit.

A separate natural-history extension uses 50 California bird species, five support histories, a shared 1,000-query panel, three encoders, and three parent families. Across the resulting nine systems, LEC improves the budget-matched scalar by 0.896 points on average while preserving the route. This extension changes the support-acquisition process and parent family, providing evidence that the controller gain is not confined to the primary TaskRes construction. Appendix E gives the scope and protocol boundaries.

Label-aware interpolation oracles in Appendix D place the routed oracle 4.220 points above the deployable selected scalar along the audited paths. This remaining headroom indicates that stronger support-only controllers can improve utility without relaxing the exact retention boundary.

Table 6: **Encoder-profile consistency.** LEC minus the budget-matched scalar in class-macro percentage points.

Profile	CLIP L/14	SigLIP2 B/16	MobileCLIP2-S0
Mixed	+1.250	+1.144	+1.440
Positive heterogeneous	+1.432	+1.475	+1.203
Near-uniform	+1.433	+1.447	+1.325
Uniform 8	+1.183	+1.176	+1.274

Table 7: **Mixed-support dataset consistency.** LEC minus comparator, averaged over three encoders and five support seeds (percentage points).

Dataset	Matched scalar	SAFEROUTE-COUNT
Cars	+1.284	+1.557
CIFAR-10	+1.111	+1.331
CIFAR-100	+1.163	+1.711
DTD	+1.706	+0.415
Flowers-102	+1.542	+2.066
Food-101	+1.072	+1.862
Pets	+1.066	+1.168

4.5 MATCHED EVIDENCE CHAIN

The experimental design uses a sequence of nested controls rather than a single weak baseline. Table 9 summarizes what changes at each step. Every retaining controller receives the same frozen and parent scores and obeys the same exact route. The budget-matched scalar uses the same legal inner splits and scalar search as LEC but cannot vary across queries. The same-loss constant additionally matches the continuous holdout objective but still cannot use query-local evidence. LEC is the only control that varies interpolation strength using local endpoint geometry and local support contrast.

This nesting supports four distinct conclusions. First, the 1.444-point gain over SAFEROUTE-COUNT shows that local competition is more effective than a vocabulary-global count reference under the same retention boundary. Second, the 1.278-point gain over the matched scalar isolates the value of query-dependent control while keeping the parent, route, support split, and scalar operating point fixed. Third, the 1.048-point gain over the same-loss constant remains after matching the continuous objective, so the result cannot be explained by replacing discrete accuracy selection with a smoother loss. Fourth, the structural audit verifies that these utility gains do not come from relaxing the protected set or from coupling to the specified noncompeting additions.

The consistency results provide a separate line of evidence. Every encoder-profile increment exceeds one point, all seven mixed-support dataset means are positive against each matched comparator, and the natural-history extension remains positive after changing both parent family and support acquisition. These tests do not prove universal transfer, but they reduce the likelihood that the aggregate result is caused by one encoder, one support profile, one dataset family, or one parent construction. Together with scalar containment, the evidence links the empirical gain specifically to the local correction inside the exact admissible route.

4.6 SCOPE AND LIMITATIONS

The retention guarantee is exact for the contract in Eq. 2; it does not certify that frozen predictions are fair, calibrated, or correct. The route may block corrections on closed-route queries where the frozen winner is wrong, and Appendix D reports oracle accounting for this cost. Conditional locality is also deliberately scoped to additions that remain outside the active competitors. If a new class enters either endpoint top-2 set, LEC recomputes its evidence rather than claiming invariance.

The primary benchmark fixes TaskRes to isolate controller behavior, while the natural-history extension broadens the study to three parent families. This does not exhaust the space of prompt, cache, feature, and residual adapters. LEC is also a six-parameter linear correction, chosen for auditability and low support demand rather than maximal expressiveness. The routed oracle gap

Table 8: **Structural audit over all 420 TaskRes episodes.** Dummy additions remain below every original endpoint score and outside the original local candidate set.

Check	Result
Protected prediction losses under mixed support	0
Dummy conditions evaluated	3,780
LEC prediction changes under specified additions	0
$\theta = 0$ differences from the matched scalar	0
Controller strengths outside $[0, 1]$	0

Table 9: **Matched controller comparison.** All retaining controls use the same frozen scores, parent scores, support labels, and exact route. The local-score columns denote endpoint margins, disagreement, residual contrast, and local support contrast as indicated.

Control	Exact route	Query-varying	Local scores	Local counts	Holdout fit
SAFEROUTE-COUNT	Yes	Yes	No	No	Count rule
Matched scalar	Yes	No	No	No	Discrete accuracy
Same-loss constant	Yes	No	No	No	Continuous CE
SAFEROUTE-LEC	Yes	Yes	Yes	Yes	Continuous CE

indicates room for stronger support-only controllers, but such extensions should preserve the matched evaluation and exact route used here.

4.7 DISCUSSION

SafeRoute changes the structure of adaptation rather than adding another parent-specific training objective. Proposition 1 characterizes the feasible decision region for every deterministic query-label-agnostic wrapper. The protection result is independent of how the parent was trained, and any adapter exposing aligned frozen and adapted scores can use the same route.

Inside this feasible region, support-only calibration selects an episode-level operating point and local evidence turns that point into a query-dependent policy. Because LEC is nested around an optimized scalar and compared with a same-loss constant, its gains are attributable to query-local control rather than to a stronger endpoint, a different route, extra labels, or a different optimization objective. The consistent improvements across profiles, encoders, datasets, and the natural-history extension support this separation.

SafeRoute is distinct from confidence thresholding and abstention. A threshold may decide when a score appears uncertain, but it does not characterize which top-1 changes are compatible with universal retention when the query label is hidden. Proposition 1 instead defines the admissible region from the observable support count of the frozen winner, and every query still receives an ordinary prediction in the fixed vocabulary. Conditional locality is intentionally narrower than invariance to arbitrary new classes: if an added class enters active competition, LEC recomputes its local evidence.

5 CONCLUSION

Mixed-shot adaptation is a constrained inference problem as well as a learning problem. SAFEROUTE provides an exact observable route that protects unsupported frozen winners and exposes residual strength as a separate design choice. SAFEROUTE-LEC replaces globally coupled count normalization with local evidence control around a matched support-selected scalar. It preserves exact endpoints, contains the scalar baseline, and is conditionally invariant to noncompeting vocabulary additions.

Across seven datasets, three encoders, and four support profiles, LEC improves the matched scalar by more than one point in every profile and encoder-profile combination while retaining every protected decision. Matched controls, dataset-level results, the structural audit, and the natural-history extension show that query-local evidence improves adaptation inside the exact route without changing parent training.

AI USE STATEMENT

The authors used generative AI tools for language polishing, structural editing, and LaTeX formatting assistance. These tools were not used to design the method, conduct experiments, generate numerical results, or draw scientific conclusions. The authors verified all manuscript content and take full responsibility for it.

REPRODUCIBILITY STATEMENT

Section 3 specifies the contract, controller, fitting rule, and structural properties. Appendices A and B provide complete proofs, support construction, parent settings, fitting details, and uncertainty definitions. Expanded results and structural tests appear in Appendices C and D.

ETHICS STATEMENT

The primary experiments use established image-classification datasets under their existing research terms. The natural-history extension uses licensed public observation images and pseudonymized user identifiers. Preserving a pretrained decision does not establish that the decision is fair, calibrated, or correct. The method should not be interpreted as a safety guarantee outside the explicitly defined retention contract.

REFERENCES

Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101–mining discriminative components with random forests. In *European conference on computer vision*, pp. 446–461. Springer, 2014.

Wei-Lun Chao, Soravit Changpinyo, Boqing Gong, and Fei Sha. An empirical study and analysis of generalized zero-shot learning for object recognition in the wild. In *European conference on computer vision*, pp. 52–68. Springer, 2016.

Dexia Chen, Qianjie Zhu, Weibing Li, Yue Yu, Tong Zhang, and Ruixuan Wang. Preserve and sculpt: Manifold-aligned fine-tuning of vision-language models for few-shot learning. In *International Conference on Learning Representations*, volume 2026, pp. 109043–109065, 2026.

Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3606–3613, 2014.

Fartash Faghri, Pavan Kumar Anasosalu Vasu, Cem Koc, Vaishaal Shankar, Alexander Toshev, Oncel Tuzel, and Hadi Pouransari. Mobileclip2: Improving multi-modal reinforced training. *arXiv preprint arXiv:2508.20691*, 2025.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pp. 1126–1135. PMLR, 2017.

Peng Gao, Shijie Geng, Renrui Zhang, Teli Ma, Rongyao Fang, Yongfeng Zhang, Hongsheng Li, and Yu Qiao. Clip-adapter: Better vision-language models with feature adapters. *International journal of computer vision*, 132(2):581–595, 2024.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International conference on machine learning*, pp. 1321–1330. PMLR, 2017.

Gabriel Ilharco, Marco Tulio Ribeiro, Mitchell Wortsman, Suchin Gururangan, Ludwig Schmidt, Hannaneh Hajishirzi, and Ali Farhadi. Editing models with task arithmetic. 2022.

Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. Scaling up visual and vision-language representation learning with noisy text supervision. In *International conference on machine learning*, pp. 4904–4916. PMLR, 2021.

Ziqi Jiang, Yanghao Wang, and Long Chen. Exploring cross-modal flows for few-shot learning. In *International Conference on Learning Representations*, volume 2026, pp. 17074–17090, 2026.

Muhammad Uzair Khattak, Hanoona Rasheed, Muhammad Maaz, Salman Khan, and Fahad Shahbaz Khan. Maple: Multi-modal prompt learning. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19113–19122. IEEE, 2023a.

Muhammad Uzair Khattak, Syed Talal Wasim, Muzammal Naseer, Salman Khan, Ming-Hsuan Yang, and Fahad Shahbaz Khan. Self-regulating prompts: Foundational model adaptation without forgetting. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15144–15154. IEEE, 2023b.

Seong-Woong Kim and Dong-Wan Choi. Better generalized few-shot learning even without base data. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 8282–8290, 2023.

Jonathan Krause, Michael Stark, Jia Deng, and Li Fei-Fei. 3d object representations for fine-grained categorization. In *Proceedings of the IEEE international conference on computer vision workshops*, pp. 554–561, 2013.

Alexey Kravets, Da Chen, and Vinay P Namboodiri. Rethinking few shot clip benchmarks: A critical analysis in the inductive setting. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1902–1911. IEEE, 2025.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Yushu Li, Yongyi Su, Adam Goodge, Kui Jia, and Xun Xu. Efficient and context-aware label propagation for zero-/few-shot training-free adaptation of vision-language model. In *International Conference on Learning Representations*, volume 2025, pp. 57322–57342, 2025.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017.

Maria-Elena Nilsback and Andrew Zisserman. Automated flower classification over a large number of classes. In *2008 Sixth Indian conference on computer vision, graphics & image processing*, pp. 722–729. IEEE, 2008.

Mateusz Ochal, Massimiliano Patacchiola, Jose Vazquez, Amos Storkey, and Sen Wang. Few-shot learning with class imbalance. *IEEE Transactions on Artificial Intelligence*, 4(5):1348–1358, 2023.

Omkar M Parkhi, Andrea Vedaldi, Andrew Zisserman, and CV Jawahar. Cats and dogs. In *2012 IEEE conference on computer vision and pattern recognition*, pp. 3498–3505. IEEE, 2012.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Manli Shu, Weili Nie, De-An Huang, Zhiding Yu, Tom Goldstein, Anima Anandkumar, and Chaowei Xiao. Test-time prompt tuning for zero-shot generalization in vision-language models. volume 35, pp. 14274–14289, 2022.

Julio Silva-Rodriguez, Sina Hajimiri, Ismail Ben Ayed, and Jose Dolz. A closer look at the few-shot adaptation of large vision-language models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition*, pp. 23681–23690, 2024.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

Hongzu Su, Jingjing Li, Zhi Chen, Lei Zhu, and Ke Lu. Distinguishing unseen from seen for generalized zero-shot learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7875–7884. IEEE, 2022.

Eleni Triantafillou, Tyler Zhu, Vincent Dumoulin, Pascal Lamblin, Utku Evci, Kelvin Xu, Ross Goroshin, Carles Gelada, Kevin Swersky, Pierre-Antoine Manzagol, et al. Meta-dataset: A dataset of datasets for learning to learn from few examples. 2019.

Michael Tschannen, Alexey Gritsenko, Xiao Wang, Muhammad Ferjad Naeem, Ibrahim Alabdulmohsin, Nikhil Parthasarathy, Talfan Evans, Lucas Beyer, Ye Xia, Basil Mustafa, et al. Siglip 2: Multilingual vision-language encoders with improved semantic understanding, localization, and dense features. *arXiv preprint arXiv:2502.14786*, 2025.

Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The inaturalist species classification and detection dataset. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 8769–8778. IEEE, 2018.

Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. Matching networks for one shot learning. *Advances in neural information processing systems*, 29, 2016.

Xucong Wang, Zhe Zhao, Zekun Wang, Xiaofeng Cao, Xu Wang, Di Wu, Pengkun Wang, and Yang Wang. Long-tailed test-time adaptation for vision-language models. In *International Conference on Learning Representations*, volume 2026, pp. 46492–46530, 2026.

Zhengbo Wang, Jian Liang, Lijun Sheng, Ran He, Zilei Wang, and Tieniu Tan. A hard-to-beat baseline for training-free clip-based adaptation. In *International Conference on Learning Representations*, volume 2024, pp. 28841–28859, 2024.

Mitchell Wortsman, Gabriel Ilharco, Jong Wook Kim, Mike Li, Simon Kornblith, Rebecca Roelofs, Raphael Gontijo Lopes, Hannaneh Hajishirzi, Ali Farhadi, Hongseok Namkoong, et al. Robust fine-tuning of zero-shot models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7949–7961. IEEE, 2022.

Yongqin Xian, Saurabh Sharma, Bernt Schiele, and Zeynep Akata. f-vaegan-d2: A feature generating framework for any-shot learning. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10267–10276. IEEE, 2019.

Wenjia Xu, Yongqin Xian, Jiuniu Wang, Bernt Schiele, and Zeynep Akata. Attribute prototype network for any-shot learning. *International Journal of Computer Vision*, 130(7):1735–1753, 2022.

Sijie Yan, Yuanjun Xiong, Kaustav Kundu, Shuo Yang, Siqi Deng, Meng Wang, Wei Xia, and Stefano Soatto. Positive-congruent training: Towards regression-free model updates. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14299–14308, 2021.

Hantao Yao, Rui Zhang, and Changsheng Xu. Visual-language prompt tuning with knowledge-guided context optimization. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6757–6767. IEEE, 2023.

Tao Yu, Zhihe Lu, Xin Jin, Zhibo Chen, and Xinchao Wang. Task residual for tuning vision-language models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10899–10909. IEEE, 2023.

Xiaohua Zhai, Xiao Wang, Basil Mustafa, Andreas Steiner, Daniel Keysers, Alexander Kolesnikov, and Lucas Beyer. Lit: Zero-shot transfer with locked-image text tuning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18102–18112. IEEE, 2022.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11941–11952. IEEE, 2023.

Renrui Zhang, Wei Zhang, Rongyao Fang, Peng Gao, Kunchang Li, Jifeng Dai, Yu Qiao, and Hongsheng Li. Tip-adapter: Training-free adaption of clip for few-shot classification. In *European conference on computer vision*, pp. 493–510. Springer, 2022.

Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Conditional prompt learning for vision-language models. In *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 16795–16804. IEEE, 2022a.

Kaiyang Zhou, Jingkang Yang, Chen Change Loy, and Ziwei Liu. Learning to prompt for vision-language models. *International journal of computer vision*, 130(9):2337–2348, 2022b.

Zhi Zhou, Ming Yang, Jiang-Xin Shi, Lan-Zhe Guo, and Yu-Feng Li. Decoop: Robust prompt tuning with out-of-distribution detection. *arXiv preprint arXiv:2406.00345*, 2024.

A Proofs and Accounting

A.1 Conditional Locality

Proof of Proposition 2. Let the appended classes have frozen and parent scores strictly below every original endpoint score. They cannot enter either endpoint top-2 set or replace either endpoint winner. The original top-2 sets, their union $U_2(x)$, and the deterministic ordering are therefore unchanged. All original support counts are fixed, so the local count contrast in Eq. 7 is unchanged. The two endpoint margins, winner-disagreement indicator, residual contrast, and bias are also unchanged. Hence $z(x)$ and $\eta_\theta(x)$ are unchanged, and Eq. 4 leaves every original output score unchanged. Each appended class is below all original classes at both endpoints and therefore remains below them for every convex interpolation coefficient in $[0, 1]$. No appended class can win, so the final prediction is unchanged. □

The proposition assumes fixed parent scores, support counts, and fitted controller parameters. It does not cover vocabulary-dependent text encoding, parent retraining, additions that enter the active candidate set, or cases in which the original vocabulary contains fewer classes than the selected top-K.

A.2 Scalar Containment and Exact Endpoints

When $\theta = 0$, Eq. 8 returns $\eta = \rho$ on Ω_+ and zero on Ω_0. The implementation uses explicit endpoint branches, so $\rho = 0$ copies s^0 exactly and $\rho = 1$ copies s^a exactly. The matched scalar and the LEC delta-zero ablation therefore have identical predictions, not merely approximately equal scores.

A.3 Benefit-Cost Identity

For a fixed comparator M, the routed predictor $R(M)$ and M agree on Ω_+. On Ω_0, their correctness difference for query x is

$$\mathbf{1}[\hat{y}^0(x) = y, \hat{y}^M(x) \neq y] - \mathbf{1}[\hat{y}^0(x) \neq y, \hat{y}^M(x) = y]. \tag{11}$$

Multiplying by fixed evaluation weights and summing gives Eq. 10. A closed-route query on which Frozen is wrong contributes to the realized cost only when M is correct. Thus the total frozen-wrong closed mass is an upper bound on blocked corrections, not the number of corrections lost by every comparator.

B Experimental and Reproduction Protocol

B.1 Datasets and Support Construction

Table 10 summarizes the seven datasets used by the primary benchmark. The natural-history extension draws licensed images from iNaturalist (Van Horn et al., 2018) and is evaluated separately from the primary aggregate.

Table 10: Primary benchmark datasets and support profiles. Class counts follow the canonical task definitions used by the source pipelines.

Dataset	Classes	Support profiles
CIFAR-10	10	mixed, positive, near-uniform, uniform 8
CIFAR-100	100	mixed, positive, near-uniform, uniform 8
DTD	47	mixed, positive, near-uniform, uniform 8
Oxford-IIIT Pets	37	mixed, positive, near-uniform, uniform 8
Flowers-102	102	mixed, positive, near-uniform, uniform 8
Stanford Cars	196	mixed, positive, near-uniform, uniform 8
Food-101	101	mixed, positive, near-uniform, uniform 8

For each dataset, a seeded class permutation assigns the cyclic mixed counts $(0, 1, 2, 4, 8)$. The positive heterogeneous profile uses $(1, 1, 2, 4, 8)$. The near-uniform profile preserves the total positive-heterogeneous budget while distributing it as evenly as possible. Uniform 8 assigns eight examples to

every class. Seeds 11 through 15 sample support without replacement. Frozen query panels are fixed across methods within an episode.

B.2 Encoders and Parent

The encoders are CLIP ViT-L/14, SigLIP2 B/16-256, and MobileCLIP2-S0. Frozen image and text features are normalized and scored at the source pipeline scale. TaskRes adds a zero-initialized class residual to the text classifier, masks unsupported residual rows, normalizes the resulting class vectors, and trains for 300 AdamW steps on the complete support batch with balanced class-macro cross-entropy. The full-support parent is identical for every controller within an episode.

B.3 Inner Holdout and Controller Fitting

Each class with at least two support examples contributes $\max(1, \mathrm{round}(n_c/4))$ inner-holdout examples. The inner parent is fitted only on the remaining support examples, and the inner count vector uses only those training labels. The matched scalar selects $\rho \in \{0, .25, .50, .75, 1\}$ by holdout class-macro accuracy, with ties toward the smaller value and fallback 0.50 when no legal split exists.

LEC uses the same inner fits and holdout labels. It fits the six coefficients in Eq. 8 by class-macro cross-entropy with $\lambda = \max(1, 20/N_{\mathrm{open}})$, coefficient bounds $[-.5, .5]$, and deterministic L-BFGS-B for at most 100 iterations. The continuous same-loss constant optimizes the same loss with the same holdout information but without query features. No controller uses final-query labels, final-query normalization, or a test-time hyperparameter sweep.

B.4 Statistics

Class-macro accuracy averages per-class query accuracy. For the seven-dataset benchmark, support seeds are averaged within each dataset and datasets are then equally weighted. The seed interval uses 5,000 paired bootstrap samples conditional on the fixed datasets and query panels. Dataset sensitivity resamples datasets and their seeds. With seven datasets, the latter is interpreted as a sensitivity analysis. No multiplicity correction is applied to exploratory subgroup analyses.

C Expanded Benchmark Results

C.1 Absolute Results by Encoder and Profile

Table 11 gives all twelve encoder-profile means in the primary benchmark. LEC exceeds the matched scalar and same-loss constant in every combination.

Table 11: TaskRes class-macro accuracy (%) by support profile and encoder. Matched is the budget-matched selected scalar; Constant is optimized under the same holdout loss.

Profile	Encoder	Frozen	SafeRoute-Count	Matched	Constant	SafeRoute-LEC
Mixed	CLIP L/14	80.901	83.581	83.643	83.833	**84.893**
Mixed	SigLIP2 B/16	85.828	86.981	87.432	87.550	**88.576**
Mixed	MobileCLIP2-S0	81.292	83.352	83.338	83.720	**84.778**
Positive	CLIP L/14	80.901	84.582	84.239	84.558	**85.671**
Positive	SigLIP2 B/16	85.828	87.575	87.488	87.858	**88.963**
Positive	MobileCLIP2-S0	81.292	83.948	84.053	84.221	**85.256**
Near-uniform	CLIP L/14	80.901	86.126	86.508	86.869	**87.941**
Near-uniform	SigLIP2 B/16	85.828	88.046	89.024	89.417	**90.471**
Near-uniform	MobileCLIP2-S0	81.292	84.732	85.688	85.922	**87.013**
Uniform 8	CLIP L/14	80.901	88.262	88.580	88.712	**89.763**
Uniform 8	SigLIP2 B/16	85.828	89.360	90.481	90.635	**91.657**
Uniform 8	MobileCLIP2-S0	81.292	86.202	87.122	87.338	**88.396**

C.2 Mixed-Support Dataset Deltas

Table 12 expands Table 7 with the same-loss constant comparison. Every dataset-level mean is positive for all three comparators.

Table 12: LEC minus comparator under mixed support, averaged over three encoders and five support seeds (percentage points).

Dataset	Matched scalar	Same-loss constant	SafeRoute-Count
Cars	+1.284	+1.117	+1.557
CIFAR-10	+1.111	+1.011	+1.331
CIFAR-100	+1.163	+0.998	+1.711
DTD	+1.706	+1.110	+0.415
Flowers-102	+1.542	+1.145	+2.066
Food-101	+1.072	+0.981	+1.862
Pets	+1.066	+0.977	+1.168

D Structural and Oracle Diagnostics

D.1 Dummy-Class Tests

The structural test holds the fitted parent, scalar, coefficients, and original scores fixed. It appends one, two, or five classes with count 0, 4, or 128. Their frozen and parent scores are below every original endpoint score, and the additions remain outside the original top-2 union. The first $\min(256, N)$ query rows are evaluated for each episode and condition.

The complete results appear in Table 8. The test verifies the specified conditional locality property. It does not claim invariance to arbitrary semantic distractors, classes that enter active competition, vocabulary-dependent text re-encoding, or retraining with an enlarged vocabulary.

D.2 Contract Cost and Interpolation Headroom

The source-bound diagnostic in Table 13 contains 108 completed fits and 870,840 query records. The pointwise interpolation oracles use final-query labels and are not deployable. The routed oracle chooses a query-specific interpolation coefficient in $[0, 1]$ while obeying the retention route. The unrestricted oracle may also change closed-route predictions.

Table 13: Source-bound label-oracle diagnostics (class-macro %). Selected is deployable; the other two optimized rows use query labels.

Condition	Selected	Routed oracle	Unrouted oracle	Closed wrong mass
Base 1	83.095	87.135	89.196	3.633
Unique near 8	82.408	87.192	89.356	3.633
Unique far 8	83.858	87.694	89.840	3.633
Mean	83.121	87.340	89.464	3.633

The 2.124-point mean difference between unrestricted and routed oracles is the realized cost of the strict route along these interpolation paths. The 4.220-point difference between the selected scalar and routed oracle is remaining controller headroom under the route. Neither number implies that support-only signals can recover the label oracle.

E Transfer Scope and Reproducibility Audit

E.1 Natural-History Extension

A separate extension uses a fixed catalog of 50 California bird species, five support histories, and a shared 1,000-query panel from disjoint users. It evaluates three encoders and three parent families.

Across the complete nine-system mean, LEC improves budget-matched scalar selection by 0.896 class-macro percentage points while the route retains protected decisions. This extension changes both the support-acquisition process and the parent family, but it remains a bounded natural-history study rather than a claim of universal transfer.

E.2 Numerical Audit

The compact audit covers 1,305 completed episodes and deterministic tie checks, 5,220 committed artifact hashes, and 211,374,360 recomputed prediction-metric rows. The maximum class-macro recount error is 4.27×10^{-14} percentage points. Independent checks produce zero baseline prediction changes and zero error in a local-η subset. These checks verify stored inputs and deterministic calculations; they are not an independent feature-extraction or data-acquisition replication.

E.3 Data Boundaries

The seven-dataset benchmark uses source-bound decoded-image identities and official class names. Exact file or decoded-pixel checks do not rule out perceptual near-duplicates or pretraining exposure. Query-mask sensitivities without parent refitting are treated only as sensitivities, not as decontaminated training. The natural-history audit checks selected image identities and user separation but does not establish geographic or demographic representativeness.