RELIABILITY-CONTROLLED COMPETITIVE ASSIGNMENT FOR CONFLICT-FREE PSEUDO-LABELING IN FEW-SHOT MULTIMODAL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Prototype-guided self-training expands a small labeled set by assigning unlabeled instances to class prototypes. Under few-shot supervision, however, overlapping class-specific candidate sets create an assignment ambiguity: independent per-class selection can admit the same instance with conflicting pseudo-labels. Confidence thresholds and class-aware admission rules regulate sample selection but do not, by themselves, resolve competition between classes. We propose reliability-controlled competitive assignment, which scores candidate assignments using a dual-constraint reliability criterion and resolves cross-class conflicts through a per-instance argmin before admission. This enforces a one-instance–one-label constraint, guaranteeing an admitted pool free of duplicate-instance label conflicts without assuming that every assigned label is correct. We integrate competitive assignment into a progressive self-training framework that combines class-balanced, confidence-guided pool expansion with momentum-smoothed pseudo-label refinement. Experiments on two few-shot multimodal classification benchmarks, MUUFL Gulfport and Trento, compare competitive assignment with independent per-class selection using matched-seed paired runs. Competitive assignment improves mean accuracy by 7.95 percentage points on MUUFL (95% confidence interval: [3.15, 12.75]) and 4.79 percentage points on Trento, although uncertainty is greater in Trento's five-split evaluation. These findings demonstrate that resolving cross-class assignment conflicts before admission can improve few-shot multimodal classification and establish assignment consistency as a useful complement to confidence-based pseudo-label selection.

1 INTRODUCTION

Pseudo-labeling pipelines that grow a labeled set from a few seeds often build one *candidate set* per class. Each class k independently decides which unlabeled instances satisfy its admissibility conditions, giving sets $\mathcal{C}_1, \ldots, \mathcal{C}_K$. For an instance i, the *claimant set* $\mathcal{A}_i = \{k : i \in \mathcal{C}_k\}$ then falls in one of three cases: $|\mathcal{A}_i| = 0$, the instance is unclaimed; $|\mathcal{A}_i| = 1$, ownership is unambiguous; $|\mathcal{A}_i| > 1$, several classes claim it and ownership is ambiguous. Two different decisions are therefore involved. *Selection* asks whether instance i is admissible for class k. *Ownership* asks, when several classes admit i, which class should own it. These are distinct: the first is a property of an instance and one class, the second of an instance and all its claimants, and a rule that answers the first says nothing about the second.

The pseudo-labeling literature has developed selection in depth. Fixed, curriculum, self-adaptive and instance-dependent confidence thresholds, and reliability filters, decide whether a candidate is trusted (Sohn et al., 2020; Zhang et al., 2021a; Wang et al., 2023; Chen et al., 2023; Li et al., 2023); class-aware and rebalancing rules decide how many instances each class may contribute (Wei et al., 2021); and label-allocation methods impose marginal constraints over a batch or dataset (Asano et al., 2020; Tai et al., 2021). What is not the object of these mechanisms is the ownership of an instance that independently constructed class-conditioned candidate sets both admit. When admission is then run per class, ownership is settled implicitly by how the per-class lists are traversed and written, that is,

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.

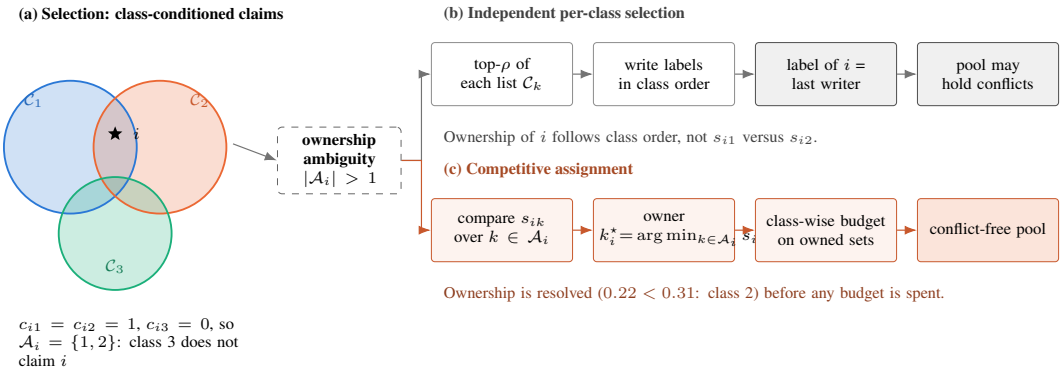


Figure 1: Selection versus ownership. (a) Selection: each class independently claims instances; instance i is claimed by classes 1 and 2, not by class 3. (b) Independent per-class selection admits from each list and writes labels class by class, so ownership of i follows class order. (c) Competitive assignment compares the claimants' reliability scores, resolves ownership, applies the class-wise budget to owned sets, and yields a conflict-free pool.

by class order and not by the competing evidence. We call this **assignment ambiguity**: independent per-class selection can admit the same instance with conflicting pseudo-labels, and the surviving label need not be the best supported one.

We formalize ownership as a separate decision and resolve it before any budget is spent. **Reliability-controlled competitive assignment** gives every claimed instance to the claimant with the lowest reliability score through a per-instance argmin and only then lets each class spend its admission budget on the instances it owns (Figure 1). This enforces a one-instance–one-label constraint. We prove that the resulting ownership is unique, disjoint, order-invariant and claimant-wise minimal, that budgets remain feasible, that the rule is the exact solution of a claim-constrained allocation problem without class marginals, and that it coincides with independent admission when the candidate sets are pairwise disjoint (Propositions 1 and 2). Its behavior therefore differs from independent selection precisely where claims overlap. The guarantee concerns consistency and does not assume that any assigned label is correct.

To evaluate the principle we perform a controlled intervention. We instantiate ownership resolution in a representative prototype-guided multimodal few-shot pipeline (PGPL; Zhang et al., 2026) and preserve its backbone, prototype construction, reliability score, candidate gates, thresholds, training schedule and progressive self-training machinery. The only change is how competing claims are resolved before budgeted admission, so any difference can be attributed to ownership resolution and not to a redesigned architecture. On MUUFL Gulfport, 42% of gate-passing pixels are claimed by two or more classes; on Trento, 4% on average but 0.01% to 12% across splits. Competitive assignment raises the purity of the initial pool by 7.1 points on MUUFL and 2.5 on Trento at a 10% budget, and the resulting configuration improves mean overall accuracy from 56.25% to 64.20% on MUUFL and from 93.20% to 97.99% on Trento.

Contributions. (1) *Problem:* we separate selection from ownership and formalize ownership ambiguity through claimant sets. (2) *Mechanism:* reliability-controlled competitive assignment resolves ownership before budgeted admission. (3) *Theory:* we establish its structural properties, its allocation-problem interpretation and its reduction to independent admission when there is no overlap. (4) *Diagnostics:* we provide measurements of when ownership ambiguity is present and consequential. (5) *Controlled evidence:* matched-seed paired runs and a 2×2 design separate the effect of ownership resolution from that of the admission budget.

2 Related Work

Whether to admit an instance. Pseudo-labeling, consistency regularization and iterative self-training underlie semi-supervised learning (Lee, 2013; Tarvainen & Valpola, 2017; Xie et al., 2020; Sohn et al., 2020), and admission thresholds have been made curriculum-based, self-adaptive, soft or

instance-dependent (Zhang et al., 2021a; Wang et al., 2023; Chen et al., 2023; Li et al., 2023). These evaluate one predicted distribution per instance and answer selection.

How much each class contributes. Class-aware rules rebalance contributions under imbalance (Wei et al., 2021); in few-shot pipelines, per-class budgets and quotas play this role. They set how many instances a class admits, not who owns a contested one.

Global label allocation. Optimal-transport and Sinkhorn approaches assign labels jointly under batch- or dataset-level marginal constraints (Asano et al., 2020; Tai et al., 2021). Their constraint is global. The ownership ambiguity studied here is local and sparse: it arises among the few classes that claim a given instance because independently generated class-conditioned eligibility sets overlap, before budgeted admission and without class marginals. In Section 4.2 we show that competitive assignment is a claim-constrained allocation with no imposed marginals; adding marginal constraints would couple instances and lead to a transport-type problem, so the two views are complementary. We do not claim novelty for the argmin; the contribution is the identification, formalization and diagnosis of ownership as a separate decision, and its resolution before budget expenditure.

Prototype-guided few-shot learning. Prototypes provide a classifier-independent notion of closeness (Snell et al., 2017) and have been used to denoise pseudo-labels (Zhang et al., 2021b). In few-shot HSI–LiDAR classification, PGPL (Zhang et al., 2026) builds an initial pool from spectral-angle and elevation gates around the seeds and expands it progressively; we use it as the control configuration of our intervention.

Multimodal HSI–LiDAR classification. Supervised fusion networks combine spectral and elevation streams through cross-modal enhancement, attention or transformers (Fang et al., 2021; Mohla et al., 2020; Ding et al., 2022; Yao et al., 2023); semi-supervised variants use adversarial, contrastive, decoupling, state-space or masked-autoencoding objectives, or auxiliary pseudo-label networks (Lu et al., 2023; Ding et al., 2024; Qu et al., 2024; Li et al., 2025; Lin et al., 2023; Wang et al., 2025).

The works above treat whether pseudo-labels should be admitted and, separately, how assignments can be globally balanced. Our setting isolates the ownership decision that arises when independently generated class-conditioned candidate sets overlap before budgeted admission.

3 Problem Formulation

3.1 Ownership under overlapping claims

Let U be a set of unlabeled instances and K the number of classes. **Candidate generation:** each class k independently produces $c_{ik} \in \{0, 1\}$ for every $i \in U$, with candidate set $\mathcal{C}_k = \{i : c_{ik} = 1\}$, and a reliability score $s_{ik} \geq 0$ for the claim, where smaller means stronger. **Claim formation:** $\mathcal{A}_i = \{k : c_{ik} = 1\}$; instances with $|\mathcal{A}_i| = 1$ are single-claimed and those with $|\mathcal{A}_i| \geq 2$ multi-claimed. **Overlap:** let $\mathcal{U}^+ = \{i : |\mathcal{A}_i| \geq 1\}$ be the claimed instances; the overlap rate is

$$\Omega = \frac{1}{|\mathcal{U}^+|} \sum_{i \in \mathcal{U}^+} \mathbb{1}[|\mathcal{A}_i| > 1]. \tag{1}$$

Independent per-class selection. With budget ratio $\rho \in (0, 1]$, class k admits $\mathcal{B}_k$, the $\lceil \rho |\mathcal{C}_k| \rceil$ lowest-score members of $\mathcal{C}_k$. When labels are written class by class, an instance in several $\mathcal{B}_k$ receives $\hat{y}(i) = \max\{k : i \in \mathcal{B}_k\}$: such instances *collide*, and their ownership is a function of the class index order. The sequence is candidate generation $\rightarrow$ claim formation $\rightarrow$ budgeted admission $\rightarrow$ label writing, and ownership is settled implicitly in the last step. Comparing s_{ik} across classes requires that scores be comparable within an instance; this is the only assumption the formulation needs.

3.2 Instantiation in prototype-guided multimodal learning

The assignment operator references only the class-conditioned candidate sets and the reliability scores supplied to it, and no backbone-specific quantity. Our experiments evaluate it in the prototype-guided HSI–LiDAR setting considered here. A scene has a hyperspectral cube and a co-registered LiDAR raster; instance i has spectrum x_i and elevation feature z_i and is classified from a $p \times p$ patch. Each class k has a labeled seed set S_k with $|S_k| = n$ ($n \in \{2, 5\}$); a label-hidden support domain U

is available for pseudo-labeling and a held-out test domain T, disjoint from $S \cup U$, is used only for evaluation. With $\theta_{ik} = \min_{s \in S_k} \arccos \frac{x_i^\top x_s}{\|x_i\| \|x_s\|}$ the spectral angle (Kruse et al., 1993) and $\delta_{ik} = \min_{s \in S_k} \|z_i - z_s\|_2$ the elevation residual to the nearest seed of class k (band-standardized spectra; elevation clipped to its 1st–99th percentile range and standardized with non-test statistics), the dual-constraint reliability criterion is

$$c_{ik} = \mathbb{1}[\theta_{ik} < \tau_{\text{sam}}] \mathbb{1}[\delta_{ik} < \tau_h], \qquad s_{ik} = \tfrac{2}{\pi} \theta_{ik} + \lambda \, \delta_{ik} / R, \tag{2}$$

with R the norm of the elevation range over valid pixels. The gate produces the candidate sets and the score is the within-instance comparable reliability.

4 METHOD

4.1 SELECTION–OWNERSHIP DECOMPOSITION AND COMPETITIVE ASSIGNMENT

We make ownership an explicit stage between claim formation and budgeted admission: (1) candidate generation, c_{ik}; (2) claim formation, $\mathcal{A}_i$; (3) ownership resolution; (4) budgeted admission; (5) pseudo-label pool construction. Stage 3 gives each claimed instance to its lowest-score claimant,

$$k_i^\star = \arg\min_{k \in \mathcal{A}_i} s_{ik} \;\; \text{(ties to the smaller class index)}, \qquad \mathcal{C}'_k = \{i : k_i^\star = k\}. \tag{3}$$

Stage 4 lets class k admit $\mathcal{B}'_k$, the $\lceil \rho |\mathcal{C}'_k| \rceil$ lowest-score members of $\mathcal{C}'_k$, and stage 5 forms the pool $\mathcal{P}_0 = S \cup \{(i,k) : i \in \mathcal{B}'_k\}$. Eligibility (stage 1) and ownership (stage 3) are separate operations, and ownership does not depend on ρ.

4.2 STRUCTURAL PROPERTIES

Proposition 1. *Let $s_{ik} \geq 0$, ties in Eq.* (3) *be broken toward the smaller class index, ties among equal scores in admission by a fixed rule that does not depend on class order, and $\rho \in (0, 1]$. (a)* Uniqueness: *every claimed instance has exactly one owner $k_i^\star$. (b)* Disjointness: *the owned sets $\mathcal{C}'_k$ partition $\mathcal{U}^+$, and $\mathcal{B}'_k$ are pairwise disjoint. (c)* Order invariance: *$\mathcal{C}'_k$, $\mathcal{B}'_k$ and every pseudo-label are the same for every ordering of the classes. (d)* Claimant-wise optimality: *$s_{ik_i^\star} = \min_{k \in \mathcal{A}_i} s_{ik}$. (e)* Conflict elimination before budget expenditure: *no instance is admitted under two labels, and ownership is defined independently of ρ. (f)* Budget feasibility: *$|\mathcal{B}'_k| \leq \lceil \rho |\mathcal{C}_k| \rceil$. (g)* Reduction: *if $\mathcal{C}_j \cap \mathcal{C}_k = \emptyset$ for all $j \neq k$, then $\mathcal{C}'_k = \mathcal{C}_k$, $\mathcal{B}'_k = \mathcal{B}_k$, and competitive assignment yields the same pool as independent per-class selection.*

Proof. (a) $\mathcal{A}_i$ is finite and non-empty for claimed i, and $\arg\min$ with a fixed tie rule returns one element. (b) each claimed i lies in exactly one $\mathcal{C}'_k$ by (a), and $\mathcal{B}'_k \subseteq \mathcal{C}'_k$. (c) $k_i^\star$ is a minimum over the unordered set $\mathcal{A}_i$; only the tie rule uses an index. (d) is the definition. (e) follows from (b) and from Eq. (3) not involving ρ. (f) $\mathcal{C}'_k \subseteq \mathcal{C}_k$ and $\lceil \rho \cdot \rceil$ is monotone. (g) if the candidate sets are pairwise disjoint then $\mathcal{A}_i = \{k\}$ for each claimed $i \in \mathcal{C}_k$, so $k_i^\star = k$ and $\mathcal{C}'_k = \mathcal{C}_k$; the budgeted sets coincide as they apply the same rule to the same set, and no instance is in two $\mathcal{B}_k$, so $\hat{y}$ equals the owner label. □

By (g), the two procedures can differ only when some instance is multi-claimed: competitive assignment changes behavior precisely where overlapping claims make ownership ambiguous.

Proposition 2. *Consider $\min_z \sum_i \sum_k z_{ik} s_{ik}$ over $z_{ik} \in \{0, 1\}$ subject to $z_{ik} = 0$ for $k \notin \mathcal{A}_i$ and $\sum_k z_{ik} = 1$ for every claimed i. The minimizer is $z_{ik} = \mathbb{1}[k = k_i^\star]$ with $k_i^\star$ from Eq.* (3)*, unique when scores have no ties.*

Proof. The objective and the constraints separate over instances, so each i independently selects a minimizer of s_{ik} over $\mathcal{A}_i$. □

The equality constraint is needed because $s_{ik} \geq 0$ would otherwise make $z = 0$ optimal. The problem imposes no class marginals; with marginal constraints $\sum_i z_{ik} = m_k$ instances would be coupled and the problem would become a transport-type allocation, the regime of the global methods above.

What is not claimed. Property (d) is optimality within one instance's claimant set, not a global optimum under marginal targets; and the guarantee is about label *consistency*: an owned label can still be wrong, which Section 5.2 measures. Property (f) implies that exclusivity can shrink the pool at equal ρ, so ownership and budget must be varied jointly to be separated (Section 5.4).

4.3 CONTROLLED INTERVENTION AND PROGRESSIVE SELF-TRAINING

We instantiate the mechanism in PGPL (Zhang et al., 2026) as a controlled intervention. *Held fixed:* backbone (a dual-stream S^{2}ENet: three Conv–BN–ReLU blocks per modality, widths 128–64–32 for HSI and 8–16–32 for LiDAR, spatial and spectral cross-modal enhancement, 1×1 fusion, global pooling, linear classifier), feature extraction, prototype construction from the seeds, the reliability score and gates of Eq. (2), thresholds, warm-up (100 epochs on $\mathcal{P}_0$), the class-balanced, confidence-guided expansion, the momentum-smoothed refinement and the fine-tuning stage. *Changed:* the resolution of competing claims before budgeted admission. Every 15 of 200 expansion epochs, the expansion step draws 300 instances from the preparation pool $V = U \setminus \{i : (i, \cdot) \in \mathcal{P}_0\}$, keeps those whose maximum softmax probability exceeds $\tau(t)$ (0.99 for $u = t/E < 0.3$, decreasing linearly to 0.92 at $u = 0.85$, then constant) and adds at most Q=100 per predicted class; the refinement step draws 300 non-seed pool members, replaces each pseudo-label by the current prediction and each stored confidence q by $\alpha \max_k p_k + (1 - \alpha)q$ with α=0.3; seeds are never modified. Fine-tuning freezes the pool for 10 epochs at learning rate 10^{-4}. Isolating one decision makes attribution possible: had several components changed at once, no difference could be assigned to ownership. Algorithm 1 (Appendix) summarizes the pipeline. We call the competitive configuration with ρ=5% **C-EPG**.

5 EXPERIMENTS

5.1 SETUP AND INFERENTIAL SCOPE

Datasets. *MUUFL Gulfport* (Gulf Park campus, University of Southern Mississippi): 325×220 pixels at 1 m, 64 spectral bands (375–1050 nm) and two LiDAR-derived channels; 53,687 labeled pixels in 11 classes. *Trento* (Trento, Italy): 600×166 pixels at 1 m, 63 spectral bands (420–990 nm, AISA Eagle) and one LiDAR channel; 30,214 labeled pixels in 6 classes. Class sizes are strongly imbalanced (23,246 versus 183 pixels for the largest and smallest MUUFL classes).

Protocol. Splits are pixel-stratified: for each class, 80% of its pixels form the label-hidden support domain U and 20% the held-out test domain T; n pixels per class are drawn from U as seeds (n=2 MUUFL, n=5 Trento), so each split fixes its seeds, U and T, and its masks are hashed in a manifest. HSI bands are z-scored and LiDAR channels clipped to their 1st–99th percentile range and z-scored with statistics from non-test pixels; the LiDAR range R of Eq. (2) is computed over all valid pixels. Each instance is classified from a patch of the zero-padded full scene (9×9 MUUFL, 7×7 Trento), so the split is not spatially disjoint and patches of T instances include neighboring pixels of U; labels are never shared, and T instances are never candidates. Prototypes are the seeds themselves: θ_{ik} and δ_{ik} are taken to the nearest seed of each class. Three repeated-split protocols are used: *MUUFL confirmation* (splits 31415, 27182, 16180, 57721, 86753 $\times$ model seeds 2026, 1337, 3407: 15 matched runs per configuration), *MUUFL development* (splits 42, 123, 2021, 3407, 9999; model seed 2026) and *Trento* (the same five split seeds; model seed 2026: 5 matched runs). The budget ρ=5% was selected on the MUUFL development splits before the confirmation runs and carried over unchanged to Trento.

Inferential scope. The unit of replication is the data split (n=5 per dataset). Model-seed replicates within a split share their data and only average out training variability; they are not independent dataset replications. For the MUUFL paired differences, the standard deviation across model seeds within a split averages 3.45 pp, versus 8.68 pp between split means. We therefore report mean paired differences with two-sided 95% t-intervals over splits, and give run-level intervals over all matched runs only where they mirror the matched-seed design (Section 5.3), labeled as such. Standardized paired effects are $d_z = 0.92$ (MUUFL) and 0.69 (Trento); a mixed-effects model is not fitted, since five splits cannot identify variance components reliably. Interpretations below are made within this scope.

Configurations and inherited settings. The control configuration, *independent per-class selection*, is the PGPL pipeline with ρ=10% (Zhang et al., 2026), re-run under our protocols; the intervention is competitive assignment. For the 2×2 design we add independent selection at ρ=5% and competitive assignment at ρ=10%. Backbone, gates, λ, seeds, schedule and initialization are shared; the thresholds $(\tau_{\text{sam}}, \tau_h) = (0.24\ \text{rad}, 2.0)$ for MUUFL and $(0.35\ \text{rad}, 0.6)$ for Trento, and $\lambda = 0.04$, are those of the PGPL reference configuration and were not re-tuned for competitive assignment, which keeps method-specific hyperparameter optimization out of the controlled comparison; ρ=5%

is the one setting selected for it, and the pair of cells at the inherited ρ=10% shows the effect without any retuning. We do not claim these values are optimal for competitive assignment. The comparison with the PGPL configuration is a controlled one and supports attribution to ownership resolution; the works in Section 2 are contextual, since their protocols differ, and no number from them is compared with ours.

Training and evaluation. Adam with weight decay 10^{-4} and cosine annealing in every stage; batch size 32; dropout 0.1; learning rates 10^{-3}, 5×10^{-4}, 10^{-4} for warm-up, expansion and fine-tuning; a single NVIDIA GeForce RTX 5090 GPU. Training is not bit-wise deterministic. We report overall accuracy (OA), average accuracy (AA), Cohen's κ, expected calibration error (ECE) and Brier score on held-out pixels, with differences paired by split (and model seed where present).

5.2 When does competitive assignment matter?

We treat the following measurements as a diagnostic framework for class-conditioned admission and ask, in the order in which the mechanism operates, whether ownership ambiguity is present and what it changes. Ground truth of the support domain is used for diagnostics only. **Overlap rate** Ω (Eq. 1) measures how often eligibility produces ownership ambiguity. **Collision rate** is the fraction of admitted slots whose instance is also admitted by another class, that is, how often class-wise budgets target competing instances. **Non-minimal rate** is the fraction of independently admitted instances whose lowest-score claimant is a different class, that is, how often sequential writing awards ownership to a claimant other than the most reliable one. **Order sensitivity** is the change in the pool under different class traversal orders. **Purity** is the fraction of admitted pseudo-labels equal to the true class.

How frequent are overlapping claims, and what does independent selection do with them? On MUUFL, $\Omega = 42.4\%$ (39.2–46.0% across splits); on Trento, 4.0% on average and 0.01–11.6% across splits (Table 1; Appendix Table 8). Pooled over the five splits, the affected set $\{i : |\mathcal{A}_i| > 1\}$ contains 52,519 MUUFL and 1,879 Trento instances. At ρ=10%, 10.8% (MUUFL) and 0.4% (Trento) of admitted slots collide, and 1,111 of 18,598 admitted MUUFL instances (6.0%) and 96 of 4,847 Trento instances (2.0%) are non-minimal. The admitted label is correct for 21.2% of the non-minimal MUUFL instances and 5.2% of the Trento ones, whereas the minimum-score owner would have been correct for 43.7% and 94.8%.

Does ownership follow the reliability scores, and does class order matter? Not under independent selection (Appendix Table 10; instances pooled over splits). On MUUFL the last-writer rule is correct for 20.3% of the affected set, below the 31.8% expected from a uniform choice among claimants and the minimum-score rule for 39.5%. On Trento, order-based rules range from 8.3% (first-writer) to 91.6% (last-writer) depending on which class is indexed first, and the minimum-score rule gives 85.5% (uniform: 49.9%). Order-based accuracy is thus set by class indexing, and can be favorable, as the last-writer order is on Trento; the minimum-score rule is invariant to it and exceeds the uniform baseline on both datasets. The two rules disagree on 49.4% (MUUFL) and 12.4% (Trento) of the affected set, and the disagreements that reach the admitted pool are the non-minimal admissions above. Pool purity under independent selection is 65.9% (ascending class order) and 64.1% (descending) on MUUFL at ρ=10% and varies by at most 0.1 points on Trento. Stratifying the affected set by the margin between the two lowest claimant scores, the minimum-score rule is most accurate at large margins on both datasets and, on Trento, the two rules disagree mostly at small margins (Appendix Table 9).

What does competition change in the pool? The competitive pool is contained in the independent pool in all 20 split–budget pairs, so competition acts mainly by excluding contested admissions and relabels few instances. At ρ=10%, 6,179 of the 18,598 independently admitted MUUFL instances (33.2%) are excluded; 51.7% of them were correctly labeled, against 73.7% of the retained instances, and 191 (1.0%) are relabeled (correct: 31.4% before, 52.9% after). On Trento 180 of 4,847 (3.7%) are excluded, 23.9% of them correct against 98.6% of the retained, and none is relabeled. At ρ=5% the excluded MUUFL instances are more often correct (69.1%), and the purity gain is smaller. Purity rises from 65.9% to 73.0% (MUUFL) and from 96.1% to 98.6% (Trento) at ρ=10%, and from 73.4% to 75.2% and 96.0% to 98.7% at ρ=5%. Competitive pools have no colliding or non-minimal admissions and do not depend on class order (Proposition 1); none is fully pure, since the guarantee is consistency, not correctness.

Table 1: Mechanism per split (MUUFL development splits, Trento splits). Ω: overlap rate; non-min.: non-minimal share of independently admitted instances; pool agr.: shared fraction of (instance, label) pairs; ΔPurity (pp) and ΔOA_ρ (pp, budget ρ): competitive minus independent; non-min., agreement and purity at ρ=10%.

Dataset	Split	Ω (%)	Non-min. (%)	Pool agr.	ΔPurity	ΔOA_{10}	ΔOA_5
MUUFL	42	41.19	8.0	63.5%	+9.9	+7.1	+4.3
MUUFL	123	42.64	5.2	67.1%	+6.4	+3.8	+0.7
MUUFL	2021	45.97	6.9	60.3%	+5.1	+0.6	+3.2
MUUFL	3407	39.19	2.6	71.3%	+6.8	+1.5	−1.7
MUUFL	9999	43.43	7.3	66.5%	+6.9	+10.1	+4.6
Trento	42	6.35	3.1	94.4%	+4.5	+6.9	+1.7
Trento	123	11.63	5.7	90.0%	+7.8	+11.3	+23.2
Trento	2021	0.01	0.0	100.0%	+0.0	−0.0	−1.6
Trento	3407	0.67	0.1	99.3%	−0.0	−0.2	−0.1
Trento	9999	0.60	0.0	99.3%	+0.2	+0.4	−1.9

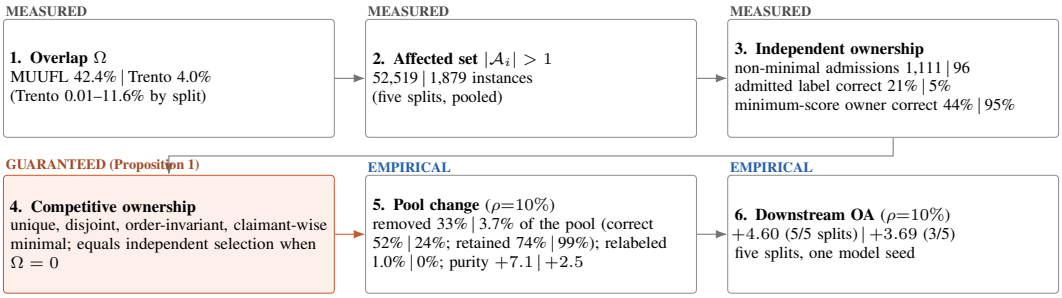


Figure 2: From overlap to downstream performance, with values for MUUFL | Trento (five splits pooled, ρ=10%). Steps 1–3 are measurements, step 4 holds by construction, steps 5–6 are empirical.

Overlap regimes and a near-zero-overlap check. The observed overlap of the two benchmarks provides complementary empirical regimes; this was observed, not designed. Proposition 1g predicts that the two procedures coincide without overlap. Trento split 2021, identified after the fact, provides an empirical check: it has one multi-claimed instance among 8,438 claimed (Ω=0.012%), the two pools are identical at both budgets (no instance excluded or relabeled), consistent with the reduction property, and OA differs by -0.03 (ρ=10%) and -1.60 (ρ=5%), which is consistent with training variability alone. On the two other low-overlap splits ($\Omega \approx 0.6\%$), 4–7 instances (at most 0.7% of the pool) are excluded and ΔOA lies between -1.9 and $+0.4$. On the two Trento splits with Ω>6%, 5.6% and 10.0% of the pool is excluded at ρ=10% and OA rises by $+6.9$ and $+11.3$ points. MUUFL is in the high-overlap regime in every split, and 28–39% of its independent pool is excluded.

Across the ten splits (two datasets pooled; descriptive), the rank association with ΔOA at ρ=10% (ρ=5%) is $+0.50$ ($+0.58$) for Ω, $+0.57$ ($+0.60$) for the collision rate, $+0.75$ ($+0.89$) for the non-minimal rate and $+0.88$ ($+0.47$) for the purity change. Within Trento, Ω orders the effect ($+0.70$, $+0.90$; n=5); within MUUFL, where Ω is nearly constant, the purity change does so at ρ=10% ($+0.80$) but not at ρ=5% ($+0.10$). The quantities closest to the mechanism thus track the downstream difference at least as closely as Ω, in line with the chain in Figure 2: overlap creates the opportunity for ownership conflicts, conflicts determine how the pool changes, and pool quality may or may not translate into accuracy.

5.3 Downstream Classification

Across the 15 matched MUUFL runs, C-EPG improves mean OA by 7.95 points (56.25% to 64.20%; run-level 95% interval [3.15, 12.75]; 12 of 15 runs won). Across the five matched splits the mean OA difference is 7.95 (interval [-2.82, $+18.72$]), with AA $+5.20$ and κ $+8.89$ (Table 2; per-split values in Appendix Figure 4); the direction was positive in four splits. On Trento, across five matched splits, OA rose by 4.79 points (93.20% to 97.99%; interval [-3.89, $+13.48$]), AA by 1.87 and κ by 6.11;

Table 2: Independent per-class selection (ρ=10%) versus C-EPG (competitive, ρ=5%). Split level: MUUFL confirmation protocol (runs averaged over model seeds within each split); Trento five splits. Δ: mean of five paired split differences (pp; κ scaled by 100) with a 95% t-interval over splits; "Won": splits won.

Dataset	Metric	PGPL	C-EPG	Δ [95% CI]	Won
MUUFL	OA	56.25	64.20	+7.95 [−2.82, +18.72]	4/5
MUUFL	AA	60.56	65.76	+5.20 [−2.58, +12.98]	4/5
MUUFL	$\kappa\times100$	47.93	56.82	+8.89 [−2.43, +20.22]	4/5
Trento	OA	93.20	97.99	+4.79 [−3.89, +13.48]	3/5
Trento	AA	95.51	97.38	+1.87 [−3.36, +7.10]	3/5
Trento	$\kappa\times100$	91.22	97.33	+6.11 [−4.91, +17.14]	3/5

Table 3: Top: OA (%) for each combination of ownership mechanism and ρ. Bottom: paired effects between cells, mean (pp) with 95% t-interval over five splits, and splits won.

Dataset	Independent ρ=10% (PGPL)	Independent ρ=5%	Competitive ρ=10%	Competitive ρ=5% (C-EPG)
MUUFL	56.19	61.81	60.79	64.03
Trento	93.20	93.72	96.89	97.99

Dataset	Effect	Δ OA, pp [95% CI over splits]	Won
MUUFL	Budget alone (independent 5% - 10%)	+5.62 [−3.02, +14.27]	4/5
MUUFL	Competition at ρ=5% (competitive - independent)	+2.22 [−1.09, +5.52]	4/5
MUUFL	Competition at ρ=10% (competitive - independent)	+4.60 [−0.34, +9.54]	5/5
MUUFL	Budget under competition (competitive 5% - 10%)	+3.24 [−3.70, +10.17]	3/5
Trento	Budget alone (independent 5% - 10%)	+0.52 [−6.23, +7.27]	2/5
Trento	Competition at ρ=5% (competitive - independent)	+4.28 [−9.00, +17.55]	2/5
Trento	Competition at ρ=10% (competitive - independent)	+3.69 [−2.74, +10.12]	3/5
Trento	Budget under competition (competitive 5% - 10%)	+1.11 [−1.79, +4.00]	3/5

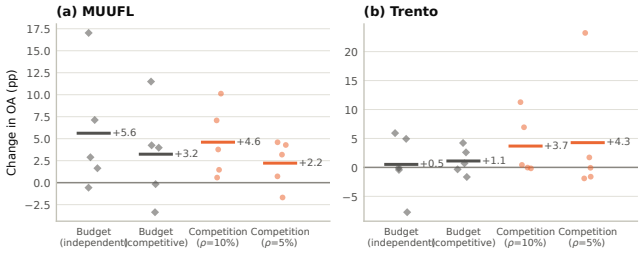


Figure 3: Per-split effects of ownership resolution (circles) and of the admission budget (diamonds) in the 2×2 design; bars are means over five splits.

the direction was positive in three splits, and in the two others independent selection was already at 99% (−1.69, −0.50). The largest differences occurred where independent selection was weakest (Trento split 123: 81.5% to 97.0%). The interval reflects substantial split-level uncertainty, wider on Trento than the MUUFL run-level interval because the latter pools three model seeds per split.

5.4 Ownership resolution versus admission budget: the 2×2 intervention

C-EPG differs from the control in two ways: the ownership mechanism and ρ. Changing both at once would confound their effects, so we cross them: **Factor A**, independent versus competitive ownership; **Factor B**, $\rho \in \{10\%, 5\%\}$ (Table 3, Figure 3; five splits, one model seed).

The pair (independent, ρ=10%) and (competitive, ρ=10%) differs only in ownership resolution and uses the inherited PGPL settings without re-tuning: OA changes by +4.60 on MUUFL (five of five splits) and +3.69 on Trento (three of five). At ρ=5% the changes are +2.22 and +4.28. Of the 20 paired split-level comparisons, ownership resolution improved OA in 14; the six decreases are no larger than 1.9 points, comparable to the training variability measured above, and the largest gains

are $+10.1$ (MUUFL) and $+23.2$ (Trento). The budget also matters on MUUFL: $\rho{=}10\% \rightarrow 5\%$ adds $+5.62$ under independent selection and $+3.24$ under competition. Averaged over the other factor, the main effects are $+3.41$ (ownership) and $+4.43$ (budget) on MUUFL and $+3.98$ and $+0.82$ on Trento. The $+7.84$-point MUUFL difference between C-EPG and the control on this protocol is thus shared between the two factors, whereas on Trento most of it is associated with ownership resolution. The effects are sub-additive on MUUFL, which the pool composition explains: at $\rho{=}5\%$ the instances competition excludes are more often correctly labeled (69.1% against 51.7% at $\rho{=}10\%$), so the purity gain is smaller ($+1.8$ against $+7.1$ points).

5.5 Class-level behavior and calibration

At matched budget ($\rho{=}10\%$, development splits; Appendix Table 12) class-level changes are heterogeneous. On MUUFL, Building shadow ($+11.1$) and Dirt and sand ($+8.1$) gain, while Sidewalk (-6.2) and Water (-4.1) lose; Water's pool shrinks from 460 to 132 instances per split. The rank association between class-wise change and the multi-claim fraction, the non-minimal rate and the reassigned fraction is -0.08, -0.41 and -0.46 on MUUFL ($n{=}11$; -0.25, -0.65, -0.85 at $\rho{=}5\%$) and $+0.43$, $+0.44$, -0.09 on Trento ($n{=}6$), so class-level ambiguity does not order class-level change in these data. A specific case is consistent with the mechanism: on Trento, the apple-tree pool falls from 35 to 6 instances per split, only 27% of them truly apple trees, and vineyard, whose ownership those instances had been contesting, gains $+10.5$ points (in the confirmation-style comparison, $+14.8$; Appendix Figure 5). ECE falls from 0.381 to 0.308 on MUUFL and from 0.062 to 0.016 on Trento, and the Brier score from 0.803 to 0.655 and from 0.127 to 0.035 (Appendix Table 13).

6 Discussion

What we learned beyond accuracy. Prototype-guided admission is not only a confidence-estimation problem: when class-conditioned candidate sets overlap it is also an allocation problem, and how much explicit allocation matters depends on the frequency and structure of the overlaps. Without overlap, competitive assignment reduces to independent selection (Proposition 1g); with overlap, independent selection resolves ownership worse than a uniform choice on MUUFL, while competition yields consistent, order-invariant, evidence-based ownership and a purer initial pool. The $2{\times}2$ design shows a positive mean effect of ownership resolution in all four cells, with the smaller budget contributing a comparable amount on MUUFL. That a cleaner pool causes the downstream difference is a hypothesis, since purity was not varied independently of the ownership rule. The diagnostics can be applied to other class-conditioned pipelines to check whether ownership resolution is needed.

Three kinds of generality. *Structural:* the ownership problem exists whenever class-conditioned candidate sets can overlap and comparable claimant scores exist. *Algorithmic:* competitive assignment uses only the candidate sets and scores it is given. *Empirical:* we evaluate it in one prototype-guided HSI–LiDAR setting, on two benchmarks with 6 and 11 classes and five splits each, against one control configuration; thresholds are inherited from that configuration and sensitivity was studied with single runs (Appendix D). The diagnostic suite (Ω, collision rate, non-minimal rate, order sensitivity, purity) needs only candidate sets and scores, so it can be computed for other class-conditioned pipelines before deciding whether explicit ownership resolution is consequential.

7 Conclusion

We identified ownership ambiguity as a decision distinct from selection in prototype-guided pseudo-label admission, formalized overlapping claims through claimant sets, and introduced reliability-controlled competitive assignment, which resolves ownership by a per-instance argmin before budgeted admission. Its ownership is unique, disjoint and order-invariant, and it reduces to independent selection without overlap. On MUUFL and Trento, resolving it purifies the initial pool and, in matched-seed paired runs, improves mean accuracy over independent per-class selection by 7.95 points (MUUFL) and 4.79 points (Trento), about 3–4 points of it associated with ownership resolution. Assignment consistency complements confidence-based selection.

REPRODUCIBILITY STATEMENT

The supplementary material contains the code, including the competitive-assignment implementation (derived from the public PGPL reference implementation and released under its Apache-2.0 license), the protocol manifests and seed files for every split, the configuration files, the run scripts for every experiment, the scripts that compute the diagnostics, tables and figures from result files, and a README listing software versions and the exact commands. Datasets are public and are prepared by the included scripts. Training is not bit-wise deterministic; results are reported over repeated matched splits, and run-to-run variation is described in Appendix A.

USE OF LARGE LANGUAGE MODELS

Large language models were used to help draft and edit the manuscript text and to write data-analysis and plotting code. The authors specified the experiments, verified every reported number against the experiment outputs, and take full responsibility for the content.

REFERENCES

Yuki M. Asano, Christian Rupprecht, and Andrea Vedaldi. Self-labelling via simultaneous clustering and representation learning. In *International Conference on Learning Representations*, 2020.

Hao Chen, Ran Tao, Yue Fan, Yidong Wang, Jindong Wang, Bernt Schiele, Xing Xie, Bhiksha Raj, and Marios Savvides. SoftMatch: Addressing the quantity-quality trade-off in semi-supervised learning. In *International Conference on Learning Representations*, 2023.

K. Ding, T. Lu, W. Fu, S. Li, and F. Ma. Global–local transformer network for HSI and LiDAR data joint classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 2022. Art. no. 5541213.

K. Ding, T. Lu, and S. Li. Uncertainty-aware contrastive learning for semi-supervised classification of multimodal remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 2024. Art. no. 5519413.

S. Fang, K. Li, and Z. Li. S^2ENet: Spatial–spectral cross-modal enhancement network for classification of hyperspectral and LiDAR data. *IEEE Geoscience and Remote Sensing Letters*, 19, 2021. Art. no. 6504205.

F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. Shapiro, P. Barloon, and A. F. H. Goetz. The spectral image processing system (SIPS) — interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2–3):145–163, 1993.

Dong-Hyun Lee. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *ICML Workshop on Challenges in Representation Learning*, 2013.

Muyang Li, Runze Wu, Haoyu Liu, Jun Yu, Xun Yang, Bo Han, and Tongliang Liu. InstanT: Semi-supervised learning with instance-dependent thresholds. In *Advances in Neural Information Processing Systems*, 2023.

Y. Li, D. Li, W. Xie, J. Ma, S. He, and L. Fang. Semi-Mamba: Mamba-driven semi-supervised multimodal remote sensing feature classification. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(10):9837–9849, 2025.

J. Lin, F. Gao, X. Shi, J. Dong, and Q. Du. SS-MAE: Spatial–spectral masked autoencoder for multisource remote sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 2023. Art. no. 5531614.

T. Lu, K. Ding, W. Fu, S. Li, and A. Guo. Coupled adversarial learning for fusion classification of hyperspectral and LiDAR data. *Information Fusion*, 93:118–131, 2023.

S. Mohla, S. Pande, B. Banerjee, and S. Chaudhuri. FusAtNet: Dual attention based SpectroSpatial multimodal fusion network for hyperspectral and LiDAR classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 92–93, 2020.

J. Qu, L. Zhang, W. Dong, N. Li, and Y. Li. Shared-private decoupling-based multilevel feature alignment semisupervised learning for HSI and LiDAR classification. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 2024. Art. no. 5537314.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. In *Advances in Neural Information Processing Systems*, 2017.

Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A. Raffel, Ekin D. Cubuk, Alexey Kurakin, and Chun-Liang Li. FixMatch: Simplifying semi-supervised learning with consistency and confidence. In *Advances in Neural Information Processing Systems*, 2020.

Kai Sheng Tai, Peter Bailis, and Gregory Valiant. Sinkhorn label allocation: Semi-supervised classification via annealed self-training. In *International Conference on Machine Learning*, 2021.

Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Advances in Neural Information Processing Systems*, 2017.

X. Wang, L. Wang, Y. Feng, and W. Cheng. Pseudo-label generation guided semi-supervised network for hyperspectral image and LiDAR data classification. *Expert Systems with Applications*, 286, 2025. Art. no. 128110.

Yidong Wang, Hao Chen, Qiang Heng, Wenxin Hou, Yue Fan, Zhen Wu, Jindong Wang, Marios Savvides, Takahiro Shinozaki, Bhiksha Raj, Bernt Schiele, and Xing Xie. FreeMatch: Self-adaptive thresholding for semi-supervised learning. In *International Conference on Learning Representations*, 2023.

C. Wei, K. Sohn, and C. Mellina. CReST: A class-rebalancing self-training framework for imbalanced semi-supervised learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

Qizhe Xie, Minh-Thang Luong, Eduard Hovy, and Quoc V. Le. Self-training with noisy student improves ImageNet classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

J. Yao, B. Zhang, C. Li, D. Hong, and J. Chanussot. Extended vision transformer (ExViT) for land use and land cover classification: A multimodal deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 2023. Art. no. 5514415.

Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. FlexMatch: Boosting semi-supervised learning with curriculum pseudo labeling. In *Advances in Neural Information Processing Systems*, 2021a.

Pan Zhang, Bo Zhang, Ting Zhang, Dong Chen, Yong Wang, and Fang Wen. Prototypical pseudo label denoising and target structure learning for domain adaptive semantic segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021b.

Yiyan Zhang, Hongmin Gao, Weiping Ding, Pedram Ghamisi, Chenkai Zhang, Zhonghao Chen, and Bing Zhang. Seed-to-semantics: Few-shot prototype-guided progressive learning for hyperspectral and LiDAR classification. *IEEE Transactions on Image Processing*, 2026. doi: 10.1109/TIP.2026.3709522.

A Implementation and reproducibility details

Split manifests (`protocol.json`) list class-wise counts and SHA-256 hashes of the support and test masks. The expansion step draws 300 instances from the preparation pool and admits at most Q=100 per predicted class; the refinement step updates 300 pool members; both run every 15 of the 200 expansion epochs (13 rounds). Classes are traversed in ascending index order when labels are written under independent selection, and ties in s_{ik} go to the smaller class index. A quota-preserving variant of competitive assignment is implemented in the code but was not evaluated. Software: Python 3.13, PyTorch 2.7.0 (CUDA 12.8). Training is not bit-wise deterministic: the same configuration on

the same Trento split (independent selection, ρ=10%, split 42, model seed 2026) reached 91.5% OA in the protocol run and 79.9% OA in a sensitivity-sweep run, so single-run differences of several points on that split are within run-to-run variation.

B ALGORITHM

Algorithm 1 Competitive assignment in progressive self-training

Require: seeds $\{S_k\}$, thresholds $(\tau_{\text{sam}}, \tau_h)$, weight λ, budget ρ, epochs E

candidates: compute c_{ik}, s_{ik}; $\mathcal{C}_k \leftarrow \{i : c_{ik} = 1\}$; **claims:** $\mathcal{A}_i \leftarrow \{k : i \in \mathcal{C}_k\}$

ownership: $k_i^\star \leftarrow \arg\min_{k \in \mathcal{A}_i} s_{ik}$; $\mathcal{C}'_k \leftarrow \{i : k_i^\star = k\}$ (Eq. 3)

admission: $\mathcal{B}'_k \leftarrow$ lowest-score $\lceil \rho |\mathcal{C}'_k| \rceil$ of $\mathcal{C}'_k$; $\mathcal{P}_0 \leftarrow S \cup \bigcup_k \mathcal{B}'_k$

train f_θ on $\mathcal{P}_0$; $V \leftarrow U \setminus \mathcal{P}_0$

for $t = 1, \ldots, E$ **do**

    every 15 epochs: confidence-guided expansion from V; momentum-smoothed refinement of pool labels

    one epoch on $\mathcal{P}$ with class-balanced sampling

end for

freeze $\mathcal{P}$; fine-tune f_θ **return** f_θ

C STATISTICAL DETAILS

Table 4 lists the per-split values behind Table 2. At the run level, the 15 MUUFL runs give a mean OA difference of $+7.95$ with 95% interval $[+3.15, +12.75]$ and 12 of 15 runs won (standard deviations of OA across runs: 11.17 independent, 14.41 competitive); the three runs within a split share data, so this interval combines training and split variability and is reported alongside the split-level interval $[-2.82, +18.72]$. The standard deviation of the paired MUUFL difference across model seeds within a split averages 3.45 pp, and the standard deviation of the five split means is 8.68 pp. On Trento, one run per split, the standard deviations of OA are 7.30 and 0.90. With five pairs the smallest attainable two-sided exact Wilcoxon p-value is 0.0625; the values for OA are 0.125 (MUUFL) and 0.3125 (Trento).

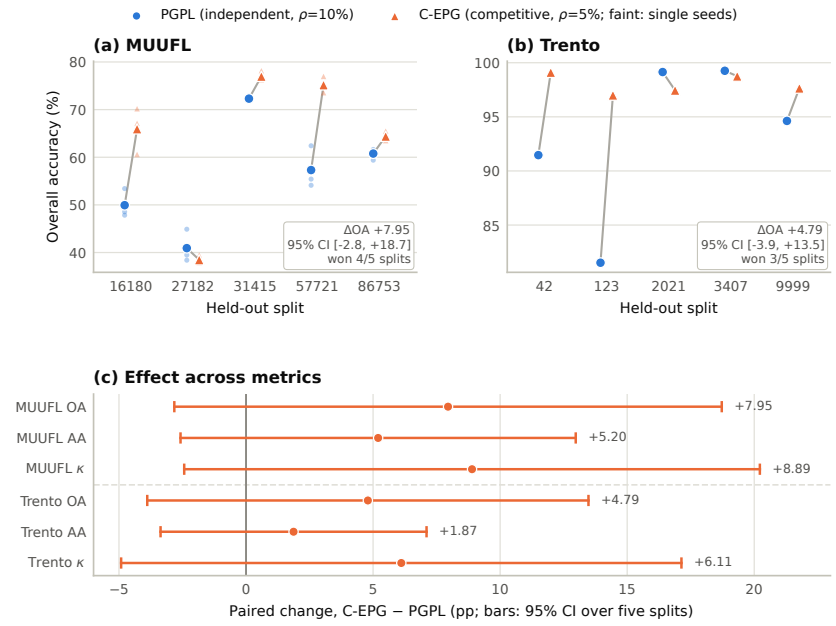


Figure 4: Paired comparison by held-out split. (a, b) OA per split; large markers are means over model seeds, faint markers individual seeds (MUUFL). (c) Mean paired change with 95% interval over five splits.

Table 4: Per-split OA (%) of independent per-class selection (PGPL) and C-EPG; MUUFL values are means over three model seeds.

Dataset	Split	PGPL	C-EPG	Δ
MUUFL	31415	72.30	76.98	+4.68
MUUFL	27182	40.92	38.47	−2.45
MUUFL	16180	49.94	65.95	+16.02
MUUFL	57721	57.31	75.19	+17.88
MUUFL	86753	60.77	64.40	+3.63
Trento	42	91.46	99.11	+7.64
Trento	123	81.52	97.01	+15.49
Trento	2021	99.14	97.45	−1.69
Trento	3407	99.26	98.76	−0.50
Trento	9999	94.62	97.65	+3.03

D DIAGNOSTIC SENSITIVITY ANALYSES (SINGLE SPLIT, SINGLE MODEL SEED)

These runs use one split per dataset (MUUFL 31415; Trento 42) and model seed 2026. They are diagnostic: single runs carry the run-to-run variation of Appendix A, and the main claims rest on the repeated matched-split experiments. The threshold sweep used independent selection at ρ=10%, the patch-size sweep competitive assignment at ρ=10%, and the budget–quota sweep competitive assignment at the stated ρ.

Table 5: Threshold diagnostic, OA (%): rows τ_{sam} (rad), columns τ_h. Left: MUUFL; right: Trento.

τ_{sam} / τ_h	1	2	3
0.18	65.6	70.9	73.6
0.24	73.2	70.9	72.2
0.3	72.0	68.2	73.8

τ_{sam} / τ_h	0.3	0.6	0.9
0.25	98.2	98.5	99.1
0.35	86.1	79.9	73.5
0.45	91.8	89.9	78.8

Table 6: Patch-size diagnostic (competitive, ρ=10%): OA / AA / $\kappa\times100$ (%).

Dataset	Patch	OA	AA	$\kappa\times100$
MUUFL	7	72.3	73.1	65.0
MUUFL	9	74.5	73.7	67.5
MUUFL	11	75.6	74.9	68.9
Trento	5	97.8	98.3	97.1
Trento	7	99.0	98.5	98.7
Trento	9	98.1	98.0	97.4

E ADDITIONAL MECHANISM AND CLASS-LEVEL ANALYSES

Ownership and score margins. Stratifying the affected set by the margin $m_i = s_{i,(2)} - s_{i,(1)}$ between the two lowest claimant scores (Appendix Table 9), the minimum-score rule is most accurate at large margins on both datasets (MUUFL 35.6%, 33.0%, 49.9% and Trento 64.7%, 93.5%, 98.2% for small, medium, large margins). On Trento, disagreement between the two rules is concentrated at small margins (30.7%, 4.5%, 2.1%), where ownership is intrinsically ambiguous; on MUUFL it occurs at every margin and the last-writer rule is least accurate at large margins (13.6% versus 49.9% for the minimum-score rule).

Feature space and predictions (qualitative). Figure 6 shows a joint t-SNE of class-balanced test-instance embeddings from the C-EPG model after warm-up, expansion and fine-tuning (MUUFL split 31415; Trento split 42). On Trento, apple trees and vineyard form one continuous structure after warm-up and separate into distinct clusters after expansion, which persist after fine-tuning. Figures 7 and 8 compare predictions (Trento split 9999; MUUFL split 31415; model seed 2026; median paired OA difference). The zoomed regions are the two windows maximizing the local excess of instances where only C-EPG is correct, so they favor C-EPG by construction; the difference maps also contain instances where only independent selection is correct. On Trento, independent selection labels parts

Table 7: Budget ρ (rows) $\times$ expansion quota Q (columns) diagnostic, OA (%), competitive. Left: MUUFL; right: Trento.

ρ / Q	25	50	100
0.05	70.2	64.5	66.5
0.1	75.2	73.4	74.7
0.2	72.1	71.9	72.4

ρ / Q	25	50	100
0.05	98.4	97.1	99.1
0.1	95.0	98.9	99.0
0.2	97.6	91.2	99.4

Table 8: Initial-pool diagnostics (means over five splits; MUUFL development splits). Independent selection $\rightarrow$ competitive assignment for pool size and purity. Multi-claimed is Ω.

Dataset	ρ	Multi-claimed / claimed pixels	Colliding slots	Non-minimal admissions	Pool size ind $\rightarrow$ comp	Purity ind $\rightarrow$ comp
MUUFL	10%	42.4%	10.8%	6.0%	3720 $\rightarrow$ 2484	65.9% $\rightarrow$ 73.0%
MUUFL	5%	42.4%	3.4%	2.9%	1935 $\rightarrow$ 1245	73.4% $\rightarrow$ 75.2%
Trento	10%	4.0%	0.4%	1.8%	969 $\rightarrow$ 933	96.1% $\rightarrow$ 98.6%
Trento	5%	4.0%	0.1%	1.4%	487 $\rightarrow$ 468	96.0% $\rightarrow$ 98.7%

of vineyard parcels as apple trees, which C-EPG largely corrects; on MUUFL it predicts water over parts of building regions, which C-EPG largely avoids.

Table 9: Affected set stratified by score margin $m_i = s_{i,(2)} - s_{i,(1)}$ (terciles within dataset, five splits pooled): number of instances, disagreement between the last-writer rule and the minimum-score owner, accuracy of the owner and of the last-writer rule.

Dataset	Margin	Instances	Rules disagree	Owner correct	Last-writer correct
MUUFL	small	17,506	50.3%	35.6%	30.0%
MUUFL	medium	17,506	36.6%	33.0%	17.4%
MUUFL	large	17,507	61.1%	49.9%	13.6%
Trento	small	626	30.7%	64.7%	84.8%
Trento	medium	626	4.5%	93.5%	93.5%
Trento	large	627	2.1%	98.2%	96.5%

Table 10: Accuracy of rules that pick one class among the claimants of a disputed instance (instances pooled over the five splits; same splits as Table 8). "Last" is independent selection, "first" its reverse, "random" a uniform choice, "min-score" competitive assignment; "true class": disputed instances whose true class is among the claimants.

Dataset	Disputed	Last	First	Random	Min-score	True class
MUUFL	52,519	20.3%	48.1%	31.8%	39.5%	74.1%
Trento	1,879	91.6%	8.3%	49.9%	85.5%	99.9%

Table 11: Purity (%) of the independent initial pool under different class-processing orders and of the competitive pool (means over five splits; random: 20 permutations).

Dataset	ρ	Ascending	Descending	Random (mean $\pm$ SD)	Competitive
MUUFL	10%	65.9	64.1	65.3 ± 1.2	73.0
MUUFL	5%	73.4	72.5	73.0 ± 0.5	75.2
Trento	10%	96.1	96.0	96.1 ± 0.1	98.6
Trento	5%	96.0	96.0	96.0 ± 0.0	98.7

Table 12: Class-level diagnostics under independent selection at ρ=10% (means over splits; MUUFL confirmation splits, Trento splits) and class-wise accuracy change of C-EPG. Multi-claimed: fraction of the class's candidate set claimed by $\geq$ 2 classes; colliding: fraction of its admitted slots also admitted by another class; non-minimal: fraction of its admitted instances whose lowest-score claimant is another class.

Dataset	Class	ΔAcc	Pool ind.→comp.	Purity ind.→comp.	Multi-claimed
MUUFL	Trees	+4.1	637→493	0.76→0.76	46%
MUUFL	Mostly grass	+2.6	429→411	0.75→0.76	19%
MUUFL	Mixed ground surface	+5.2	186→145	0.92→0.93	33%
MUUFL	Dirt and sand	+8.1	273→169	0.78→0.83	82%
MUUFL	Road	+5.4	365→317	0.74→0.73	24%
MUUFL	Water	−4.1	460→132	0.66→1.00	97%
MUUFL	Building shadow	+11.1	708→507	0.55→0.57	86%
MUUFL	Building	+6.2	237→125	0.72→0.82	47%
MUUFL	Sidewalk	−6.2	243→121	0.37→0.40	89%
MUUFL	Yellow curb	+3.8	108→33	0.37→0.56	68%
MUUFL	Cloth panels	−1.9	73→32	0.78→0.96	31%
Trento	Apple trees	+0.3	35→6	0.27→0.29	48%
Trento	Buildings	+0.2	192→190	0.97→0.97	2%
Trento	Ground	−0.4	6→6	1.00→1.00	5%
Trento	Wood	+0.0	88→88	1.00→1.00	0%
Trento	Vineyard	+10.5	442→438	1.00→1.00	7%
Trento	Roads	−0.3	206→204	1.00→1.00	2%

Table 13: Calibration (lower is better), split-level means; last column: splits with lower ECE under C-EPG.

Dataset	ECE PGPL	ECE C-EPG	Brier PGPL	Brier C-EPG	Splits with lower ECE
MUUFL	0.381	0.308	0.803	0.655	4/5
Trento	0.062	0.016	0.127	0.035	3/5

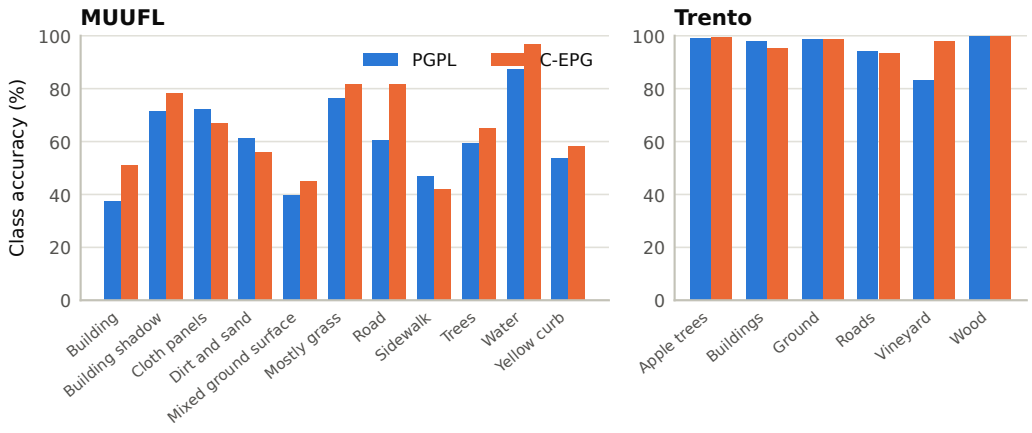


Figure 5: Class-wise test accuracy of independent per-class selection (PGPL) and C-EPG, averaged over splits (MUUFL: 15 runs; Trento: 5 runs).

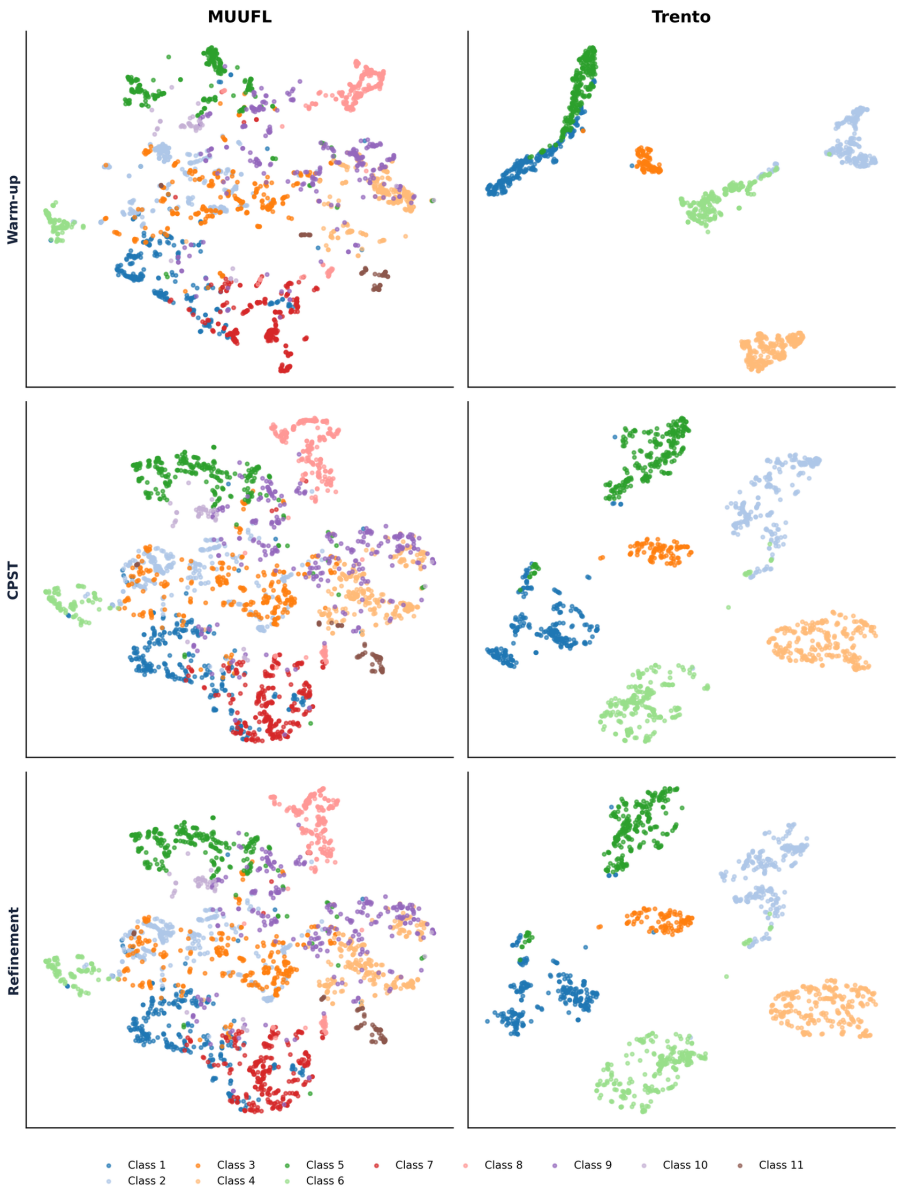


Figure 6: Joint t-SNE of test-instance embeddings from the C-EPG model after warm-up, expansion and fine-tuning (MUUFL split 31415; Trento split 42).

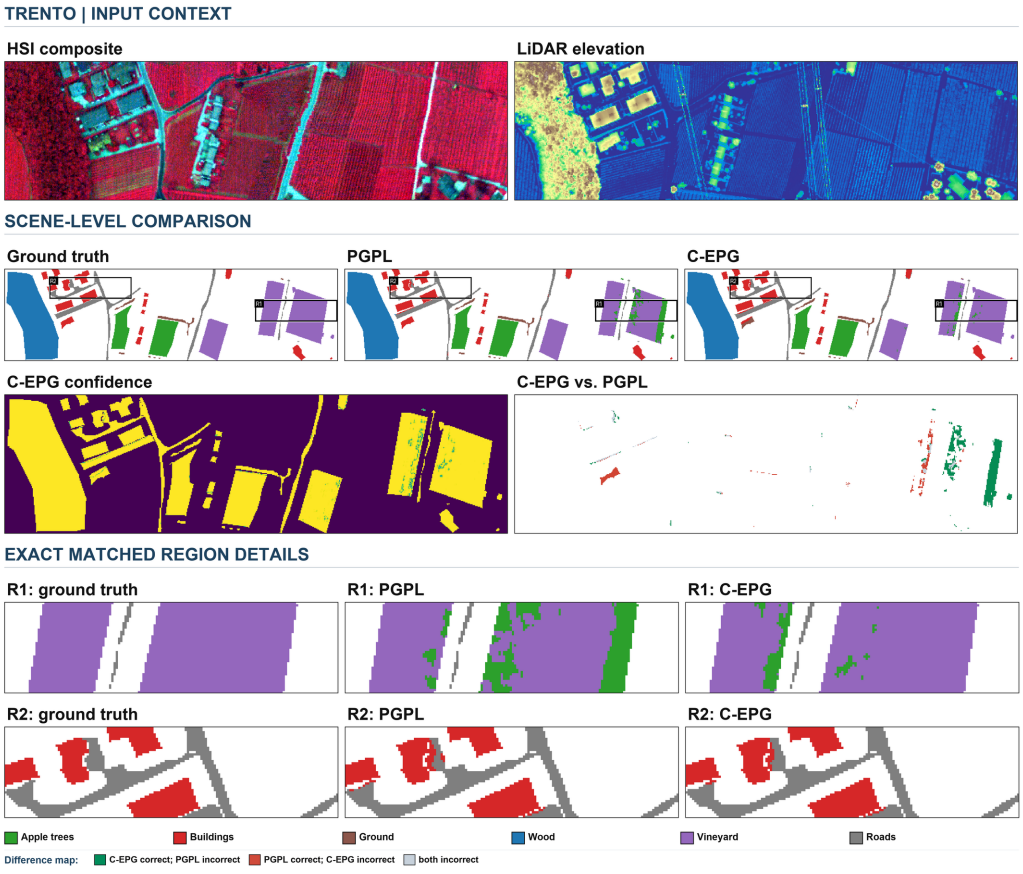


Figure 7: Trento, split 9999. Ground truth, independent-selection and C-EPG predictions, C-EPG confidence, and instances where exactly one method is correct; regions R1 and R2 are illustrative, not representative.

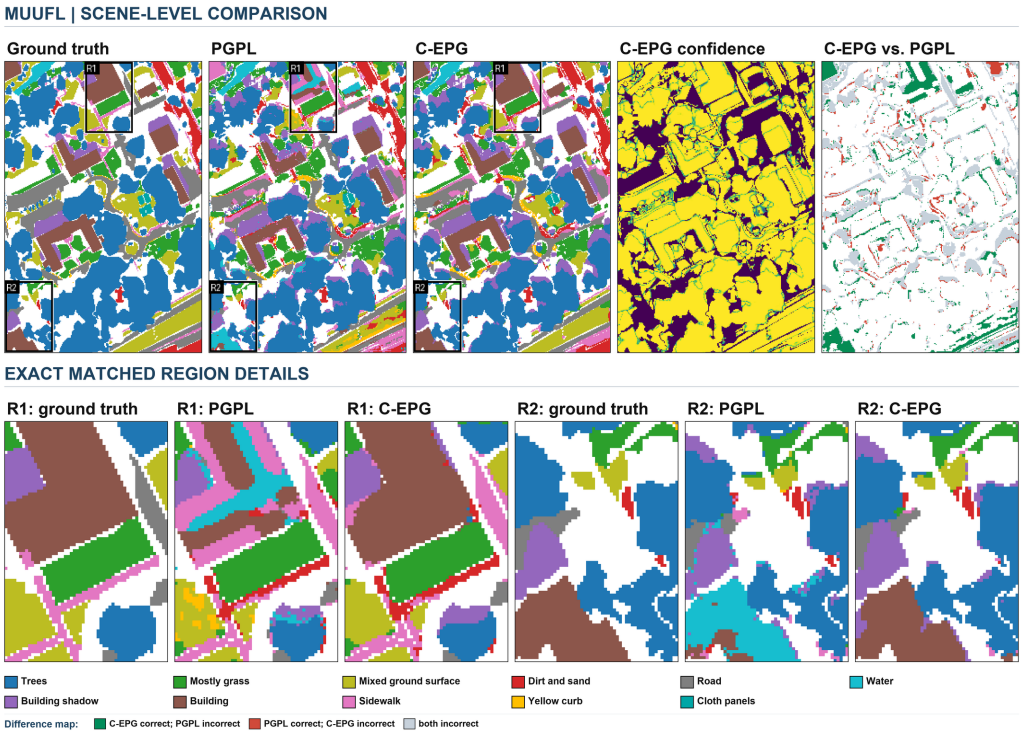


Figure 8: MUUFL, split 31415, model seed 2026. Same layout and region-selection rule as Figure 7.