3D GENERALIZED CATEGORY DISCOVERY

Anonymous authors
Paper under double-blind review

ABSTRACT

Generalized category discovery (GCD) aims to recognize known categories and discover novel ones in unlabeled data, using supervision from known classes. 3D generalized category discovery (3DGCD) extends this setting to point clouds. Our analysis of pretrained 3D features reveals two problems. Feature neighborhoods contain objects from different categories and change across augmented views. Even when neighborhoods are relatively pure, known objects can dominate clustering and leave novel categories without distinct groups. We propose Geo3D to address both problems by coupling geometric relations between objects with category-level guidance. Geometric compatibility and agreement across views help identify reliable neighbors. Geo3D combines their relational evidence with a memory of category structure to construct discovery targets. Objects confidently assigned to known categories are anchored to their predicted classes, reducing their competition for novel groups during training. The resulting supervision supports coherent grouping while allowing novel categories to form distinct clusters. Extensive experiments on synthetic shapes and real-world scans demonstrate the effectiveness of the proposed method.

1 INTRODUCTION

A vase can share the rounded body of a bowl and the narrow neck of a bottle, yet belong to neither category (Figure 1). A model should recognize the familiar categories while discovering that such unfamiliar objects form a category of their own. Generalized category discovery (GCD) studies this problem using labels from known categories and an unlabeled collection containing both known and novel categories (Vaze et al., 2022; Pu et al., 2023; Chiaroni et al., 2023). Its 3D counterpart, 3D generalized category discovery (3DGCD), operates on point clouds, where shared geometry offers cues for interpreting unfamiliar shapes. Category discovery has also been studied in point-cloud segmentation (Riz et al., 2023; Xu et al., 2024; Kim et al., 2026). DNIK (Weng et al., 2023) transfers knowledge through shared part concepts and their spatial relations. We study a complementary challenge at the object level: deciding which relations between shapes are reliable and how they should support the formation of novel groups. Similarity to familiar objects can aid discovery, but it can also draw novel objects into known categories. Successful 3DGCD therefore requires a grouping process that uses shared geometry while preserving distinctions between categories.

Local geometric learning (Qi et al., 2017b; Wang et al., 2019; Zhao et al., 2021) and contrastive pre-training (Xie et al., 2020; Zhang et al., 2021; Afham et al., 2022) provide a strong basis for learning transferable shape features. Masked shape modeling (Yu et al., 2022; Pang et al., 2022; Zhang et al., 2022a; 2023a) and multimodal pre-training (Xue et al., 2023; Huang et al., 2023; Xue et al., 2024; Zhou et al., 2024) offer complementary sources of supervision. Yet our analysis of frozen pretrained features reveals two obstacles to discovery (Table 2). Some feature neighborhoods contain many objects from other categories and change substantially across augmented views. Neighborhood-based discovery methods use such relations as training supervision (Zhong et al., 2021a; Zhang et al., 2023b; Banerjee et al., 2024; Choi et al., 2024). An unreliable neighbor can therefore repeatedly pull an object toward the wrong group. A second obstacle persists even when neighborhoods are relatively pure. Diagnostic clustering still produces groups dominated by known objects and leaves novel categories poorly covered. This failure cannot be diagnosed from neighborhood purity alone. Prediction bias is also a central concern in parametric GCD (Wen et al., 2023; Guo et al., 2025). Our observations show why category assignment must be considered alongside local feature structure. The two problems require different decisions. Neighborhood reliability determines which objects

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.

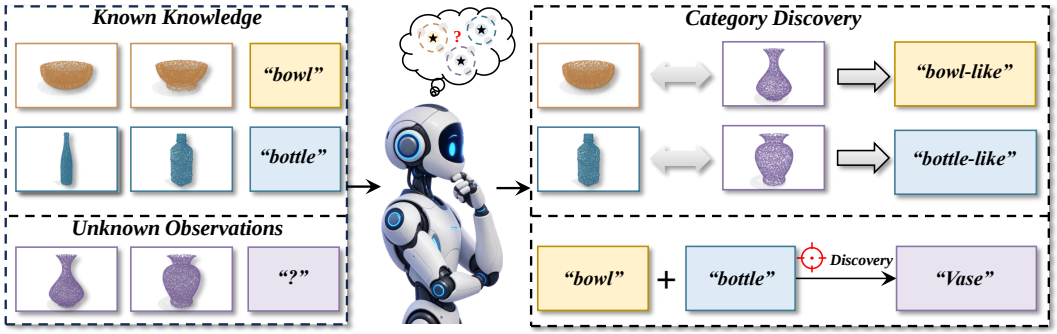


Figure 1: 3D generalized category discovery. Known shapes offer cues for interpreting novel objects: a vase can resemble both a bowl and a bottle. The goal is to discover the new category while recognizing known categories in the same unlabeled pool.

should inform one another, while category assignment determines whether this evidence supports distinct novel groups. Improving the first alone does not resolve the second.

We propose Geo3D to address both problems by coupling geometric neighborhood refinement with Memory-Guided Category Assignment. Candidate neighbors are evaluated using pairwise shape compatibility and agreement across augmented views. Neighborhood agreement also controls aggregation strength, so unstable neighborhoods exert less influence on feature updates. The refined relations guide discovery targets together with stored object features and category prototypes. To limit known-class competition for novel groups, Geo3D anchors objects confidently assigned to known categories to known prototypes. Agreement between the learned model and a frozen reference helps identify these objects, with assignment thresholds calibrated on labeled examples. The remaining unlabeled objects receive targets that combine neighborhood evidence with the evolving category structure. Consistency across views and training epochs controls the influence of these targets, while prototype matching and momentum updates maintain continuity as features change. The same refined graph supports local aggregation and discovery supervision, linking neighbor selection to category assignment.

Experiments on multiple public 3D benchmark datasets evaluate the effectiveness of Geo3D and investigate the contributions of its key design choices. The ablations follow the patterns identified by the feature diagnostics. Removing geometric refinement causes the largest loss in novel-category accuracy on the benchmark with the least reliable neighborhoods. Removing memory-guided assignment produces its largest drop in novel-category accuracy where known-class dominance is strongest. Replacing graph evidence with constant or shuffled alternatives also reduces discovery accuracy, showing that the gains depend on the relations encoded by the graph. Representation diagnostics and 3D point-cloud visualizations further examine the resulting category structure. Extensive experiments demonstrate the strong performance of Geo3D on standard 3D category discovery benchmarks.

In summary, our contributions are as follows:

- We identify unreliable neighborhoods and known-class dominance in clustering as two obstacles to 3DGCD. Our diagnostics show that locally coherent features can coexist with poor coverage of novel categories.
- We propose Geo3D, which uses a shared geometric graph for neighborhood refinement and memory-guided discovery. Known-class anchors regulate category assignments, reducing competition for novel groups.
- We demonstrate consistent gains in overall and novel-category accuracy. Component ablations, graph interventions, representation diagnostics, and 3D visualizations examine how the proposed design improves discovery.

2 Related Work

2.1 3D Representation Learning

3D representation learning aims to derive discriminative, geometry-aware, and transferable features from point clouds, voxels, and multi-view observations. Point clouds preserve geometric structure directly and support efficient processing. Early point-based architectures enabled direct learning on unordered point sets. PointNet demonstrated that permutation-invariant set functions support point cloud recognition, while PointNet++ introduced hierarchical neighborhood aggregation for multi-scale local geometry Qi et al. (2017a;b). PointCNN, DGCNN, A-CNN, and KPConv strengthened local modeling and relational reasoning through learnable point convolutions, dynamic graphs, and geometric neighborhood operators Li et al. (2018); Wang et al. (2019); Komarichev et al. (2019); Thomas et al. (2019). Walk in the Cloud, Point Transformer, PointMLP, and PointNeXt further explored curve-based aggregation, self-attention, and streamlined backbone design Xiang et al. (2021); Zhao et al. (2021); Ma et al. (2022); Qian et al. (2022).

Recent research has shifted toward self-supervised and multimodal pre-training. PointContrast and DepthContrast learn transferable 3D features through contrastive pre-training without dense manual annotation, while OcCo uses occlusion completion as a geometry-aware pretext task Xie et al. (2020); Zhang et al. (2021); Wang et al. (2021). Inspired by masked modeling in language and vision, Point-BERT, Point-MAE, Point-M2AE, and I2P-MAE improve feature quality and transferability through masked reconstruction and cross-modal transfer Yu et al. (2022); Pang et al. (2022); Zhang et al. (2022a; 2023a). Multimodal methods connect point clouds with vision–language priors. PointCLIP transfers these priors to point clouds, ULIP aligns language, image, and point cloud features in a unified space, and Uni3D extends this direction toward scalable foundation-level 3D representation learning Zhang et al. (2022b); Xue et al. (2023); Zhou et al. (2024). These approaches provide transferable features for downstream 3D recognition and discovery.

2.2 Generalized Category Discovery

Generalized Category Discovery (GCD) learns from labeled known classes and an unlabeled mixture of known and novel classes. Unlike conventional semi-supervised learning, the labeled and unlabeled sets need not share the same label space; unlike Novel Category Discovery (NCD), the unlabeled set is not restricted to novel classes. Early NCD methods include ranking statistics, Neighborhood Contrastive Learning, joint representation learning for single- and multi-modal data, UNO, and OpenMix. These methods improve transferable representations, pseudo-labeling, and supervised–unsupervised coupling Han et al. (2020); Zhong et al. (2021a); Jia et al. (2021); Fini et al. (2021); Zhong et al. (2021b). ORCA formulated open-world semi-supervised learning with unknown classes in the unlabeled pool Cao et al. (2021). Vaze *et al.* introduced GCD and established a contrastive-learning-plus-clustering framework Vaze et al. (2022).

Subsequent GCD methods mainly address representation quality, clustering robustness, and classifier design. Divide-and-Conquer uses compositional experts to characterize known and novel subsets, while DCCL exploits dynamic conceptual relations for structure-aware representation learning Yang et al. (2022); Pu et al. (2023). SimGCD attributes weak parametric classification primarily to prediction bias and unreliable pseudo-labels rather than classifier form. PIM formulates GCD through parametric information maximization and achieves strong performance across diverse settings Wen et al. (2023); Chiaroni et al. (2023). OpenGCD, AGCD, TRAILER, AllGCD, and HypCD extend category discovery toward broader open-world settings, including open-world recognition, active labeling, targeted representation alignment, and fuller use of unlabeled data Gao et al. (2023); Ma et al. (2024); Xiao et al. (2024); Cao et al. (2025); Liu et al. (2025). CryoGCD studies GCD in 3D cryo-ET density volumes using auxiliary semantic prototypes Nair et al. (2025). These works highlight the need to preserve known-class discriminability while discovering novel-category structure from partially supervised data.

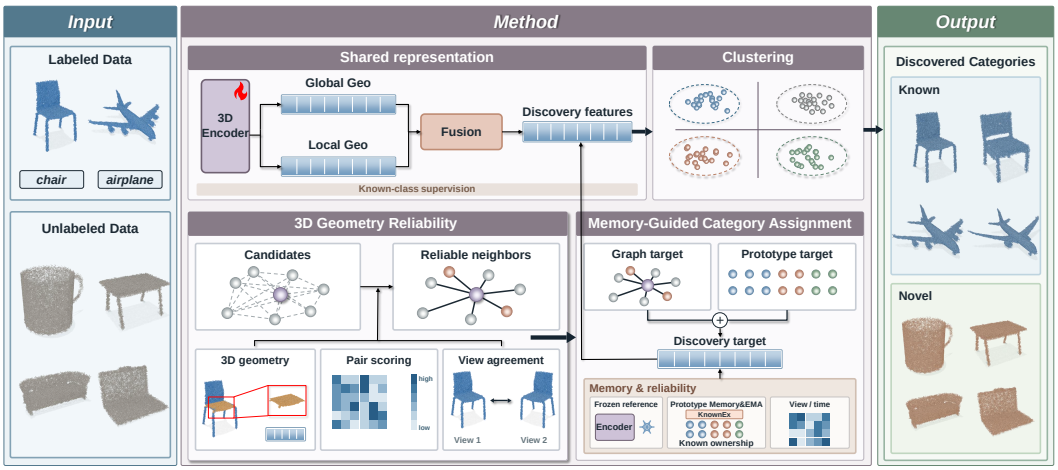


Figure 2: Overview of Geo3D. 3DGR selects reliable neighbors, and MGCA constructs memory-guided discovery targets to guide feature learning.

3 METHOD

3.1 PROBLEM DEFINITION

We study generalized category discovery from 3D point clouds. Training uses a labeled set and a mixed unlabeled set:

$$\mathcal{D}_l = \{(X_i, y_i)\}_{i=1}^{n_l}, \qquad \mathcal{D}_u = \{X_i\}_{i=n_l+1}^{n_l+n_u}, \tag{1}$$

where $X_i \in \mathbb{R}^{N\times 3}$ contains N sampled points, and n_l and n_u denote the two set sizes. Labels are available only for known categories $\mathcal{Y}_k$. The unlabeled set contains both known and novel categories, with $\mathcal{Y} = \mathcal{Y}_k \cup \mathcal{Y}_n$ and $\mathcal{Y}_k \cap \mathcal{Y}_n = \varnothing$. The total category count $K = |\mathcal{Y}|$ is given. The goal is to recognize known categories and discover novel groups within this joint category space.

As illustrated in Figure 2, Geo3D couples reliable neighborhood construction with category-level guidance. The 3D Geometry Reliability (3DGR) module evaluates relations between objects using geometry and cross-view agreement. Memory-Guided Category Assignment (MGCA) combines the selected relations with category memory to construct discovery targets. These targets guide feature learning during training.

3.2 SHARED GLOBAL AND LOCAL REPRESENTATION

A shared encoder processes two augmented views of each point cloud. With the default Uni3D backbone, the classification token supplies the global feature, while averaged point-group tokens supply the local feature. A shared projection head maps both into the discovery space. We suppress the view index and write the fused representation as

$$z_i \propto (1-\beta_i) z_i^g + \beta_i z_i^l, \tag{2}$$

where $\beta_i \in [0,1]$ controls the local contribution. Projected features and prototypes are unit-normalized; $v \propto u$ denotes $v = u/\|u\|_2$. We fix $\beta_i = 0.3$ on ModelNet40 and ShapeNetCore. On ScanObjectNN, a geometry-aware reliability network predicts β_i and a routing coefficient from the global and local features and point-cloud statistics.

3.3 RELIABLE GEOMETRIC NEIGHBORHOOD CONSTRUCTION

3DGR retrieves candidate neighbors from a memory of augmented-view features refreshed each epoch. All stored views of the query object are excluded. Let $\mathcal{C}_i$ denote the candidate indices, $\bar{z}_j$ the detached feature of memory entry j, and $o(j)$ its object identity. A multilayer perceptron predicts a pair score ℓ_{ij} from feature similarity and geometric descriptor comparisons. The descriptors

summarize covariance eigenvalues, radial statistics, and local geometry at two scales. We refine the candidate scores as

$$e_{ij} = \frac{z_i^\top \bar{z}_j}{\tau} + \delta_{ij} + \log s_{ij},$$
$$\delta_{ij} = \gamma \left(\tanh \ell_{ij} - \frac{1}{|\mathcal{C}_i|} \sum_{v \in \mathcal{C}_i} \tanh \ell_{iv} \right), \qquad j \in \mathcal{C}_i. \tag{3}$$

Here, τ is the similarity temperature and γ is the geometric correction weight, including its activation ramp. The centered, bounded residual adjusts the ranking within each candidate set. The factor s_{ij} equals one when $o(j)$ appears among the other query view's top-k distinct objects ranked by raw feature similarity, and a positive floor otherwise. The top-k entries under e_{ij} define $\mathcal{N}_i$. The pair scorer receives binary supervision from pairs with labeled endpoints and gradients through the selected edge weights.

For each query view, we also select the top-k entries in $\mathcal{C}_i$ using $z_i^\top \bar{z}_j/\tau + \delta_{ij}$, before adding $\log s_{ij}$. The cross-view shift confidence is the fraction of distinct objects in this neighborhood that also appear among the other query view's top-k distinct objects ranked by raw feature similarity. Each object contributes once, regardless of how many stored views it has. This confidence controls the subsequent feature-shift strength.

3.4 Memory-Guided Category Assignment

MGCA maintains object features, category prototypes, and assignment histories. At each epoch, labeled class means provide known prototypes. Known exclusion, denoted KnownEx, retains unlabeled candidates with lower maximum similarity to known prototypes. Weighted clustering of these candidates provides novel prototypes. Novel slots are matched across epochs before prototype exponential moving averaging, denoted Prototype EMA. At this stage, we cache softmax probabilities over all K slots for each augmented-view memory entry. This cache is used to construct the auxiliary neighborhood-consensus target and remains fixed within the epoch. MGCA then aligns the updated prototype bank with its own prototype memory and applies a second momentum update.

A frozen pretrained encoder supplies reference features. An unlabeled object is accepted as known when the student and reference predict the same known class and both exceed class-specific similarity thresholds calibrated on labeled objects. The student's known score must also lie within a margin of its best novel score. We denote the accepted class by $\hat{y}_i$. Novel slots that duplicate known prototypes are refreshed from unaccepted objects with low known-class similarity. The resulting bank $\{\mu_c\}_{c=1}^K$ is used for MGCA targets and current prototype-based predictions. In the auxiliary consensus loss, current-view predictions use this final bank, while the target aggregates probabilities cached before the MGCA update. MGCA maintains its object-level assignment memory separately from this augmented-view probability cache.

For each training view, MGCA computes current assignments q_{ic} from $z_i^\top \mu_c/\tau_p$, where τ_p is the MGCA assignment temperature. Accepted known objects use a softmax over known slots, with zero mass on novel slots. Other unlabeled objects use all K slots. Labeled assignments are fixed to their observed classes. KnownEx therefore controls the samples used to construct novel prototypes, while known acceptance controls subsequent assignments. MGCA retains the entries selected by 3DGR and normalizes their weights as

$$w_{ij} = \frac{a_{ij} \exp(e_{ij})}{\sum_{v \in \mathcal{N}_i} a_{iv} \exp(e_{iv})}, \qquad j \in \mathcal{N}_i. \tag{4}$$

The positive factor a_{ij} combines frozen-reference similarity, agreement between stored MGCA assignments, and endpoint reliability, with a floor for uncertain edges. Let m_o be the normalized mean of the stored features for object o. The graph and prototype targets are

$$g_i \propto \sum_{j \in \mathcal{N}_i} w_{ij} m_{o(j)}, \qquad p_i \propto \sum_{c=1}^{K} \mathrm{sg}(q_{ic}) \mu_c, \tag{5}$$

where sg stops gradients. Memory features and prototypes remain fixed within each epoch; the selected graph weights remain differentiable. Object reliability $r_i \in [0, 1]$ combines student and reference view stability, their assignment agreement, temporal consistency, and assignment confidence

through a geometric mean. With $b_i \propto (g_i + p_i)/2$, the discovery target is

$$t_i \propto \begin{cases} \mathrm{sg}(z_i), & \text{labeled,} \\ p_i, & \text{unlabeled, accepted as known,} \\ r_i b_i + (1 - r_i)\,\mathrm{sg}(z_i), & \text{otherwise.} \end{cases} \tag{6}$$

Reliable objects receive stronger structural guidance. Uncertain objects retain more of their current feature direction. Accepted known objects receive known-only prototype targets, limiting their competition for novel slots.

3.5 Training Objective

Discovery targets guide feature shifting during training. Let h_i aggregate $\bar{z}_j$ using softmax weights from e_{ij} over $\mathcal{N}_i$. We blend this aggregate with t_i:

$$\begin{aligned} t_i^{\mathrm{shift}} &\propto (1 - \rho) h_i + \rho t_i, \\ \widetilde{z}_i &\propto (1 - \alpha_i) z_i + \alpha_i t_i^{\mathrm{shift}}. \end{aligned} \tag{7}$$

The coefficient ρ increases from zero to one after warm-up. The shift confidence defined in Section 3.3 controls α_i on all datasets. On ScanObjectNN, the learned routing coefficient and attention concentration further limit this strength. Target alignment uses t_i directly.

The representation objective is $\mathcal{L}_{\mathrm{rep}} = \lambda_s \mathcal{L}_{\mathrm{sup}} + \lambda_c \mathcal{L}_{\mathrm{cons}} + \mathcal{L}_{\mathrm{aux}}$. Supervised contrastive loss $\mathcal{L}_{\mathrm{sup}}$ trains shared labeled features and a residual adapter with detached inputs. Cross-view contrastive loss $\mathcal{L}_{\mathrm{cons}}$ acts on $\widetilde{z}_i$. The auxiliary term $\mathcal{L}_{\mathrm{aux}}$ collects representation constraints, including neighborhood consensus, with additional gate supervision on ScanObjectNN. Neighborhood consensus compares a detached aggregate of cached neighbor probabilities with the current-view prediction using Jensen–Shannon divergence. For this loss, the current-view prediction is a softmax over all K slots using the final MGCA bank. The pair scorer is trained by $\mathcal{L}_{\mathrm{pair}}$, a weighted binary cross-entropy on labeled pair scores ℓ_{ij}.

The memory objective is $\mathcal{L}_{\mathrm{m}} = \mathcal{L}_{\mathrm{align}} + \lambda_a \mathcal{L}_{\mathrm{anchor}} + \mathcal{L}_{\mathrm{reg}}$. For unlabeled objects not accepted as known, $\mathcal{L}_{\mathrm{align}}$ minimizes the mode-weighted cosine distance between z_i and t_i. For accepted known objects, $\mathcal{L}_{\mathrm{anchor}}$ aligns z_i with $\mu_{\hat{y}_i}$. The regularizer $\mathcal{L}_{\mathrm{reg}}$ constrains assignment consistency, novel-slot coverage, and agreement with reference relations. We combine these objectives as

$$\mathcal{L}_{\mathrm{total}} = \mathcal{L}_{\mathrm{rep}} + \lambda_p \mathcal{L}_{\mathrm{pair}} + \rho \lambda_m \mathcal{L}_{\mathrm{m}}. \tag{8}$$

The λ coefficients weight the corresponding terms. The same ρ controls target blending and the memory-loss contribution. On the shared encoder and projection head, we remove discovery-gradient components that oppose the supervised gradient. Detailed loss definitions and training settings are provided in Appendices A.3 and A.2.

4 Experiments

4.1 Experimental Setup

We evaluate ModelNet40 (Wu et al., 2015), ScanObjectNN (Uy et al., 2019), and ShapeNetCore (Chang et al., 2015), using 20/20, 8/7, and 28/27 known/novel classes. We label 50% of known-class training objects; the remaining objects form the mixed unlabeled pool. All, Old, and New report test accuracy under one global Hungarian matching. Geo3D uses fixed-K Ward clustering; frozen representations use K-Means. Training uses 1,024 points per object and 100 epochs. We fine-tune the final block of Uni3D (Zhou et al., 2024) with SGD and batch size 16. Learning rates are 10^{-5} for the backbone and 0.01 for the heads. Eight neighbors are selected from 32 candidates. After five warm-up epochs, γ and ρ are linearly ramped up over ten epochs. Seed 42 is the default. Appendix A gives the full settings.

4.2 Diagnosing the Discovery Problem

Frozen features expose two obstacles (Table 2). ScanObjectNN has the lowest neighborhood purity, 53.7%, and the smallest cross-view graph overlap, 0.333. ShapeNetCore has 75.9% purity, yet

Table 1: Comparison with 3D representation, open-world semi-supervised, and category-discovery methods. Accuracy (%) is reported as All/Old/New. Frozen representations are evaluated using K-Means. Best and second-best results are shown in bold and underlined, respectively.

Method	ModelNet40			ScanObjectNN			ShapeNetCore		
	All	Old	New	All	Old	New	All	Old	New
3D Representation Methods									
DGCNN Wang et al. (2019)	52.39	56.91	48.10	41.33	44.17	36.99	34.19	32.86	36.99
Point-BERT Yu et al. (2022)	51.74	53.33	50.24	32.51	35.38	28.13	31.83	30.06	35.56
ULIP Xue et al. (2023)	62.93	66.14	59.87	47.02	41.82	54.95	43.09	37.30	55.31
Uni3D Zhou et al. (2024)	59.76	69.05	50.95	50.24	47.85	53.90	43.76	40.16	51.35
Open-World Semi-Supervised									
OwMatch Niu et al. (2024)	66.37	91.01	42.97	61.28	72.43	44.26	68.25	80.50	42.41
TRAILER Xiao et al. (2024)	65.36	87.52	44.31	61.31	**84.15**	26.47	70.58	85.56	38.99
LPS Ye et al. (2024)	68.07	84.03	52.92	62.66	74.21	45.05	72.86	**89.74**	37.26
Category Discovery Methods									
UNO Fini et al. (2021)	62.28	87.27	38.55	60.31	65.54	52.32	72.96	87.17	42.99
GCD Vaze et al. (2022)	64.75	72.05	57.82	57.25	57.67	<u>56.62</u>	48.48	51.39	42.35
SimGCD Wen et al. (2023)	71.03	86.69	56.16	56.84	60.02	51.97	63.15	68.36	52.17
DNIK Weng et al. (2023)	63.17	<u>94.18</u>	33.73	35.36	38.31	30.85	72.56	<u>89.31</u>	37.23
CMS Choi et al. (2024)	72.00	92.35	52.69	61.73	71.51	46.80	72.33	87.43	40.50
SelEx Rastegar et al. (2024)	64.67	68.55	<u>60.98</u>	57.36	59.91	53.46	53.81	54.66	52.01
ProtoGCD Ma et al. (2025)	<u>73.22</u>	**95.42**	52.13	<u>65.20</u>	75.99	48.73	<u>77.86</u>	84.91	<u>62.98</u>
PartCo Cendra & Han (2025)	67.99	93.09	44.15	63.29	81.73	35.14	70.72	86.64	37.17
CryoGCD Nair et al. (2025)	57.11	78.40	36.89	52.74	65.53	33.22	75.65	86.81	52.11
BIA Tang et al. (2026)	64.22	93.43	36.49	61.55	<u>83.92</u>	27.43	70.70	88.68	32.81
Ours	**78.40**	90.77	**66.67**	**70.68**	75.53	**63.28**	**80.61**	85.65	**69.98**

Table 2: Dataset-level diagnostics for 3D generalized category discovery. Different datasets exhibit distinct challenges in class imbalance, representation reliability, and old–novel structure. Ground-truth labels are used only for offline diagnosis.

Dataset	Composition		Representation Reliability			Discovery Structure		
	IR	U-Novel	kNN@10	Flip	Graph-J	Known AUC	Old-modes	Novel Cov.
ModelNet40	13.9×	64.9	74.9	21.8	0.485	0.896	30.0	60.0
ScanObjectNN	5.3×	59.4	53.7	36.2	0.333	0.814	28.6	71.4
ShapeNetCore	150.3×	49.8	75.9	18.9	0.459	0.892	55.6	25.9

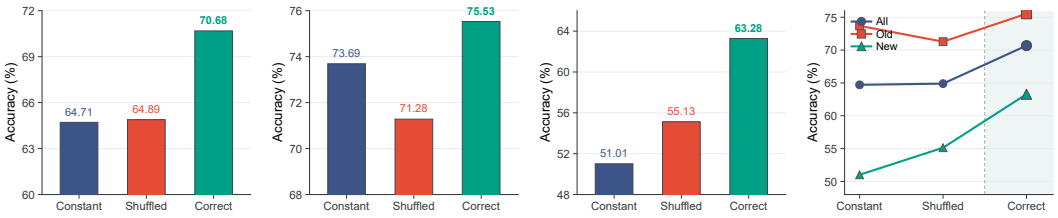
known objects dominate 55.6% of diagnostic modes and novel coverage is only 25.9%. Locally coherent features therefore do not ensure distinct novel groups. These patterns motivate reliable neighbor selection and control of category assignments. The diagnostic clusters the mixed pool into $|\mathcal{Y}_n|$ modes to expose this competition; the main evaluation uses all K categories. Appendix A.1 defines the protocol.

4.3 Testing the Proposed Corrections

The component ablation follows these diagnosed difficulties (Table 3). Removing 3DGR lowers ScanObjectNN New accuracy from 63.28% to 50.22%, its largest drop across datasets. Removing MGCA has its largest effect on ShapeNetCore, costing 7.79 points in New accuracy. Prototype

Table 3: Leave-one-out ablation of the Geo3D. Removing individual components generally degrades overall accuracy, while the changes in old- and novel-category performance reveal their complementary roles in category discovery.

Variant	ModelNet40			ScanObjectNN			ShapeNetCore		
	All	Old	New	All	Old	New	All	Old	New
w/o 3DGR	77.39	89.10	66.27	66.07	**76.45**	50.22	77.97	84.31	64.59
w/o MGCA	77.84	89.93	66.35	69.50	73.69	63.10	78.83	**86.72**	62.19
w/o KnownEx	75.32	87.60	63.67	68.32	70.53	**64.94**	78.90	84.84	66.37
w/o Proto. EMA	76.46	**92.01**	61.69	68.56	75.88	57.41	79.18	85.24	66.40
Full	**78.40**	90.77	**66.67**	**70.68**	75.53	63.28	**80.61**	85.65	**69.98**



(a) All accuracy (b) Old accuracy (c) New accuracy (d) Overall trend

Figure 3: Effect of geometric graph evidence on ScanObjectNN. We compare correct, constant, and shuffled evidence while keeping the remaining training pipeline unchanged.

EMA improves New accuracy on every dataset, and each component removal lowers All accuracy. Known exclusion improves the joint result even when it trades some novel accuracy for known-class retention, as on ScanObjectNN. The incremental configurations in Appendix C.3 examine these interactions further.

The loss ablations test how supervision supports these corrections (Table 10). Removing pair supervision or contrastive consistency lowers ScanObjectNN New accuracy by 6.14 or 20.42 points. Removing the MGCA loss costs 8.18 New points on ShapeNetCore. Target alignment and known-prototype anchoring each benefit novel discovery on all three datasets. These are subterms of the MGCA loss; removing that loss retains target construction, unlike removing the whole module. Appendix A.3 specifies the loss hierarchy.

We evaluate the geometric graph evidence using constant and shuffled controls (Figure 3). The graph encodes geometric relations between objects. Correct evidence achieves 63.28% New accuracy, compared with 51.01% for constant evidence and 55.13% for shuffled evidence, and also yields higher All and Old accuracy. These results support the contribution of object-level geometric relations to category discovery.

4.4 Representation and 3D Visualization

We compare Baseline and Full using independently extracted diagnostic features. On ScanObjectNN, novel-neighbor purity increases from 62.42% to 71.44%, while novel-to-known absorption decreases from 27.34% to 22.79% (Table 9). Novel–old cosine margins increase on all three datasets (Figure 7). On ShapeNetCore, novel-neighbor purity changes only slightly, from 77.39% to 77.86%, while New accuracy rises from 65.34% to 69.98%. Among the component removals, removing MGCA causes the largest drop in New accuracy on this dataset. These results show that neighborhood purity alone does not capture changes in the final grouping.

Figure 4 compares Baseline and Full using a shared color scale, ranging from blue at the confuser end, through a neutral midpoint, to red at the ground-truth category end. Full shows local shifts toward the ground-truth end around the vase rim and along the lamp body. Figure 5 shows three queries with their ground-truth labels and the category predictions of each method. Geo3D correctly identifies the vase, sofa, and tent, while several baselines confuse the vase with a cup or the sofa with a bed.



Figure 4: Point-cloud visualizations of Baseline and Full.

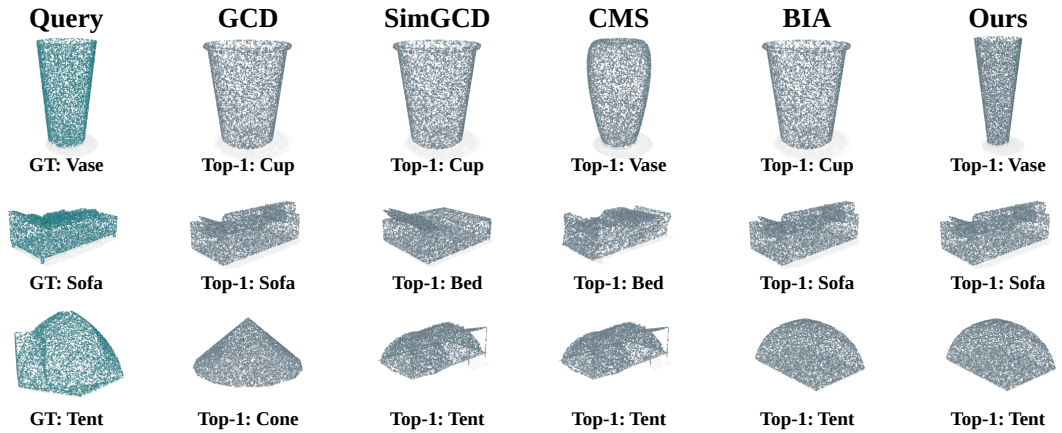


Figure 5: Qualitative category predictions for three query objects.

4.5 Discovery Accuracy and Robustness

Geo3D achieves the highest All and New accuracy among the compared methods (Table 1). New gains over the best competing scores are 5.69, 6.66, and 7.00 points; All gains are 5.18, 5.48, and 2.75 points. Across four seeds, mean New accuracy is 64.70 ± 1.16, 62.41 ± 0.87, and 68.55 ± 0.96. Every run exceeds the best competing All and New scores. Geo3D also leads both metrics under four additional category partitions. Appendix B reports these results and the cross-backbone comparison; Appendix E examines training settings and label availability.

5 Conclusion

We identified unreliable neighborhoods and known-class dominance in clustering as two obstacles to 3D generalized category discovery. To address them, we proposed Geo3D, which combines geometric neighborhood refinement with memory-guided category assignment to guide feature learning. Known-category anchoring reduces competition for novel groups. Experiments on synthetic shapes and real-world scans demonstrate improved overall and novel-category accuracy while preserving known-category recognition.

AI Use Statement

Generative AI tools were used to assist with language polishing, translation, improving the clarity and readability of the manuscript, and code implementation and debugging. They were not used to develop the core research methodology, design experiments, generate experimental results, or independently interpret the results. All AI-assisted text and code were reviewed, tested, and revised by the authors. The authors independently verified all technical content, experimental results, citations, and claims, and take full responsibility for the final content of this work.

REFERENCES

Mohamed Afham, Isuru Dissanayake, Dinithi Dissanayake, Amaya Dharmasiri, Kanchana Thilakarathna, and Ranga Rodrigo. Crosspoint: Self-supervised cross-modal contrastive learning for 3d point cloud understanding. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9892–9902. IEEE, 2022.

Anwesha Banerjee, Liyana Sahir Kallooriyakath, and Soma Biswas. Amend: Adaptive margin and expanded neighborhood for efficient generalized category discovery. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 2090–2099. IEEE, 2024.

Kaidi Cao, Maria Brbic, and Jure Leskovec. Open-world semi-supervised learning. *arXiv preprint arXiv:2102.03526*, 2021.

Xinzi Cao, Ke Chen, Feidiao Yang, Xiawu Zheng, Yonghong Tian, and Yutong Lu. AllGCD: Leveraging all unlabeled data for generalized category discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3293–3303, 2025.

Fernando Julio Cendra and Kai Han. Partco: Part-level correspondence priors enhance category discovery. *arXiv preprint arXiv:2509.22769*, 2025.

Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

Florent Chiaroni, Jose Dolz, Ziko Imtiaz Masud, Amar Mitiche, and Ismail Ben Ayed. Parametric information maximization for generalized category discovery. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1729–1739, 2023.

Sua Choi, Dahyun Kang, and Minsu Cho. Contrastive mean-shift learning for generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23094–23104, 2024.

Enrico Fini, Enver Sangineto, Stéphane Lathuilière, Zhun Zhong, Moin Nabi, and Elisa Ricci. A unified objective for novel class discovery. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9264–9272. IEEE, 2021.

Fulin Gao, Weimin Zhong, Zhixing Cao, Xin Peng, and Zhi Li. OpenGCD: Assisting open world recognition with generalized category discovery. *arXiv preprint arXiv:2308.06926*, 2023.

Pengcheng Guo, Yonghong Song, and Boyu Wang. Towards debiased generalized category discovery. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 5271–5279, 2025. doi: 10.24963/ijcai.2025/587.

Kai Han, Sylvestre-Alvise Rebuffi, Sebastien Ehrhardt, Andrea Vedaldi, and Andrew Zisserman. Automatically discovering and learning new visual categories with ranking statistics. *arXiv preprint arXiv:2002.05714*, 2020.

Tianyu Huang, Bowen Dong, Yunhan Yang, Xiaoshui Huang, Rynson WH Lau, Wanli Ouyang, and Wangmeng Zuo. Clip2point: Transfer clip to point cloud classification with image-depth pretraining. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22100–22110. IEEE, 2023.

Xuhui Jia, Kai Han, Yukun Zhu, and Bradley Green. Joint representation learning and novel category discovery on single-and multi-modal data. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 590–599. IEEE, 2021.

Minseok Kim, Jiyong Boo, and Kuk-Jin Yoon. Generalized category discovery for lidar semantic segmentation. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 8416–8426. IEEE, 2026.

Artem Komarichev, Zichun Zhong, and Jing Hua. A-cnn: Annularly convolutional neural networks on point clouds. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7413–7422. IEEE, 2019.

Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31, 2018.

Yuanpei Liu, Zhenqi He, and Kai Han. Hyperbolic category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9891–9900, 2025.

Shijie Ma, Fei Zhu, Zhun Zhong, Xu-Yao Zhang, and Cheng-Lin Liu. Active generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16890–16900, 2024.

Shijie Ma, Fei Zhu, Xu-Yao Zhang, and Cheng-Lin Liu. Protogcd: Unified and unbiased prototype learning for generalized category discovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(7):6022–6038, 2025.

Xu Ma, Can Qin, Haoxuan You, Haoxi Ran, and Yun Fu. Rethinking network design and local geometry in point cloud: A simple residual MLP framework. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

Ashwin Nair, Xingjian Li, Bhupendra Solanki, Souradeep Mukhopadhyay, Ankit Jha, Mostofa Rafid Uddin, Mainak Singha, Biplab Banerjee, and Min Xu. Towards molecular structure discovery from cryo-et density volumes via modelling auxiliary semantic prototypes. *Briefings in Bioinformatics*, 26(1):bbae570, 2025.

Shengjie Niu, Lifan Lin, Jian Huang, and Chao Wang. Owmatch: Conditional self-labeling with consistency for open-world semi-supervised learning. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Yatian Pang, Wenxiao Wang, Francis EH Tay, Wei Liu, Yonghong Tian, and Li Yuan. Masked autoencoders for point cloud self-supervised learning. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*, pp. 604–621. Springer, 2022.

Nan Pu, Zhun Zhong, and Nicu Sebe. Dynamic conceptional contrastive learning for generalized category discovery. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7579–7588. IEEE, 2023.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017a.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017b.

Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Advances in neural information processing systems*, 35:23192–23204, 2022.

Sarah Rastegar, Mohammadreza Salehi, Yuki M. Asano, Hazel Doughty, and Cees G. M. Snoek. Selex: Self-expertise in fine-grained generalized category discovery. In *European Conference on Computer Vision*, pp. 440–458, 2024.

Luigi Riz, Cristiano Saltori, Elisa Ricci, and Fabio Poiesi. Novel class discovery for 3d point cloud semantic segmentation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9393–9402. IEEE, 2023.

Luyao Tang, Kunze Huang, Chaoqi Chen, and Cheng Chen. Bures-isotropy alignment: Manifold learning of generalized category discovery. In *International Conference on Learning Representations*, 2026.

Hugues Thomas, Charles R Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J Guibas. Kpconv: Flexible and deformable convolution for point clouds. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6411–6420, 2019.

Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1588–1597. IEEE, 2019.

Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7492–7501, June 2022.

Hanchen Wang, Qi Liu, Xiangyu Yue, Joan Lasenby, and Matt J Kusner. Unsupervised point cloud pre-training via occlusion completion. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9762–9772. IEEE, 2021.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)*, 38(5): 1–12, 2019.

Xin Wen, Bingchen Zhao, and Xiaojuan Qi. Parametric classification for generalized category discovery: A baseline study. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 16544–16554. IEEE, 2023.

Tingyu Weng, Jun Xiao, and Haiyong Jiang. Decompose novel into known: Part concept learning for 3d novel class discovery. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1912–1920, 2015.

Tiange Xiang, Chaoyi Zhang, Yang Song, Jianhui Yu, and Weidong Cai. Walk in the cloud: Learning curves for point clouds shape analysis. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 895–904. IEEE, 2021.

Ruixuan Xiao, Lei Feng, Kai Tang, Junbo Zhao, Yixuan Li, Gang Chen, and Haobo Wang. Targeted representation alignment for open-world semi-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23072–23082, 2024.

Saining Xie, Jiatao Gu, Demi Guo, Charles R Qi, Leonidas Guibas, and Or Litany. Pointcontrast: Unsupervised pre-training for 3d point cloud understanding. In *European conference on computer vision*, pp. 574–591. Springer, 2020.

Ruijie Xu, Chuyu Zhang, Hui Ren, and Xuming He. Dual-level adaptive self-labeling for novel class discovery in point cloud segmentation. In *European Conference on Computer Vision*, pp. 288–305. Springer, 2024.

Le Xue, Mingfei Gao, Chen Xing, Roberto Martín-Martín, Jiajun Wu, Caiming Xiong, Ran Xu, Juan Carlos Niebles, and Silvio Savarese. Ulip: Learning a unified representation of language, images, and point clouds for 3d understanding. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1179–1189. IEEE, 2023.

Le Xue, Ning Yu, Shu Zhang, Artemis Panagopoulou, Junnan Li, Roberto Martín-Martín, Jiajun Wu, Caiming Xiong, Ran Xu, Juan Carlos Niebles, et al. Ulip-2: Towards scalable multimodal pre-training for 3d understanding. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 27081–27091. IEEE, 2024.

Muli Yang, Yuehua Zhu, Jiaping Yu, Aming Wu, and Cheng Deng. Divide and conquer: Compositional experts for generalized novel class discovery. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14248–14257. IEEE, 2022.

Bo Ye, Kai Gan, Tong Wei, and Min-Ling Zhang. Bridging the gap: Learning pace synchronization for open-world semi-supervised learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 5362–5370, 2024.

Xumin Yu, Lulu Tang, Yongming Rao, Tiejun Huang, Jie Zhou, and Jiwen Lu. Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 19291–19300. IEEE, 2022.

Renrui Zhang, Ziyu Guo, Peng Gao, Rongyao Fang, Bin Zhao, Dong Wang, Yu Qiao, and Hongsheng Li. Point-m2ae: multi-scale masked autoencoders for hierarchical point cloud pre-training. *Advances in neural information processing systems*, 35:27061–27074, 2022a.

Renrui Zhang, Ziyu Guo, Wei Zhang, Kunchang Li, Xupeng Miao, Bin Cui, Yu Qiao, Peng Gao, and Hongsheng Li. Pointclip: Point cloud understanding by clip. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8542–8552. IEEE, 2022b.

Renrui Zhang, Liuhui Wang, Yu Qiao, Peng Gao, and Hongsheng Li. Learning 3d representations from 2d pre-trained models via image-to-point masked autoencoders. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 21769–21780. IEEE, 2023a.

Sheng Zhang, Salman Khan, Zhiqiang Shen, Muzammal Naseer, Guangyi Chen, and Fahad Shahbaz Khan. Promptcal: Contrastive affinity learning via auxiliary prompts for generalized novel category discovery. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3479–3488. IEEE, 2023b.

Zaiwei Zhang, Rohit Girdhar, Armand Joulin, and Ishan Misra. Self-supervised pretraining of 3d features on any point-cloud. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10232–10243. IEEE, 2021.

Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 16259–16268, 2021.

Zhun Zhong, Enrico Fini, Subhankar Roy, Zhiming Luo, Elisa Ricci, and Nicu Sebe. Neighborhood contrastive learning for novel class discovery. *arXiv preprint arXiv:2106.10731*, 2021a.

Zhun Zhong, Linchao Zhu, Zhiming Luo, Shaozi Li, Yi Yang, and Nicu Sebe. Openmix: Reviving known knowledge for discovering novel visual categories in an open world. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9457–9465. IEEE, 2021b.

Junsheng Zhou, Jinsheng Wang, Baorui Ma, Yu-Shen Liu, Tiejun Huang, and Xinlong Wang. Uni3d: Exploring unified 3d representation at scale. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*, 2024.

A More Experimental Settings

A.1 Dataset Setup

We evaluate synthetic CAD models and real-world scans using three datasets. For each dataset, known and novel categories are disjoint. We label 50% of known-class training objects; the remaining known objects and all novel objects form the mixed unlabeled pool. The main comparison uses the same class partition and labeled-sample selection across methods, with seed 42 as the default. Every object is represented by 1,024 sampled points. Table 4 gives the unlabeled-pool composition.

ModelNet40 (Wu et al., 2015) contains 12,311 CAD models from 40 categories. We use 20 known and 20 novel categories. ScanObjectNN (Uy et al., 2019) contains 15 categories from scanned indoor scenes, with clutter, occlusion, and acquisition noise. We use the perturbed split with 8 known and 7 novel categories. ShapeNetCoreV2 (Chang et al., 2015) contains 55 categories; we use 28 known and 27 novel categories. Its larger class-frequency imbalance provides a different challenge from the noisy scanned objects.

Table 4: Statistics of the unlabeled training subsets. Numbers and known/novel composition are reported at the sample level.

Dataset	Labeled Ratio	Total	Known (Ratio)	Novel (Ratio)	Novel Examples
ModelNet40	50%	7,292	2,563 (35.1%)	4,729 (64.9%)	*laptop, mantel*
ScanObjectNN	50%	8,122	3,295 (40.6%)	4,827 (59.4%)	*table, shelf*
ShapeNetCore	50%	23,789	11,935 (50.2%)	11,854 (49.8%)	*lamp, mug*

Table 2 uses a common diagnostic protocol to distinguish data composition from representation quality. IR is the ratio of the largest to the smallest training class, and U-Novel is the fraction of novel samples in the unlabeled pool. Representation diagnostics use frozen Uni3D features from a class-balanced diagnostic bank. kNN@10 reports class-averaged same-category neighbor purity. Flip measures how often strong augmentation changes the nearest clean class-centroid prediction. It measures prediction stability, not classification error. Graph-J is the mean Jaccard overlap between neighborhoods rebuilt from clean and strongly augmented features. Known AUC measures old–novel separation using maximum similarity to labeled known-class centroids. Natural class frequencies are restored by sample weights for this score.

To measure mode allocation, we cluster the mixed unlabeled diagnostic pool into as many groups as there are novel categories. Weighted K-Means preserves the natural class composition. A mode is old-dominated when its total known-class mass exceeds its novel-class mass. Old-modes reports the fraction of nonempty modes satisfying this condition. Novel coverage counts the novel categories that become the largest individual class in at least one novel-dominated mode. The diagnostic labels reveal these patterns after feature extraction; they do not supervise Geo3D.

A.2 Training Parameters and Details

We initialize the Uni3D (Zhou et al., 2024) backbone with its publicly available pretrained weights. We fine-tune the last Transformer block, while freezing the remaining backbone parameters. The projection head and method-specific networks are trainable. The head is a two-layer MLP with a 1,024-dimensional hidden layer, a GELU activation, and a 1,024-dimensional output. Training lasts 100 epochs with seed 42; the multi-seed study also uses seeds 1, 2, and 3.

All three datasets use 1,024 points per object. Training, memory-bank inference, and evaluation use a batch size of 16. We use SGD with momentum 0.9 and weight decay 5×10^{-5}. The initial learning rate is 0.01 for the projection head and method-specific networks. The trainable backbone parameters use a multiplier of 10^{-3}, giving an initial learning rate of 10^{-5}. Cosine decay reduces each parameter group's learning rate to 0.1% of its initial value by the end of training. All experiments use mixed-precision training.

Each point cloud produces two independently augmented views. ModelNet40 and ScanObjectNN use random rotation around the vertical axis, scaling in $[0.8, 1.2]$, and Gaussian jitter with standard deviation 0.01 clipped to 0.05. For ShapeNetCore, the weak view uses limited rotation and scaling in $[0.9, 1.1]$. The strong view uses full vertical rotation, scaling in $[0.7, 1.3]$, and Gaussian jitter. The supervised contrastive weight is $\lambda_s = 0.35$, the contrastive temperature is 0.5, and the base feature-shift coefficient is 0.5.

3DGR selects $k = 8$ memory entries from 32 candidates using feature similarity, geometric pair scores, and cross-view support. We use an outlier-robust geometric descriptor on all datasets; the dynamic reliability gate is enabled only on ScanObjectNN. The fusion coefficient is fixed at $\beta_i = 0.3$ on ModelNet40 and ShapeNetCore. Cross-view support reranks candidate edges. The shift confidence measures the fraction of distinct objects in the geometry-corrected neighborhood that also appear in the other view's raw-feature neighborhood, as defined in Section 3.3. This confidence scales the feature-shift strength. After a five-epoch warm-up, the geometric correction coefficient γ increases linearly from zero to 0.25 over ten epochs. The pair-supervision loss weight is 0.1.

We maintain cross-batch category prototypes with an EMA momentum of 0.8. Prototypes and MGCA memories are updated from the first epoch. MGCA uses the eight memory entries selected by 3DGR, with loss weight $\lambda_m = 0.1$. After five warm-up epochs, ρ increases linearly from zero to one over ten epochs, controlling target blending and the MGCA loss contribution. The known-prototype anchoring weight is 0.75, and the duplicate-mode similarity threshold is 0.95. The uniform coverage mixture and mode-balancing exponent are 0.05 and 0.25, respectively. KnownEx retains unlabeled candidates with lower maximum similarity to known prototypes when constructing novel prototypes. The shared-space supervision weight and its lower bound are both 0.5.

The auxiliary representation objective $\mathcal{L}_{\text{aux}}$ includes prototype supervision, global–local consistency, neighborhood consensus, local-prototype alignment, and mode distillation, with weights 0.3, 0.1, 0.2, 0.15, and 0.2, respectively. These terms are retained in the loss ablations. Class-prior debiasing uses weight 0.4 and prior momentum 0.95. The auxiliary representation losses contribute to the total objective through $\mathcal{L}_{\text{rep}}$ in Eq. (8).

For auxiliary neighborhood consensus, we cache softmax probabilities over all K slots for augmented-view memory entries after Prototype EMA and before the MGCA prototype update. The cache remains fixed within each epoch. The loss uses Jensen–Shannon divergence between a detached aggregate of cached neighbor probabilities and the current-view prediction over all K slots using the final MGCA bank. ScanObjectNN uses neighborhood attention weights for aggregation, whereas ModelNet40 and ShapeNetCore use uniform averaging. MGCA uses separate graph weights w_{ij} and object-level assignment memory.

A.3 Loss Definitions and Ablation Settings

Table 10 ablates four principal objectives: supervised contrastive learning $\mathcal{L}_{\text{sup}}$, binary supervision for labeled candidate pairs $\mathcal{L}_{\text{pair}}$, contrastive consistency on shifted views $\mathcal{L}_{\text{cons}}$, and the full MGCA loss $\mathcal{L}_{\text{m}}$. It also evaluates the target-alignment and known-prototype anchoring terms within $\mathcal{L}_{\text{m}}$. The four principal loss weights are 0.35, 0.1, 0.65, and 0.1, respectively, with the scheduled ramp applied to the MGCA contribution. The pair loss uses a positive-class weight determined by the numbers of positive and negative labeled pairs. Removing it retains the geometric scorer and graph construction.

The auxiliary neighborhood-consensus loss is included in $\mathcal{L}_{\text{aux}}$. It is distinct from the shifted-feature contrastive loss $\mathcal{L}_{\text{cons}}$ and the assignment-consistency term within $\mathcal{L}_{\text{m}}$.

MGCA combines target alignment, assignment consistency across views, novel-mode coverage, reference-relation consistency, and known-prototype anchoring:

$$\mathcal{L}_{\text{m}} = \mathcal{L}_{\text{align}} + 0.25\mathcal{L}_{\text{view}} + 0.10\mathcal{L}_{\text{coverage}} + 0.10\mathcal{L}_{\text{relation}} + 0.75\mathcal{L}_{\text{anchor}}. \tag{9}$$

Target alignment minimizes the mode-weighted cosine distance to the discovery target for unlabeled objects not accepted as known. Known-prototype anchoring minimizes cosine distance to the predicted known prototype for accepted unlabeled objects. The view term matches the two assignment distributions to their detached mean. The coverage term matches the mean novel-slot probabilities to the target prior. The relation term uses a reliability-weighted smooth-ℓ_1 penalty between stu-

dent and reference similarities on selected edges. The view and relation terms are distinct from the contrastive consistency loss on shifted features.

Removing $\mathcal{L}_{\mathrm{m}}$ sets its outer weight to zero while retaining the controller, its memories, and its targets for feature shifting. Removing $\mathcal{L}_{\mathrm{align}}$ or $\mathcal{L}_{\mathrm{anchor}}$ sets only the corresponding internal coefficient to zero. These two rows therefore test subterms of the same loss. Removing the MGCA component disables its target controller and the associated shared-space supervision pathway. Removing $\mathcal{L}_{\mathrm{sup}}$ also disables the gradient projection that uses its shared-space supervised gradient; the other supervision terms remain active. The supervised and contrastive-consistency weights are otherwise unchanged in their respective ablations.

In the component tables, MGCA denotes memory-guided category assignment, KnownEx denotes known exclusion, and EMA denotes prototype exponential moving averaging. Prototype EMA smooths the prototypes used for prototype supervision across epochs, after matching novel slots. MGCA also maintains its own prototype and temporal memories with momentum 0.8. Disabling prototype EMA therefore differs from removing the entire MGCA controller. Known exclusion filters the candidate pool used to construct novel prototypes according to similarity to known prototypes. This differs from restricting accepted known objects to known assignment slots and from the known-prototype anchoring loss.

A.4 Training and Inference Algorithm

Algorithm 1 summarizes the training and inference procedure of Geo3D. At each epoch, we refresh the augmented-view feature memory. Known prototypes are computed from labeled class means, while weighted clustering of KnownEx-filtered candidates constructs novel prototypes. Novel slots are matched across epochs before Prototype EMA. We then cache softmax probabilities over all K slots for each augmented-view memory entry. MGCA aligns the bank with its own prototype memory and applies a second momentum update. It also updates known acceptance, replaces duplicated novel slots using unaccepted objects with low known-class similarity, and refreshes object-level memories and reliability. Prototype and memory updates begin in the first epoch.

During minibatch training, stored features, MGCA assignment memory, the auxiliary consensus cache, and the final prototype bank remain fixed. Selected graph weights remain differentiable. MGCA recomputes q_{ic} for each unlabeled view using the final bank. Accepted known objects use a softmax over known slots, with zero mass on novel slots; other unlabeled objects use all K slots. Labeled assignments are fixed to their observed classes.

Auxiliary neighborhood consensus uses Jensen–Shannon divergence between a detached aggregate of cached neighbor probabilities and the current-view prediction over all K slots using the final bank. ScanObjectNN uses neighborhood attention weights for aggregation; ModelNet40 and ShapeNetCore use uniform averaging. MGCA maintains separate object-level assignment memory and graph weights w_{ij}.

The aggregate h_i combines $\bar{z}_j$ using softmax weights from e_{ij} over $\mathcal{N}_i$. The cross-view shift confidence defined in Section 3.3 controls α_i on all datasets. On ScanObjectNN, the learned routing coefficient and attention concentration further limit this strength.

B Robustness and Efficiency

B.1 Alternative Category Partitions

Table 5 changes the number of known classes while retaining the mixed known–novel discovery setting. Geo3D achieves the highest All and New accuracy under all four partitions. With 10 known and 30 novel ModelNet40 classes, it reaches 75.57% All accuracy and 66.67% New accuracy. Its New score exceeds the best competing result by 5.01 points. With 30 known classes, the New score remains 67.00%, 1.17 points above the strongest competitor. The result is therefore not confined to the equal 20/20 class split used in the main experiment.

The advantage is larger on ShapeNetCore. For the 20/35 and 40/15 splits, Geo3D improves New accuracy over the best competing results by 13.79 and 15.33 points. In the 40/15 setting, it obtains 78.80% Old and 79.70% New accuracy. Geo3D retains strong novel discovery even when known

Algorithm 1 Training and inference for Geo3D

Require: $\mathcal{D}_l$, $\mathcal{D}_u$, K, pretrained Uni3D, neighbor count k, and test point clouds
Ensure: Category assignments for the test point clouds

Initialize trainable components and a frozen reference encoder.
for each epoch **do**
  Set $\rho = \gamma = 0$ during warm-up; afterward update them using the ramps in Section A.2.
  Refresh augmented-view feature memory and frozen-reference features.
  Construct known prototypes from labeled class means and novel prototypes by weighted clustering of KnownEx-filtered candidates.
  Match novel slots across epochs and apply Prototype EMA.
  Cache softmax probabilities over all K slots for each augmented-view memory entry using this bank.
  Perform the MGCA refresh, including prototype alignment, the second momentum update, known acceptance, duplicate-slot refresh, and object-level memory updates; obtain $\{\mu_c\}_{c=1}^{K}$, m_o, and r_i.
  for each training batch **do**
    Encode and project paired views to obtain unit-normalized z_i^g, z_i^l.
    On ScanObjectNN, predict β_i and the routing coefficient using the reliability network; otherwise set $\beta_i = 0.3$.
    $z_i \propto (1-\beta_i) z_i^g + \beta_i z_i^l$.
    Retrieve $\mathcal{C}_i$, excluding all stored views of the query object; compute ℓ_{ij} from feature similarity and comparisons of outlier-robust geometric descriptors.
    $$\delta_{ij} \leftarrow \gamma \left(\tanh \ell_{ij} - \frac{1}{|\mathcal{C}_i|} \sum_{v \in \mathcal{C}_i} \tanh \ell_{iv} \right).$$
    Set $s_{ij} = 1$ if $o(j)$ appears among the other query view's top-k distinct objects ranked by raw feature similarity; otherwise use a positive floor.
    For each query view, select the top-k entries in $\mathcal{C}_i$ using $z_i^\top \bar{z}_j / \tau + \delta_{ij}$.
    Compute shift confidence as the fraction of distinct selected objects that appear among the other view's top-k distinct objects ranked by raw feature similarity.
    $e_{ij} \leftarrow z_i^\top \bar{z}_j / \tau + \delta_{ij} + \log s_{ij}$; select the final top-$k$ entries in $\mathcal{C}_i$ as $\mathcal{N}_i$.
    Fix labeled assignments to observed classes; compute unlabeled q_{ic} from $z_i^\top \mu_c / \tau_p$ using the MGCA slot restrictions above.
    Obtain a_{ij} from frozen-reference similarity, agreement between stored MGCA assignments, and endpoint reliability, with a floor for uncertain edges.
    $$w_{ij} \leftarrow \frac{a_{ij} \exp(e_{ij})}{\sum_{v \in \mathcal{N}_i} a_{iv} \exp(e_{iv})}.$$
    $$g_i \propto \sum_{j \in \mathcal{N}_i} w_{ij} m_{o(j)}, \qquad p_i \propto \sum_{c=1}^{K} \mathrm{sg}(q_{ic}) \mu_c, \qquad b_i \propto (g_i + p_i)/2.$$
    $$t_i \propto \begin{cases} \mathrm{sg}(z_i), & \text{labeled}, \\ p_i, & \text{unlabeled, accepted as known}, \\ r_i b_i + (1-r_i)\,\mathrm{sg}(z_i), & \text{otherwise}. \end{cases}$$
    Compute h_i as in Section 3.5; obtain α_i from shift confidence, with additional routing and attention-concentration controls on ScanObjectNN.
    $t_i^{\mathrm{shift}} \propto (1-\rho) h_i + \rho t_i, \quad \widetilde{z}_i \propto (1-\alpha_i) z_i + \alpha_i t_i^{\mathrm{shift}}$.
    Form the auxiliary consensus target from cached neighbor probabilities; compute its current-view prediction over all K slots using $\{\mu_c\}_{c=1}^{K}$.
    Compute $\mathcal{L}_{\mathrm{rep}}$, $\mathcal{L}_{\mathrm{pair}}$, and $\mathcal{L}_{\mathrm{m}}$ as in Section 3.5, with auxiliary consensus in $\mathcal{L}_{\mathrm{aux}}$.
    $\mathcal{L}_{\mathrm{total}} \leftarrow \mathcal{L}_{\mathrm{rep}} + \lambda_p \mathcal{L}_{\mathrm{pair}} + \rho \lambda_m \mathcal{L}_{\mathrm{m}}$.
    On the shared encoder and projection head, remove discovery-gradient components opposing the supervised gradient; update trainable parameters.
  end for
end for
Extract normalized fused test features $\{z_i\}$ as in Section 3.2.
return category assignments from Ward clustering of $\{z_i\}$ into K clusters.

categories occupy most of the category set. The comparison within each partition shows that its gains extend across different amounts of known-category coverage.

Table 5: Comparison under different known/novel class splits on ModelNet40 and ShapeNetCore.

Method	ModelNet40						ShapeNetCore					
	10/30			30/10			20/35			40/15		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New
GCD	65.76	75.92	61.66	68.11	68.84	65.83	42.08	37.43	49.91	41.11	37.32	57.06
SimGCD	69.12	94.08	59.04	81.89	**91.92**	50.67	74.07	86.86	52.52	77.67	80.83	64.37
BIA	62.20	**98.73**	47.44	78.81	91.65	38.83	69.08	**89.72**	34.32	76.92	**85.66**	40.15
PartCo	64.30	98.03	50.68	78.28	88.65	46.00	69.57	89.34	36.28	76.88	84.80	43.55
Ours	**75.57**	97.61	**66.67**	**82.74**	87.79	**67.00**	**77.56**	84.23	**66.31**	**78.97**	78.80	**79.70**

Table 6: Joint comparison of category discovery performance and computational efficiency. Results are reported with PointNet++ and Uni3D backbones on ModelNet40. All, Old, and New denote clustering accuracy (%), while Trainable, Head, and FLOPs indicate optimization parameters and computational cost.

Method	PointNet++						Uni3D					
	Performance			Efficiency			Performance			Efficiency		
	All	Old	New	Train	Head	FLOPs	All	Old	New	Train	Head	FLOPs
SimGCD	58.07	72.78	44.11	7.59	6.86	1.692	71.03	86.69	56.16	13.95	6.86	87.69
ProtoGCD	64.30	85.52	44.15	7.59	6.86	1.692	73.22	95.42	52.13	13.95	6.86	87.69
BIA	51.74	77.04	27.73	2.08	1.36	1.681	64.22	93.43	36.49	8.45	1.36	88.48
Ours	**67.59**	81.03	**54.82**	3.36	2.63	1.686	**78.40**	90.77	**66.67**	9.72	2.63	**87.68**

B.2 Results with Different Backbones

Table 6 tests whether the gains depend on the Uni3D backbone. With PointNet++, Geo3D reaches 67.59% All and 54.82% New accuracy on ModelNet40. These scores exceed ProtoGCD by 3.29 and 10.67 points. With Uni3D, the gains are 5.18 and 14.54 points. Geo3D therefore improves discovery with both backbone choices.

The parameter counts help distinguish the source of this gain. Geo3D trains 3.36 million parameters with PointNet++ and 9.72 million with Uni3D. Both ProtoGCD and SimGCD train 7.59 million parameters with PointNet++ and 13.95 million with Uni3D. With Uni3D, this is a 30.3% reduction in trainable parameters at nearly unchanged reported FLOPs. Geo3D thus improves the accuracy–parameter trade-off relative to these two methods. BIA remains smaller, but its New accuracy is lower with both backbones.

B.3 Variation across Random Seeds

Table 7: Multi-seed evaluation across three benchmarks. We report clustering accuracy (%) over four random seeds. Seed 42 corresponds to the main result reported in the paper.

Dataset	Seed 1			Seed 2			Seed 3			Seed 42			Mean±Std		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New	All	Old	New
ModelNet40	76.70	89.85	64.22	77.51	91.51	64.22	76.05	89.10	63.67	78.40	90.77	66.67	77.17±0.88	90.31±0.91	64.70±1.16
ScanObjectNN	69.73	73.98	63.25	70.12	75.61	61.74	69.40	74.67	61.35	70.68	75.53	63.28	69.98±0.48	74.95±0.67	62.41±0.87
ShapeNetCore	78.78	84.12	67.53	79.43	84.93	67.84	79.39	84.39	68.84	80.61	85.65	69.98	79.55±0.66	84.77±0.58	68.55±0.96

Table 7 repeats Geo3D with seeds 1, 2, 3, and 42. Mean All accuracy is 77.17 ± 0.88 on ModelNet40, 69.98 ± 0.48 on ScanObjectNN, and 79.55 ± 0.66 on ShapeNetCore. Every run exceeds the strongest baseline All score in Table 1. Mean New accuracy is 64.70%, 62.41%, and 68.55%, with standard deviations of 1.16, 0.87, and 0.96 points. The advantage is therefore observed beyond the default

seed. All four runs also exceed the best competing New score on each benchmark, so the discovery advantage persists across the evaluated seeds.

B.4 Computational Complexity

Table 8: Computational efficiency comparison with representative GCD methods. All methods are evaluated with the same Uni3D backbone on ModelNet40. Trainable parameters, FLOPs, training memory, and estimated training computation are reported under the same evaluation protocol.

Method	Trainable (M)	FLOPs (G)	AMP Train Mem. (MiB)	100e Est. PFLOPs
BIA	8.45	88.48	589.2	522.6
ProtoGCD	13.95	87.69	629.8	517.8
SimGCD	13.95	87.69	630.0	517.8
Ours	9.72	87.68	596.8	517.8

Table 8 reports the computational profile with Uni3D on ModelNet40. Geo3D uses 9.72 million trainable parameters and 596.8 MiB of reported mixed-precision training memory. ProtoGCD and SimGCD use 13.95 million parameters and about 630 MiB. FLOPs remain close across the three methods: 87.68 G for Geo3D and 87.69 G for ProtoGCD and SimGCD. The gain in discovery accuracy is therefore obtained with fewer trainable parameters at a similar reported forward computation cost. BIA has the smallest parameter and memory counts in this comparison, making the accuracy–cost trade-off explicit.

The "100e Est. PFLOPs" column reports the estimated training computation over a common 100-epoch accounting horizon, consistent with the training schedule used in our experiments. These measurements describe the model profile. Memory refresh, frozen-reference extraction, prototype updates, and final clustering introduce additional computation beyond the reported forward FLOPs.

C Extended Mechanism Analysis

C.1 Representation and Assignment Diagnostics

Table 9 compares Baseline and Full in category accuracy, neighborhood structure, and category assignment.

Novel kNN@10 measures same-category purity among the ten nearest neighbors of true novel samples. The old-neighbor ratio measures the proportion of known samples in these neighborhoods. N→O is the proportion of true novel samples assigned to known categories after global matching. Representation–decision mismatch restricts this proportion to true novel samples whose ten neighbors contain a strict majority of true novel samples, counting all novel categories together.

Table 9: Representation and assignment diagnostics for Baseline and Full (%).

Dataset	**Method**	**All ↑**	**Old ↑**	**New ↑**	**Novel kNN@10 ↑**	**Old-neighbor ratio ↓**	**N→O ↓**	**Rep.–decision mismatch ↓**
ModelNet40	Baseline	74.43	90.93	58.77	73.73	14.74	20.77	10.14
	Full	78.40	90.77	66.67	79.12	13.32	19.91	10.79
ScanObjectNN	Baseline	66.52	72.66	57.14	62.42	22.26	27.34	13.95
	Full	70.68	75.53	63.28	71.44	17.41	22.79	10.17
ShapeNetCore	Baseline	78.30	84.44	65.34	77.39	12.53	15.84	8.60
	Full	80.61	85.65	69.98	77.86	12.24	14.06	7.67

On ScanObjectNN, novel kNN@10 purity increases from 62.42% to 71.44%, while the old-neighbor ratio decreases from 22.26% to 17.41%. Novel-to-old absorption falls from 27.34% to 22.79%, and representation–decision mismatch decreases from 13.95% to 10.17%. These changes indicate more category-consistent neighborhoods and lower novel-to-known assignment rates.

On ModelNet40, novel kNN@10 purity increases from 73.73% to 79.12%, and the old-neighbor ratio decreases from 14.74% to 13.32%. Novel-to-old absorption falls from 20.77% to 19.91%, but representation–decision mismatch increases from 10.14% to 10.79%. Improved local purity thus coexists with a small increase in conditional assignment error.

On ShapeNetCore, New accuracy improves from 65.34% to 69.98%.

C.2 Interpreting the Loss Ablations

Table 10: Loss ablation on three benchmarks. $\mathcal{L}_{\text{sup}}$, $\mathcal{L}_{\text{pair}}$, $\mathcal{L}_{\text{cons}}$, and $\mathcal{L}_{\text{m}}$ denote supervised contrastive, pair supervision, contrastive consistency, and MGCA losses, respectively. $\mathcal{L}_{\text{align}}$ and $\mathcal{L}_{\text{anchor}}$ are the target-alignment and known-prototype anchoring terms within $\mathcal{L}_{\text{m}}$.

		Configurations						ModelNet40			ScanObjectNN			ShapeNetCore		
ID	**Variant**	$\mathcal{L}_{\text{sup}}$	$\mathcal{L}_{\text{pair}}$	$\mathcal{L}_{\text{cons}}$	$\mathcal{L}_{\text{m}}$	$\mathcal{L}_{\text{align}}$	$\mathcal{L}_{\text{anchor}}$	**All**	**Old**	**New**	**All**	**Old**	**New**	**All**	**Old**	**New**
(a)	w/o $\mathcal{L}_{\text{sup}}$	✗						**78.61**	87.35	**70.30**	62.07	67.32	54.08	56.50	51.28	67.49
(b)	w/o $\mathcal{L}_{\text{pair}}$		✗					75.73	87.27	64.77	69.36	**77.37**	57.14	78.44	84.45	65.77
(c)	w/o $\mathcal{L}_{\text{cons}}$			✗				68.84	81.86	56.48	57.70	67.43	42.86	79.17	85.53	65.77
(d)	w/o $\mathcal{L}_{\text{m}}$				✗			76.26	90.93	62.32	68.91	76.68	57.06	78.15	**85.91**	61.80
(e)	w/o $\mathcal{L}_{\text{align}}$					✗		77.31	**92.10**	63.27	68.67	76.39	56.88	77.93	84.40	64.28
(f)	w/o $\mathcal{L}_{\text{anchor}}$						✗	77.07	90.77	64.06	70.02	75.88	61.09	78.69	85.65	64.01
(g)	**Full**	✓	✓	✓	✓	✓	✓	78.40	90.77	66.67	**70.68**	75.53	**63.28**	**80.61**	85.65	**69.98**

Table 10 examines the training objectives separately from the component removals in the main text. Supervised contrastive learning is especially important on ScanObjectNN and ShapeNetCore. Removing it lowers All accuracy by 8.61 and 24.11 points. On ShapeNetCore, Old accuracy falls from 85.65% to 51.28%, showing the role of labeled anchors in maintaining known-category structure. On ModelNet40, the same removal slightly raises All accuracy to 78.61%, but reduces Old accuracy to 87.35%. Supervision therefore has a dataset-dependent trade-off; its clearest benefit appears in the more difficult joint discovery settings.

The pair and consistency losses supervise the neighborhood relations and shifted views. Without pair supervision, New accuracy falls by 1.90, 6.14, and 4.21 points across the three benchmarks. Removing contrastive consistency is most damaging on ScanObjectNN, where New accuracy decreases from 63.28% to 42.86%. Removing the MGCA loss lowers ShapeNetCore New accuracy to 61.80%, close to the failure pattern seen when the module is removed. Target alignment and known-prototype anchoring each improve New accuracy on all three datasets. On ShapeNetCore, their contributions are 5.70 and 5.97 points. The two MGCA subterms act on different unlabeled subsets: target alignment guides objects not accepted as known, while known-prototype anchoring stabilizes accepted known objects.

C.3 Incremental Component Ablation

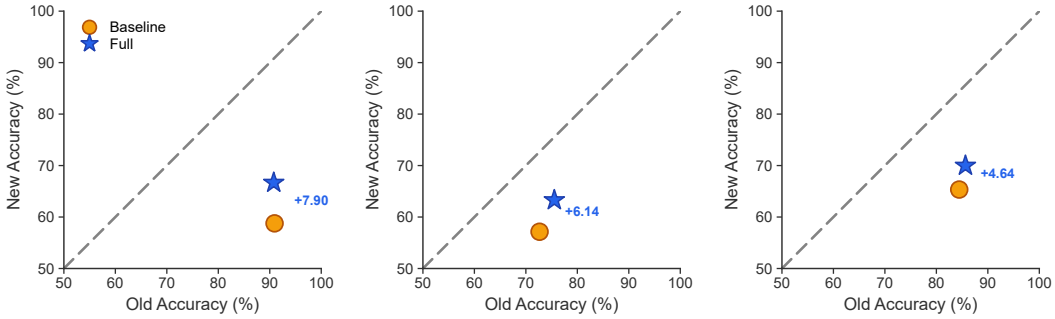
Table 11: Incremental component ablation of Geo3D. We progressively introduce the proposed components to the baseline model. All results are reported using the same evaluation protocol.

Variant	**Components**					ModelNet40			ScanObjectNN			ShapeNetCore		
	Baseline	**3DGR**	**MGCA**	**KnownEx**	**EMA**	**All**	**Old**	**New**	**All**	**Old**	**New**	**All**	**Old**	**New**
Baseline	✓					74.43	90.93	58.77	66.52	72.66	57.14	78.30	84.44	65.34
+ 3DGR	✓	✓				78.00	87.85	**68.64**	68.32	74.15	59.42	80.04	**86.75**	65.89
+ MGCA	✓	✓	✓			75.93	85.69	66.67	66.79	74.27	55.39	76.90	84.47	60.95
+ KnownEx	✓	✓	✓	✓		76.46	**92.01**	61.69	68.56	**75.88**	57.41	79.18	85.24	66.40
Full	✓	✓	✓	✓	✓	**78.40**	90.77	66.67	**70.68**	75.53	**63.28**	**80.61**	85.65	**69.98**

Table 11 builds Geo3D from the baseline, denoted Baseline in the table. Adding 3DGR first improves All accuracy on every dataset. On ModelNet40, New accuracy rises from 58.77% to 68.64%; on ScanObjectNN, it rises from 57.14% to 59.42%. Adding MGCA before known-category exclusion lowers ShapeNetCore New accuracy from 65.89% to 60.95%. Introducing exclusion recovers

it to 66.40%, and the full configuration reaches 69.98%. This sequence is consistent with the need to constrain ownership when the unlabeled pool includes known objects.

The final model has the highest All accuracy on all three datasets, exceeding baseline by 3.97, 4.16, and 2.31 points. Incremental rows describe the behavior of successive configurations, so their differences should not be treated as independent estimates of each component's effect. In particular, the full ScanObjectNN setting also includes its dataset-specific reliability gates. The leave-one-out results isolate removals from that full setting and are the basis for attributing the prototype-momentum gains in the main text. Together, the two ablations show both the value of each component in the complete model and the interactions that matter when constructing it.

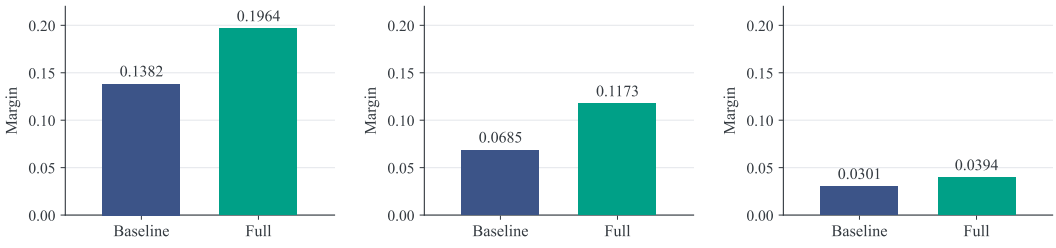


(a) ModelNet40 (b) ScanObjectNN (c) ShapeNetCore

Figure 6: **Old–new accuracy balance from baseline to Full.** Each panel shows the movement from the baseline model to the Full model in Old–New accuracy space. The dashed diagonal indicates balanced performance between old and novel categories.

Figure 6 summarizes the net movement from baseline to Full. New accuracy increases by 7.90 points on ModelNet40, 6.14 on ScanObjectNN, and 4.64 on ShapeNetCore. The Old–New gap contracts from 32.16 to 24.10 points, from 15.52 to 12.25 points, and from 19.10 to 15.67 points, respectively. On ScanObjectNN and ShapeNetCore, both Old and New accuracy improve. On ModelNet40, Old accuracy changes by only -0.16 points while New accuracy rises substantially. Geo3D therefore reduces the imbalance through stronger novel discovery without a large loss in known-class recognition relative to baseline.

C.4 Original-Space Geometry

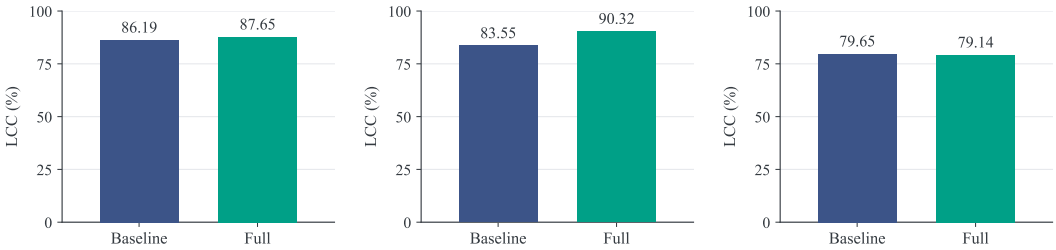


(a) ModelNet40 (b) ScanObjectNN (c) ShapeNetCore

Figure 7: Novel–old cosine margin in the original feature space. Full increases the margin over Baseline on all three datasets, indicating improved separation between novel and known categories.

Figure 7 measures separation in the original feature space. The novel–old cosine margin increases from 0.1382 to 0.1964 on ModelNet40, from 0.0685 to 0.1173 on ScanObjectNN, and from 0.0301 to 0.0394 on ShapeNetCore. The consistent direction supports improved separation between known and novel categories. The gain is largest in absolute margin on ModelNet40, while ScanObjectNN shows the largest accompanying improvement in connectivity.

Figure 8 asks whether samples of the same category remain connected in a 10-NN graph. The largest connected component ratio rises from 86.19% to 87.65% on ModelNet40 and from 83.55%

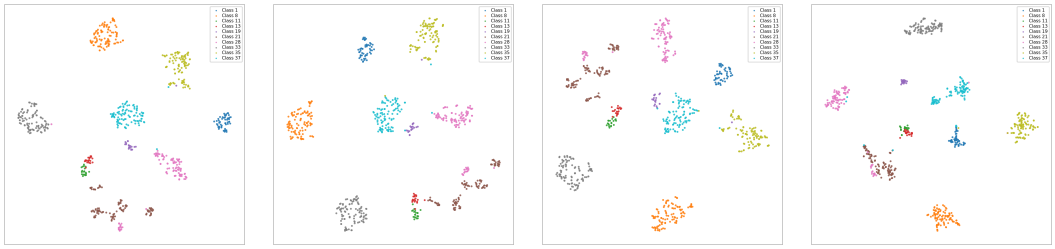


(a) ModelNet40 (b) ScanObjectNN (c) ShapeNetCore

Figure 8: Largest connected component (LCC) ratio in same-class 10-NN graphs constructed in the original feature space. Full improves connectivity on ModelNet40 and ScanObjectNN, with a slight decrease on ShapeNetCore.

to 90.32% on ScanObjectNN. The larger ScanObjectNN gain agrees with the improvement in novel-neighbor purity. ShapeNetCore changes from 79.65% to 79.14%, despite its higher novel accuracy and cosine margin. Geo3D's benefit is therefore not equivalent to increasing within-class connectivity everywhere.

C.5 Embedding Visualization



(a) SimGCD (b) ProtoGCD (c) BIA (d) Ours

Figure 9: t-SNE visualization of feature embeddings for 10 selected ModelNet40 categories using Uni3D as the backbone. The four panels compare the intra-class grouping and inter-class overlap produced by SimGCD, ProtoGCD, BIA, and our method. Colors indicate categories and are used only for visualization.

Figure 9 visualizes ten selected ModelNet40 categories. Geo3D forms compact groups for several categories, with residual overlap among some neighboring groups. Each panel is a separate t-SNE projection. The original-space diagnostics in Figures 7 and 8 provide complementary evidence for the Baseline-to-Full comparison.

D Remaining Assignment Errors

Table 12: Recoverable information in the learned representation. Oracle-based variants use ground-truth old–novel membership only for offline diagnosis.

Dataset	Old Categories			Novel Categories	
	Shared	L-Centroid	Oracle	Shared	Oracle
ModelNet40	90.77	96.59	94.51	66.67	73.30
ScanObjectNN	75.53	81.91	80.70	63.28	68.36
ShapeNetCore	85.65	89.68	87.61	69.98	75.89

Table 12 asks how much category information remains recoverable when old–novel membership is supplied for diagnosis. The labeled-centroid variant classifies the true old subset using centroids

estimated from up to 64 labeled training examples per class. The oracle clustering variant separates the true old and novel subsets before clustering. Both use unavailable membership information to define the evaluation problem, so neither is an inference alternative used for the main results.

Under this diagnostic separation, New accuracy rises from 66.67% to 73.30% on ModelNet40, from 63.28% to 68.36% on ScanObjectNN, and from 69.98% to 75.89% on ShapeNetCore. Known-class accuracy also improves when the old subset is evaluated separately. The labeled-centroid scores reach 96.59%, 81.91%, and 89.68%. These results indicate that the shared partition does not fully recover the category information accessible under the diagnostic settings.

Table 13: Error profile under joint clustering. We report the micro–macro gap for novel categories and the main cross-partition and within-old clustering errors. Ground-truth labels are used only for offline diagnosis.

Dataset	**Novel Accuracy**			**Clustering Errors**		
	Micro	**Macro**	**Gap**	**N→O**	**O→N**	**O→O**
ModelNet40	66.67	58.07	8.59	19.91	5.49	3.74
ScanObjectNN	63.28	63.46	−0.19	22.79	11.09	13.38
ShapeNetCore	69.98	44.73	25.25	14.06	10.62	3.74

Table 13 separates frequency effects from cross-group assignment errors. ModelNet40 has an 8.59-point gap between novel micro and macro accuracy, whereas ScanObjectNN has nearly equal scores, with a gap of -0.19 points. ShapeNetCore has the largest gap: 69.98% micro accuracy versus 44.73% macro accuracy, a difference of 25.25 points. Strong sample-level discovery on this dataset thus coexists with uneven accuracy across novel categories, highlighting balanced category coverage as a remaining challenge.

Novel-to-old absorption is 19.91% on ModelNet40, 22.79% on ScanObjectNN, and 14.06% on ShapeNetCore. On ScanObjectNN, old-to-novel and old-to-other-old errors are 11.09% and 13.38%, respectively. Discovery errors therefore arise both across the old–novel boundary and within the known group. The rates use the corresponding source group as the denominator: novel samples for N→O and old samples for O→N or O→O, where O→O denotes assignment to an incorrect known category. All rates use the same globally matched category assignment. Gap is computed from unrounded micro and macro accuracies.

E Sensitivity to Training Settings

E.1 Number of Neighbors

Table 14: Sensitivity to the number of neighbors k. We evaluate different neighborhood sizes while keeping other configurations fixed. The default setting $k = 8$ is highlighted and adopted in the main experiments.

k	**ModelNet40**			**ScanObjectNN**			**ShapeNetCore**		
	All	Old	New	All	Old	New	All	Old	New
4	76.58	90.77	63.11	57.46	63.87	47.68	79.53	85.49	66.98
8	78.40	90.77	66.67	70.68	75.53	63.28	80.61	85.65	69.98
16	75.41	89.77	61.77	69.57	76.34	59.25	78.64	82.50	70.49

Table 14 varies the number of selected neighbors while keeping the candidate pool fixed. The default $k = 8$ gives the highest All accuracy on every dataset. ScanObjectNN is the most sensitive to a small neighborhood: $k = 4$ yields 57.46% All accuracy, compared with 70.68% at $k = 8$. Increasing the neighborhood to 16 gives 69.57%. This pattern favors enough neighbors to support stable targets, while retaining selective aggregation.

On ModelNet40, $k = 8$ improves All accuracy by 1.82 points over $k = 4$ and by 2.99 points over $k = 16$. On ShapeNetCore, $k = 16$ gives a slightly higher New score, 70.49% versus 69.98%, but lowers Old accuracy from 85.65% to 82.50%. The default yields the better joint result.

E.2 Geometric Refinement Schedule

Table 15: Sensitivity to the geometric refinement start epoch. We vary the start epoch while keeping other settings fixed. The default setting is highlighted and adopted in the main experiments.

Refinement start	ModelNet40			ScanObjectNN			ShapeNetCore		
	All	Old	New	All	Old	New	All	Old	New
0	76.18	90.52	62.56	54.72	56.81	51.53	68.16	71.97	60.13
5	78.40	90.77	66.67	70.68	75.53	63.28	80.61	85.65	69.98
10	76.30	89.43	63.82	70.40	76.68	60.82	77.85	84.91	62.95

Table 15 varies the epoch at which geometric refinement begins. Starting at epoch 5 gives the highest All accuracy on all three datasets. Activating it immediately is particularly harmful on ScanObjectNN and ShapeNetCore, where All accuracy falls to 54.72% and 68.16%. A short warm-up allows the shared features to establish useful candidate neighborhoods before geometric refinement shapes the targets. Waiting until epoch 10 also reduces New accuracy, from 66.67% to 63.82% on ModelNet40 and from 69.98% to 62.95% on ShapeNetCore. The five-epoch warm-up balances early feature adaptation with sufficient training under the refined graph.

E.3 Geometric Score-Correction Weight

Table 16: Sensitivity to the geometric score-correction weight. We vary the correction weight while keeping other settings fixed. The default value is 0.25.

Correction weight	ModelNet40			ScanObjectNN			ShapeNetCore		
	All	Old	New	All	Old	New	All	Old	New
0.10	77.80	91.43	64.85	70.75	78.00	59.68	78.63	85.16	64.86
0.25	78.40	90.77	66.67	70.68	75.53	63.28	80.61	85.65	69.98
0.50	75.77	87.60	64.53	70.26	77.54	59.16	78.50	83.74	67.46

Table 16 varies the geometric score-correction weight from 0.1 to 0.5. The default 0.25 gives the highest New accuracy on all three datasets and the highest All accuracy on ModelNet40 and ShapeNetCore. On ScanObjectNN, 0.1 slightly increases All accuracy from 70.68% to 70.75%, but reduces New accuracy from 63.28% to 59.68%. Increasing the weight from 0.25 to 0.5 lowers both All and New accuracy on every dataset.

E.4 Labeled Fraction

Table 17: Sensitivity to the labeled ratio. We evaluate different labeled ratios while keeping other configurations fixed. The default setting is highlighted and adopted in the main experiments.

Label ratio	ModelNet40			ScanObjectNN			ShapeNetCore		
	All	Old	New	All	Old	New	All	Old	New
0.10	76.62	88.02	65.80	65.54	72.09	55.57	57.86	52.86	68.40
0.25	77.67	89.93	66.03	64.16	67.55	58.98	71.14	72.10	69.13
0.50	78.40	90.77	66.67	70.68	75.53	63.28	80.61	85.65	69.98

Table 17 reduces the labeled fraction of known-class training samples from 50% to 25% and 10%. On ModelNet40, All accuracy remains 76.62% at the 10% ratio, and New accuracy remains 65.80%, close to 66.67% at the default ratio. The model can therefore preserve much of its discovery performance with fewer labels on this dataset. ScanObjectNN benefits more from additional labels: New accuracy rises from 55.57% to 58.98% and then 63.28% across the three ratios.

ShapeNetCore exposes a different dependency. At the 10% ratio, New accuracy is 68.40%, while Old accuracy is 52.86%. Increasing the ratio to 50% raises these scores to 69.98% and 85.65%. The main change is therefore stronger known-category recognition, not a comparable rise in novel accuracy. Alongside the supervision ablation, this result shows the importance of labeled anchors for the known portion of a severely imbalanced category space. Additional labels strengthen joint recognition even when the novel representation already supports useful grouping.

F Simulation with Real-Scanned Objects

We evaluate category-guided robot indication in simulation using real-scanned objects. Blocks, cans, and cups are scanned with a handheld scanner (Figure 10). Their reconstructed geometry is imported into the simulated scene.



(a) Block (Old) (b) Can (Novel) (c) Cup (Old)

Figure 10: Handheld scanning of a block, a can, and a cup. The reconstructed geometry is imported into the simulation.

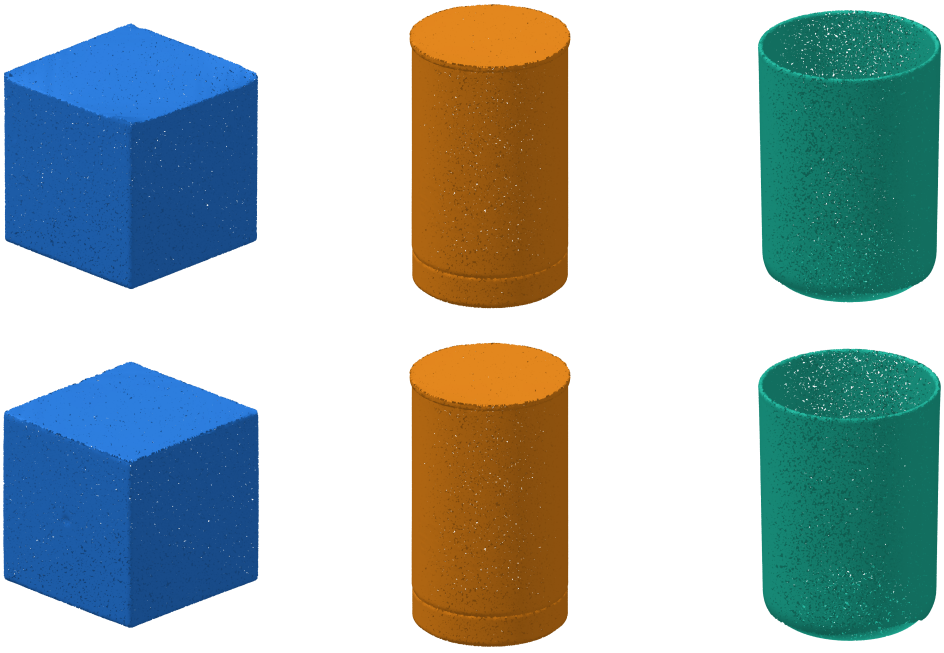
Blocks and cups serve as known reference categories, while cans are treated as novel. The scene contains six object instances, with two per category. Figure 11 shows their point clouds. The two can instances share the same source scan.

Geo3D assigns the block and cup instances to their respective known groups and places the cans in a separate novel cluster. The robot arm uses these assignments to select the indication target associated with each group (Figure 12). The complete simulation video is included in the supplementary material.

G Limitations and Future Work

Geo3D assumes a known category count and a fixed set of known and novel classes. Real-world environments may introduce new categories over time. Handling this setting requires estimating category counts and updating both the prototypes and the final partition. Student and reference predictions can also agree on incorrect relations. Their agreement may therefore be insufficient for identifying reliable neighbors under changes in sensors or environments. The ShapeNetCore diagnostic shows a gap between novel micro and macro accuracy. This indicates uneven performance across categories. Improving rare-category coverage remains important for deployment in environments where some objects appear much less frequently than others.

The current robot demonstration uses real-scanned objects in simulation. Future work will evaluate Geo3D on physical robots using point clouds captured during operation. We will examine how occlusion, incomplete geometry, and sensor noise affect category discovery. Further experiments will test whether the predicted groups can guide object selection, grasping, and sorting. Longer deployments will also require the system to preserve known-category recognition while incorporating newly encountered categories.



(a) Blocks (Old) (b) Cans (Novel) (c) Cups (Old)

Figure 11: Point clouds of the six scene instances, with two instances per category.



(a) Block (Old) (b) Can (Novel) (c) Cup (Old)

Figure 12: Robot indication guided by Geo3D. The arm moves to the targets associated with the block, can, and cup groups in turn. Dashed lines highlight the selected group.