PartCo Part-level correspondence priors enhance Category Discovery

Fernando Julio Cendra & Kai Han
Visual AI Lab
School of Computing and Data Science
The University of Hong Kong

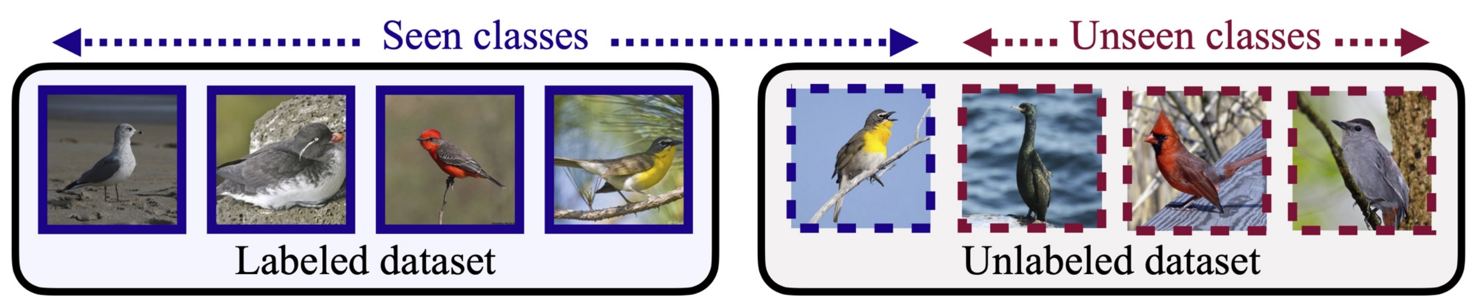
Project page

ICML
International Conference
On Machine Learning

Generalized Category Discovery (GCD)

The focus of this paper is on **GCD**. Given a labeled subset contains seen classes, the task is to categorize the unlabeled images, which may belong to seen or unseen classes.

Our contribution: a novel framework that improves GCD by incorporating part-level visual correspondences priors.



Motivation

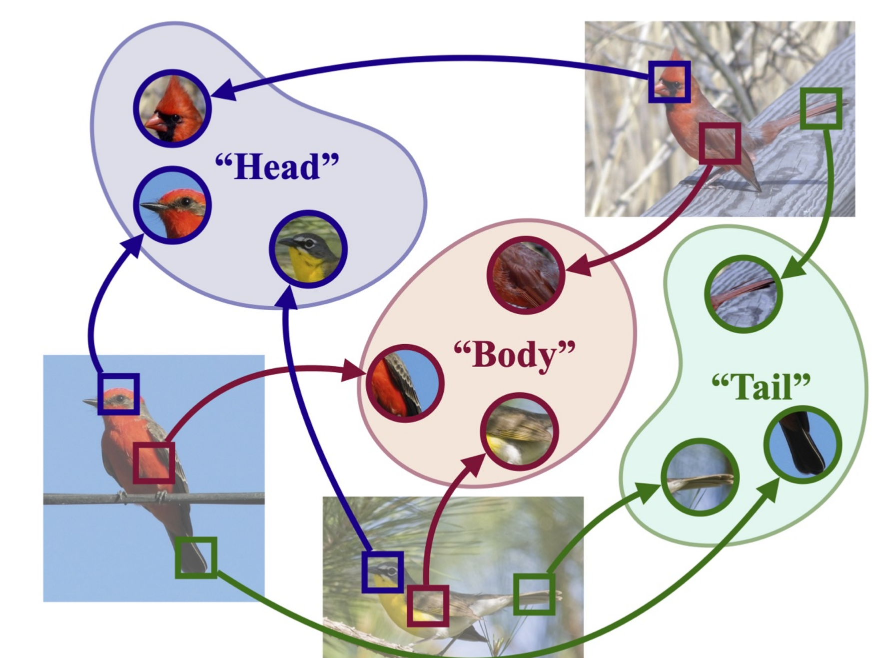
A growing body of literature in GCD, emphasizes the significance of object parts as effective conduits for transferring knowledge between "seen" and "unseen" categories.

Object parts encapsulate fine-grained visual features whose correspondence are often shared across different categories, **facilitating generalization priors to novel class**.

Proposition

DINO-based vision foundation models excel at capturing global context and fine-grained patch details.

Crucially, **these patch features inherently supply part-level correspondence signals**, offering ideal supervision robust to scale, occlusion, and part number variability.

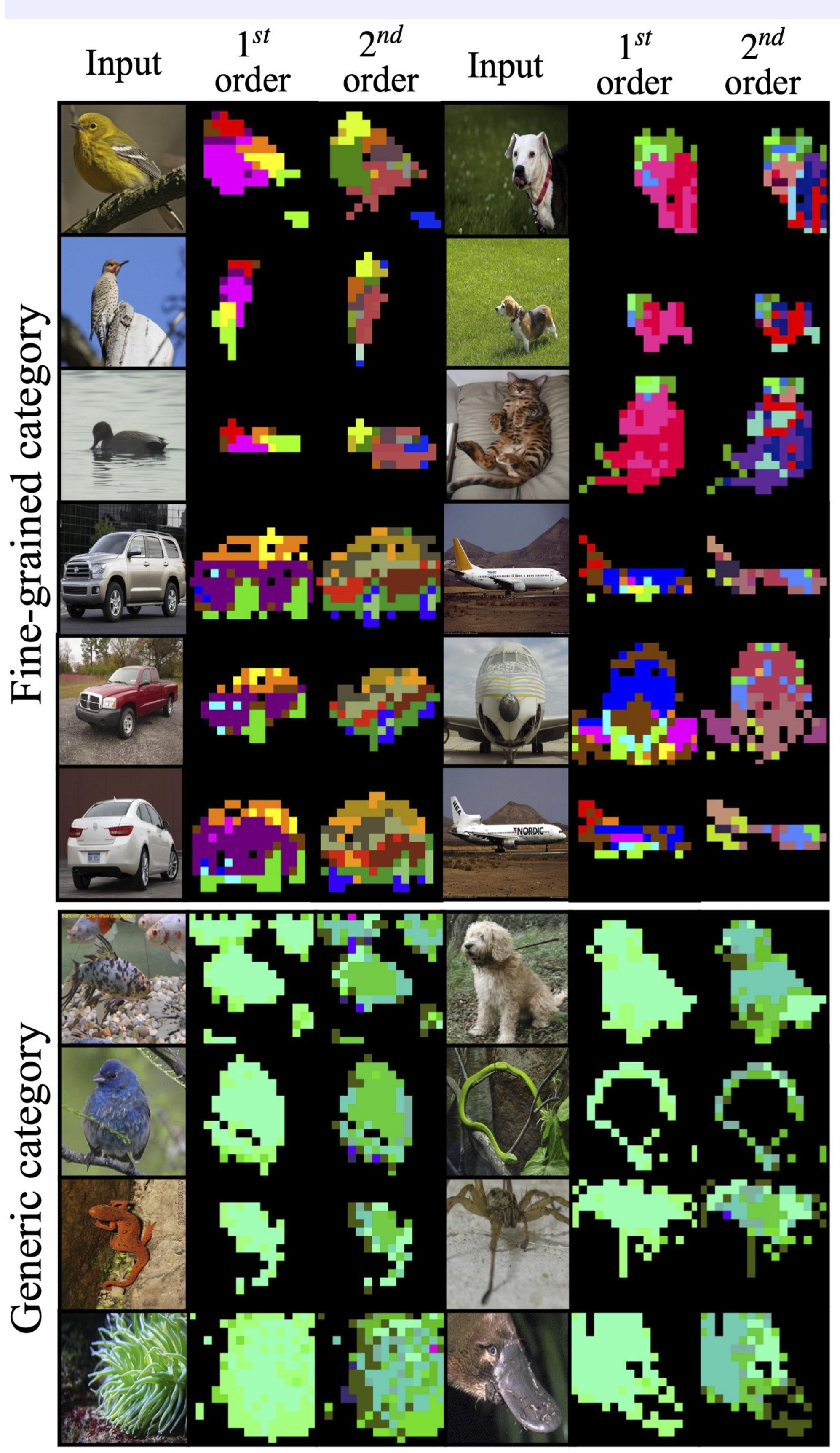


Method

Label construction

We start by applying PCA projection on DINO's patch features to extract foreground object and then another PCA projection on the masked features to extract the fine-grained part-level correspondence information.

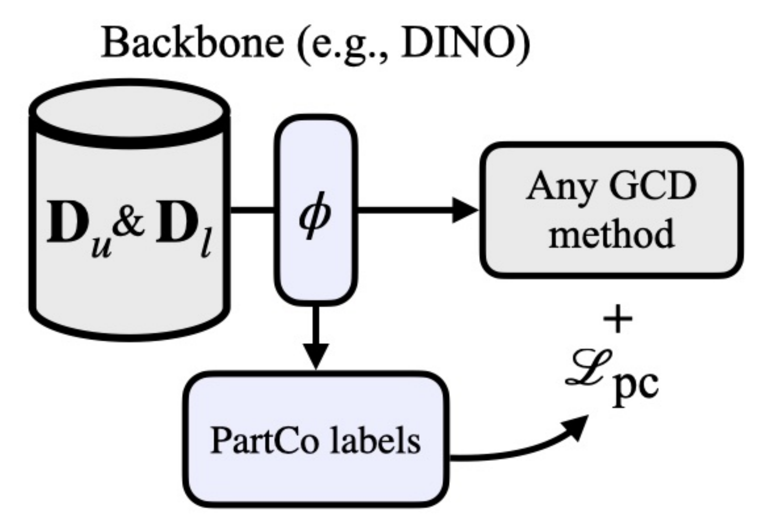
These projections are then applied to the entire labeled set to construct **part-level Correspondence labels.**



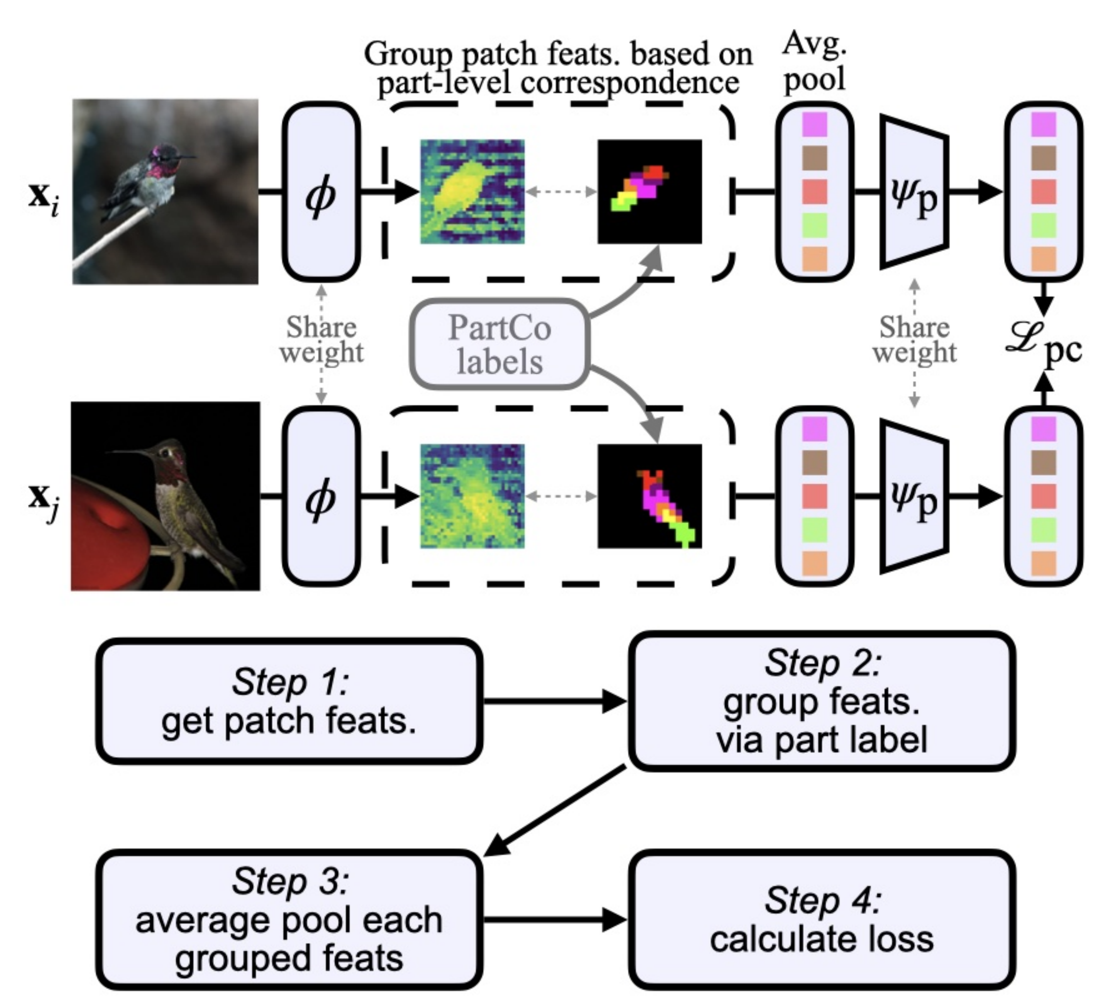
PartCo framework

(a) PartCo framework: a plug-and-play module that incorporates **part-level correspondence labels** to enhance GCD methods. *PartCo adds no inference cost

(b) Part-level correspondence loss: the loss encourages patch features of the same part type and class to be close while separating different parts and / or classes, thereby enhancing models with part-level correspondence priors.



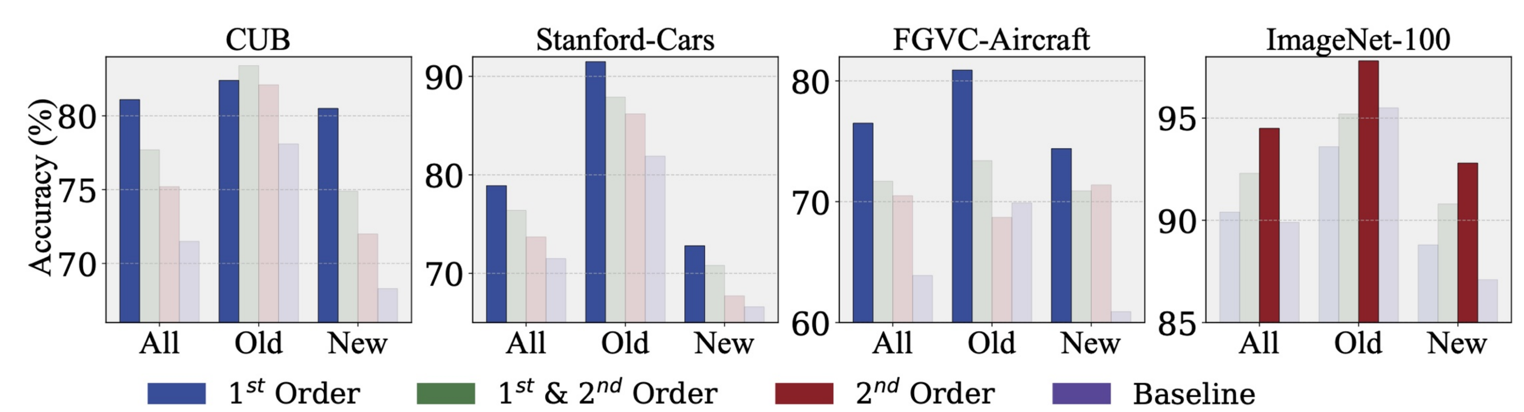
(a) PartCo framework



(b) Part-level correspondent loss ($\mathcal{L}_{pc}$)

Impact of different order levels in PartCo labels

We conduct an ablation study within our PartCo framework to compare 1st and 2nd order labels, their combination, and a baseline. The results show that 1st order labels consistently achieve highest accuracy on fine-grained datasets. While, 2nd order labels outperform 1st order by providing more fine-grained information on generic datasets



Attention map visualization from PartCo

The attention map visualizations from PartCo show that our framework **consistently focuses on discriminative object parts**, across both seen and novel classes, highlighting regions essential for category discovery

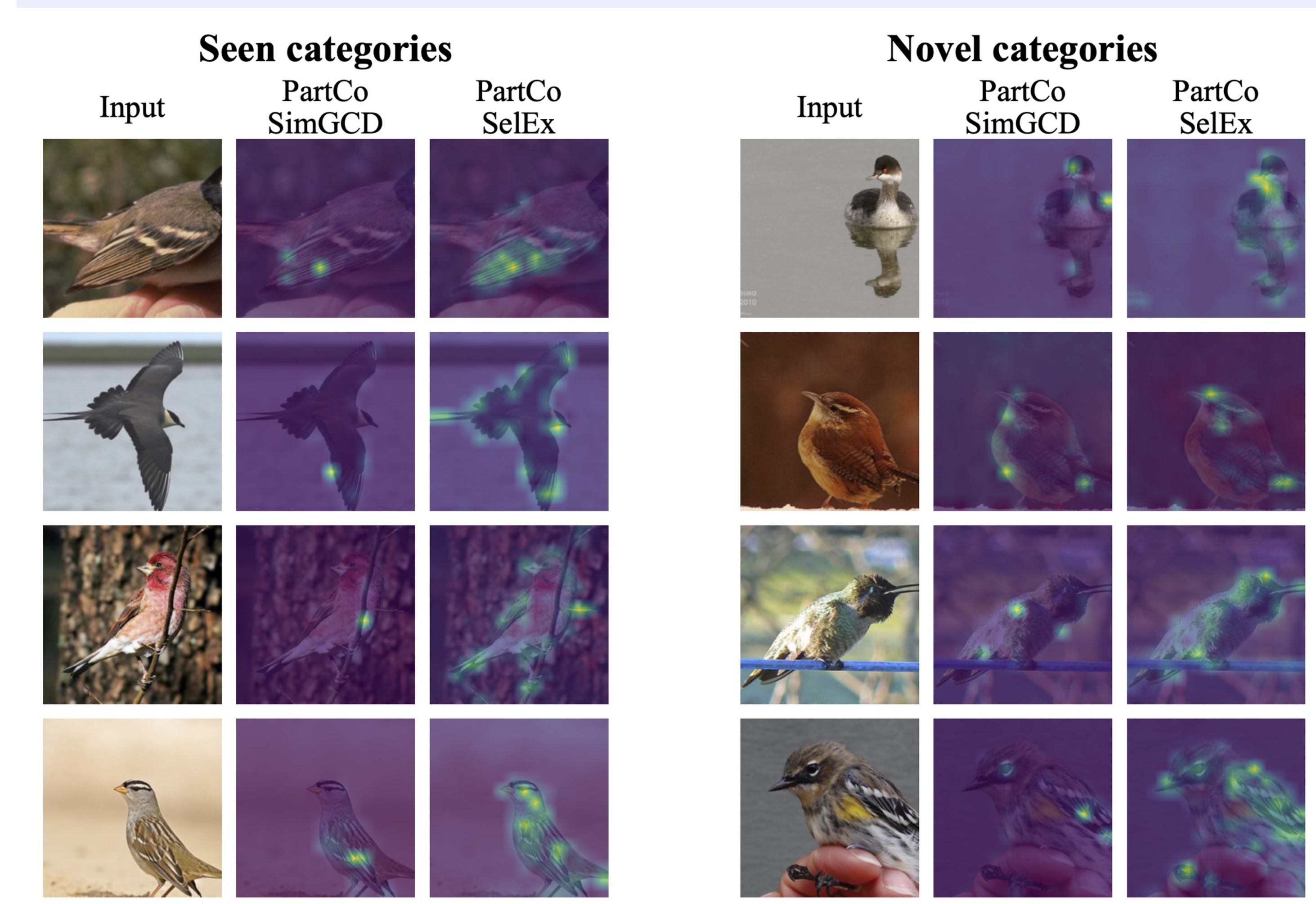


Table 1. Comparison of GCD methods on the SSB benchmark. Results are reported in *ACC* across the 'All', 'Old' and 'New' categories.

			CUB			Stanford-Cars			FGVC-Aircraft			Average		
Method	Venue	Backbone	All	Old	New	All	Old	New	All	Old	New	All	Old	New
k-means	-	DINOv2	67.6	60.6	71.1	29.4	24.5	31.8	18.9	16.9	19.9	38.6	34.0	40.0
GCD	CVPR'22	DINOv2	71.9	71.2	72.3	65.7	67.8	64.7	55.4	47.9	59.2	64.3	62.3	65.4
μGCD	NeurIPS'23	DINOv2	74.0	75.9	73.1	76.1	91.0	68.9	66.3	68.7	65.1	72.1	78.6	69.0
CiPR	TMLR	DINOv2	78.3	73.4	80.8	66.7	77.0	61.8	-	-	-	-	-	-
ProtoGCD	TPAMI	DINOv2	75.7	81.5	72.9	77.6	90.5	71.5	71.1	76.3	68.5	74.8	82.7	71.0
APL	CVPR'25	DINOv2	75.1	79.1	73.2	73.4	87.6	66.7	68.8	74.1	66.6	72.4	80.3	68.8
AFGCD	ICCV'25	DINOv2	76.5	77.3	76.0	75.5	86.2	70.3	68.1	75.9	64.1	73.4	79.8	70.1
DebGCD	ICLR'25	DINOv2	77.5	80.8	75.8	75.4	87.7	69.5	71.9	76.0	69.8	74.9	81.5	71.7
MOS	CVPR'25	DINOv2	81.1	82.1	80.6	75.5	89.6	68.7	69.7	78.1	65.5	75.4	83.3	71.6
PartGCD	TMM	DINOv2	77.6	80.6	76.1	78.2	88.7	73.1	71.1	75.6	68.8	75.6	81.6	72.7
SEAL	NeurIPS'25	DINOv2	76.7	78.3	75.9	77.7	88.7	72.4	74.6	73.2	75.3	76.3	80.1	74.5
RLCD	ICML'25	DINOv2	78.7	79.5	78.3	79.5	91.8	73.5	72.6	77.3	70.3	76.9	82.9	74.0
AllGCD	ICCV'25	DINOv2	78.4	82.8	76.2	76.2	88.3	70.4	-	-	-	-	-	-
Hyp-SimGCD	CVPR'25	DINOv2	77.6	77.9	77.4	82.5	85.8	81.0	76.4	70.3	79.4	78.8	78.0	79.3
ConGCD-SelEx	ICCV'25	DINOv2	86.3	87.4	85.8	79.8	93.1	73.3	81.7	83.3	81.0	82.6	87.9	80.0
NCGCD-SelEx	NeurIPS'25	DINOv2	87.9	85.3	89.2	81.1	90.9	79.3	83.1	88.3	82.5	84.0	88.2	83.7
Hyp-SelEx	CVPR'25	DINOv2	90.7	85.3	93.4	83.8	93.3	79.2	83.4	82.0	84.1	86.0	86.9	85.5
SimGCD	ICCV'23	DINOv2	71.5	78.1	68.3	71.5	81.9	66.6	63.9	69.9	60.9	69.0	76.6	65.3
PartCo-SimGCD	**Ours**	DINOv2	**81.1**	**82.4**	**80.5**	**78.9**	**91.5**	**72.8**	**76.5**	**80.9**	**74.4**	**78.8**	**84.9**	**75.9**
SPTNet	ICLR'24	DINOv2	76.3	79.5	74.6	72.3	82.0	67.5	68.0	75.2	60.9	72.2	78.9	67.6
PartCo-SPTNet	**Ours**	DINOv2	**82.6**	**82.3**	**81.8**	**80.1**	**92.0**	**73.5**	**78.6**	**77.6**	**79.0**	**80.4**	**84.0**	**78.1**
FlipClass	NeurIPS'24	DINOv2	79.3	80.7	78.5	78.0	88.0	73.2	71.1	75.1	69.1	76.1	81.3	73.6
PartCo-FlipClass	**Ours**	DINOv2	**85.2**	**86.3**	**84.7**	**80.5**	**92.7**	**74.6**	**77.1**	**78.5**	**76.4**	**80.9**	**85.8**	**78.6**
SelEx	ECCV'24	DINOv2	87.4	**85.1**	88.5	82.2	**93.7**	76.7	79.8	82.3	78.6	83.1	**87.0**	81.3
PartCo-SelEx	**Ours**	DINOv2	**90.6**	84.5	**93.2**	**82.5**	91.8	**78.0**	**83.4**	**83.6**	**83.3**	**85.5**	86.6	**84.8**
k-means	-	DINOv3	69.8	64.7	72.3	59.0	50.8	63.1	40.3	35.3	42.8	56.4	50.3	59.4
SimGCD	ICCV'23	DINOv3	75.9	**83.8**	72.0	73.9	79.4	71.3	71.4	**84.4**	65.0	73.7	82.5	69.4
PartCo-SimGCD	**Ours**	DINOv3	**78.8**	83.4	**76.6**	**78.5**	**86.5**	**74.6**	**75.3**	81.6	**72.2**	**77.5**	**83.8**	**74.5**
SelEx	ECCV'24	DINOv3	83.5	78.1	86.2	78.8	90.3	73.3	73.9	77.3	72.4	78.7	81.9	77.3
PartCo-SelEx	**Ours**	DINOv3	**86.1**	**84.4**	**87.0**	**81.7**	**92.1**	**76.6**	**76.9**	**82.2**	**74.2**	**81.6**	**86.2**	**79.3**

Table 2. Comparison of GCD methods on the generic benchmark. Results are reported in *ACC* across the 'All', 'Old' and 'New' categories.

			CIFAR10			CIFAR100			ImageNet-100			Average		
Method	Venue	Backbone	All	Old	New	All	Old	New	All	Old	New	All	Old	New
k-means	-	DINOv2	94.9	95.2	94.8	70.9	70.8	72.1	78.3	80.5	77.2	81.4	82.2	81.4
GCD	CVPR'22	DINOv2	97.8	99.0	97.1	79.6	84.5	69.9	78.5	89.5	73.0	85.3	91.0	80.0
AMEND	WACV'24	DINOv2	97.7	96.6	98.3	83.5	83.0	84.5	87.3	95.1	83.4	89.5	91.6	88.7
CiPR	TMLR	DINOv2	99.0	98.7	99.2	90.3	89.0	93.1	88.2	87.6	88.5	92.5	91.8	93.6
SPTNet	ICLR'24	DINOv2	98.9	99.1	98.8	89.0	91.5	79.2	90.1	96.1	87.1	92.7	95.6	88.4
FlipClass	NeurIPS'24	DINOv2	99.0	98.2	99.4	91.7	90.4	94.2	91.0	96.3	88.3	93.9	95.0	94.0
DebGCD	ICLR'25	DINOv2	98.9	97.5	99.6	90.1	90.9	88.6	93.2	97.0	91.2	94.1	95.1	93.1
SEAL	NeurIPS'25	DINOv2	98.9	98.1	99.3	89.8	90.4	89.5	91.3	93.3	90.3	93.3	93.9	93.0
RLCD	ICML'25	DINOv2	99.0	98.9	99.1	91.2	91.2	91.2	92.1	96.2	90.0	94.1	95.4	93.4
Hyp-SimGCD	CVPR'25	DINOv2	98.9	97.7	99.5	91.5	90.0	94.6	91.9	96.2	89.8	94.1	94.6	94.6
Hyp-SelEx	CVPR'25	DINOv2	98.6	98.1	98.9	88.6	91.5	82.8	92.3	96.4	90.2	93.2	95.3	90.6
SimGCD	ICCV'23	DINOv2	98.7	96.7	**99.7**	88.5	89.2	**87.2**	89.9	**95.5**	87.1	92.4	93.8	**91.3**
PartCo-SimGCD	**Ours**	DINOv2	**99.0**	**98.7**	99.2	**89.0**	**92.0**	83.0	**90.1**	92.0	**89.2**	**92.7**	**94.2**	90.4
SelEx	ECCV'24	DINOv2	98.5	98.8	98.5	87.7	90.8	81.5	90.9	96.2	88.3	92.4	95.3	89.4
PartCo-SelEx	**Ours**	DINOv2	**99.2**	**99.4**	**98.9**	**90.0**	**92.8**	**84.3**	**94.5**	**97.8**	**92.8**	**94.6**	**96.7**	**92.0**
k-means	-	DINOv3	94.1	95.1	93.6	65.5	66.4	63.5	78.3	78.3	78.3	79.3	79.9	78.5
SimGCD	ICCV'23	DINOv3	98.4	**98.7**	98.2	84.3	88.3	74.3	92.2	96.5	90.0	91.6	**94.8**	87.5
PartCo-SimGCD	**Ours**	DINOv3	**98.9**	98.2	**99.3**	**85.0**	**88.9**	**77.1**	**93.7**	**96.5**	**92.3**	**92.5**	94.5	**89.6**
SelEx	ECCV'24	DINOv3	98.2	99.0	97.7	87.7	88.7	85.7	93.4	**96.9**	91.6	93.1	94.9	91.6
PartCo-SelEx	**Ours**	DINOv3	**98.3**	**99.1**	**97.9**	**90.2**	**91.7**	**87.0**	**93.8**	96.6	**92.4**	**94.1**	**95.8**	**92.4**

Experiments

We integrate our PartCo framework with several GCD baselines: SimGCD, SPTNet, FlipClass and SOTA GCD baseline: SelEx, employing DINO-variants backbones.

Overall, PartCo integration consistently boost GCD methods, yielding new SOTA results across both fine-grained and generic datasets.