模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

Generalized Category Discovery

Sagar Vaze* Kai Han† Andrea Vedaldi* Andrew Zisserman*

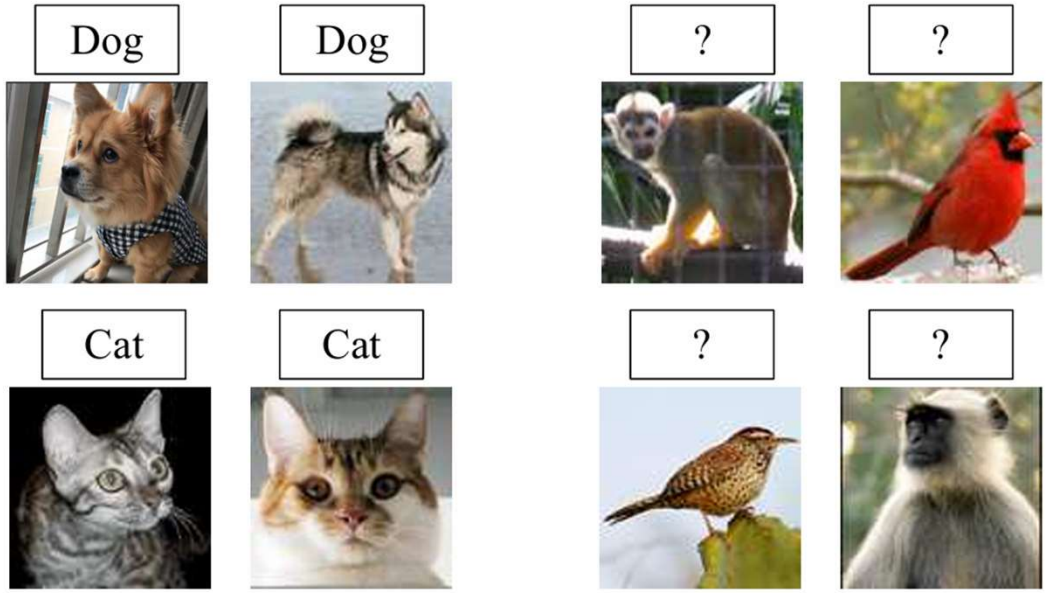
*Visual Geometry Group, Department of Engineering Science, University of Oxford

†The University of Hong Kong

{sagar,vedaldi,az}@robots.ox.ac.uk kaihanx@hku.hk

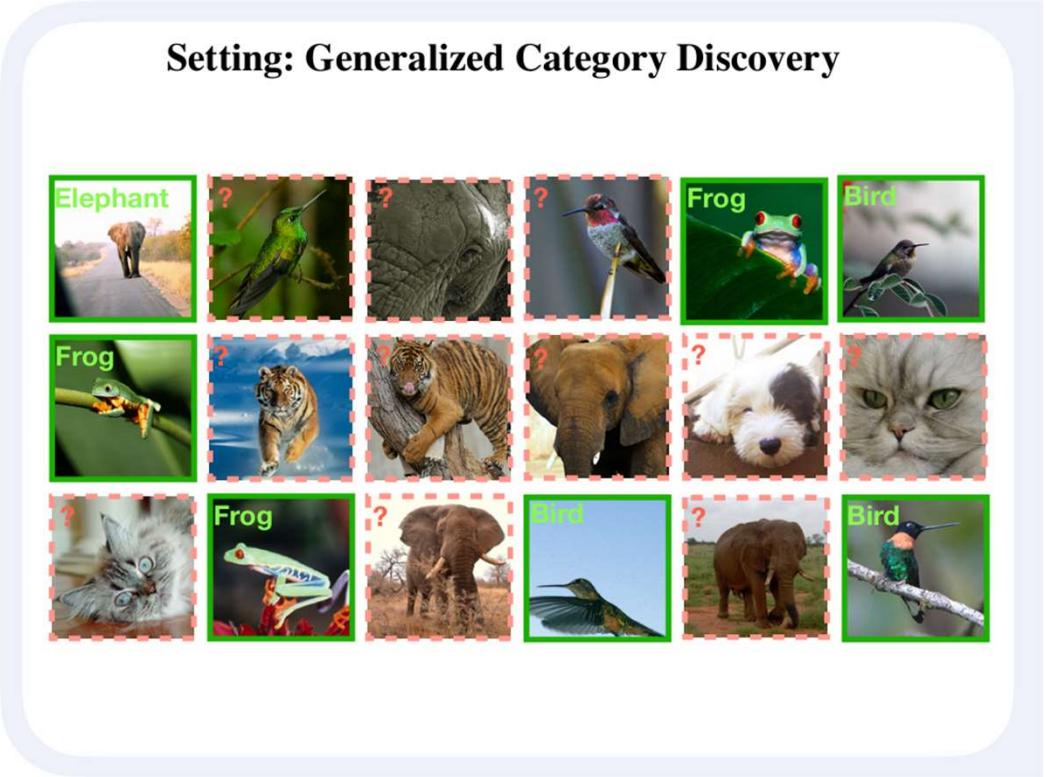
CVPR 2022

Novel category discovery (NCD)



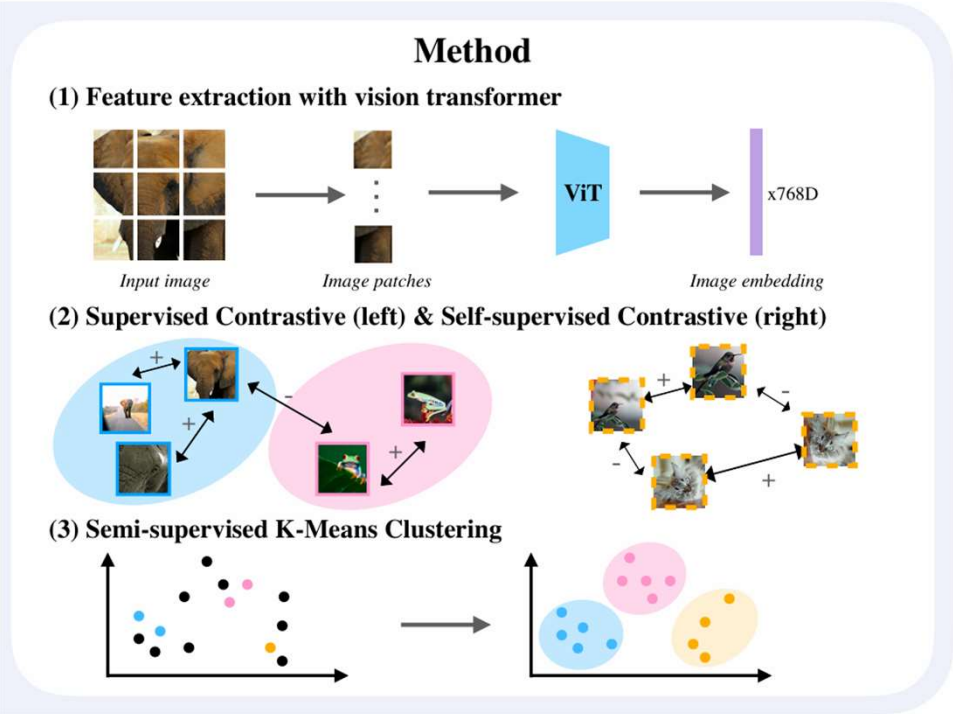
In Novel Category Discovery (NCD), methods learn from labelled and unlabelled images to discover new categories in the unlabelled set, but they assume that all unlabelled images come from new categories, which is often unrealistic.

Generalized Category Discovery (GCD)



The Generalized Category Discovery (GCD) problem refers to the task of assigning category labels to unlabelled images in an image dataset, where some images are already labelled, and the unlabelled images may belong to new categories not observed in the labelled set.

Method



The approach in this work adapts to image recognition in an open-world setting by removing the need for parametric classification heads and performing clustering directly in the feature space of a deep network. In real-world applications, models are typically initialized with large-scale pretrained weights (such as ImageNet pretraining) to optimize performance. However, to avoid conflicts with the experimental setup, which assumes a finite labelled set, self-supervised ImageNet weights are used, and the representation is fine-tuned on the target data, combining both supervised and unsupervised contrastive learning methods.

Method

Non-Parametric Clustering: They modify the classic k-means algorithm by adding constraints that force instances from the same class in the labelled dataset ($\mathcal{DL}$) to be assigned to the same cluster, while unlabelled instances (from $\mathcal{DU}$) can be freely assigned to any cluster.

Estimating the Number of Classes: The authors address the challenge of estimating the number of classes in unlabelled data. They perform k-means clustering on the entire dataset ($\mathcal{D}$), then evaluate clustering accuracy on the labelled subset ($\mathcal{DL}$). The number of clusters, k, is treated as a parameter that is optimized using the clustering accuracy as a "black box" scoring function.

Brent's Algorithm: Instead of exhaustively testing all possible values of k, they optimize the number of clusters using Brent’s algorithm, which allows the method to scale well with datasets containing many categories.

Experiment

Table 1. Datasets used in our experiments. We show the number of classes in the labelled and unlabelled sets ($|\mathcal{Y}_{\mathcal{L}}|, |\mathcal{Y}_{\mathcal{U}}|$), as well as the number of images ($|\mathcal{D}_{\mathcal{L}}|, |\mathcal{D}_{\mathcal{U}}|$).

	CIFAR10	CIFAR100	ImageNet-100	CUB	SCars	Herb19
$\lvert\mathcal{Y}_{\mathcal{L}}\rvert$	5	80	50	100	98	341
$\lvert\mathcal{Y}_{\mathcal{U}}\rvert$	10	100	100	200	196	683
$\lvert\mathcal{D}_{\mathcal{L}}\rvert$	12.5k	20k	31.9k	1.5k	2.0k	8.9k
$\lvert\mathcal{D}_{\mathcal{U}}\rvert$	37.5k	30k	95.3k	4.5k	6.1k	25.4k

Experiment

Table 2. Results on generic image recognition datasets.

	CIFAR10			CIFAR100			ImageNet-100		
Classes	All	Old	New	All	Old	New	All	Old	New
k-means [30]	83.6	85.7	82.5	52.0	52.2	50.8	72.7	75.5	**71.3**
RankStats+	46.8	19.2	60.5	58.2	77.6	19.3	37.1	61.6	24.8
UNO+	68.6	**98.3**	53.8	69.5	**80.6**	47.2	70.3	**95.0**	57.9
Ours	**91.5**	97.9	**88.2**	**73.0**	76.2	**66.5**	**74.1**	89.8	66.3

Table 3. Results on SSB [45] and Herbarium19 [42].

	CUB			Stanford Cars			Herbarium19		
Classes	All	Old	New	All	Old	New	All	Old	New
k-means [30]	34.3	38.9	32.1	12.8	10.6	13.8	12.9	12.9	12.8
RankStats+	33.3	51.6	24.2	28.3	61.8	12.1	27.9	**55.8**	12.8
UNO+	35.1	49.0	28.1	35.5	**70.5**	18.6	28.3	53.7	14.7
Ours	**51.3**	**56.6**	**48.7**	**39.0**	57.6	**29.9**	**35.4**	51.0	**27.0**

Table 4. Estimation of the number of classes in unlabelled data.

	CIFAR10	CIFAR100	ImageNet-100	CUB	SCars	Herb19
Ground truth	10	100	100	200	196	683
Ours	9	100	109	231	230	520
Error	10%	0%	9%	16%	15%	28%

Experiment

Table 5. Ablation study on the different components of our approach.

	ViT Backbone	Contrastive Loss	Sup. Contrastive Loss	Semi-Sup k-means	CIFAR100			Herbarium19		
					All	Old	New	All	Old	New
(1)	✗	✗	✗	✗	34.0	34.8	32.4	12.1	12.5	11.9
(2)	✓	✗	✗	✗	52.0	52.2	50.8	12.9	12.9	12.8
(3)	✓	✓	✗	✗	54.6	54.1	53.7	14.3	15.1	13.9
(4)	✓	✗	✓	✗	60.5	72.2	35.0	17.8	22.7	15.4
(5)	✓	✓	✓	✗	71.1	**78.3**	56.6	28.7	32.1	26.9
(6)	✓	✓	✓	✓	**73.0**	76.2	**66.5**	**35.4**	**51.0**	**27.0**

模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

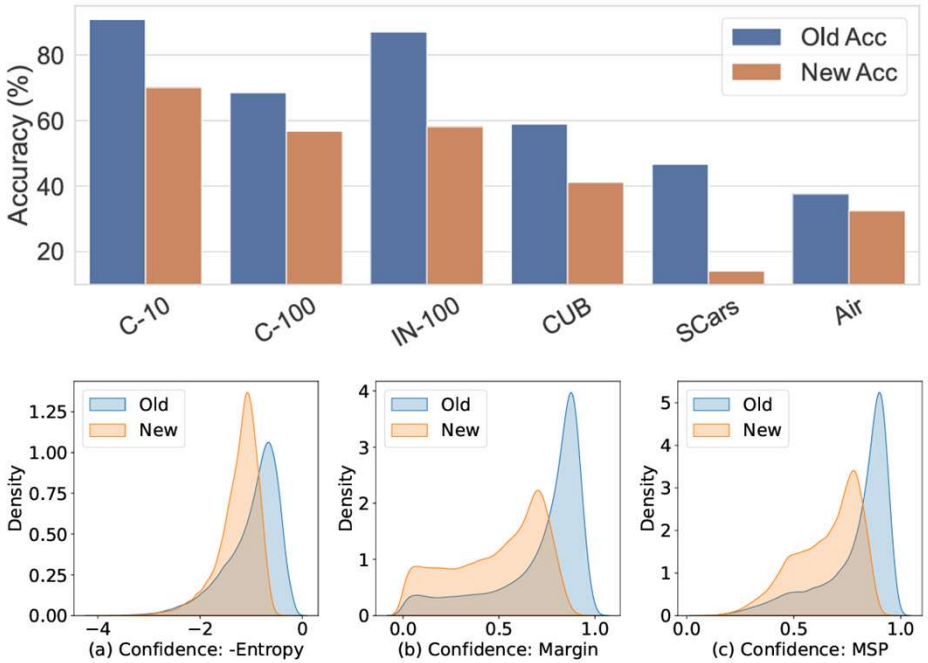
ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

Active Generalized Category Discovery

Shijie Ma[1,2], Fei Zhu[3], Zhun Zhong[4,5], Xu-Yao Zhang[1,2*], Cheng-Lin Liu[1,2]
[1]MAIS, Institute of Automation, Chinese Academy of Sciences, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, China
[3]Centre for Artificial Intelligence and Robotics, HKISI-CAS, China
[4]School of Computer Science and Information Engineering, Hefei University of Technology, China
[5]School of Computer Science, University of Nottingham, NG8 1BB Nottingham, UK
{mashijie2021, zhufei2018}@ia.ac.cn, zhunzhong007@gmail.com, {xyz, liucl}@nlpr.ia.ac.cn

CVPR 2024

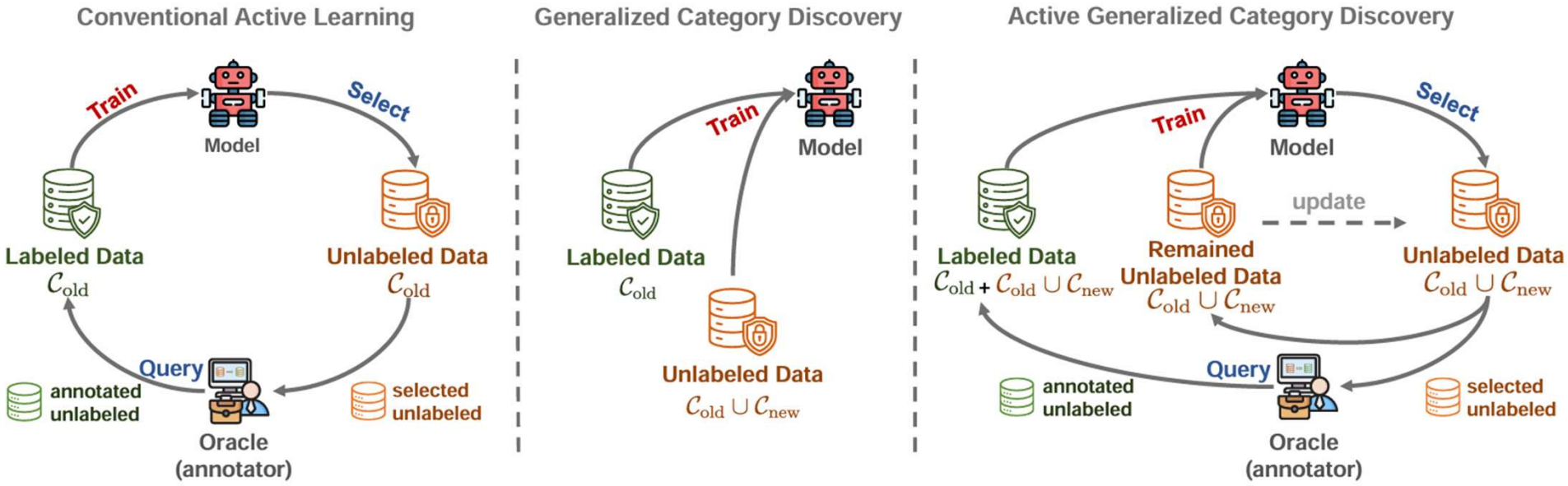
Background



GCD still faces intractable problems, including imbalanced accuracy and inconsistent confidence between old and new classes, especially in low-labeling regimes. In essence, these issues arise from the nature of the GCD task itself.

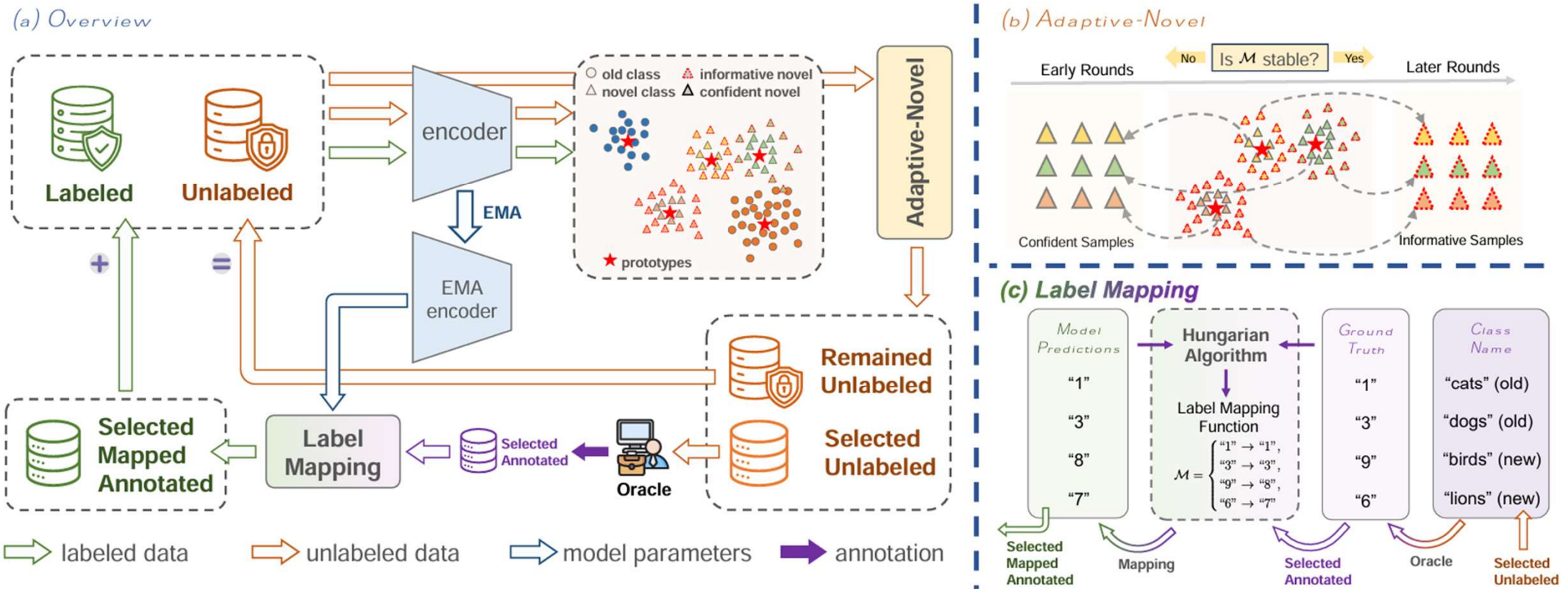
Can deep learning models actively select a small number of unlabeled samples for labeling to remarkably enhance category discovery?

Background

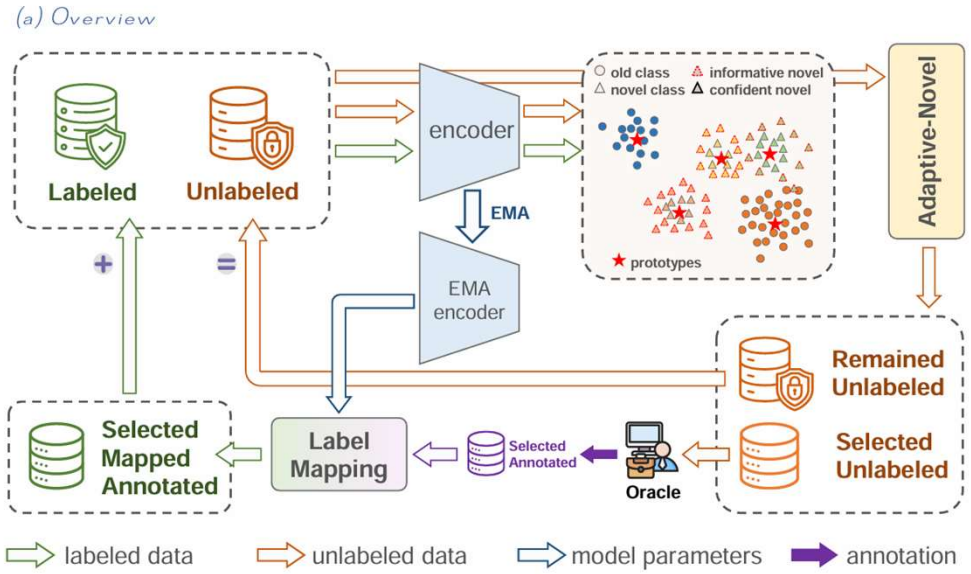


- Conventional AL methods do not take novel categories into consideration, which makes them not applicable to AGCD and leads to sub-optimal results. → Adaptive-Novel strategy
- Considering the clustering nature of GCD, the queried ground truth labels could not be directly used by parametric classifiers due to the different ordering of indices. → perform label mapping on the queried samples

Method


(a) Overview
Labeled
Unlabeled
encoder
EMA
EMA encoder
old class
novel class
informative novel
confident novel
prototypes
Adaptive-Novel
Remained Unlabeled
Selected Unlabeled
Oracle
Selected Annotated
Label Mapping
Selected Mapped Annotated
labeled data
unlabeled data
model parameters
annotation
(b) Adaptive-Novel
Early Rounds
No
Is M stable?
Yes
Later Rounds
Confident Samples
Informative Samples
(c) Label Mapping
Model Predictions
"1"
"3"
"8"
"7"
Hungarian Algorithm
Label Mapping Function
M = {"1" → "1", "3" → "3", "9" → "8", "6" → "7"}
Ground Truth
"1"
"3"
"9"
"6"
Class Name
"cats" (old)
"dogs" (old)
"birds" (new)
"lions" (new)
Selected Mapped Annotated
Mapping
Selected Annotated
Oracle
Selected Unlabeled

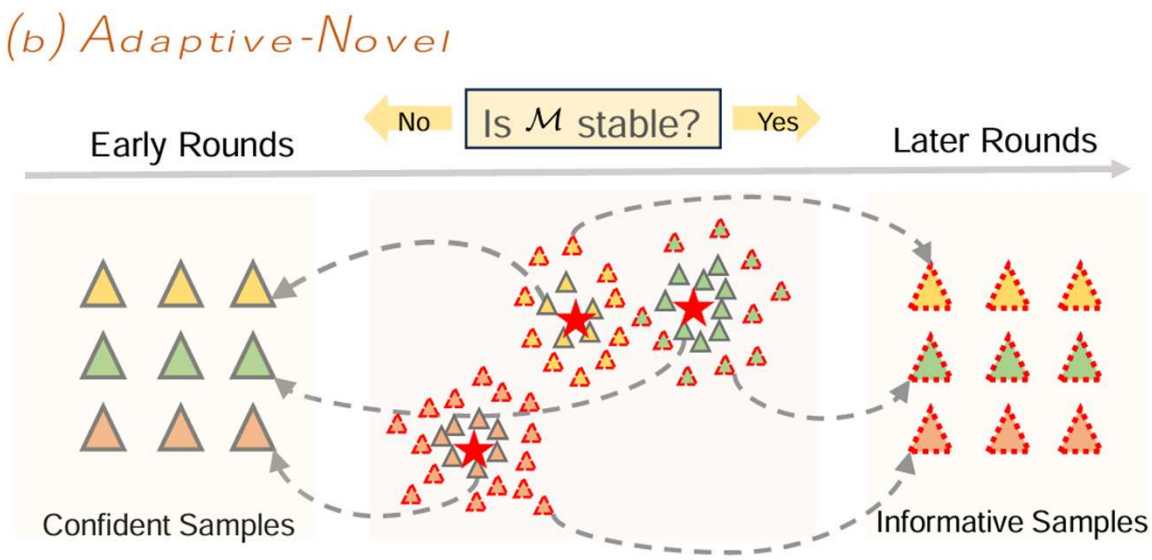
Method



$$ACC = \max_{p \in \mathcal{P}(\mathcal{Y}_u)} \frac{1}{M} \sum_{i=1}^{M} \mathbb{1}(y_i = p(\hat{y}_i))$$

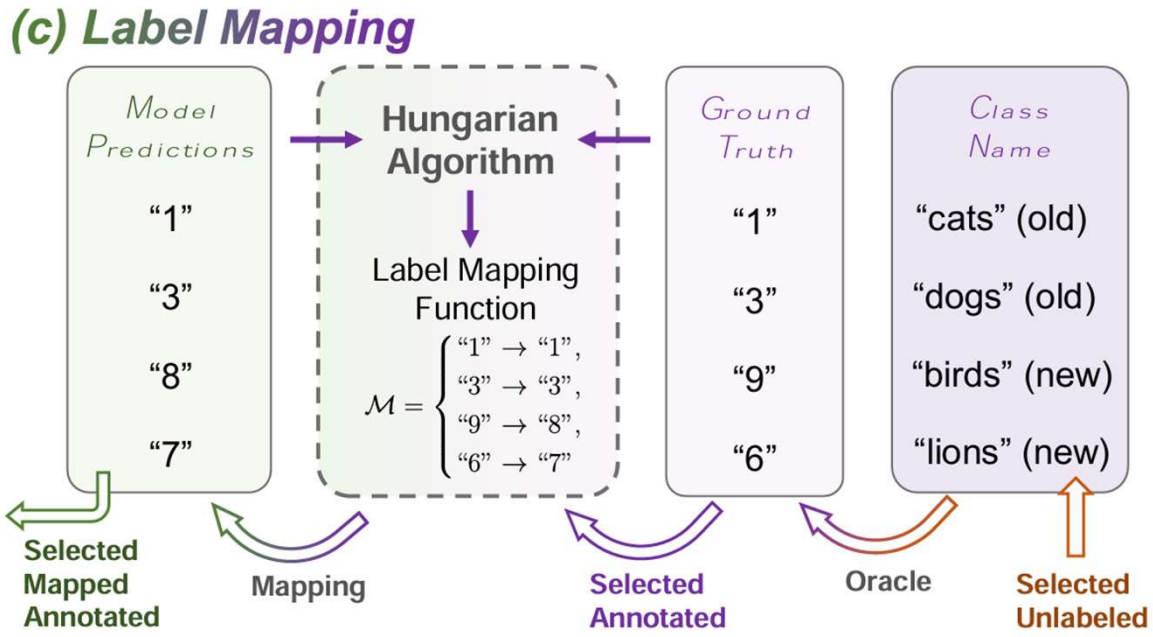
The model is initially trained on labeled data D_l^0 and unlabeled data D_u^0 using SimGCD, with the initial split similar to GCD .After this, AGCD runs for multiple rounds. In round t, the model selects a batch of b samples from D_u^{t-1} , queries their labels, and updates the datasets as $D_l^t = D_l^{t-1} \cup D_q^t$ and $D_u^t = D_u^{t-1} \backslash D_q^t$. The model is then trained on $D_l^t \cup D_u^t$ using SimGCD. Initially, D_l^0 contains only old classes, but after querying, D_l^t may include new classes. The total budget is $b \times n$. For queried data, the model is trained with losses $\mathcal{L}_{con}^l$ and $\mathcal{L}_{cls}^l$.

Method



(1) For the aspect of novelty, as the initial labeling condition is severely imbalanced, we should give priority to selecting samples from new classes. Models' predictions $\widehat{y_i}$ are proxies of samples' novelty.
(2) For the aspect of diversity, we uniformly select samples from novel classes, *i.e.*, at each round, we select $\lfloor b/K_{new} \rfloor$ samples in each new class based on the model's prediction.
(3) For the aspect of informativeness, we choose Margin as the uncertainty metric.

Method



$$\mathcal{M}^t = \underset{m \in \mathcal{P}(\mathcal{C}_{all})}{\arg\max} \frac{1}{|\mathcal{D}_l^t|} \sum_{i \in \mathcal{D}_l^t} \mathbb{1}(m(y_i) = \hat{y}_i^{ema})$$

Experiment

Table 2. Comparative results of various methods with 5 rounds of active category discovery on generic datasets. Our method outperforms several uncertainty-based (Unc.) and representative/diversity-based (Rep./Div.) methods. Mean results over three runs are reported.

Type	AL Strategies	CIFAR10			CIFAR100			ImageNet-100		
		All	Old	New	All	Old	New	All	Old	New
Baseline	w/o AGCD	74.22	90.80	70.07	62.62	68.46	56.78	72.56	87.00	58.12
	Random	82.74	93.05	80.16	67.28	74.52	60.04	79.16	89.40	68.92
Unc.	Entropy [53]	76.25	95.55	71.43	64.59	73.94	55.24	75.96	91.04	60.88
	LeastConf [53]	78.32	**96.00**	73.90	65.63	**76.74**	54.52	76.82	91.92	61.72
	Margin [40]	92.34	94.35	91.84	69.08	75.58	62.58	80.46	92.40	68.52
Rep./Div.	KMeans [32]	91.18	93.10	90.70	66.70	72.66	60.74	78.18	90.08	66.28
	CoreSet [42]	85.51	94.95	83.15	65.72	77.64	53.80	78.08	91.92	64.24
	BADGE [3]	92.31	94.75	91.70	67.22	73.70	60.74	81.48	**92.68**	70.28
Ours	Adaptive-Novel	**93.15**	94.55	**92.80**	**71.25**	75.72	**66.78**	**83.34**	90.20	**76.48**

Experiment

Table 3. Comparative results of various methods with 5 rounds of active category discovery on fine-grained datasets. Our method outperforms several uncertainty-based (Unc.) and representative/diversity-based (Rep./Div.) methods. Mean results over three runs are reported.

Type	Query Strategies	CUB			Stanford Cars			FGVC-Aircraft		
		All	Old	New	All	Old	New	All	Old	New
Baseline	w/o AGCD	50.17	58.95	41.18	30.12	46.71	14.12	35.01	37.53	32.49
	Random	62.74	64.88	60.62	44.12	53.44	35.13	50.41	51.38	49.43
Unc.	Entropy [53]	62.82	**69.52**	56.19	42.40	53.75	31.44	43.89	51.92	35.86
	LeastConf [53]	61.48	66.12	56.87	45.82	55.32	36.65	44.91	50.42	39.40
	Margin [40]	65.08	68.41	61.79	46.03	57.67	34.79	51.37	**52.46**	50.27
Rep./Div.	KMeans [32]	61.30	68.27	54.40	40.79	52.99	29.03	51.58	51.08	52.07
	CoreSet [42]	63.44	65.95	60.96	42.52	52.00	33.37	45.03	51.68	38.38
	BADGE [3]	65.84	69.00	62.71	45.82	54.41	37.53	52.03	51.68	52.37
Ours	Adaptive-Novel	**66.62**	66.54	**66.70**	**48.36**	**57.73**	**39.34**	**53.74**	51.50	**55.98**

Experiment

Table 4. Novelty metrics of all the selected data over 5 rounds on CIFAR100 and Stanford Cars.

AL Strategies	CIFAR100				Stanford Cars			
	Nov-C	Nov-R	Nov-U	Nov-I	Nov-C	Nov-R	Nov-U	Nov-I
Random	**1.00**	0.52	0.97	0.50	0.93	0.57	0.96	0.55
Entropy	0.90	0.44	0.91	0.40	0.85	0.64	0.92	0.59
Margin	0.96	0.63	0.95	0.60	0.90	0.66	0.93	0.61
CoreSet	0.96	0.61	0.94	0.57	0.89	**0.69**	0.94	0.65
BADGE	**1.00**	0.63	**0.98**	0.62	0.95	0.64	**0.97**	0.62
Ours	**1.00**	**0.71**	**0.98**	**0.70**	**0.98**	**0.69**	**0.97**	**0.67**

$$\text{Nov-C} = |\mathcal{C}_{new,select}|/K_{new}$$

$$\text{Nov-R} = \sum_{i=1}^{N_{select}} \frac{\mathbb{1}(y_i \in \mathcal{C}_{new})}{N_{select}}$$

$$\text{Nov-U} = -\sum_{c=1}^{K_{new}} \frac{N_{new,i}}{N_{select}} \log \frac{N_{new,i}}{N_{select}} / \log K_{new}$$

$$\text{Nov-I} = \text{Nov-R} \times \text{Nov-U}$$

Experiment

Table 5. Ablations on three key factors, *i.e.*, novelty, informativeness and diversity for sample selection in AGCD.

ID	Novelty	Informativeness	Diversity	CIFAR100			CUB		
				All	Old	New	All	Old	New
(a)	✗	✗	✗	67.28	74.52	60.04	62.74	64.88	60.62
(b)	✓	✗	✗	69.33	72.76	65.90	63.58	63.31	63.85
(c)	✓	✓	✗	69.65	75.56	63.74	64.28	65.33	63.23
(d)	✓	✓	✓	**71.25**	**75.72**	**66.78**	**66.62**	**66.54**	**66.70**

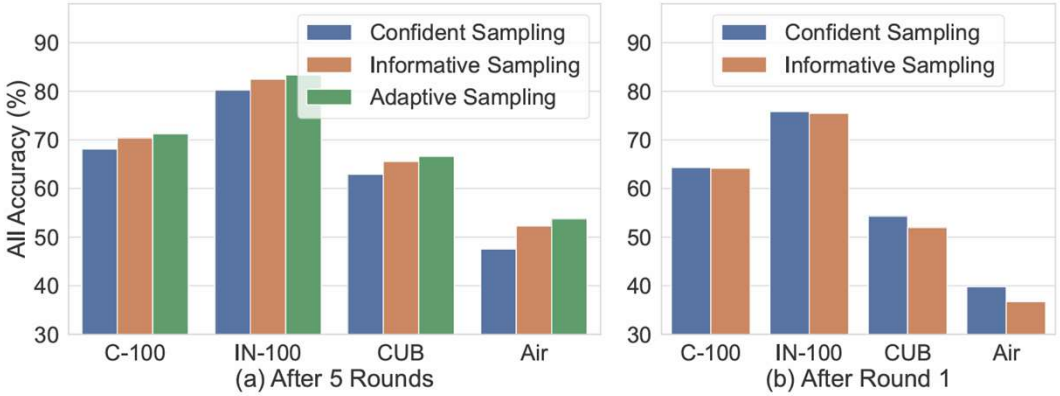


Figure 6. Ablation on adaptive sampling.

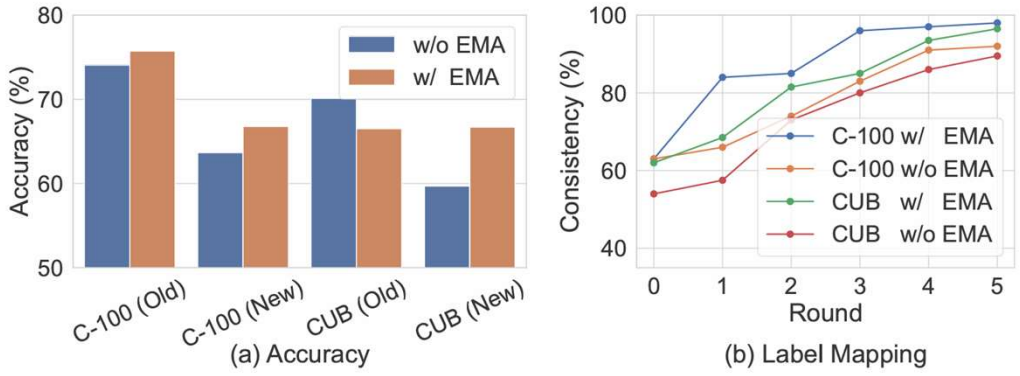


Figure 7. Ablation on model EMA.

Experiment

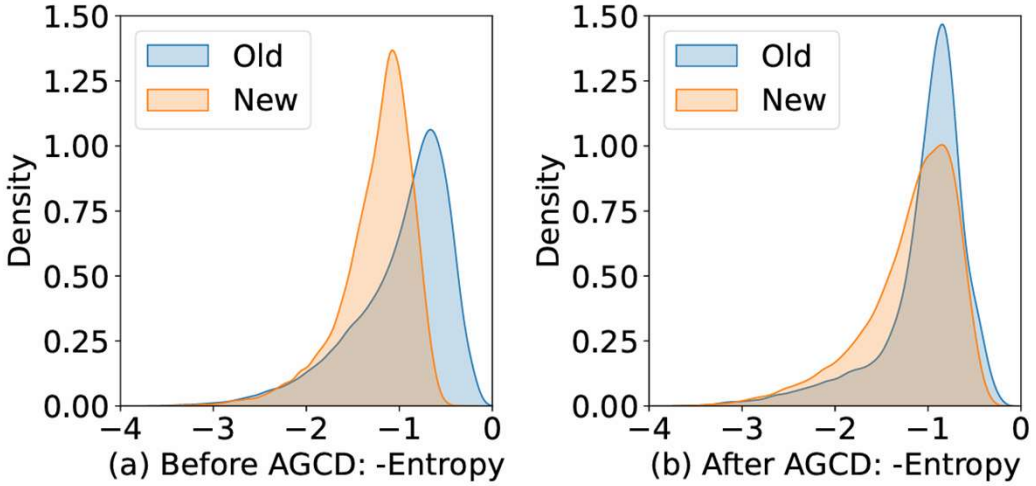


Figure 8. Confidence distribution before and after AGCD.

Table 6. Results of All Acc on CUB of various initial label ratios.

label ratio	0	0.01	0.05	0.1	0.2	0.3
w/o AGCD	14.15	18.12	27.68	37.49	50.17	58.49
Random	31.46	32.00	45.88	53.16	62.74	66.45
Entropy	32.02	36.95	46.32	53.78	62.82	60.55
Ours	**33.36**	**38.73**	**46.88**	**55.30**	**66.62**	**69.47**

Table 7. Results on CUB with various budge sizes per round b. Models are trained with 3 AGCD rounds.

b	30	50	100	300	500
Random	52.92	56.44	58.54	66.78	72.68
Entropy	52.90	54.47	58.87	68.69	70.87
Ours	**53.40**	**56.50**	**59.86**	**69.59**	**73.56**

Thanks