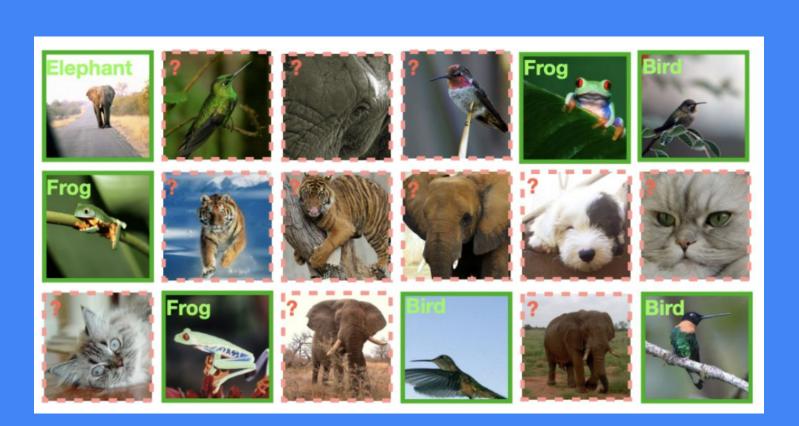
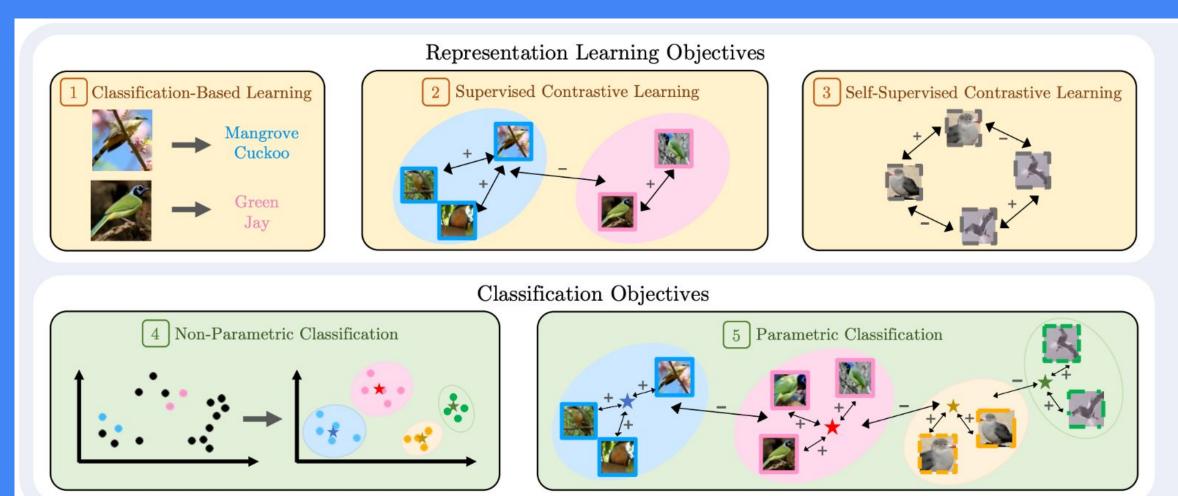
# Parametric Classification for Generalized Category Discovery: A Baseline Study by Xin Wen\* (HKU), Bingchen Zhao\* (UoE), and Xiaojuan Qi (HKU) ICCV23

## 1. Setting and Background

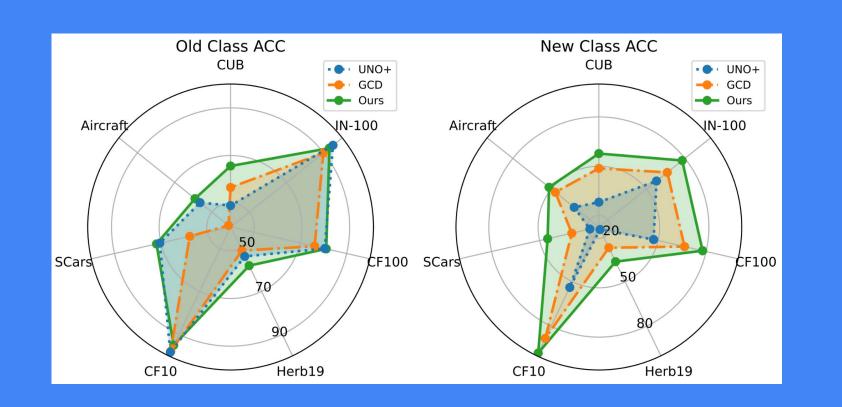
Generalized Category Discovery aims to **recognise novel** categories from unlabelled data using knowledge learned from labelled samples.



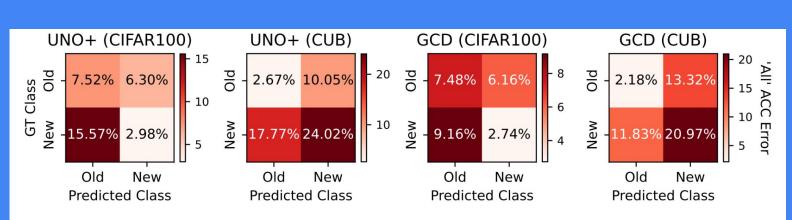
Overview of current works: SOTA is still semi-supervised k-means. And we target on **parametric classification**.



Prior parametric SOTA (UNO+) suffers from over-fitting to seen ('Old') categories. But why? (GOTO next col.)



#### 3) So what's wrong with UNO+'s pseudo labels? The devil is in the biased predictions.



Prediction bias between 'Old'/'New' classes. We simplify the setting to binary classification and categorise the errors in 'All' ACC into four types. Both works, especially UNO+, are prone to make "False Old" predictions. In other words, the predictions are biased towards 'Old' classes, and many samples corresponding to 'New' classes are misclassified as an 'Old' class.

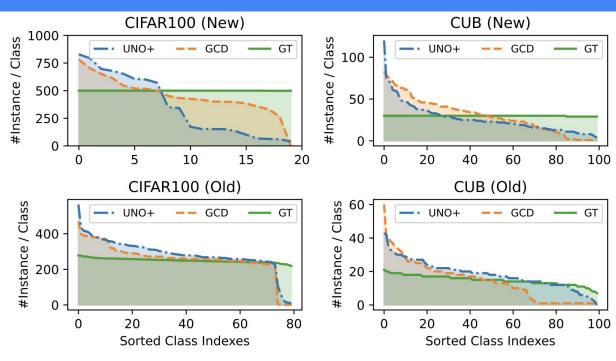
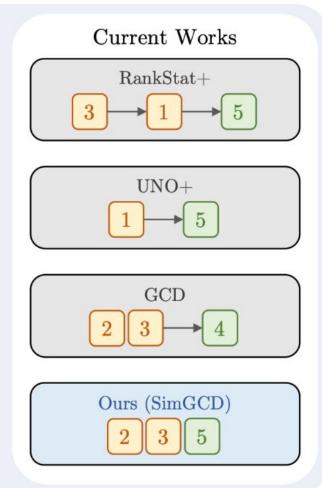
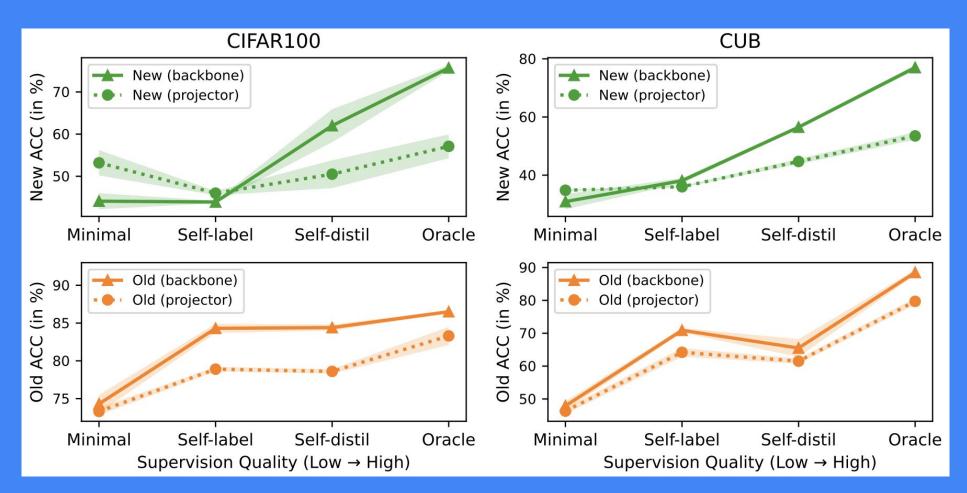


Figure 6. Prediction bias across 'Old'/'New' classes. We show the per-class prediction distributions. Both works, especially UNO+, are prone to make long-tailed predictions. In other words, across all classes, the predictions are unexpectedly long-tailed and biased towards the head classes.

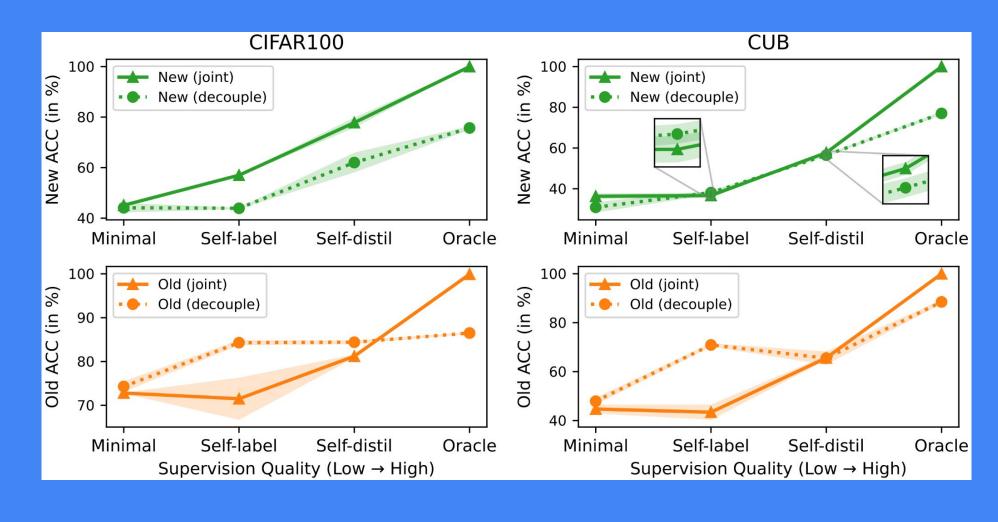


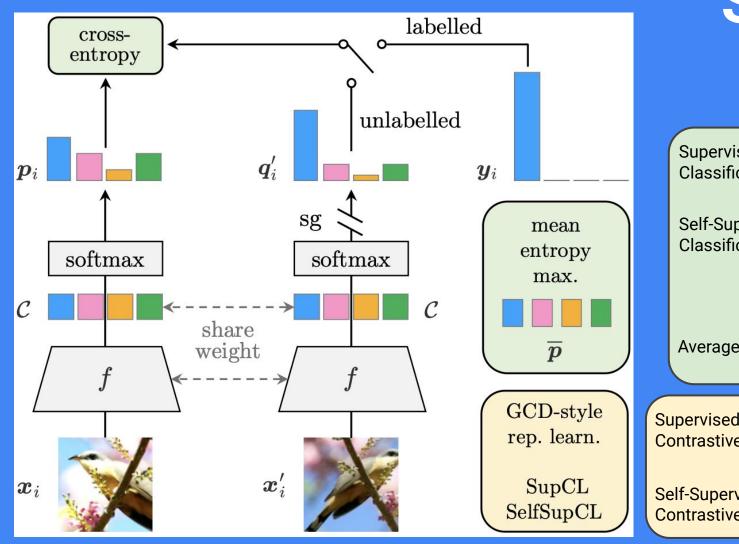
### 2. On the Failure of Parametric Classification

1) Which feature space to build your classifier? The *post-backbone* representations consistently benefit classification performance.



### 2) Decoupled or joint representation learning? Guiding rep. learning with cls. objective can be helpful, only when high-quality sup. is available.







# 香港大學 THE UNIVERSITY OF HONG KONG

### **3. A Frustratingly** Simple Yet Strong Baseline

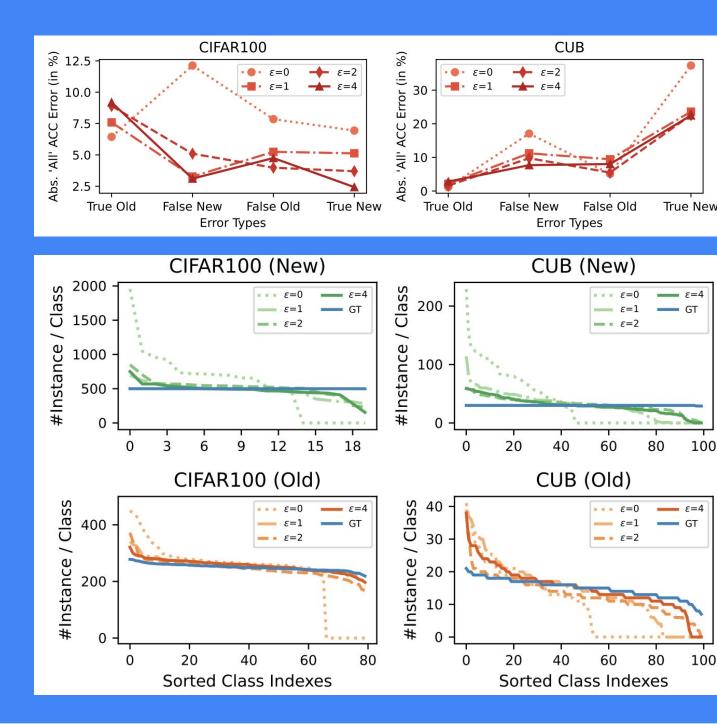
vised fication Objective:	$\mathcal{L}_{\mathrm{cls}}^{s} = \frac{1}{ B^{l} } \sum_{i \in B^{l}} \ell(\boldsymbol{y}_{i}, \boldsymbol{p}_{i}) \qquad \text{Entropy}$
Ipervised fication Objective:	$\mathcal{L}_{ ext{cls}}^u = rac{1}{ B } \sum_{i \in B} \ell(oldsymbol{q}_i',oldsymbol{p}_i) {-}arepsilon H(oldsymbol{ar{p}})$
e Predictions:	$\overline{oldsymbol{p}} = rac{1}{2 B } \sum_{i \in B} \left(oldsymbol{p}_i + oldsymbol{p}_i' ight)$
d ve Learning: $\mathcal{L}_{rep}^s$	$= \frac{1}{ B^l } \sum_{i \in B^l} \frac{1}{ \mathcal{N}_i } \sum_{q \in \mathcal{N}_i} -\log \frac{\exp\left(\boldsymbol{z}_i^\top \boldsymbol{z}_q' / \tau_c\right)}{\sum_i^{i \neq n} \exp\left(\boldsymbol{z}_i^\top \boldsymbol{z}_n' / \tau_c\right)}$
rvised $\mathcal{L}_{rep}^{u}$	$= \frac{1}{ B } \sum_{i \in B} -\log \frac{\exp\left(\boldsymbol{z}_i^\top \boldsymbol{z}_i' / \tau_u\right)}{\sum_i^{i \neq n} \exp\left(\boldsymbol{z}_i^\top \boldsymbol{z}_n' / \tau_u\right)}$

### 4. Experiments

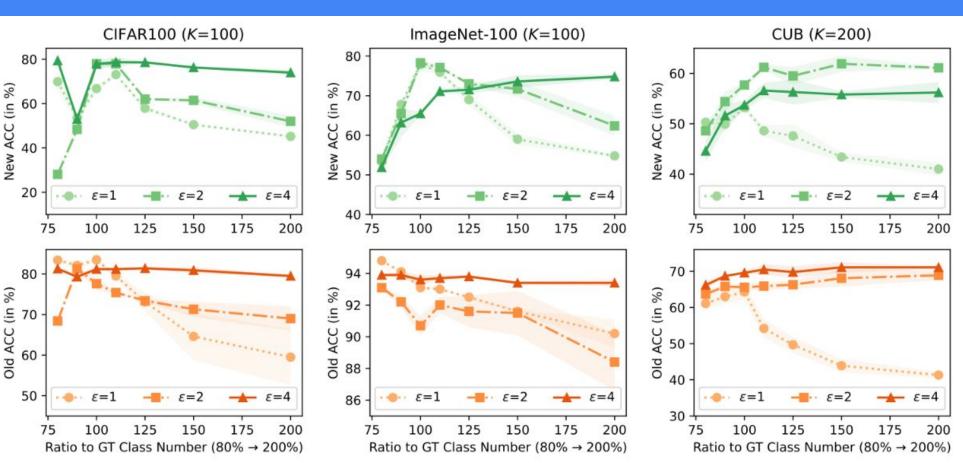
**SOTA ( 1 ~10%)** over all current GCD benchmarks that are

	CUB			Sta	anford C	FGVC-Aircra			
Methods	All	Old	New	All	Old	New	All	Old	
k-means [28] RS+ [20] UNO+ [16] ORCA [6]	34.3 33.3 35.1 35.3	38.9 51.6 49.0 45.6	32.1 24.2 28.1 30.2	12.8 28.3 35.5 23.5	10.6 61.8 70.5 50.1	13.8 12.1 18.6 10.7	16.0 26.9 40.3 22.0	14.4 36.4 56.4 31.8	
$\begin{array}{c} \text{GCD} [37] \\ \text{SimGCD} \\ \Delta \end{array}$	51.3 60.3 +9.0	56.6 65.6 +9.0	48.7 57.7 +9.0	39.0 <b>53.8</b> <b>+14.8</b>	57.6 <b>71.9</b> <b>+14.3</b>	29.9 <b>45.0</b> +15.1	45.0 <b>54.2</b> <b>+9.2</b>	41.1 <b>59.1</b> <b>+18.0</b>	

Entropy regularisation shows notable benefit in alleviating the prediction biases between/within seen/novel classes.



Entropy regularisation also enforces robustness to unknown class numbers, but over-regularisation could harm recognising 'New' classes under GT class numbers.







# THE UNIVERSITY of EDINBURGH

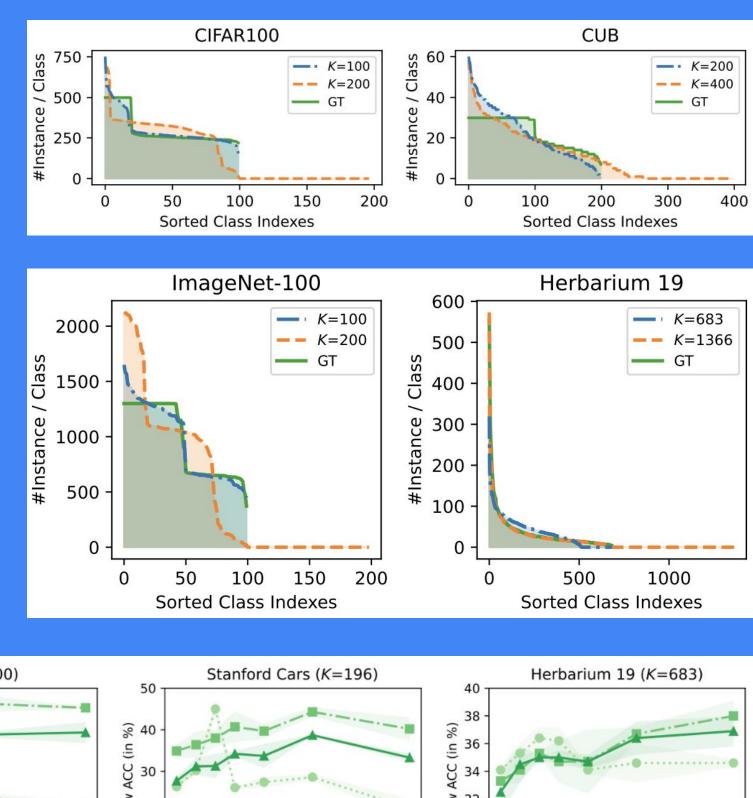


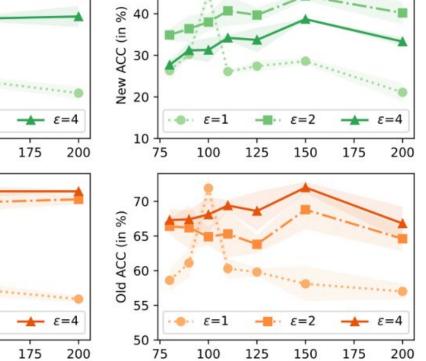
## coarse/fine-grained, balanced/long-tailed, or small/large-scale.

	CIFAR10		CIFAR100			ImageNet-100			
Methods	All	Old	New	All	Old	New	All	Old	Nev
k-means [28] RS+ [20] UNO+ [16] ORCA [6]	83.6 46.8 68.6 81.8	85.7 19.2 <b>98.3</b> 86.2	82.5 60.5 53.8 79.6	52.0 58.2 69.5 69.0	52.2 77.6 80.6 77.4	50.8 19.3 47.2 52.0	72.7 37.1 70.3 73.5	75.5 61.6 <b>95.0</b> 92.6	71. 24. 57. 63.
GCD [ <mark>37</mark> ] SimGCD Δ	91.5 <b>97.1</b> + <b>5.6</b>	97.9 95.1 <b>-2.8</b>	88.2 98.1 +9.9	73.0 <b>80.1</b> +7.1	76.2 <b>81.2</b> +5.0	66.5 77.8 +11.3	74.1 <b>83.0</b> +8.9	89.8 93.1 <b>+3.3</b>	66. 77. +11
	k-means [28] RS+ [20] UNO+ [16] ORCA [6] GCD [37] SimGCD	Methods       All         k-means [28]       83.6         RS+ [20]       46.8         UNO+ [16]       68.6         ORCA [6]       81.8         GCD [37]       91.5         SimGCD       97.1	Methods       All       Old         k-means [28]       83.6       85.7         RS+ [20]       46.8       19.2         UNO+ [16]       68.6       98.3         ORCA [6]       81.8       86.2         GCD [37]       91.5       97.9         SimGCD       97.1       95.1	Methods       All       Old       New         k-means [28]       83.6       85.7       82.5         RS+ [20]       46.8       19.2       60.5         UNO+ [16]       68.6       98.3       53.8         ORCA [6]       81.8       86.2       79.6         GCD [37]       91.5       97.9       88.2         SimGCD       97.1       95.1       98.1	MethodsAllOldNewAll $k$ -means [28]83.685.782.552.0RS+ [20]46.819.260.558.2UNO+ [16]68.6 <b>98.3</b> 53.869.5ORCA [6]81.886.279.669.0GCD [37]91.597.988.273.0SimGCD <b>97.1</b> 95.1 <b>98.180.1</b>	Methods       All       Old       New       All       Old         k-means [28]       83.6       85.7       82.5       52.0       52.2         RS+ [20]       46.8       19.2       60.5       58.2       77.6         UNO+ [16]       68.6       98.3       53.8       69.5       80.6         ORCA [6]       81.8       86.2       79.6       69.0       77.4         GCD [37]       91.5       97.9       88.2       73.0       76.2         SimGCD       97.1       95.1       98.1       80.1       81.2	Methods         All         Old         New         All         Old         New           k-means [28]         83.6         85.7         82.5         52.0         52.2         50.8           RS+ [20]         46.8         19.2         60.5         58.2         77.6         19.3           UNO+ [16]         68.6         98.3         53.8         69.5         80.6         47.2           ORCA [6]         81.8         86.2         79.6         69.0         77.4         52.0           GCD [37]         91.5         97.9         88.2         73.0         76.2         66.5           SimGCD         97.1         95.1         98.1         80.1         81.2         77.8	Methods       All       Old       New       All       Old       New       All         k-means [28]       83.6       85.7       82.5       52.0       52.2       50.8       72.7         RS+ [20]       46.8       19.2       60.5       58.2       77.6       19.3       37.1         UNO+ [16]       68.6       98.3       53.8       69.5       80.6       47.2       70.3         ORCA [6]       81.8       86.2       79.6       69.0       77.4       52.0       73.5         GCD [37]       91.5       97.9       88.2       73.0       76.2       66.5       74.1         SimGCD       97.1       95.1       98.1       80.1       81.2       77.8       83.0	Methods         All         Old         New         All         Old         New         All         Old         New         All         Old           k-means [28]         83.6         85.7         82.5         52.0         52.2         50.8         72.7         75.5           RS+ [20]         46.8         19.2         60.5         58.2         77.6         19.3         37.1         61.6           UNO+ [16]         68.6 <b>98.3</b> 53.8         69.5         80.6         47.2         70.3 <b>95.0</b> ORCA [6]         81.8         86.2         79.6         69.0         77.4         52.0         73.5         92.6           GCD [37]         91.5         97.9         88.2         73.0         76.2         66.5         74.1         89.8           SimGCD <b>97.1</b> 95.1 <b>98.1 80.1 81.2 77.8 83.0</b> 93.1

	He	erbarium	19	ImageNet-1K			
Methods	All	Old	New	All	Old	New	
<i>k</i> -means [28]	13.0	12.2	13.4	-	-	_	
RS+ [20]	27.9	55.8	12.8	-	-	-	
UNO+ [ <mark>16</mark> ]	28.3	53.7	14.7	-	-	-	
ORCA [6]	20.9	30.9	15.5	-	-	-	
GCD [37]	35.4	51.0	27.0	52.5	72.5	42.2	
SimGCD	<b>44.0</b>	58.0	36.4	57.1	77.3	46.9	
$\Delta$	+8.6	+7.0	+9.4	+4.6	+4.8	+4.7	

#### The model tends to **keep the number of active** prototypes close to the GT class number.





Ratio to GT Class Number ( $80\% \rightarrow 200\%$ )

