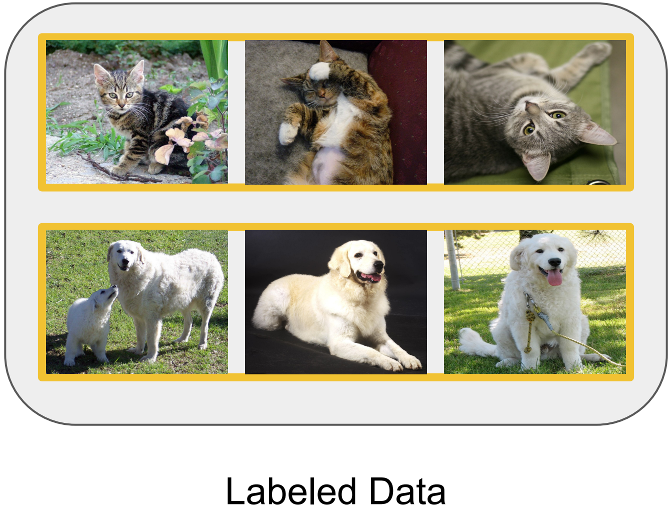
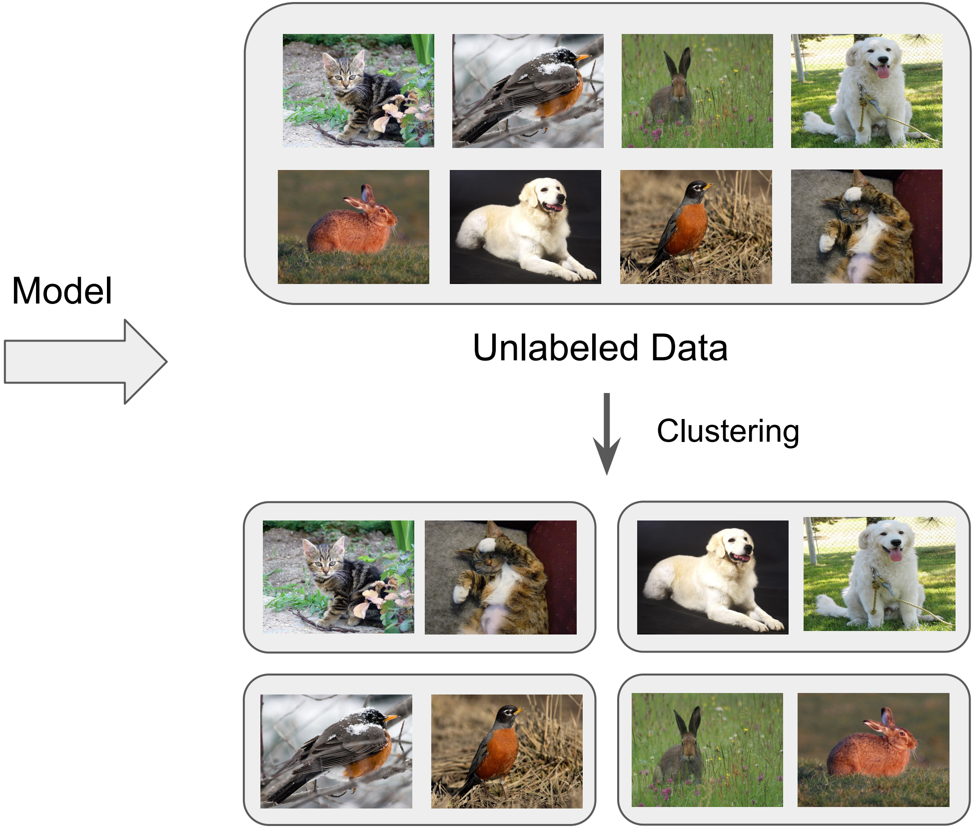
XCon: Learning with Experts for Fine-grained Category Discovery

Yixin Fei[1], Zhongkai Zhao[1], Siwei Yang[1,3], Bingchen Zhao[2,3]

[1]Tongji University [2]University of Edinburgh [3]LunarAI

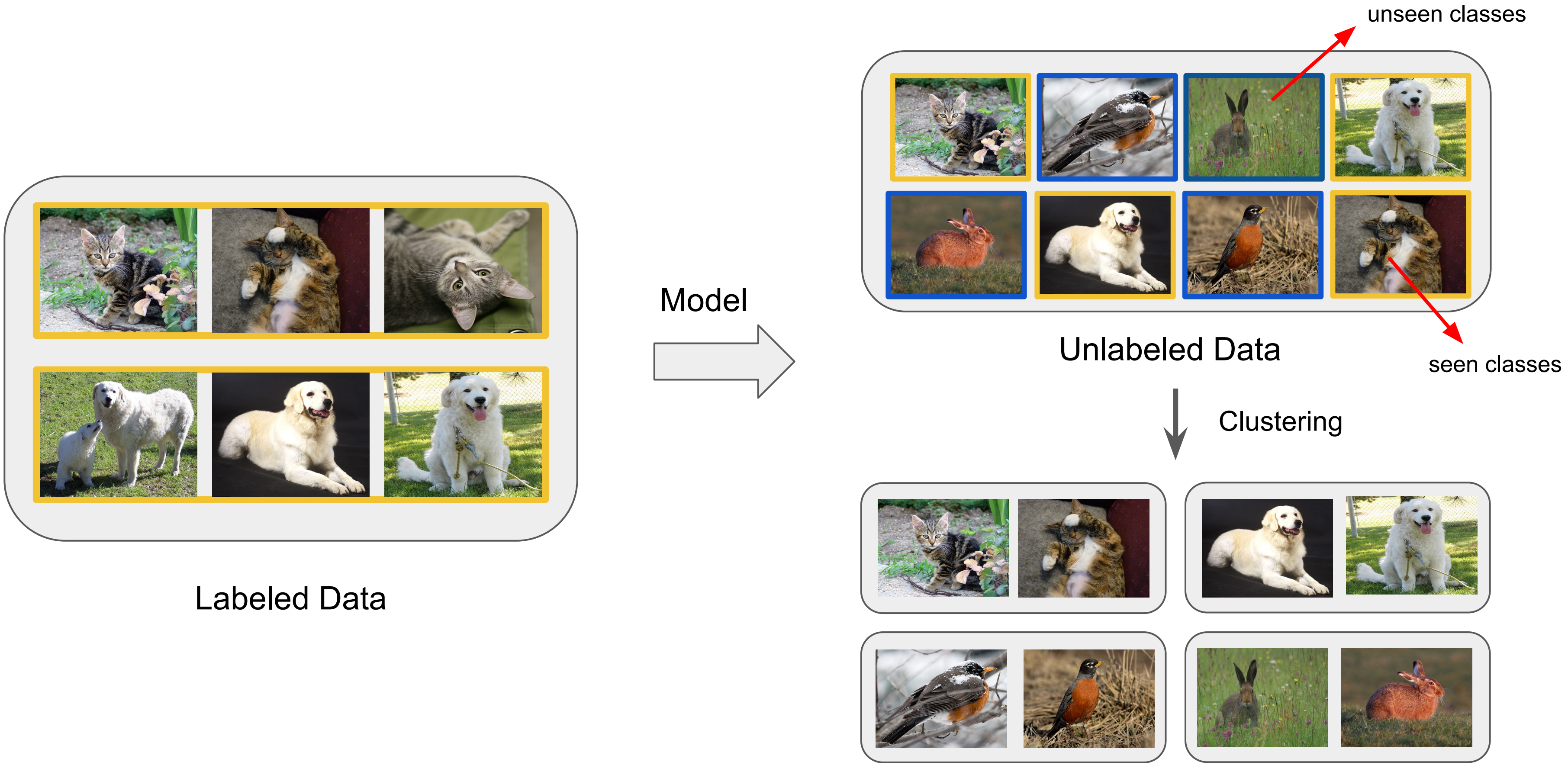
Problem definition: Generalized Category Discovery



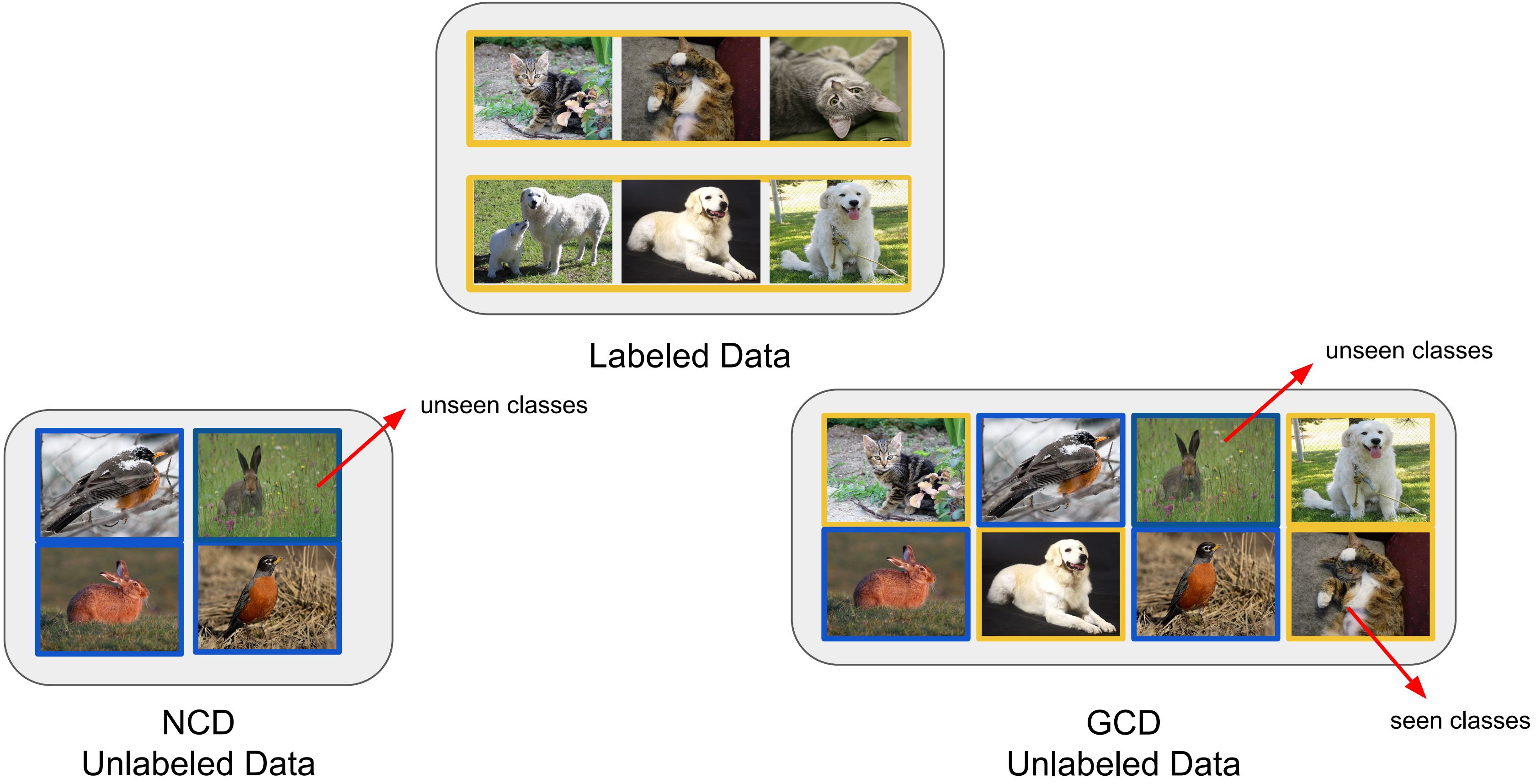


[1] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. CVPR 2022

Problem definition: Generalized Category Discovery



Problem definition: NCD & GCD



Fine-grained category discovery

- contain categories from the same entry level classes, e.g., birds, cars, aircrafts, and pets
- the large inter-class similarity and the intra-class variance

Motivation

- learn subtle discriminative cues between categories to be able to distinguish
- learn more class-relevant features, not class-irrelevant ones


tabby cat


castle


swing

ImageNet-100


Bobolink


Green Jay


Forsters Tern

CUB-200

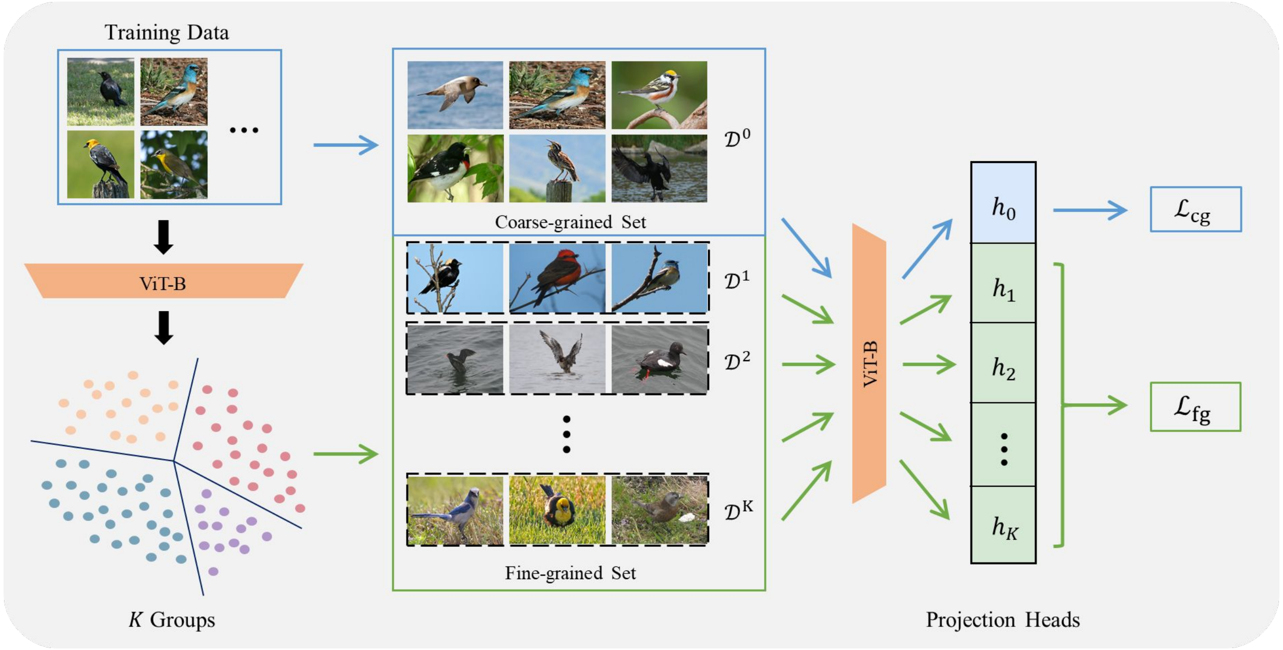
[1] Bingchen Zhao and Kai Han. Novel visual category discovery with dual ranking statistics and mutual knowledge distillation. NeurIPS 2021

Self-supervised representation

Cluster the data based on class irrelevant cues such as the object pose or the background

[1] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. CVPR 2022

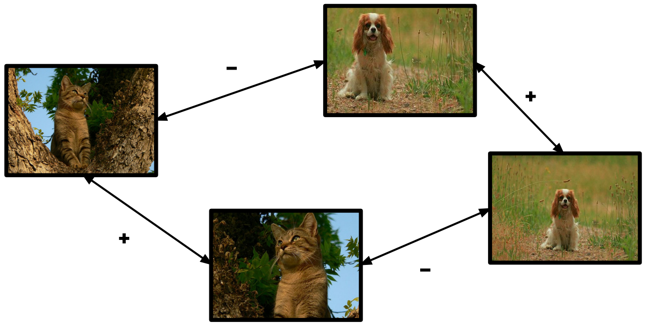
Method Overview



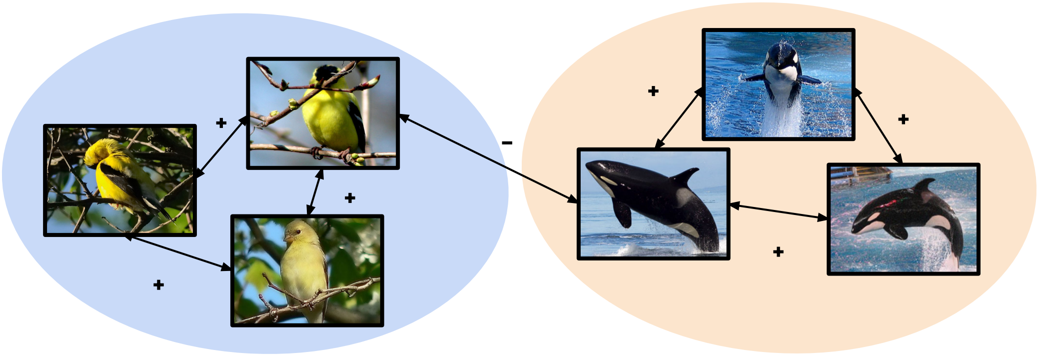
- Dataset partitioning
- Learning discriminative representations

Preliminary: Contrastive learning

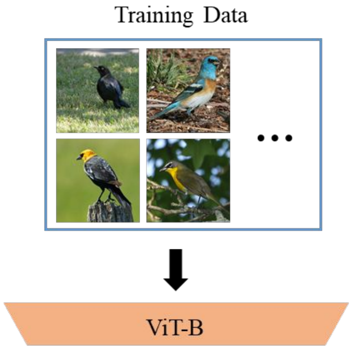
- unsupervised contrastive learning on **all the data**



- supervised contrastive learning on **the labeled data**

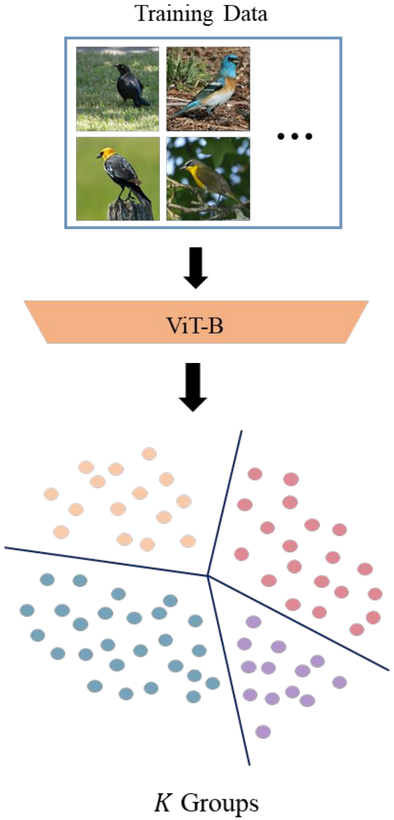


Dataset partitioning



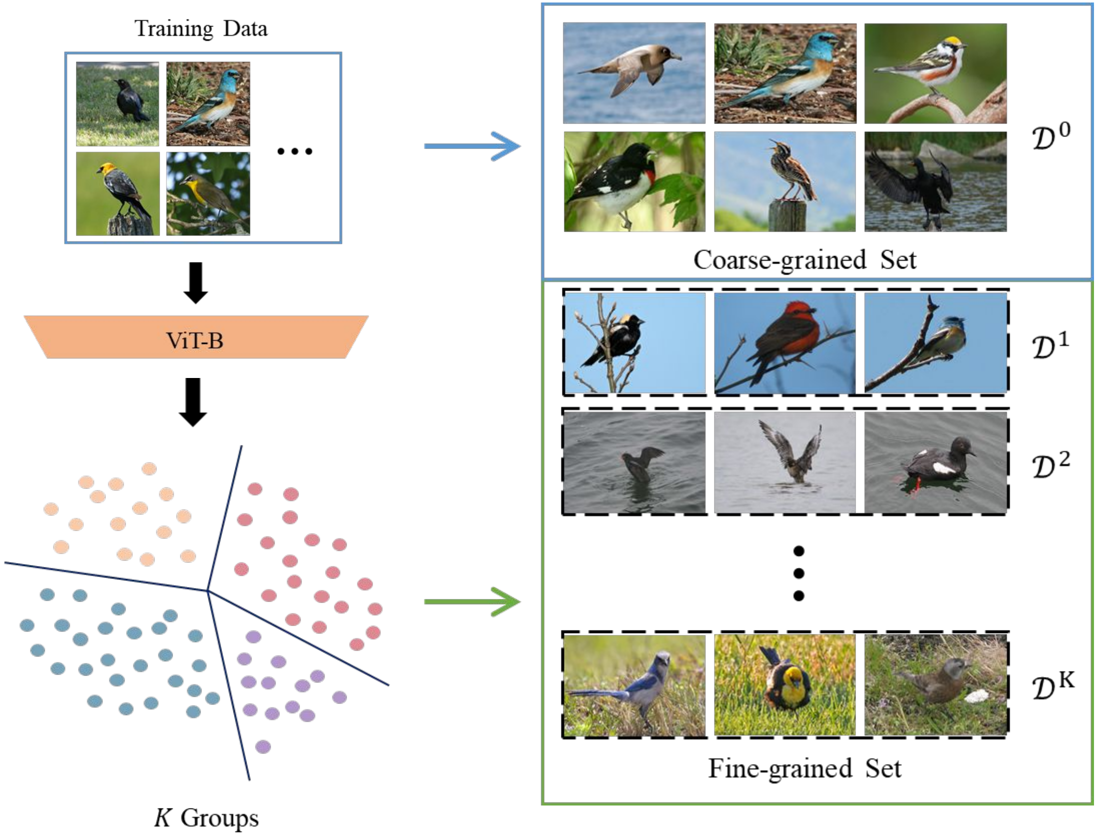
- features extacted by ViT-B

Dataset partitioning



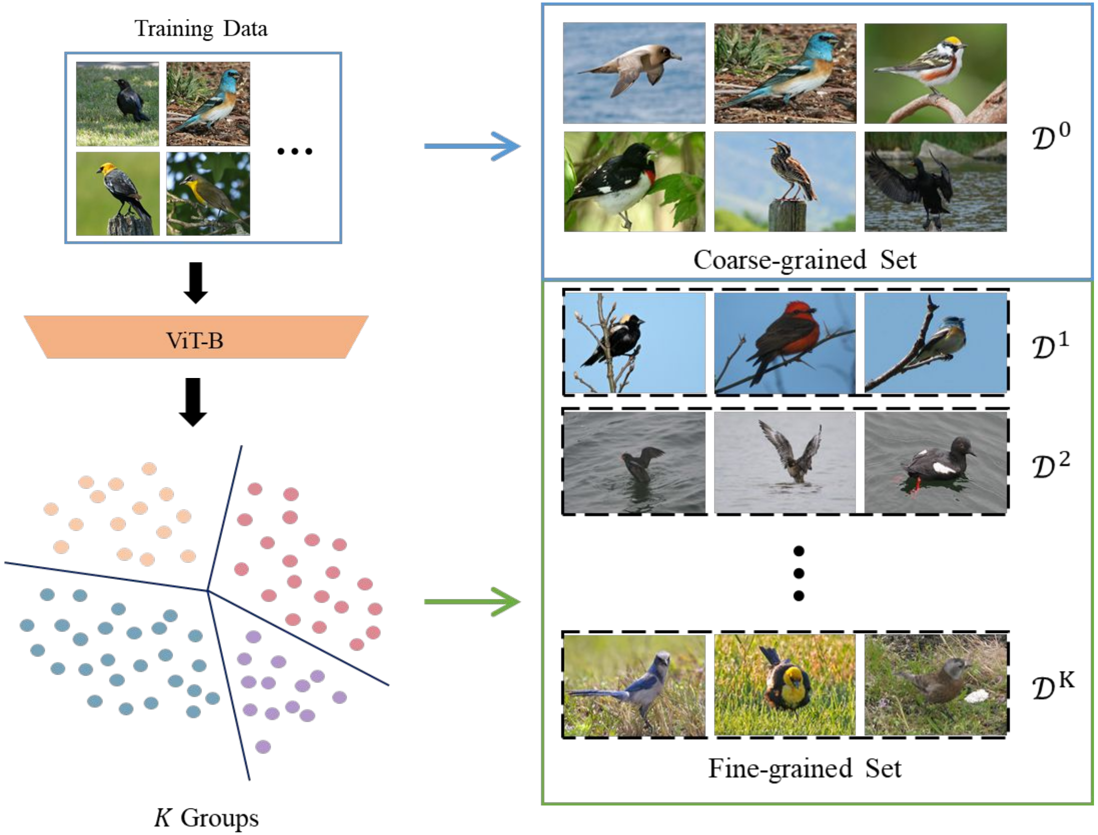
- features extacted by ViT-B
- features clustering into K groups by k-means

Dataset partitioning



- features extacted by ViT-B
- features clustering into K groups by k-means
- the whole dataset partitioned into K sub datasets

Dataset partitioning

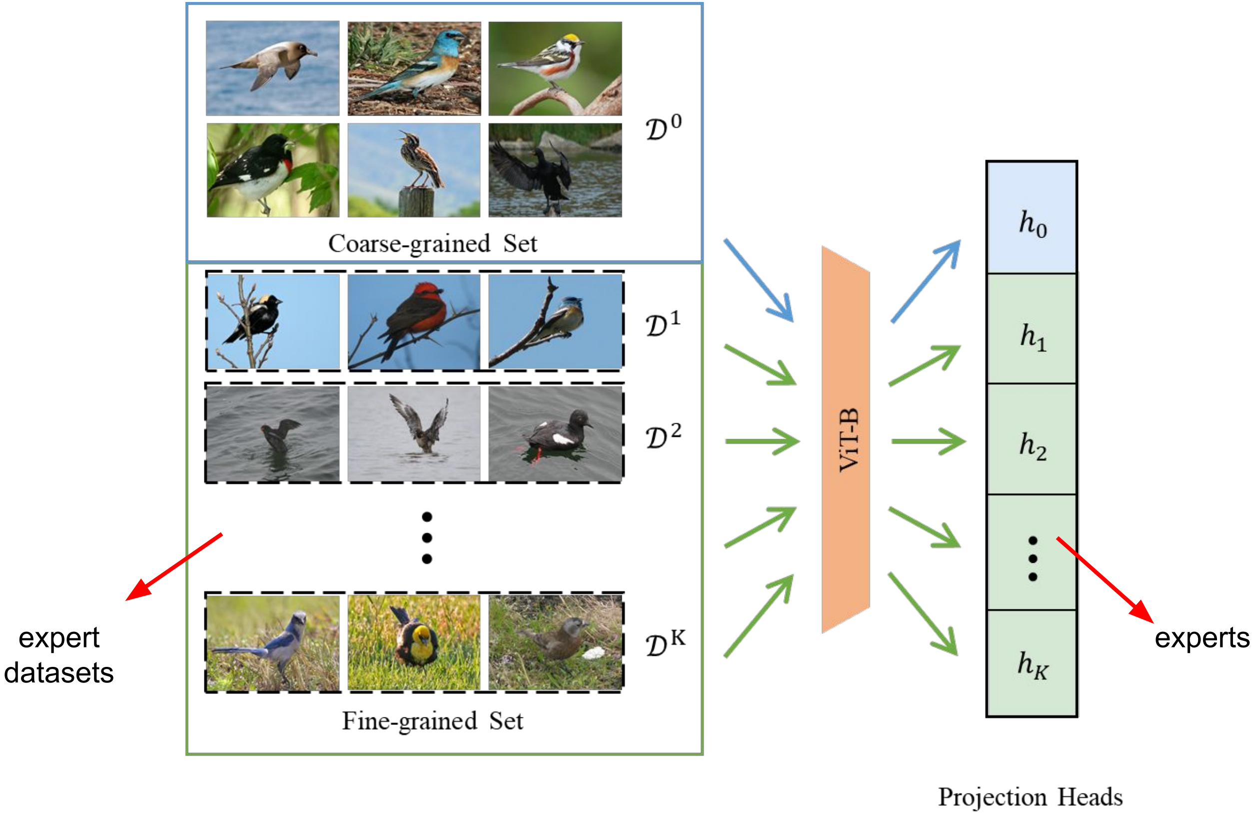


- features extacted by ViT-B
- features clustering into K groups by k-means
- the whole dataset partitioned into K sub datasets

a strong prior

Learning discriminative representations

- project features

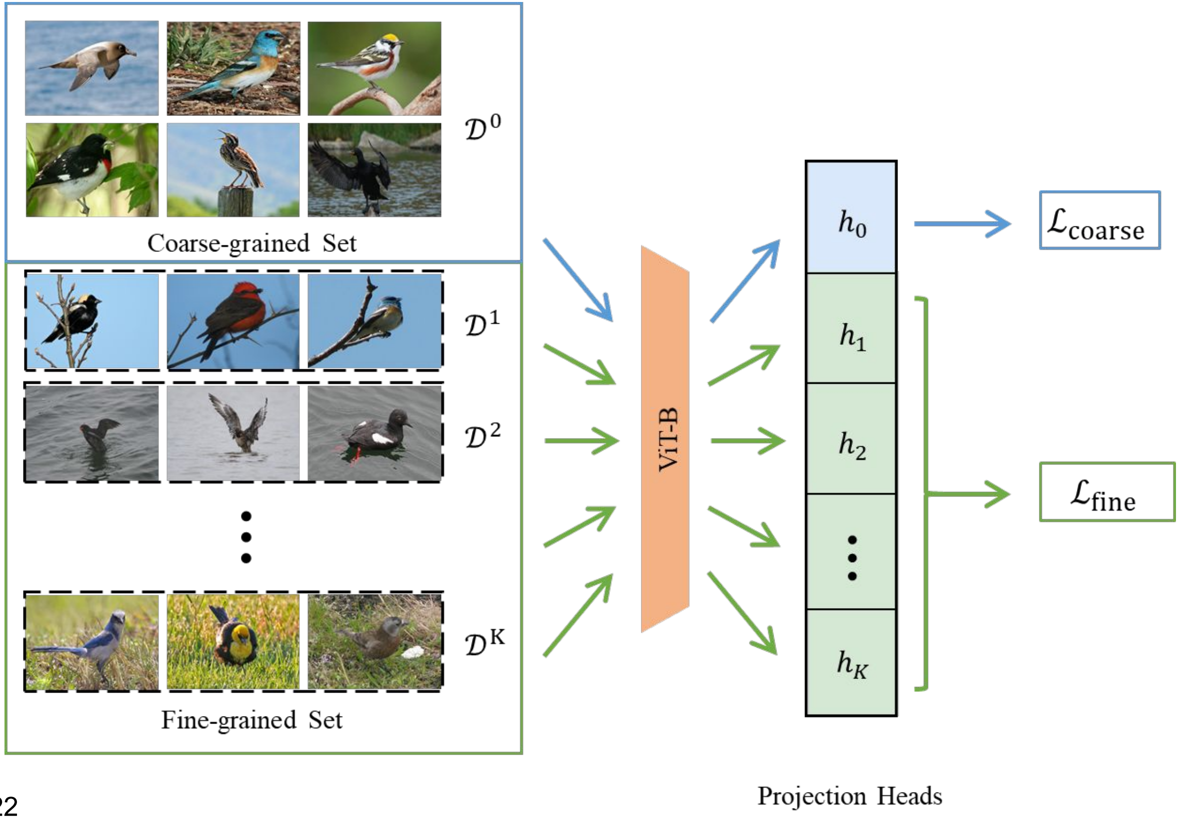


Learning discriminative representations

- project features
- apply contrastive loss [1]

$$\mathcal{L}^{u}_{\text{fine}} = -\sum_{k=1}^{K} \frac{1}{|\mathcal{B}^k|} \sum_{i \in \mathcal{B}^k} \log \frac{\exp(h_k(\mathbf{v}_i) \cdot h_k(\hat{\mathbf{v}}_i)/\tau)}{\sum_j \mathbb{1}_{[j \neq i]} \exp(h_k(\mathbf{v}_i) \cdot h_k(\mathbf{v}_j)/\tau)}$$

$$\mathcal{L}^{l}_{\text{fine}} = -\sum_{k=1}^{K} \frac{1}{|\mathcal{B}^k|} \sum_{i \in \mathcal{B}^k} \frac{1}{|\mathcal{N}(i)|} \sum_{q \in \mathcal{N}(i)} \log \frac{\exp(h_k(\mathbf{v}_i) \cdot h_k(\mathbf{v}_q)/\tau)}{\sum_j \mathbb{1}_{[j \neq i]} \exp(h_k(\mathbf{v}_i) \cdot h_k(\mathbf{v}_j)/\tau)}$$



[1] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. CVPR 2022

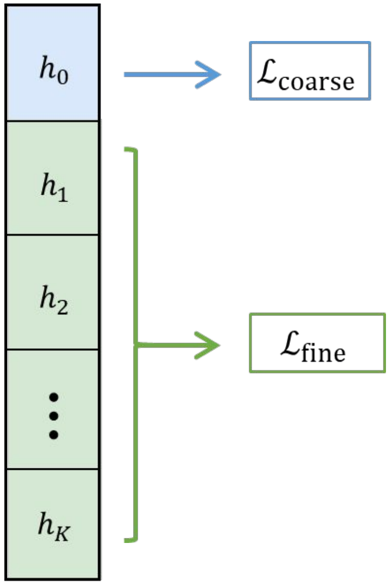
Learning discriminative representations

- project features
- apply contrastive loss [1]
- the optimization objective

$$\mathcal{L}_{\text{coarse}} = (1-\lambda) \sum_{i \in \mathcal{B}_{\mathcal{U}} \cup \mathcal{B}_{\mathcal{L}}} \mathcal{L}_i^u + \lambda \sum_{i \in \mathcal{B}_{\mathcal{L}}} \mathcal{L}_i^s$$

$$\mathcal{L} = \mathcal{L}_{\text{coarse}} + \alpha \mathcal{L}_{\text{fine}}$$

$$\mathcal{L}_{\text{fine}} = (1-\lambda)\mathcal{L}_{\text{fine}}^u + \lambda \mathcal{L}_{\text{fine}}^l$$



Projection Heads

[1] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Generalized category discovery. CVPR 2022

Experiments

- Datasets: generic image classification datasets + fine-grained datasets
- Evaluation metric: clustering accuracy (ACC) on the unlabeled set
  - All: the entire unlabeled set
  - Old: instances in unlabeled set belonging to classes in labeled set
  - New: instances in unlabeled set belonging to classes not in labeled set

Table 1: Our dataset splits in the experiments.

Dataset		CIFAR10	CIFAR100	ImageNet-100	CUB-200	SCars	Aircraft	Pet
Labelled	Classes	5	80	50	100	98	50	19
	Images	12.5k	20k	31.9k	1498	2000	1666	942
Unlabelled	Classes	10	100	100	200	196	100	37
	Images	37.5k	30k	95.3k	4496	6144	5001	2738

Experiments: generic datasets

Table 2: Results on generic datasets.

	CIFAR10			CIFAR100			ImageNet-100		
Method	All	Old	New	All	Old	New	All	Old	New
k-means [18]	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
GCD [24]	91.5	97.9	88.2	73.0	76.2	**66.5**	74.1	89.8	66.3
XCon	**96.0**	97.3	**95.4**	**74.2**	**81.2**	60.3	**77.6**	93.5	69.7

Experiments: fine-grained datasets

Table 3: Results on fine-grained datasets.

Method	CUB-200			Stanford-Cars			FGVC-Aircraft			Oxford-Pet		
	All	Old	New	All	Old	New	All	Old	New	All	Old	New
k-means [18]	34.3	38.9	32.1	12.8	10.6	13.8	16.0	14.4	16.8	77.1	70.1	80.7
RankStats+	33.3	51.6	24.2	28.3	61.8	12.1	26.9	36.4	22.2	-	-	-
UNO+	35.1	49.0	28.1	35.5	**70.5**	18.6	40.3	**56.4**	32.2	-	-	-
GCD [24]	51.3	**56.6**	48.7	39.0	57.6	29.9	45.0	41.1	46.9	80.2	85.1	77.6
XCon	**52.1**	54.3	**51.0**	**40.5**	58.8	**31.7**	**47.7**	44.4	**49.4**	**86.7**	**91.5**	**84.1**

Experiments: ablation study

Table 4: Ablation study of fine-grained loss and coarse-grained loss.

$\mathcal{L}_{\text{fine}}$	$\mathcal{L}_{\text{coarse}}$	CUB-200			Stanford-Cars		
		All	Old	New	All	Old	New
✓		48.0	50.5	46.8	21.3	30.6	16.8
	✓	49.9	53.4	48.2	37.1	57.9	27.0
✓	✓	**51.8**	**53.8**	**50.8**	**41.0**	**59.1**	**32.2**

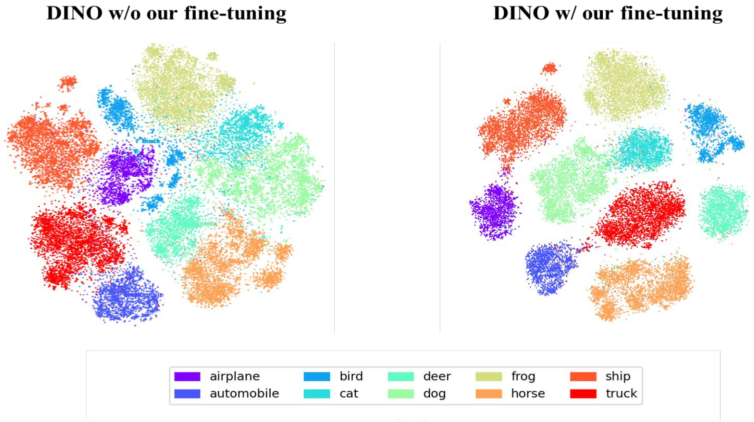
Table 5: Ablation study on the weight α of loss. $\alpha = 0$ is the baseline(Vaze *et al*. [24]).

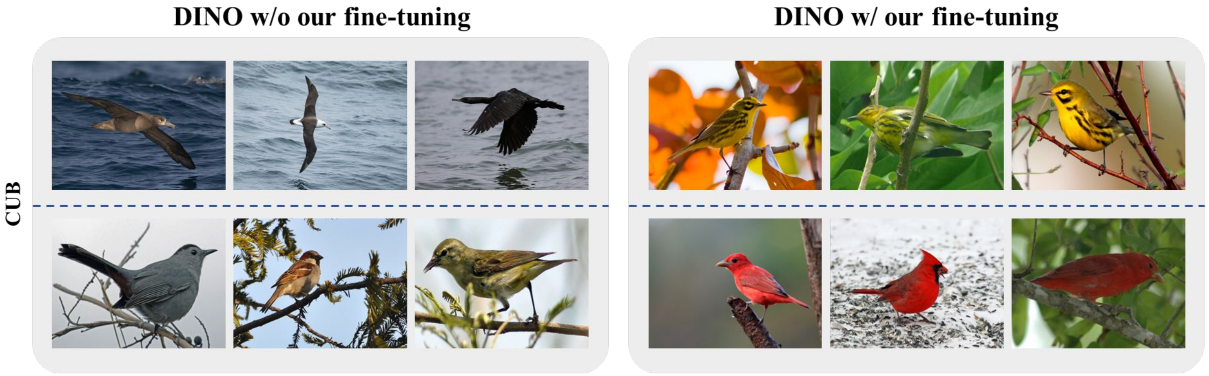
α	CUB-200			Stanford-Cars		
	All	Old	New	All	Old	New
0	49.9	53.4	48.2	37.1	57.9	27.0
0.1	51.8	53.8	50.8	41.0	59.1	32.2
0.2	51.6	54.5	50.2	**42.4**	**63.0**	**32.4**
0.4	**53.4**	**58.6**	**50.9**	41.1	61.2	31.4

Table 6: Ablation study on the number K of split sub-groups.

K	CUB-200			Stanford-Cars		
	All	Old	New	All	Old	New
1	49.9	53.4	48.2	37.1	57.9	27.0
2	51.4	**59.3**	47.4	40.9	**61.0**	31.1
4	51.7	54.6	50.2	39.8	55.3	32.3
6	50.3	51.9	49.5	**42.1**	60.7	**33.1**
8	**51.8**	53.8	**50.8**	41.0	59.1	32.2

Experiments





Summary

- We address the problem of generalized category discovery.
- We observed that self-supervised representations can group the data based on class irrelevant cues.
- A method that can learn discriminative features for fine-grained category discovery by partitioning the data into k sub-datasets is proposed.
- A new state-of-the-art performance on seven tested generalized category discovery benchmarks.

Our code is available on GitHub

https://github.com/YiXXin/XCon

Thanks for listening!