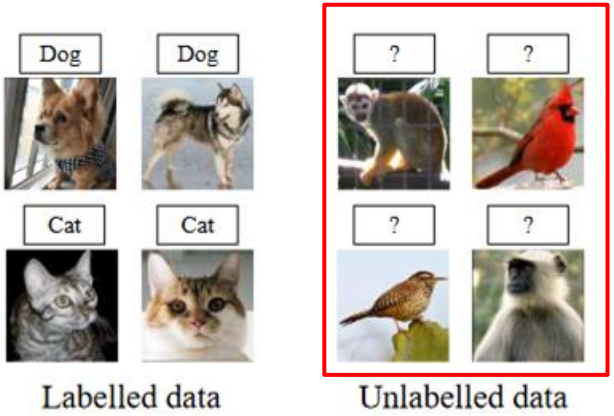
Generalized Category Discovery

CVPR 2022

Sagar Vaze Kai Han Andrea Vedaldi Andrew Zisserman

- **Problem/Objective**
  - NCD의 unrealistic problem
- **Contribution/Key Idea**
  - GCD

- NCD(Novel Class Discovery)

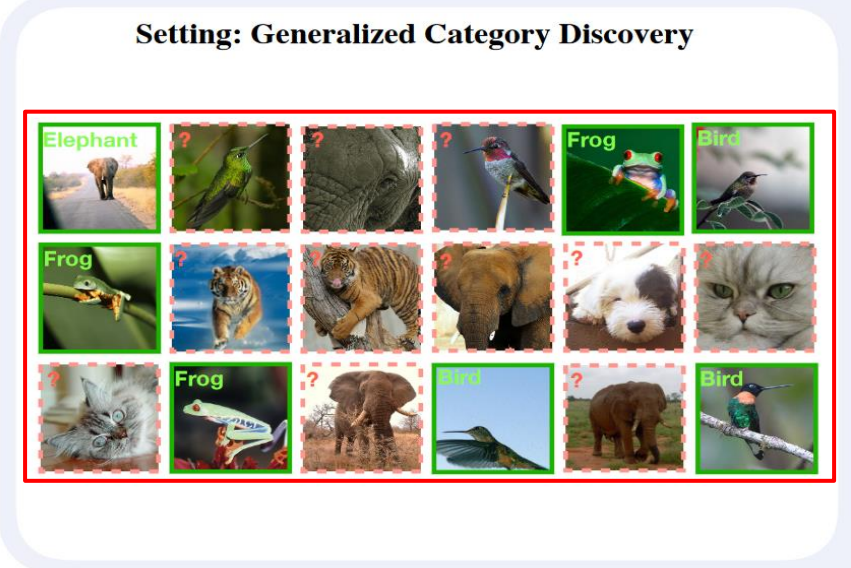


$$\mathcal{D}_{\mathcal{L}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N} \in \mathcal{X} \times \mathcal{Y}_{\mathcal{L}}$$

$$\mathcal{D}_{\mathcal{U}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M} \in \mathcal{X} \times \mathcal{Y}_{\mathcal{U}}$$

$$\mathcal{Y}_{\mathcal{L}} \cap \mathcal{Y}_{\mathcal{U}} = \emptyset$$

- GCD(Generalized Category Discovery)



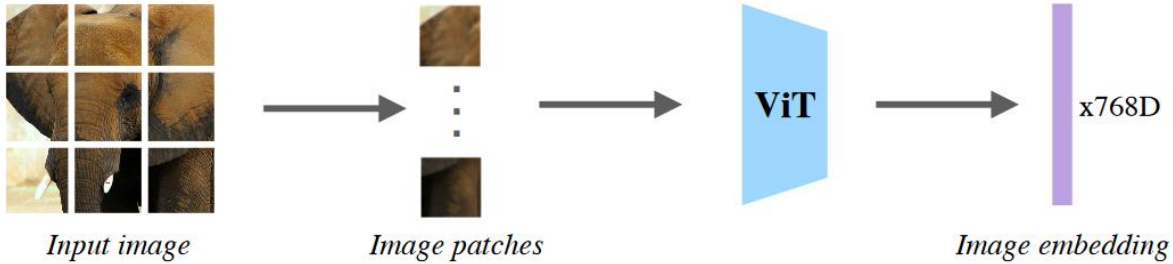
$$\mathcal{D}_{\mathcal{L}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N} \in \mathcal{X} \times \mathcal{Y}_{\mathcal{L}}$$

$$\mathcal{D}_{\mathcal{U}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{M} \in \mathcal{X} \times \mathcal{Y}_{\mathcal{U}}$$

$$\mathcal{Y}_{\mathcal{L}} \subset \mathcal{Y}_{\mathcal{U}}$$

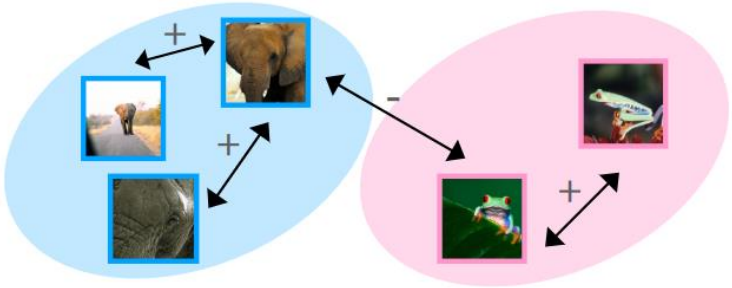
1. Feature extraction with vision transformer

Backbone : ViT-B-16 pretrained with DINO self-supervision on (unlabeled) ImageNet.

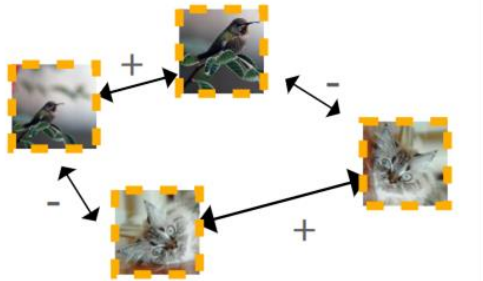


2. Finetune the representation using contrastive learning

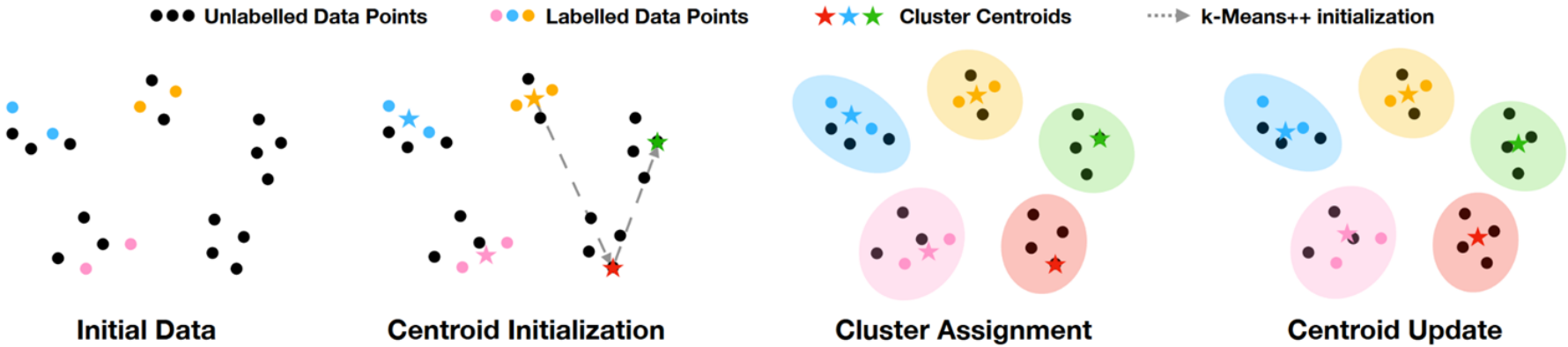
Supervised contrastive learning on labeled data.



Self-supervised contrastive learning on all data.



3. Semi-supervised K-Means Clustering



- Experiments

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

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**

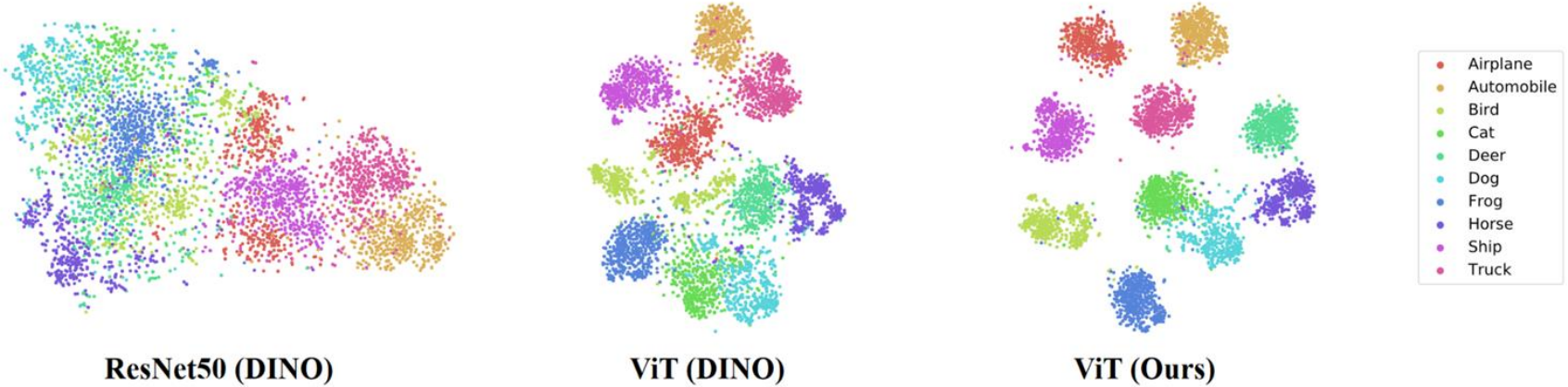
Table 4. Estimation of the number of classes in unlabelled data.

	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%

- Ablation Study

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**



ResNet50 (DINO) **ViT (DINO)** **ViT (Ours)**