Novel Class Discovery without Forgetting

K J Joseph[1,2] Sujoy Paul[1] Gaurav Aggarwal[1] Soma Biswas[3] Piyush Rai[1,4] Kai Han[1,5] Vineeth N Balasubramanian[2]

1 Google Research 2 భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్ भारतीय प्रौद्योगिकी संस्थान हैदराबाद Indian Institute of Technology Hyderabad 3 INDIAN INSTITUTE OF SCIENCE भारतीय विज्ञान संस्थान 4 INDIAN INSTITUTE OF TECHNOLOGY KANPUR 5 香 港 大 學 THE UNIVERSITY OF HONG KONG

{cs17m18p100001,vineethnb}@iith.ac.in, {sujoyp,gauravaggarwal}@google.com, somabiswas@iisc.ac.in, piyush@cse.iitk.ac.in, kaihanx@hku.hk

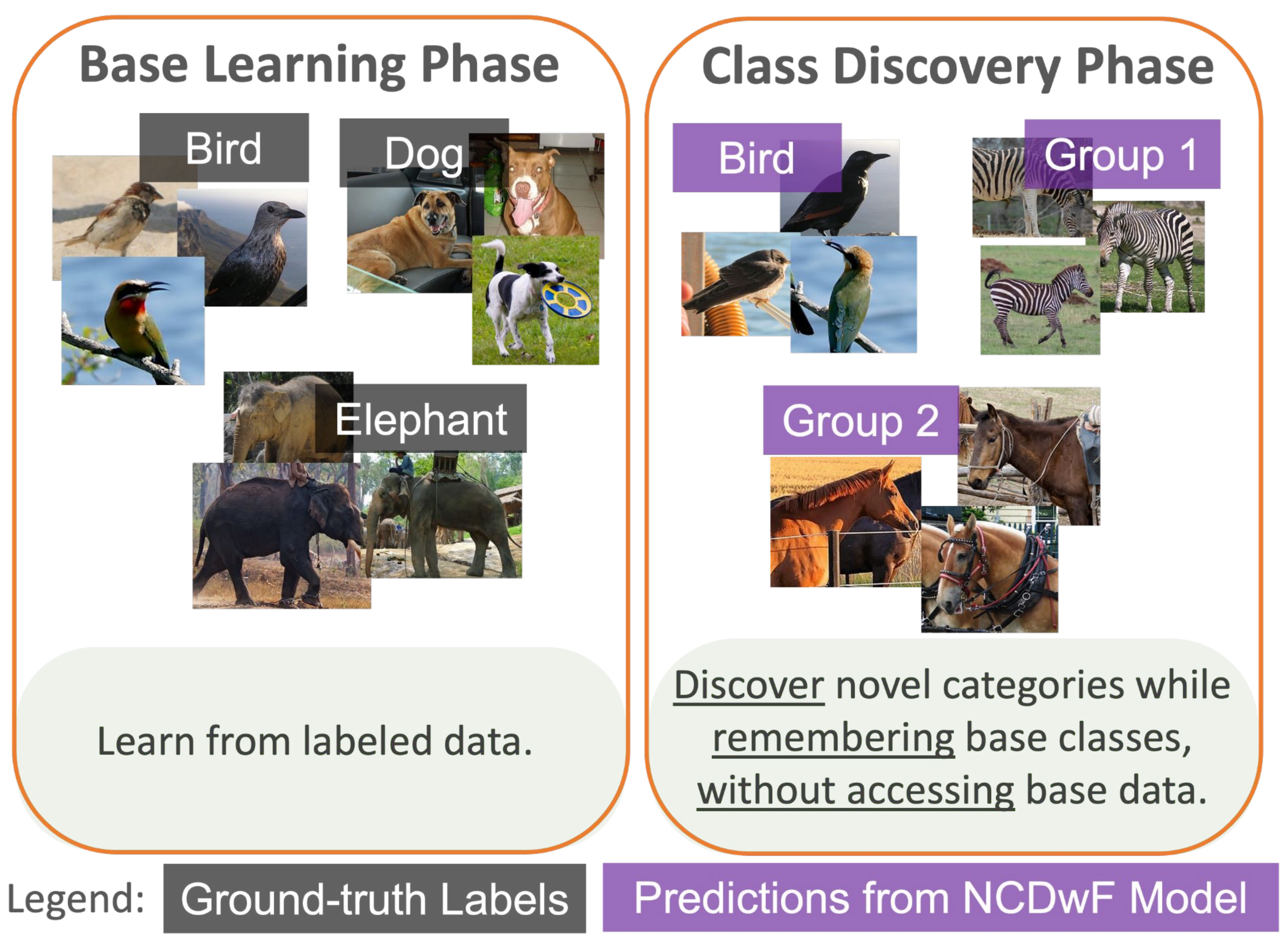
Thought Experiment: What are these?

Our existing knowledge about birds helps us to easily identify **two groups** in these images, which are **different** from those birds that we are familiar with, without:

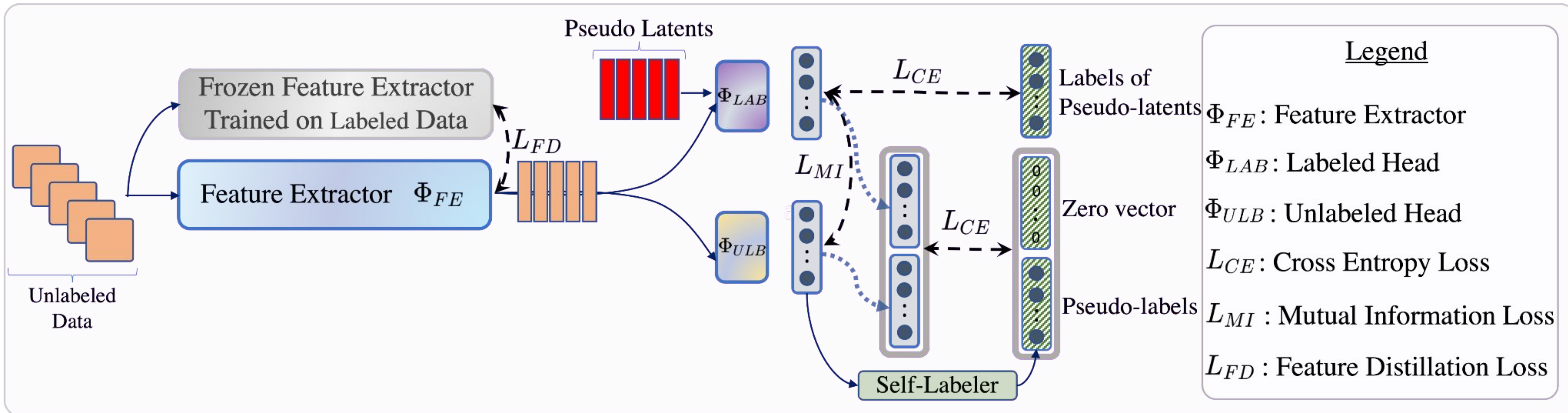
- Degrading the performance on earlier knowledge.
- Any extra information other than these images.

Novel Class Discovery without Forgetting (NCDwF) tasks an ML model to have these traits!

Problem Setting



Our Proposed Method



Pseudo Latent Generation

Step 1: Sample z. $z_i \sim \mathcal{N}(0, \mathcal{I})$

Step 2: Choose the class to invert. Φ_{LAB}

Step 3: Modify z to maximize the score of the selected logit.

$z_{i+1} = z_i + \nabla p_i[c]$

$p_i = \Phi_{LAB}(z_i)$ $\quad p_i[c]$ $\quad \nabla p_i[c]$

Step 4: Latent Mix-up with the class mean.

z_i $\quad z_p$ $\quad z_\mu^c$ $\quad \alpha \quad 1-\alpha$

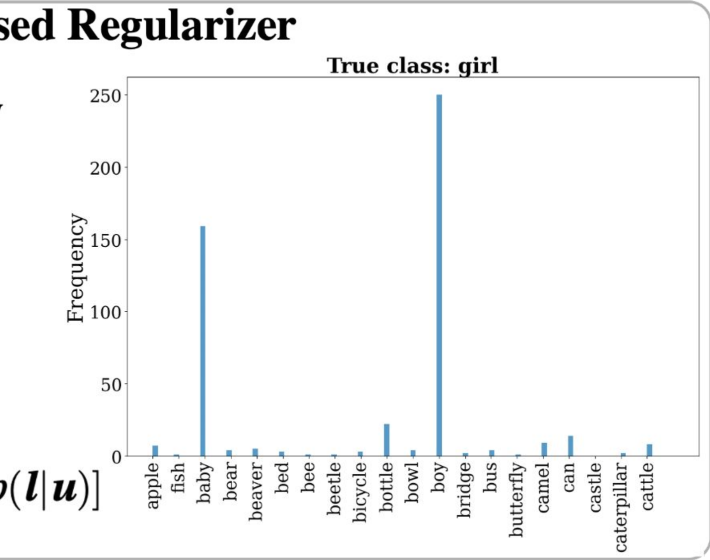
Mutual Information based Regularizer

Observation: An unlabeled instance gets wrongly classified into semantically closest base class.

For the logits: $\boldsymbol{l} = \phi_{LAB} \circ \phi_{FE}(\boldsymbol{x})$
$\boldsymbol{u} = \phi_{ULB} \circ \phi_{FE}(\boldsymbol{x})$

MI based loss: $L_{MI} = -I(\boldsymbol{l}; \boldsymbol{u})$

$$I(\boldsymbol{l}; \boldsymbol{u}) = H(\boldsymbol{l}) - H(\boldsymbol{l}|\boldsymbol{u}) = -\mathbb{E}_{\boldsymbol{l}}[\log p(\boldsymbol{l})] + \mathbb{E}_{\boldsymbol{l},\boldsymbol{u}}[\log p(\boldsymbol{l}|\boldsymbol{u})]$$



Results

Settings (→)	CIFAR-10-5-5			CIFAR-100-80-20			CIFAR-100-50-50			CIFAR-100-20-80			ImageNet-1000-882-30		
Methods (↓)	Lab	Unlab	All	Lab	Unlab	All	Lab	Unlab	All	Lab	Unlab	All	Lab	Unlab	All
Task Aware Evaluation															
RS (ICLR 20)	20.00	84.48	52.24	44.1	55.7	49.9	18.14	32.56	25.35	13.05	11.5	12.28	3.34	24.54	13.94
NCL (CVPR 21)	20.00	59.96	39.98	13.59	57.9	35.75	10.14	12.18	11.16	12.65	4.73	8.69	1.52	11.45	6.49
UNO (ICCV 21)	33.16	**93.22**	63.19	2.01	72.78	37.39	1.76	53.85	27.81	7.95	48.7	28.33	0.75	63.4	32.08
Ours	**92.72**	90.32	**91.52**	**65.03**	**77.03**	**71.03**	**73.18**	**55.66**	**64.42**	**84.8**	**49.67**	**67.24**	**27.46**	**79.07**	**53.27**
Generalized Evaluation															
UNO (ICCV 21)	0	71.36	35.68	0	58.15	29.08	0	34.22	17.11	0	41.61	20.81	0	68.34	34.17
Ours	**79.68**	**73.66**	**76.67**	**53.23**	**60.6**	**56.92**	**62.76**	**36.42**	**49.59**	**57.85**	**42.18**	**50.02**	**21.32**	**70.99**	**46.16**

Key References

- Kai Han et al. "Automatically discovering and learning new visual categories with ranking statistics." ICLR 2020.
- Zhun Zhong et al. "Neighborhood contrastive learning for novel class discovery." CVPR 2021.
- Enrico Fini et al. "A unified objective for novel class discovery." ICCV 2021.