PromptCCD:
Learning Gaussian Mixture Prompt Pool for Continual Category Discovery

visual-ai.github.io/promptccd/

Visual AI Lab

Fernando Julio Cendra[1], Bingchen Zhao[2], Kai Han[1]
[1]Visual AI Lab, The University of Hong Kong
[2]The University of Edinburgh

Accepted at ECCV 2024 in Milan, Italy

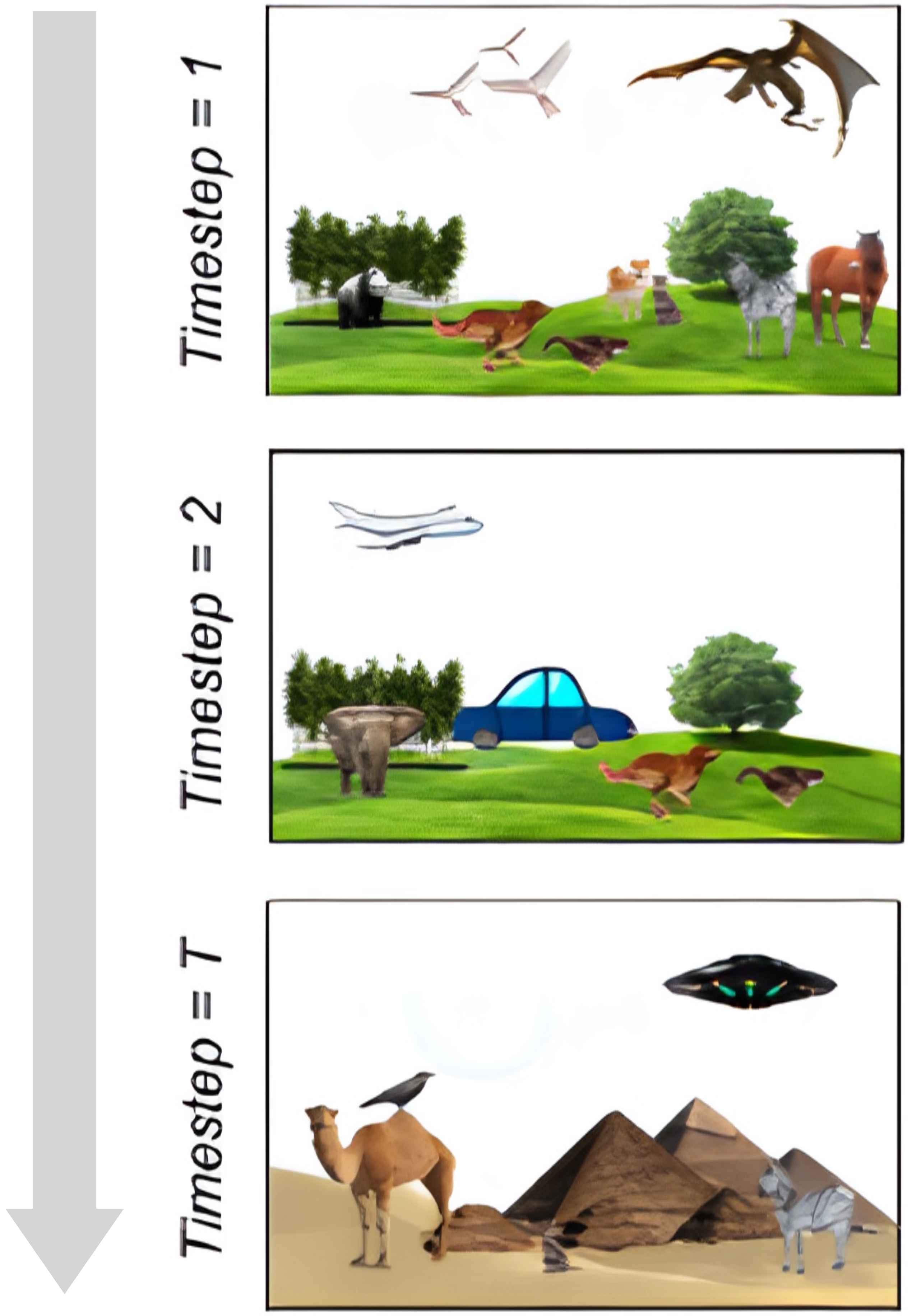
香 港 大 學
THE UNIVERSITY OF HONG KONG

THE UNIVERSITY of EDINBURGH

Contents

- [*Intro.*] What is Continual Category Discovery (CCD)?
- [*Related work*] Current state of CCD solutions.
- [*PromptCCD*] Our proposed solution.
- [*Analysis*] Experiment on CCD benchmarks & model analysis.
- Conclusion

[Intro.] Our visual world is open and dynamic



CNN

A legless lizard and hundreds of other new species were discovered in 2023

Hundreds of species were newly discovered in 2023, including a spiny-throated reed frog named Hyperolius ukaguruensis. Found in Tanzania's...

29 Dec 2023

CNN

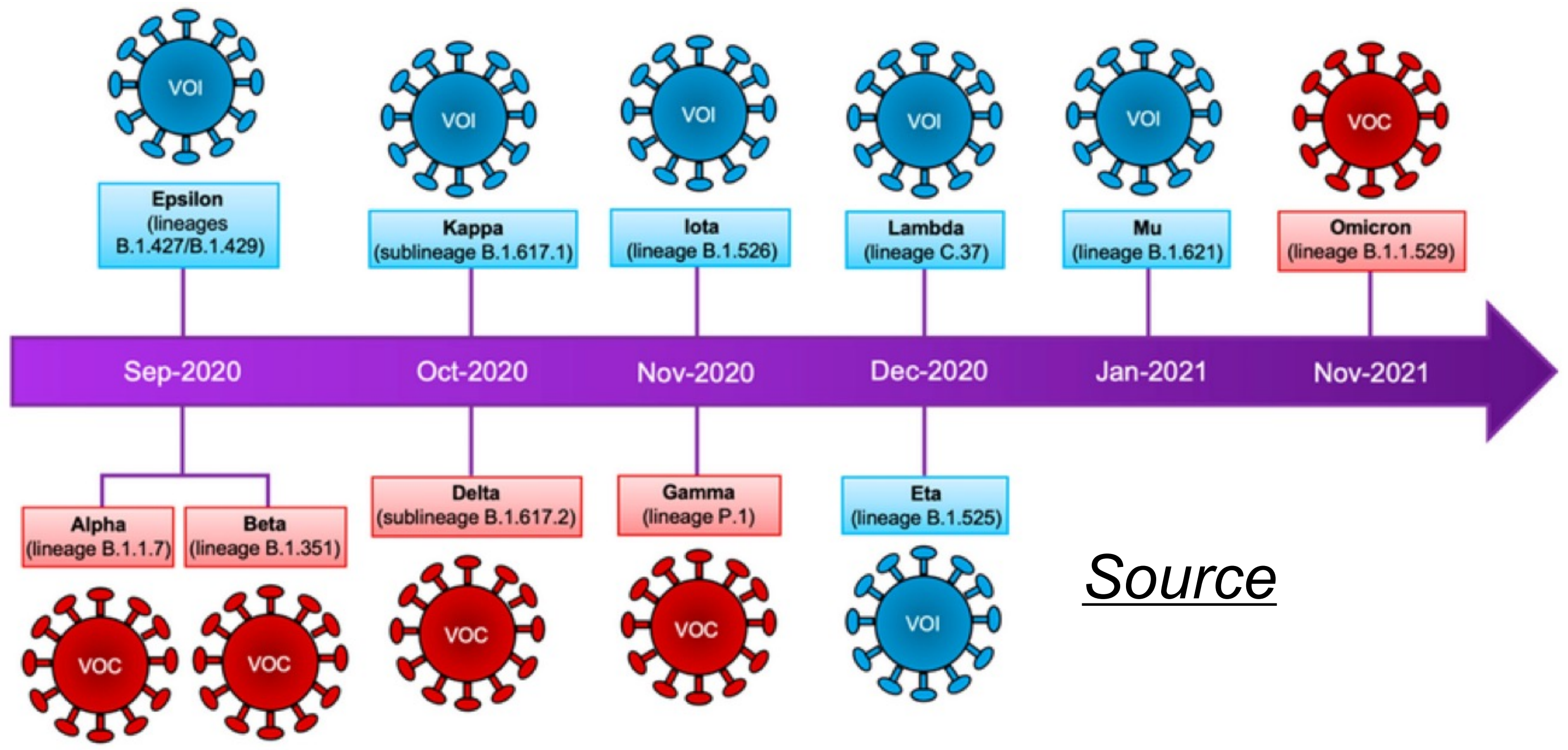
Scientists discover 100 potential new deep-sea species, including mystery creature

Scientists reported they found about 100 potential new deep-sea species — including one mystery creature — during an expedition off the...

12 Mar 2024

} Newly discovered species

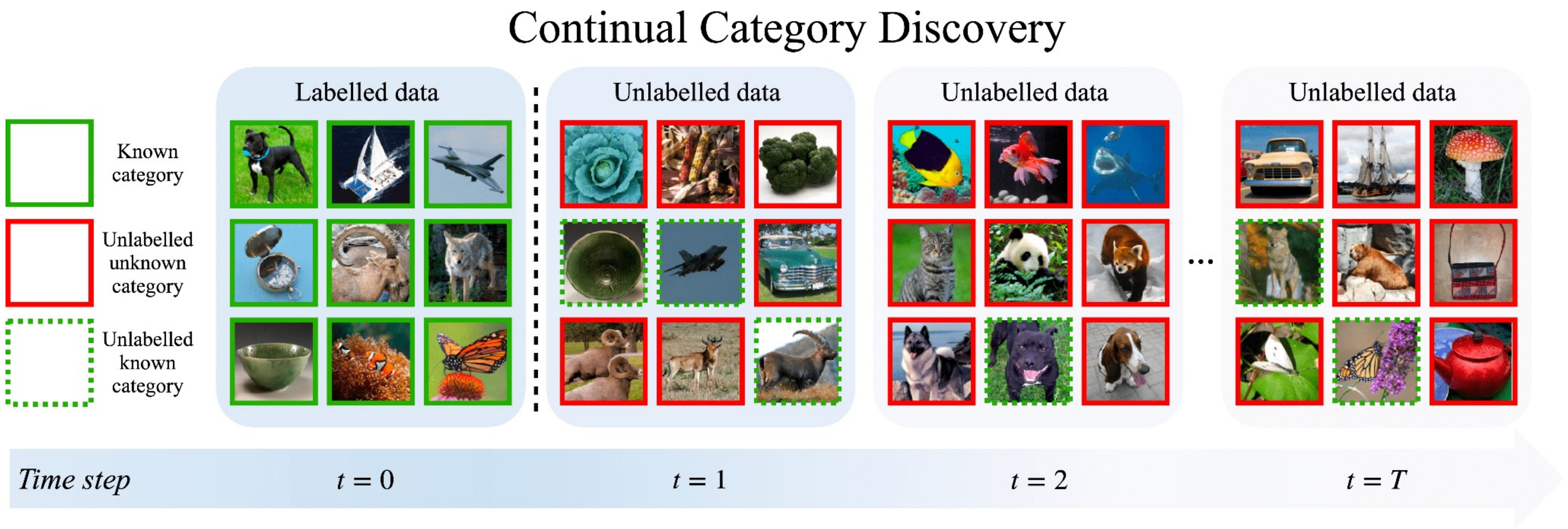
The emergence of COVID variants {



Source

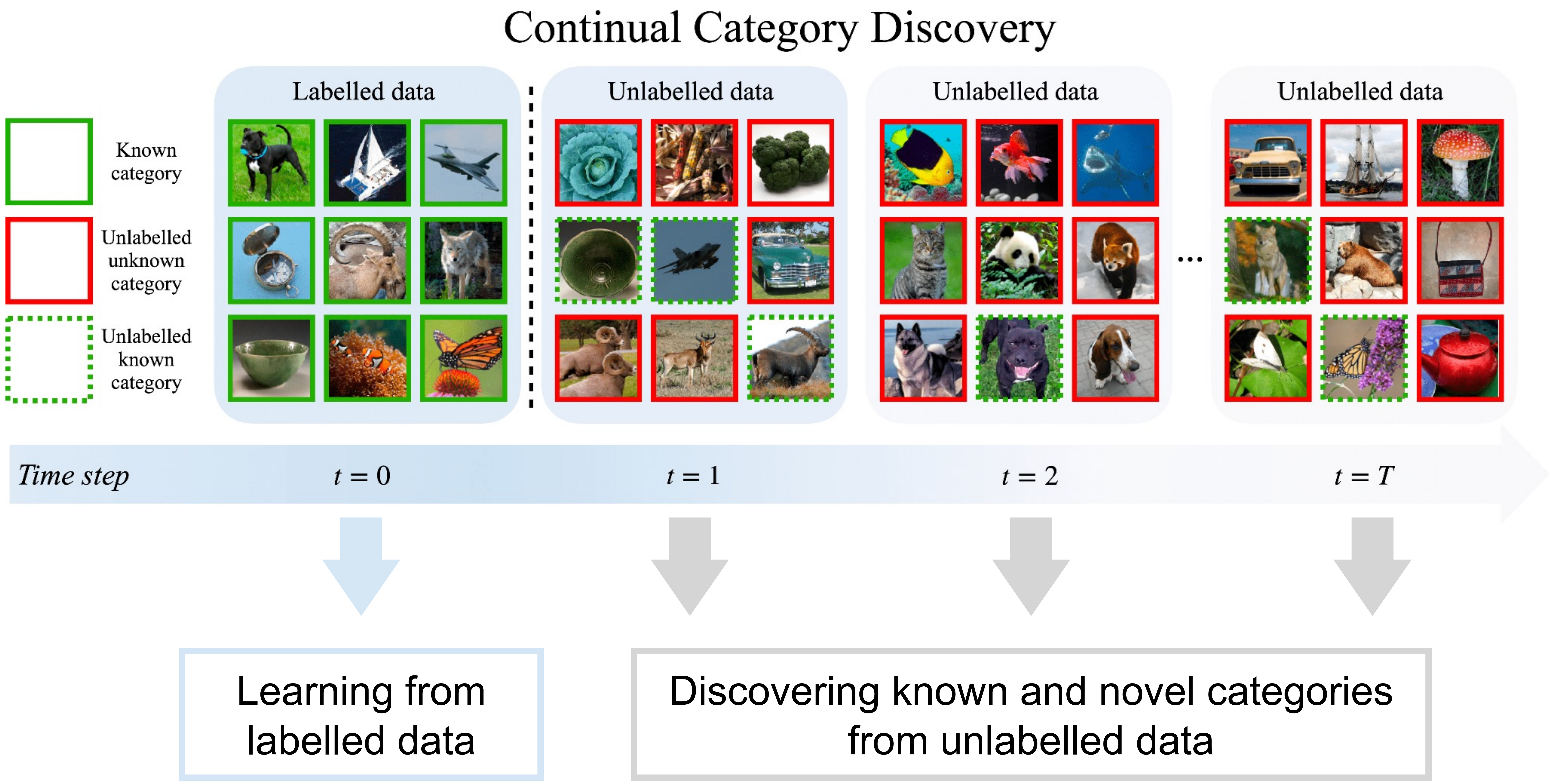
Intelligent perception systems should be **dynamic** and **open** (able to handle OOD samples). Additionally, they need to be **parameter efficient** to ensure easy adaptability.

[Intro.] Continual Category Discovery (CCD)



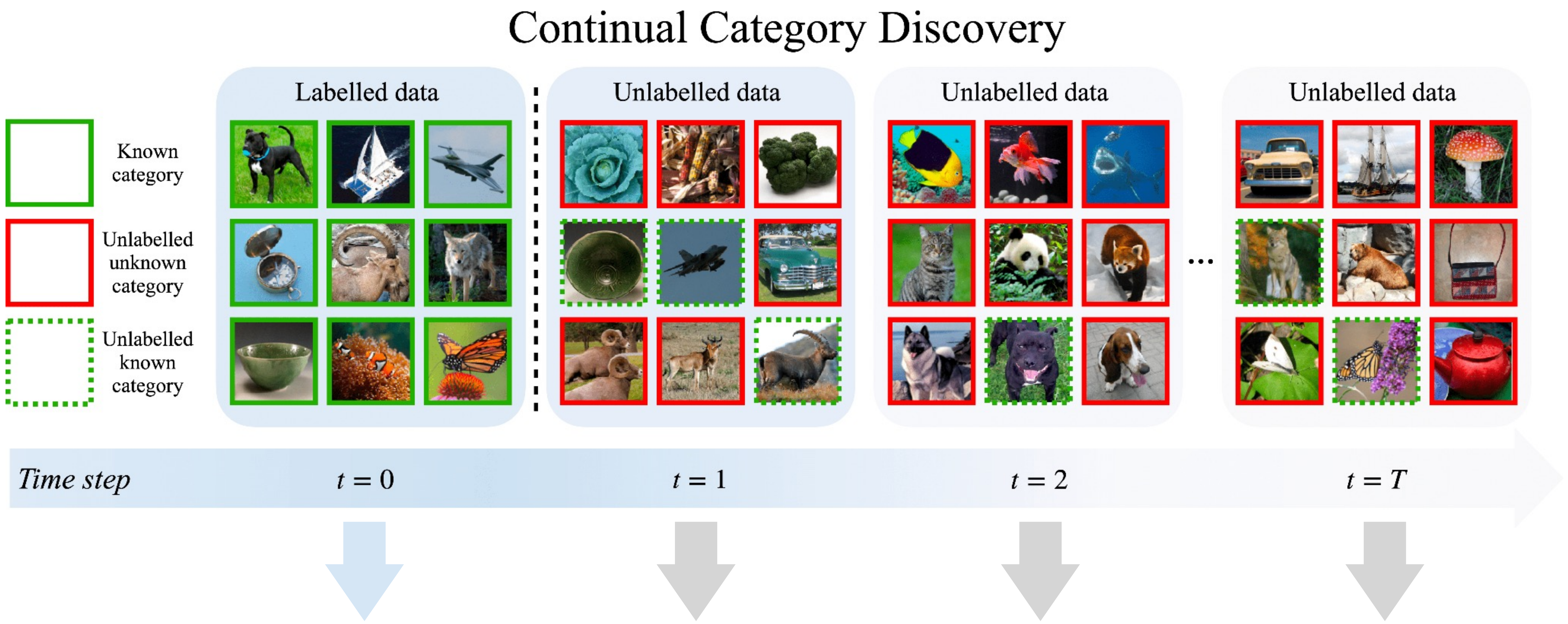
Continual Category Discovery task mimics our visual world where…

[Intro.] Continual Category Discovery (CCD)



Continual Category Discovery task mimics our visual world where…

[Intro.] Continual Category Discovery (CCD)



Learning from labelled data

Discovering known and novel categories from unlabelled data

Challenges

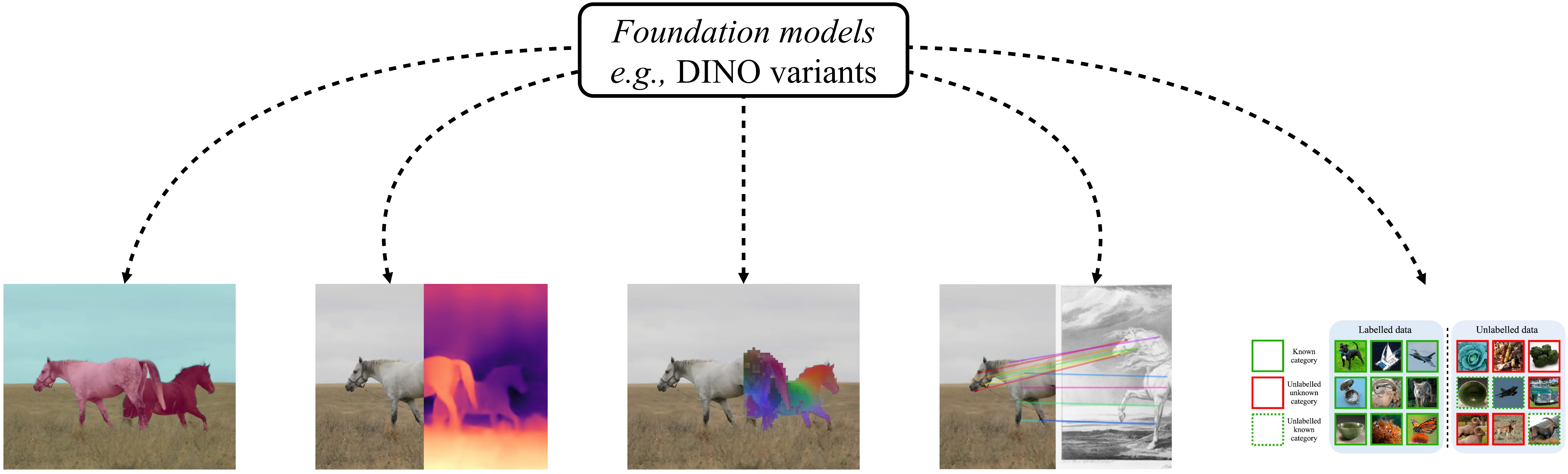
1. Category discovery

2. Catasthrophic forgetting

3. Category number estimation

[Intro.] Motivation

Recently, vision foundation models have shown remarkable performance on multiple tasks. Thus, we **aim to** unleash the potential of such models for CCD.



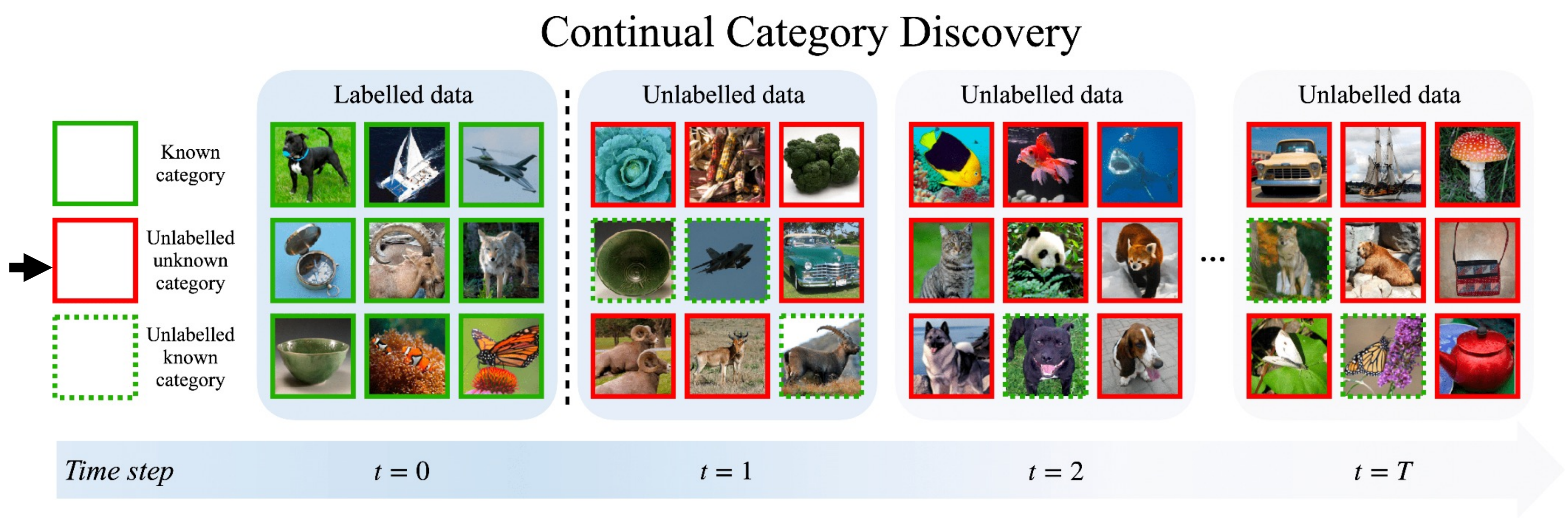
Task A Task B Task C Task D CCD task (?)

Some images are taken from https://dinov2.metademolab.com/

[Intro.] Motivation

Preliminary results by directly using DINO model for CCD task

Foundation models
e.g., DINO variants



CCD results using frozen DINO

	CIFAR100			ImageNet-100		
Method	*All*	*Old*	*New*	*All*	*Old*	*New*
Frozen DINO	56.45	68.10	53.21	67.25	73.25	65.44

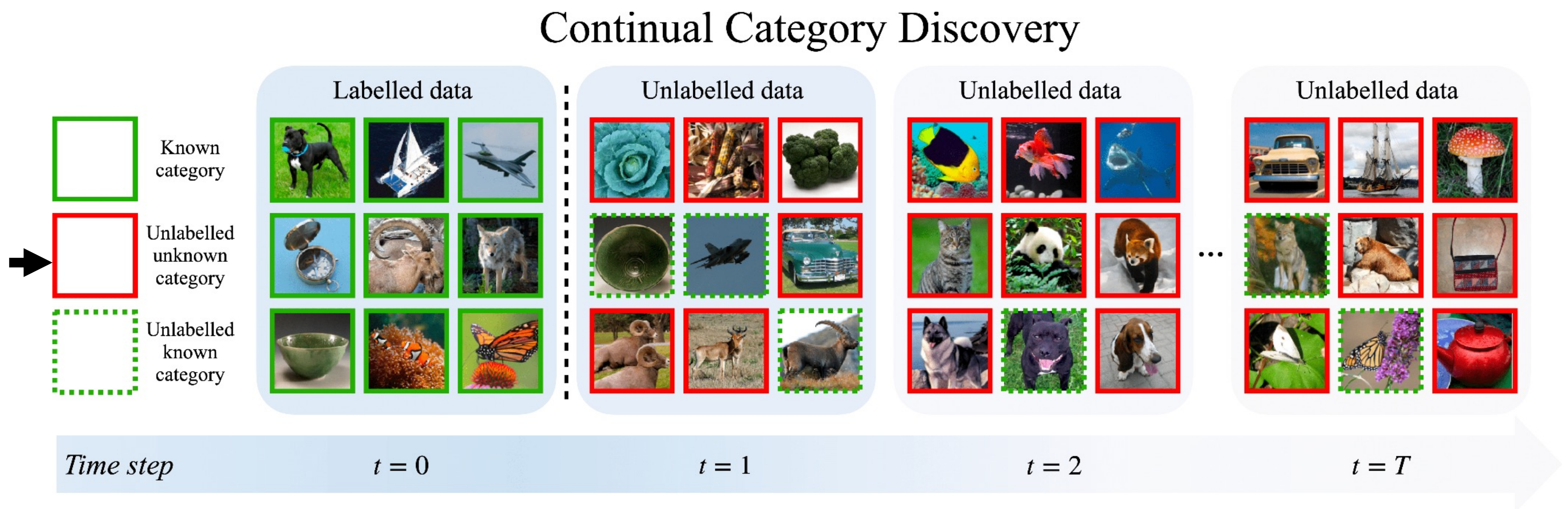
	TinyImageNet			CUB		
Method	*All*	*Old*	*New*	*All*	*Old*	*New*
Frozen DINO	49.25	59.21	45.94	39.21	68.29	30.86

***Experiment details will be explained later**

[Intro.] Motivation

Is there any potential method to further enhance foundation model for CCD?

Foundation models
e.g., DINO variants



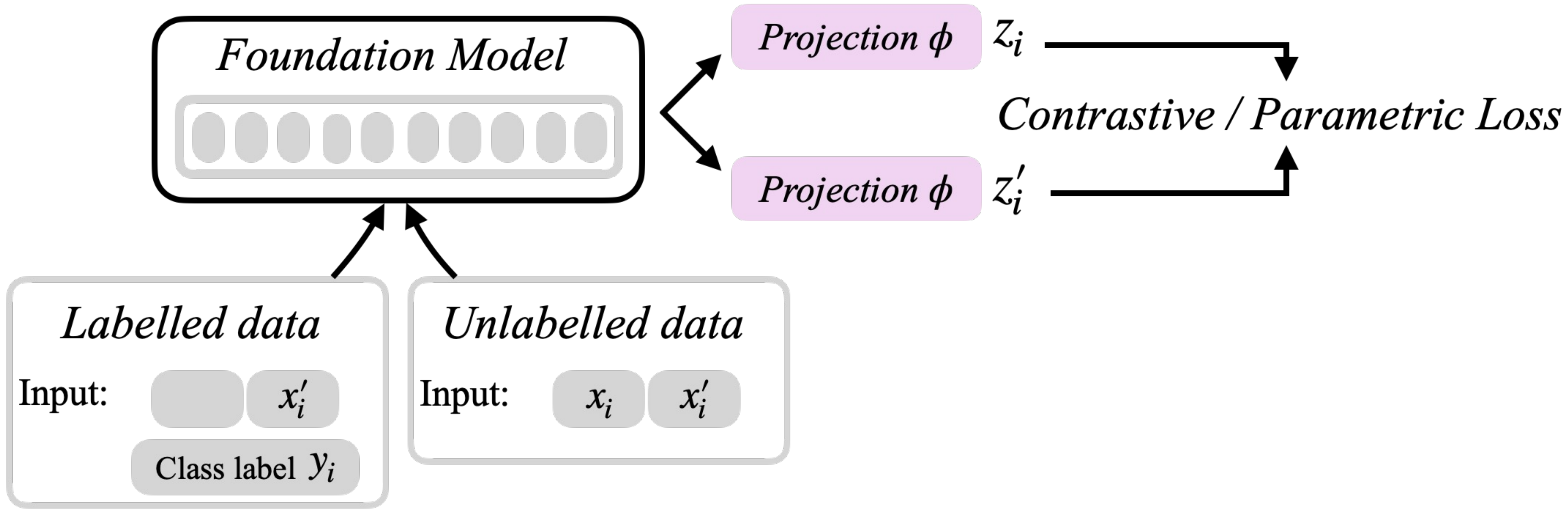
[Intro.] Motivation

Before presenting our proposed method,
let’s review some of the works that inspired it.

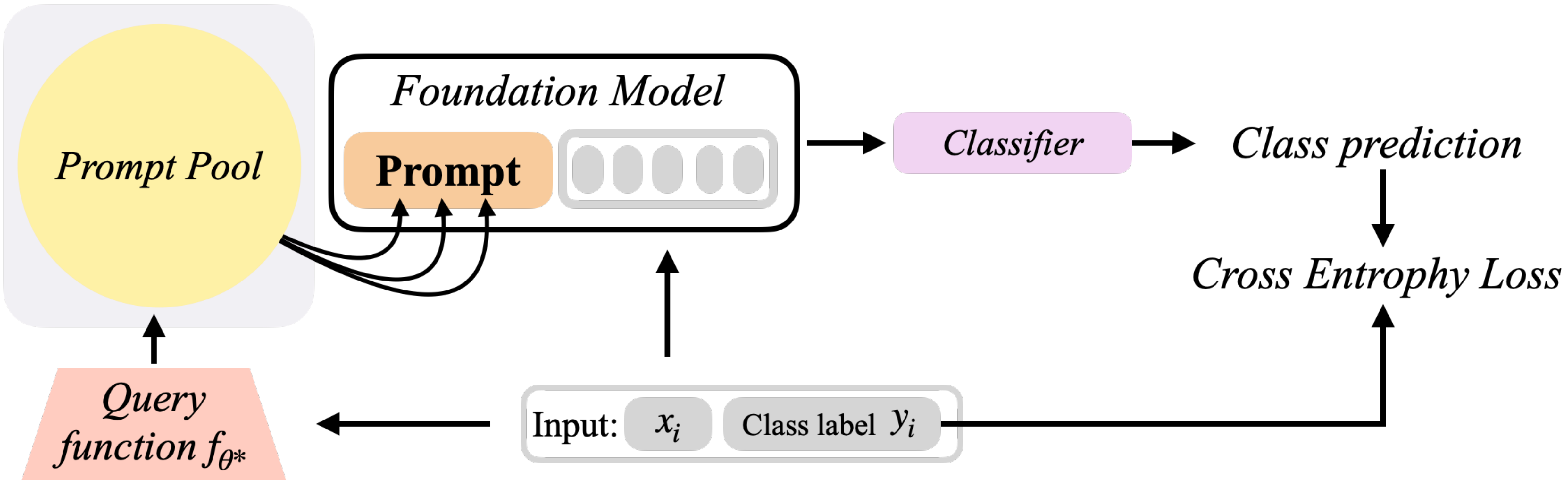
[Intro.] Motivation - background

Generalized Category Discovery (GCD)

- In GCD, given a dataset, a subset of which has class labels, the model is tasked to categorize all unlabelled images.



Prompt-based Supervised Continual Learning

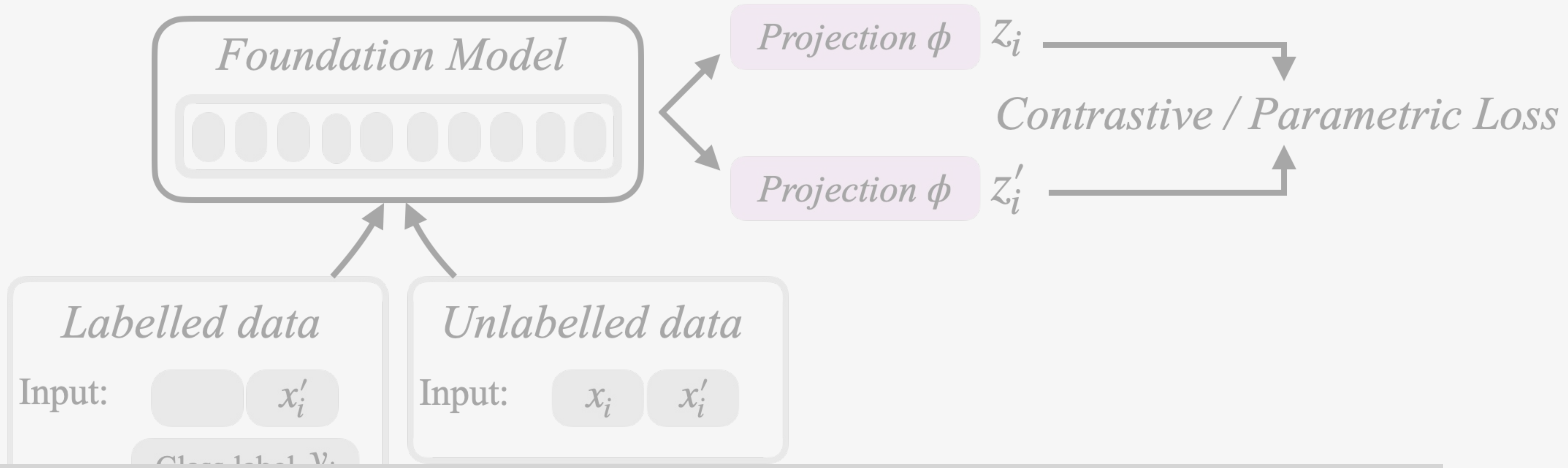


- The model leverages a prompt pool (e.g., L2P, Dual Prompt) to guide a foundation model in supervised continual learning.
- It extracts feature queries to retrieve top-k relevant prompts from the pool, which, along with class labels, are used to supervised the model.

[Intro.] Motivation - background

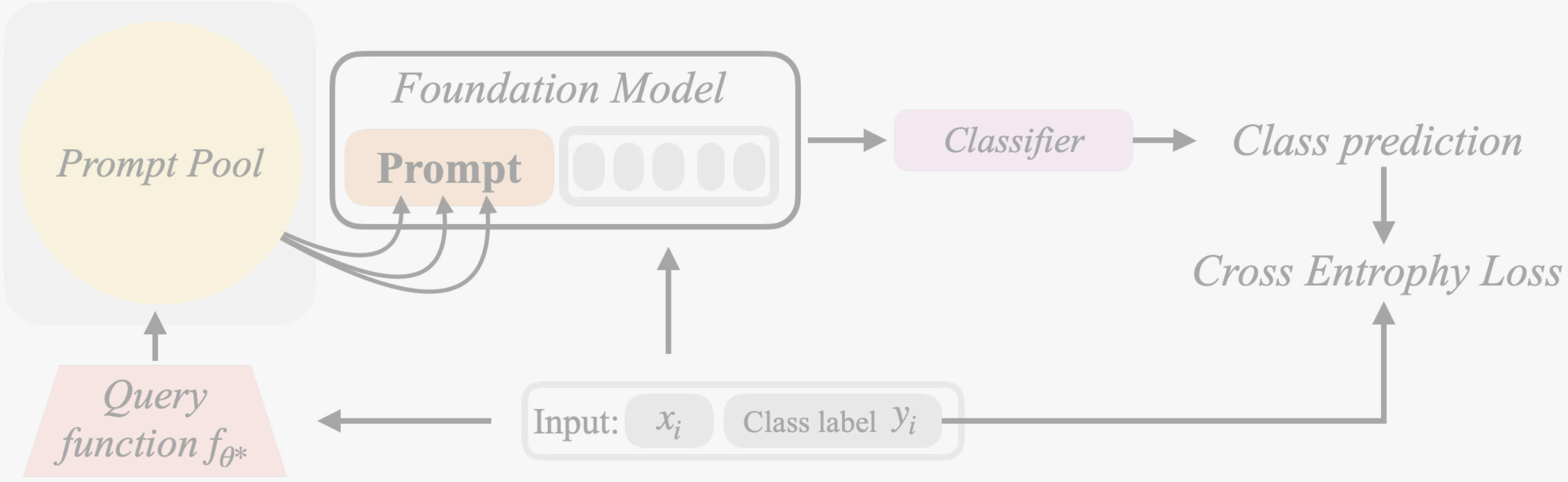
Generalized Category Discovery (GCD)

- In GCD, given a dataset, a subset of which has class labels, the model is tasked to categorize all unlabelled images.



<u>Neither method is effective in CCD</u> because only unlabelled data is available during the discovery stage. Furthermore, the unlabelled data may contain categories not present in the initial labelled set.

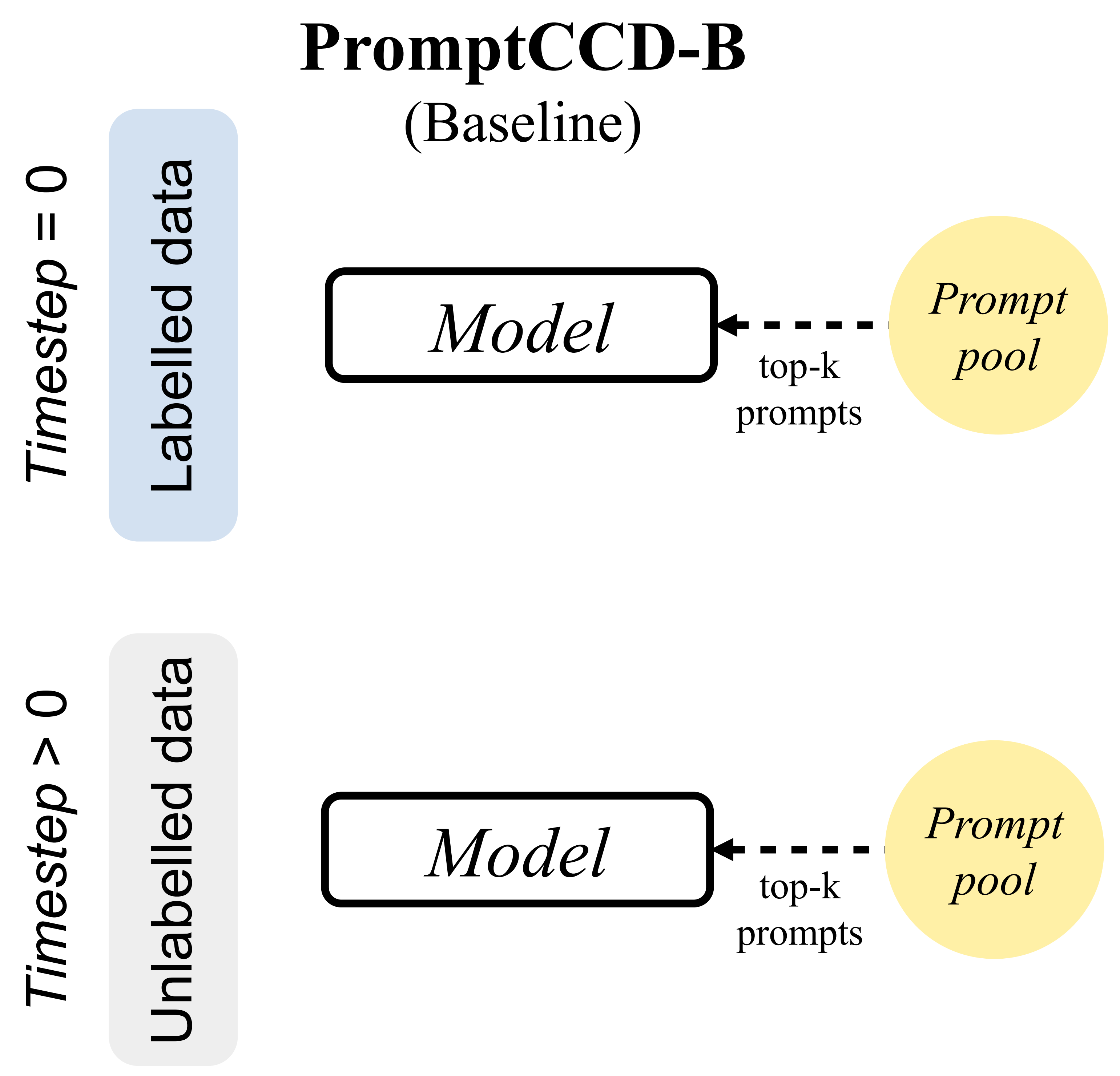
Prompt-based Supervised Continual Learning



- The model leverages a prompt pool (e.g., L2P, Dual Prompt) to guide a foundation model in supervised continual learning.
- It extracts feature queries to retrieve top-k relevant prompts from the pool, which, along with class labels, are used to supervised the model.

[Intro.] Motivation

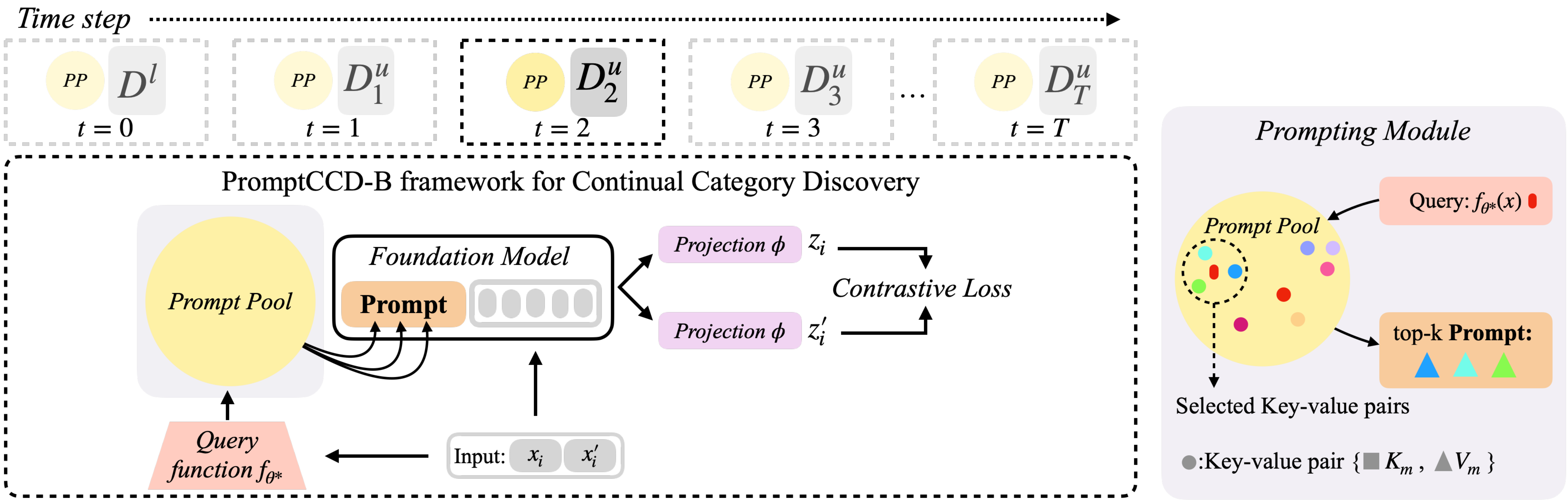
Therefore, we propose a baseline model for CCD, named, PromptCCD-B (baseline) to adapt vision foundation model with prompt pool modules, e.g., L2P & DP for CCD task:



[Intro.] Motivation

Therefore, we propose a baseline model for CCD, named, PromptCCD-B (baseline) to adapt vision foundation model with prompt pool modules, e.g., L2P & DP for CCD task:

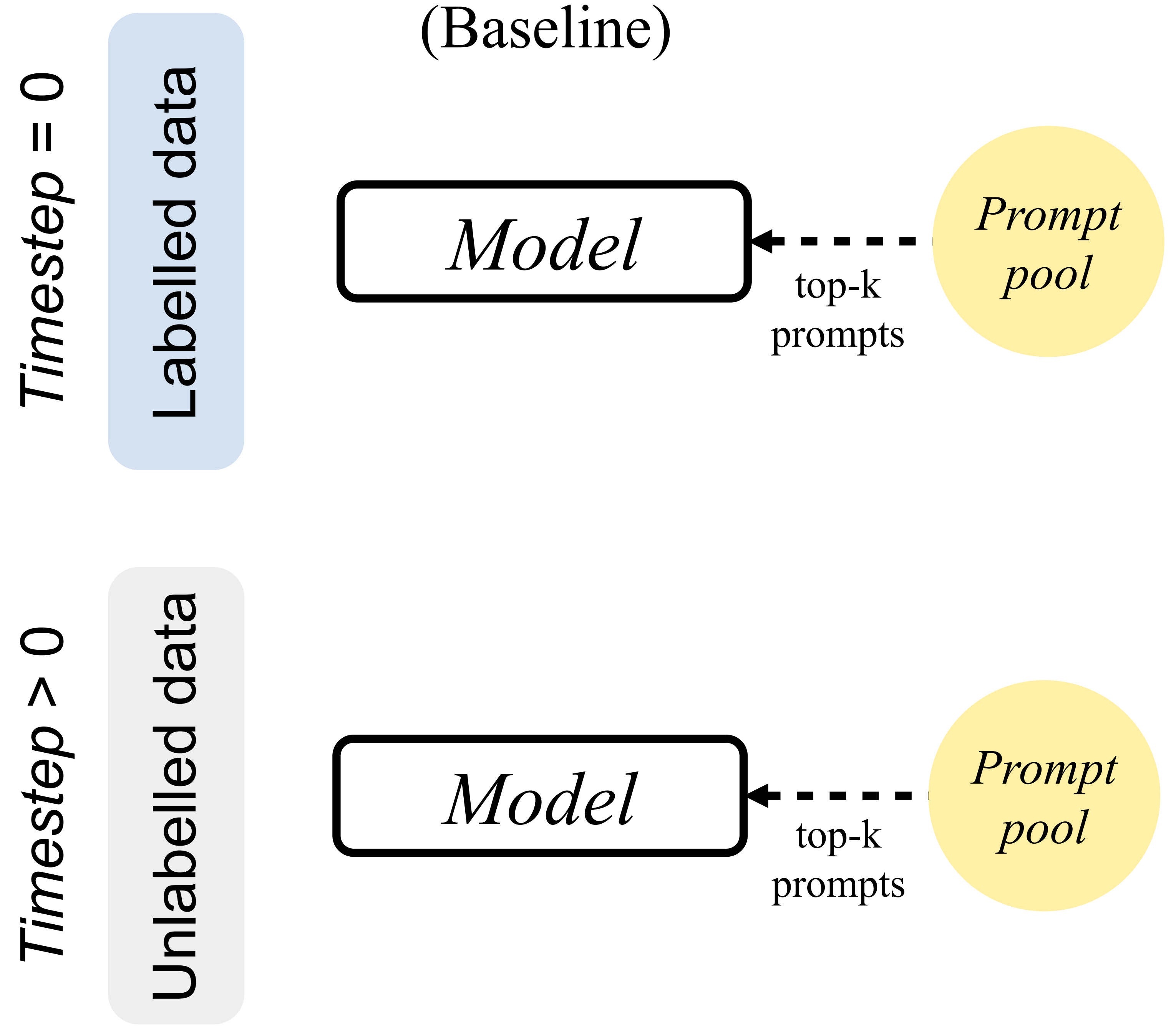
PromptCCD-B
(Baseline)



[Intro.] Motivation

Preliminary results of PromptCCD-B (baseline) by repurposing DINO model with recent prompt pool modules, e.g., L2P & DP for CCD task:

PromptCCD-B
(Baseline)



CCD comparison results between frozen DINO & the PromptCCD-B with different variants of prompt pool designs.

Method	CIFAR100			ImageNet-100		
	All	*Old*	*New*	*All*	*Old*	*New*
Frozen DINO	56.45	68.10	53.21	67.25	73.25	65.44
PromptCCD-B *w*/L2P (Ours)	51.59 ± 6.3	67.27 ± 8.7	46.14 ± 6.1	66.14 ± 2.3	81.05 ± 1.5	61.36 ± 3.2
PromptCCD-B *w*/DP (Ours)	**59.60** ± 1.2	78.93 ± 1.3	54.14 ± 1.6	**70.64** ± 1.3	83.46 ± 0.4	67.24 ± 1.8

Method	TinyImageNet			CUB		
	All	*Old*	*New*	*All*	*Old*	*New*
Frozen DINO	49.25	59.21	45.94	39.21	68.29	30.86
PromptCCD-B *w*/L2P (Ours)	56.66 ± 0.4	66.05 ± 0.8	53.69 ± 0.4	51.31 ± 1.0	72.43 ± 1.0	44.27 ± 1.4
PromptCCD-B *w*/DP (Ours)	**58.61** ± 1.5	66.61 ± 0.6	55.84 ± 1.7	**56.30** ± 1.1	78.64 ± 1.7	48.91 ± 1.1

[Intro.] Motivation

Limitation of current baseline

- ***Lack of explicit guidance*** (labels) may introduce bias when discovering categories from the continuous unlabelled data stream.
- ***Fixed-size prompt pool limits scalability*****.** The ablation study shows that changing the number of pool size does not significantly impact the CCD performance, even when the size matches the total classes in the dataset.
- There is ***no mechanism to dynamically estimate the number of categories***, which is a crucial and under-explored challenge in CCD task.

Ablation study on the prompt pool size using CUB datasets

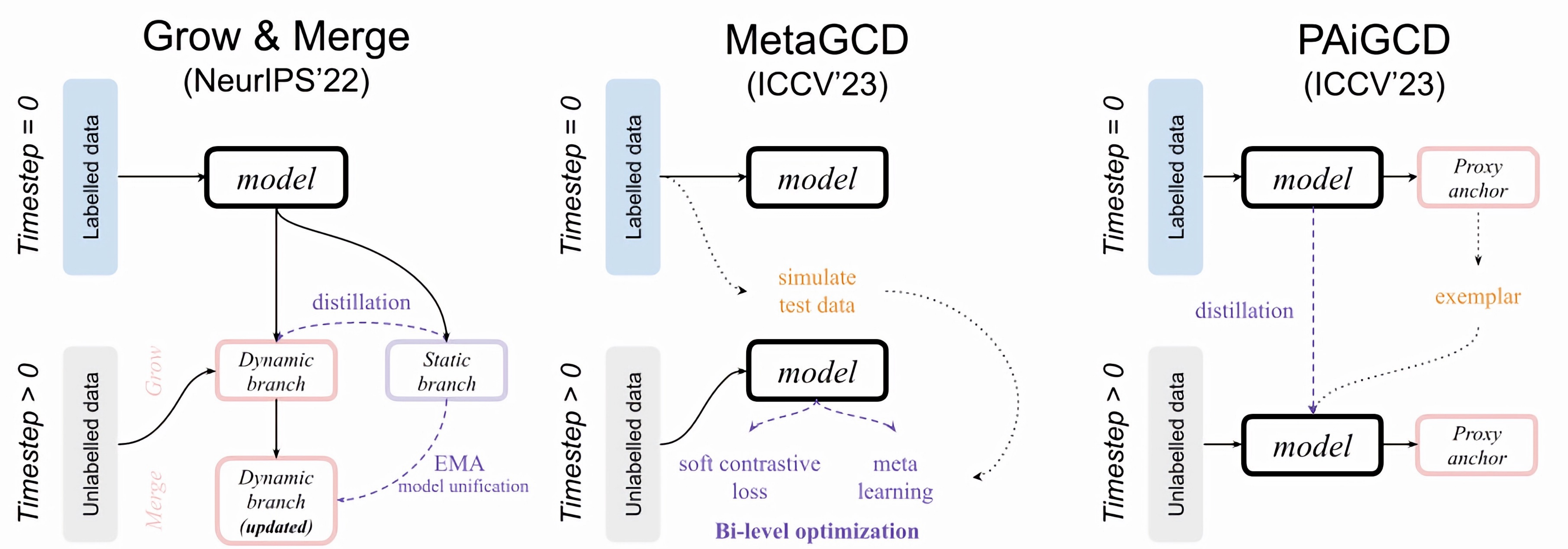
	w/ L2P			w/ DP		
Pool Size	*All*	*Old*	*New*	*All*	*Old*	*New*
5	48.69	70.12	41.51	55.54	77.78	48.21
10	50.57	73.22	43.28	55.21	77.24	48.04
20	49.32	70.60	42.29	55.41	76.31	48.18
40	48.59	69.26	41.43	53.97	77.38	46.26
100	51.84	73.09	44.39	55.12	77.26	47.82
200	49.40	71.79	41.94	54.54	79.05	46.29

[Intro.] Motivation

Goal:

To propose a ***parameter efficient prompt pool*** module specifically designed for adapting foundation models to address Continual Category Discovery.

[Related work] Current CCD solutions



Key idea:

Introduce dynamic and static branches where dynamic branch is used to learn unlabelled data incrementally (**growing phase**) while static branch is used to maintain prev. Knowledge by distillation and model unification (**merging phase**).

Key idea:

Introduce a meta-learning framework and leveraged offline labelled data to simulate testing incremental learninig process. Moreover, a soft contrastive neighbourhood loss is introduced to learn better feature representations.

Key idea:

Introduce a framework which makes use of proxy anchors module to retain knowledge from labelled data and generates high quality examplars from it.

[*PromptCCD*] Our solution

What solution did we come up with?

[*PromptCCD*] Our solution

Our simple solution: Gaussian Mixture Prompting Module (**GMP**)

[PromptCCD] Our solution

Our simple solution: Gaussian Mixture Prompting Module (**GMP**)

Our GMP learns a parameter-efficient, learnable GMM as a pool of prompts, leading to a new framework, called **PromptCCD**.

[PromptCCD] Our solution

GMP vs other prompt pools

- Unlike other prompt pools, GMP's prompt serves dual roles, i.e,. (1) **prompt to Instruct** and (2) **prompt as a class prototype** (explicit guidance, which does not exist on other prompt pool designs).

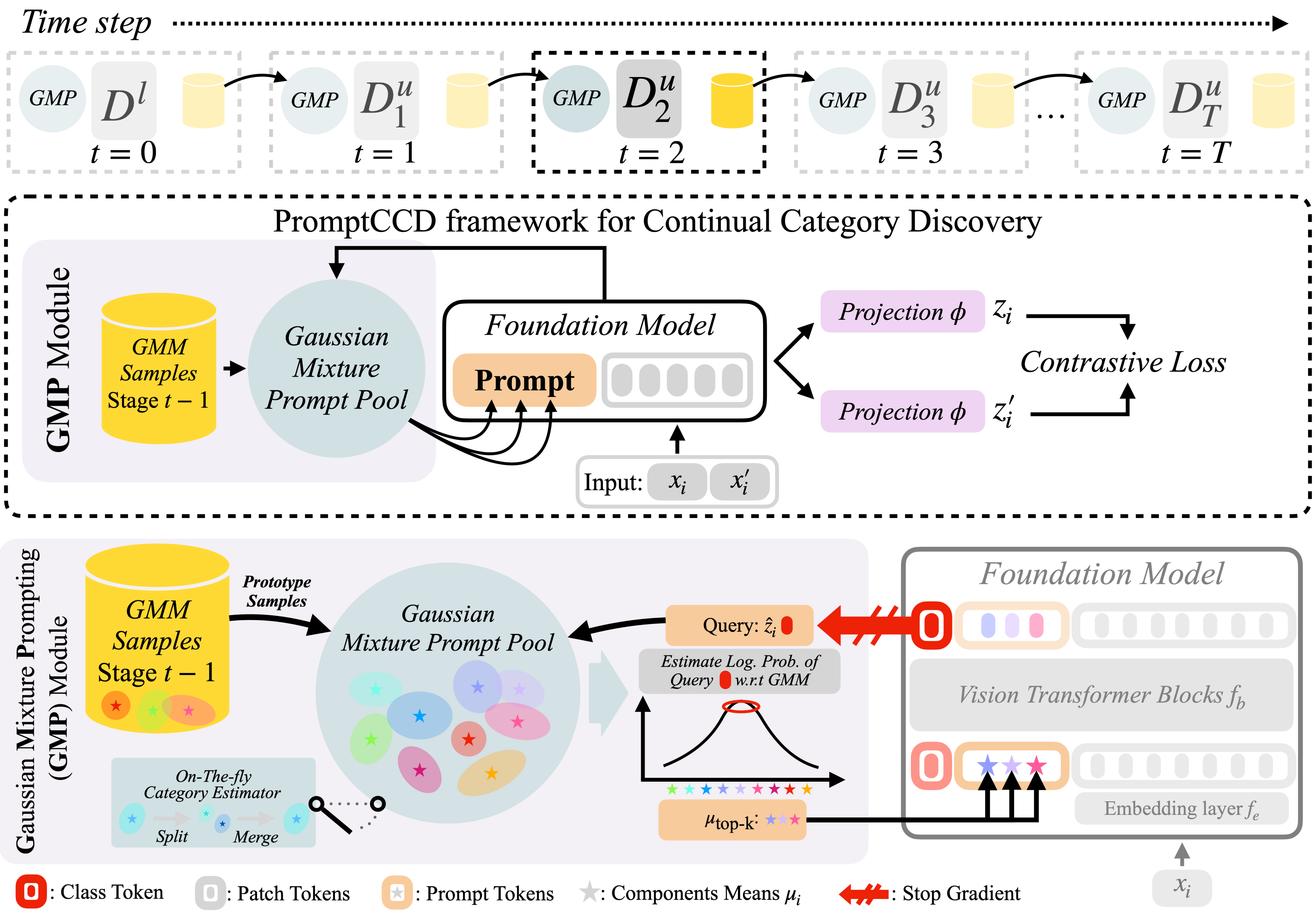
- GMP module **generates samples** to preserved the learned class prototypes for next incremental stage.

- GMP module enables **on-the-fly** category number estimation during discovery.

[PromptCCD] Our solution

How can we integrated GMP module to our PromptCCD framework?

[PromptCCD] Overall framework



[Intro.] Continual Category Discovery (CCD)

Experiment Details

[Intro.] Continual Category Discovery (CCD)

Stages	C100 [28]		IN-100 [43]		Tiny [30]		C-101 [12]		Aircraft [35]		SCars [27]		CUB [48]	
	C	#	C	#	C	#	C	#	C	#	C	#	C	#
Stage 0 (D^l)	70	$30.45K$	70	$77.46K$	140	$60.90K$	71	$4.70K$	70	$1.98K$	130	$4.62K$	140	$3.65K$
Stage 1 (D_1^u)	80	$5.95K$	80	$15.14K$	160	$11.90K$	81	$0.73K$	80	$0.37K$	152	$0.98K$	160	$0.71K$
Stage 2 (D_2^u)	90	$6.55K$	90	$16.66K$	180	$13.10K$	91	$0.65K$	90	$0.43K$	174	$1.13K$	180	$0.79K$
Stage 3 (D_3^u)	100	$7.05K$	100	$17.94K$	200	$14.10K$	101	$1.12K$	100	$0.55K$	196	$1.38K$	200	$0.85K$

CCD datasets & splits

- Above table shows the statistics of the CCD benchmarks datasets.
- While the table on right side shows how we split the classes for each stages.

Class splits	D^l	D_1^u	D_2^u	D_3^u
$\{y_i \mid y_i \leq 0.7 * \lvert\mathcal{Y}\rvert\}$	87%	7%	3%	3%
$\{y_i \mid 0.7 * \lvert\mathcal{Y}\rvert < y_i \leq 0.8 * \lvert\mathcal{Y}\rvert\}$	0%	70%	20%	10%
$\{y_i \mid 0.8 * \lvert\mathcal{Y}\rvert < y_i \leq 0.9 * \lvert\mathcal{Y}\rvert\}$	0%	0%	90%	10%
$\{y_i \mid 0.9 * \lvert\mathcal{Y}\rvert < y_i \leq \lvert\mathcal{Y}\rvert\}$	0%	0%	0%	100%

[Intro.] Continual Category Discovery (CCD)

Continual Accuracy (*cACC*) metric

- To evaluate the perfomance on CCD task, we make use of semi-supervised K-means algorithm (SS-k-means).

- SS-k-means uses labelled data from both the initial stage and the unlabelled data with assigned labels from the previous stage to assist the clustering algorithm in identifying both known and novel categories.

- High quality label assignments facilitate the subsequent category discovery while low quality label assignments accumulate errors for the sebsequent category discovery.

Algorithm 1 *Continual ACC* (*cACC*) evaluation metric

Input: Models $\{f_\theta^t \mid t = 1, \ldots, T\}$ and datasets $\{D^l, D^u\}$.
Output: $cACC$ value.
Require: SS-k-MEANS(Model, Labelled set, Unlabelled set).
Require: Initialize set $\mathbb{A}^L \leftarrow D^l$.
for $t \in \{1, \cdots, T\}$ **do**
    $ACC_t\ ,\ D_t^{u^*} \leftarrow$ SS-k-MEANS(f_θ^t, $\mathbb{A}^L$, D_t^u)
    $\mathbb{A}^L \leftarrow \mathbb{A}^L \cup D_t^{u^*}$ *// append $D_t^{u^*}$ (w/ assigned labels) to $\mathbb{A}^L$*
$ACCs \leftarrow \{ACC_t \mid t = 1, \ldots, T\}$
$cACC \leftarrow$ AVERAGE($ACCs$)
return $cACC$

[*Analysis*] PromptCCD variants

Variants of PromptCCD

- ***Comparison with PromptCCD variants and frozen DINO:*** This table further highlights the need to adapt foundation models for CCD.
- ***Performance:*** Our PromptCCD w/GMP consistently outperforms other models.

CCD results comparison between fronzen DINO, PromptCCD-B, and our proposed PromptCCD w/ GMP.

	CIFAR100			ImageNet-100		
Method	*All*	*Old*	*New*	*All*	*Old*	*New*
Frozen DINO	56.45	68.10	53.21	67.25	73.25	65.44
PromptCCD-B w/L2P (Ours)	51.59 ± 6.3	67.27 ± 8.7	46.14 ± 6.1	66.14 ± 2.3	81.05 ± 1.5	61.36 ± 3.2
PromptCCD-B w/DP (Ours)	59.60 ± 1.2	78.93 ± 1.3	54.14 ± 1.6	70.64 ± 1.3	83.46 ± 0.4	67.24 ± 1.8
PromptCCD w/GMP (Ours)	**63.97** ± 1.4	76.67 ± 2.6	60.01 ± 1.7	**75.38** ± 0.7	81.16 ± 0.7	73.71 ± 0.8

	TinyImageNet			CUB		
Method	*All*	*Old*	*New*	*All*	*Old*	*New*
Frozen DINO	49.25	59.21	45.94	39.21	68.29	30.86
PromptCCD-B w/L2P (Ours)	56.66 ± 0.4	66.05 ± 0.8	53.69 ± 0.4	51.31 ± 1.0	72.43 ± 1.0	44.27 ± 1.4
PromptCCD-B w/DP (Ours)	58.61 ± 1.5	66.61 ± 0.6	55.84 ± 1.7	56.30 ± 1.1	78.64 ± 1.7	48.91 ± 1.1
PromptCCD w/GMP (Ours)	**61.15** ± 1.0	66.29 ± 2.0	58.83 ± 1.0	**56.65** ± 1.0	79.88 ± 2.5	48.96 ± 0.8

[*Analysis*] CCD benchmarks

CCD benchmark results

- Comparison with other representative CCD and GCD methods for CCD task.
- Experiment on different datasets (generic & finegrained) and foundation (pretrained) models.
- Evaluate using *cACC* metric.
- Overall, our PromptCCD w/GMP outperforms other approaches across all datasets.

Method	Pretrained Model	CIFAR100			ImageNet-100			TinyImageNet			Caltech-101		
		All	*Old*	*New*	*All*	*Old*	*New*	*All*	*Old*	*New*	*All*	*Old*	*New*
ORCA	DINO	60.91	66.61	58.33	40.29	45.85	35.40	54.71	63.13	51.93	76.77	82.80	73.20
GCD	DINO	58.18	72.27	52.83	69.41	81.56	65.65	55.20	65.87	51.61	78.27	86.60	72.92
SimGCD	DINO	25.56	38.76	20.43	31.38	40.47	27.44	33.40	29.11	34.74	33.65	37.53	31.62
GCD *w*/replay	DINO	49.93	73.15	41.47	72.04	83.75	69.01	56.33	67.54	52.60	76.51	86.14	72.48
SimGCD *w*/replay	DINO	40.13	66.72	30.91	47.53	67.86	39.18	37.45	58.15	30.36	49.38	52.72	47.99
Grow & Merge	DINO	57.43	63.68	55.31	67.84	75.10	66.60	52.14	59.68	49.96	75.75	83.66	71.59
MetaGCD	DINO	55.49	69.38	48.98	66.41	80.54	60.65	55.26	66.12	50.79	80.75	89.02	75.86
PA-CGCD	DINO	58.25	87.11	49.04	64.79	91.15	57.83	51.13	74.95	43.52	77.96	94.75	69.66
PromptCCD *w*/GMP (Ours)	DINO	**64.17**	75.57	60.34	**76.16**	81.76	74.35	**61.84**	66.54	60.26	**82.44**	89.08	79.72
GCD	DINOv2	65.35	77.06	60.46	71.58	83.02	68.05	59.05	77.44	53.41	83.00	88.65	79.80
MetaGCD	DINOv2	52.10	79.64	43.13	70.20	82.62	64.66	56.15	74.69	49.37	83.05	88.08	80.89
PA-CGCD	DINOv2	54.36	79.19	45.65	74.82	88.20	72.02	52.10	68.07	46.32	83.06	94.07	77.55
PromptCCD *w*/GMP (Ours)	DINOv2	**69.73**	78.01	66.16	**76.28**	82.61	74.53	**68.20**	75.56	65.23	**83.86**	87.93	81.42

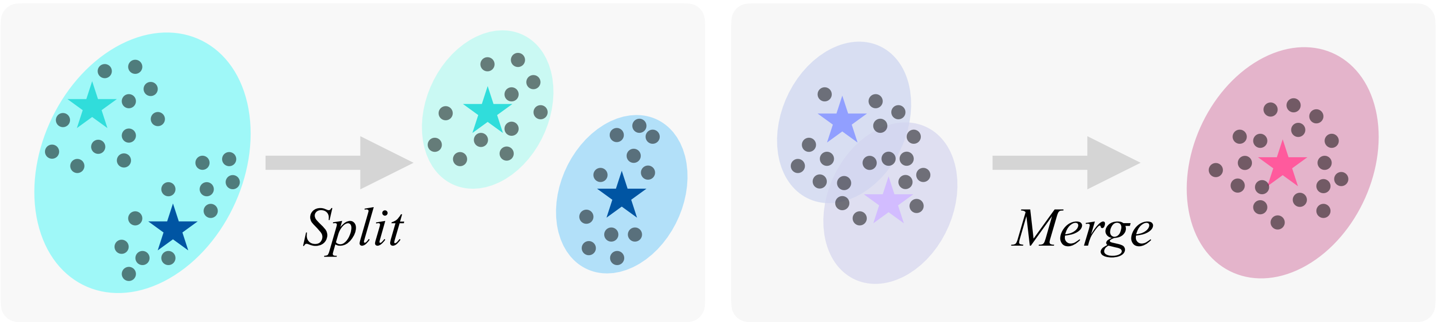
Method	Pretrained Model	Aircraft			Stanford Cars			CUB			**Avg. results**		
		All	*Old*	*New*	*All*	*Old*	*New*	*All*	*Old*	*New*	*All*	*Old*	*New*
ORCA	DINO	30.77	25.71	32.44	20.79	33.40	17.60	41.73	66.19	34.14	46.57	54.81	43.29
GCD	DINO	47.37	61.43	42.53	39.21	58.29	33.45	54.98	75.47	48.15	57.52	71.64	52.45
SimGCD	DINO	29.03	35.72	25.61	21.01	40.93	16.48	39.89	59.25	33.75	30.56	40.25	27.15
GCD *w*/replay	DINO	45.63	62.38	39.89	39.87	58.18	33.89	54.66	74.64	47.81	56.42	72.25	51.02
SimGCD *w*/replay	DINO	37.44	61.43	28.96	22.76	49.04	16.65	42.08	72.65	31.92	39.54	61.22	32.28
Grow & Merge	DINO	31.06	33.33	30.78	21.90	35.29	18.17	38.87	65.00	30.29	49.28	59.39	46.10
MetaGCD	DINO	44.63	59.05	39.39	35.98	56.97	29.96	44.59	74.40	35.40	54.73	70.78	48.72
PA-CGCD	DINO	48.24	73.09	40.60	43.88	80.43	33.54	52.48	77.26	44.74	56.68	82.68	48.42
PromptCCD *w*/GMP (Ours)	DINO	**52.64**	60.48	50.23	**44.07**	66.36	36.83	**55.45**	75.48	48.56	**62.40**	73.61	58.62
GCD	DINOv2	57.87	63.80	55.39	58.52	71.65	53.80	66.70	83.33	60.81	66.01	77.85	61.67
MetaGCD	DINOv2	54.90	64.29	52.08	57.16	71.87	52.01	62.19	82.50	55.13	62.25	77.67	56.75
PA-CGCD	DINOv2	58.15	77.62	51.08	64.91	89.64	57.84	66.88	92.62	58.48	64.90	84.20	58.42
PromptCCD *w*/GMP (Ours)	DINOv2	**62.71**	68.33	60.82	**65.08**	76.60	60.75	**67.81**	81.55	62.81	**70.52**	78.66	67.39

[*Analysis*] CCD benchmarks

CCD benchmark results
(when category number is unknown)

	Est. method	CIFAR100			ImageNet-100			TinyImageNet			CUB		
Category discovery at stage --→		1	2	3	1	2	3	1	2	3	1	2	3
Estimated category C	GPC	85	100	115	83	98	113	155	170	185	161	180	198
Ground truth category C	–	80	90	100	80	90	100	160	180	200	160	180	200
Methods		*All*	*Old*	*New*	*All*	*Old*	*New*	*All*	*Old*	*New*	*All*	*Old*	*New*
GCD	GPC	53.78	74.05	46.37	68.55	82.05	63.96	55.28	65.04	52.15	50.69	72.43	43.16
Grow & Merge	GPC	53.33	66.64	49.61	66.40	74.52	64.01	52.40	57.87	51.00	38.12	62.21	30.00
MetaGCD	GPC	47.55	70.79	38.57	63.48	80.82	56.28	56.21	68.33	50.99	44.30	70.69	35.83
PA-CGCD	GPC	55.66	90.21	44.99	66.74	91.28	58.97	50.55	72.44	43.38	52.27	76.38	44.24
PromptCCD-U w/GMP (Ours)	GPC	**59.12**	77.62	53.70	**70.12**	81.84	66.12	**57.76**	64.57	55.37	**55.20**	73.19	48.82

- Considering the continual nature of CCD, it would be better to estimate category number *on-the-fly*.



- As our GMP module is based on Gaussian Mixture Model (GMM), we incorporate the GMM-based category number estimation method in our framework. Specifically, the idea is to automatically **splitting** and **merging** the GMM's clusters during learning by assessing the cluster's compactness and separability using MCMC algorithm.

- For the sake comparison, we also compare our method with others by directly using the estimated category number from our model.

[*Analysis*] PromptCCD model analysis

PromptCCD	CIFAR100 Avg. *ACC* *All*	*Old*	*New*	ImageNet-100 Avg. *ACC* *All*	*Old*	*New*
w/o GMP	58.18	72.27	52.83	69.41	81.56	65.65
GMP (random-k)	59.98	73.81$^{+1.54}$	55.68$^{+2.85}$	68.30	80.09$^{-1.47}$	63.50$^{-2.15}$
GMP (top-k) (Ours)	**64.17**	75.57$^{+3.30}$	60.34$^{+7.51}$	**76.16**	81.76$^{+0.20}$	74.35$^{+8.70}$
PromptCCD	**TinyImageNet Avg. *ACC* *All***	***Old***	***New***	**CUB Avg. *ACC* *All***	***Old***	***New***
w/o GMP	55.20	65.87	51.61	54.98	75.47	48.15
GMP (random-k)	55.69	63.95$^{-1.92}$	52.52$^{+0.91}$	51.46	73.10$^{-2.37}$	43.90$^{-4.25}$
GMP (top-k) (Ours)	**61.84**	66.54$^{+0.67}$	60.26$^{+8.65}$	**55.45**	75.48$^{+0.01}$	48.56$^{+0.41}$

top-k Prompts	GMM Samples	C100 Avg. *ACC* *All*	*Old*	*New*	CUB Avg. *ACC* *All*	*Old*	*New*
0	0	58.18	72.27	52.83	54.98	75.47	48.15
5	0	61.48	74.68	57.55	53.54	74.28	46.47
5	20	62.21	75.71	57.90	54.37	74.88	46.41
5	200	61.00	72.46	57.08	51.67	73.33	44.08
2	100	61.39	73.04	57.64	53.36	73.45	46.04
<u>5</u>	<u>100</u>	**64.17**	75.57	60.34	**55.45**	75.48	48.56
10	100	61.03	72.91	56.97	52.76	71.67	46.02

Method	Prompt	*All*	*Old*	*New*
G&M	w/o GMP	57.43	63.68	55.31
G&M	$w/$GMP	**61.14**	**64.94**	**59.10**
MetaGCD	w/o GMP	55.49	69.38	48.98
MetaGCD	$w/$GMP	**58.99**	**69.69**	**54.29**
PromptCCD	w/o GMP	58.18	72.27	52.83
PromptCCD	$w/$GMP	**64.17**	**75.57**	**60.34**

Model analysis

- ***Left table:*** to validate the effectiveness of using top-k prompts.
- ***Middle table:*** The choices of number for top-k prompts and GMM samples
- ***Right table:*** Improving other CCD methods with our GMP.

Conclusion

- **PromptCCD:** A novel prompt learning framework for CCD, adapting vision foundation models to address CCD.
- **Gaussian Mixture Prompting (GMP):** A specialized prompting module for CCD utilizing learnable GMM as a pool of prompts. Additionally, GMP enables on-the-fly category estimation making it well suited for handling CCD task.
- **Performance:** PromptCCD achieves state of the art results in CCD benchmark.

Visual AI Lab

香 港 大 學
THE UNIVERSITY OF HONG KONG

THE UNIVERSITY of EDINBURGH

Q&A

visual-ai.github.io/promptccd/

Project page

(Including Paper and code)