PromptCCD: Learning Gaussian Mixture Prompt Pool for Continual Category Discovery

Fernando Julio Cendra[1], Bingchen Zhao[2], Kai Han[1]

[1]Visual AI Lab, The University of Hong Kong
[2]The University of Edinburgh

Visual AI Lab

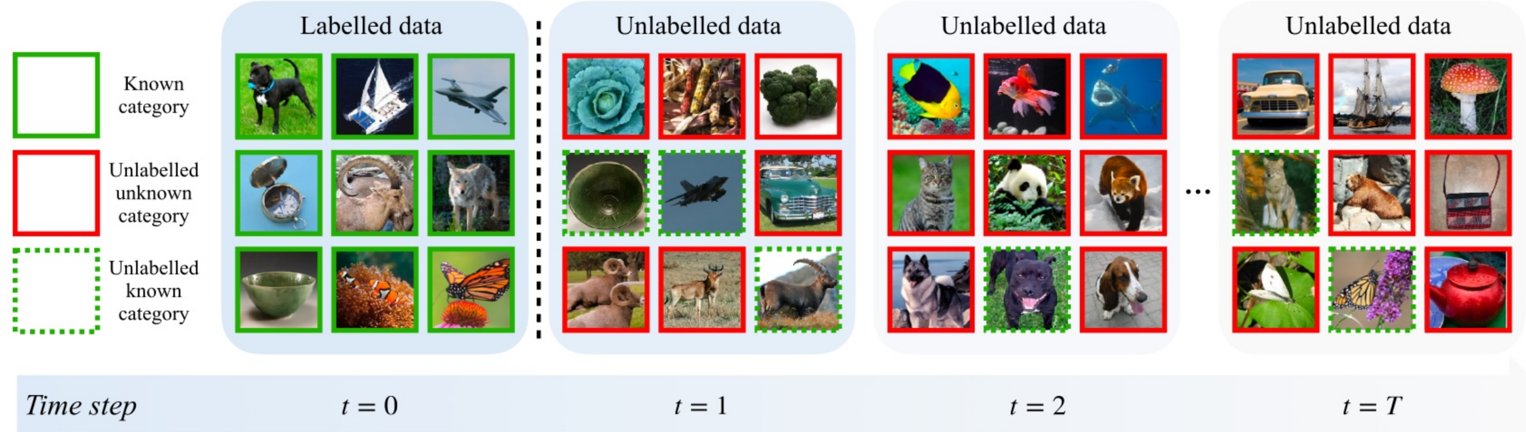
香 港 大 學
THE UNIVERSITY OF HONG KONG

THE UNIVERSITY of EDINBURGH

ECCV

Project Page

Problem Formulation & Motivation



- In ***Continual Category Discovery (CCD)***, the model receives the labelled set at the initial stage and is tasked to discover categories from the continuous unlabelled data stream in the subsequent stages. There are two major challenges in CCD:
  - ***Category Discovery*** of known and novel categories in the unlabelled stream of data.
  - ***Catastrophic forgetting***, a well-known issue even in conventional continual learning.
- Recently, ***vision foundation models*** have shown remarkable performance on various vision task. Here, we aim to unleash the potential of such models for CCD.

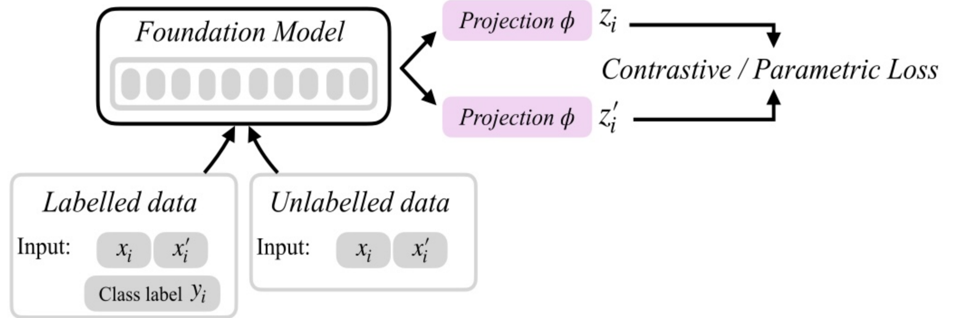
Contributions

- **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.

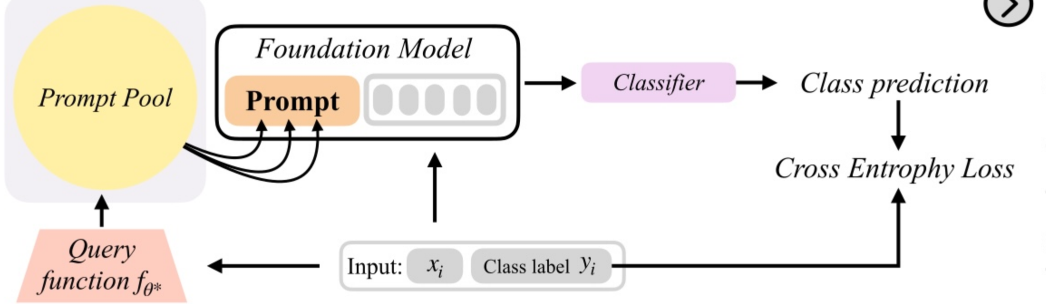
Related Work

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



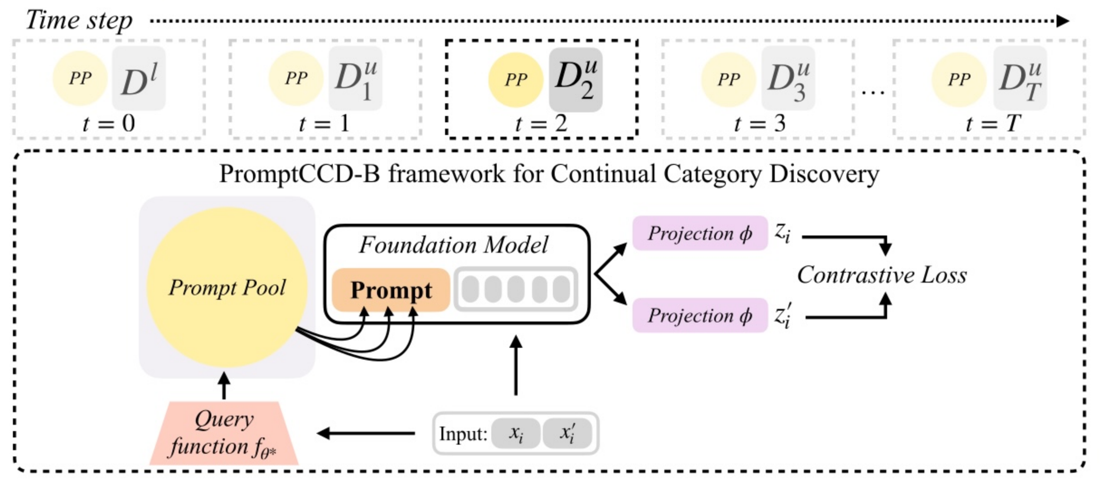
- The model leverages a prompt pool (e.g., L2P, Dual Prompt) to guide a foundation model in supervised continual learning.

  The model extracts query feature to retrieve top-k most relevant prompts from the pool, which, along with class labels, are used to supervise the model.

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

PromptCCD

PromptCCD-B (*Baseline*): Learning Prompt Pool for CCD

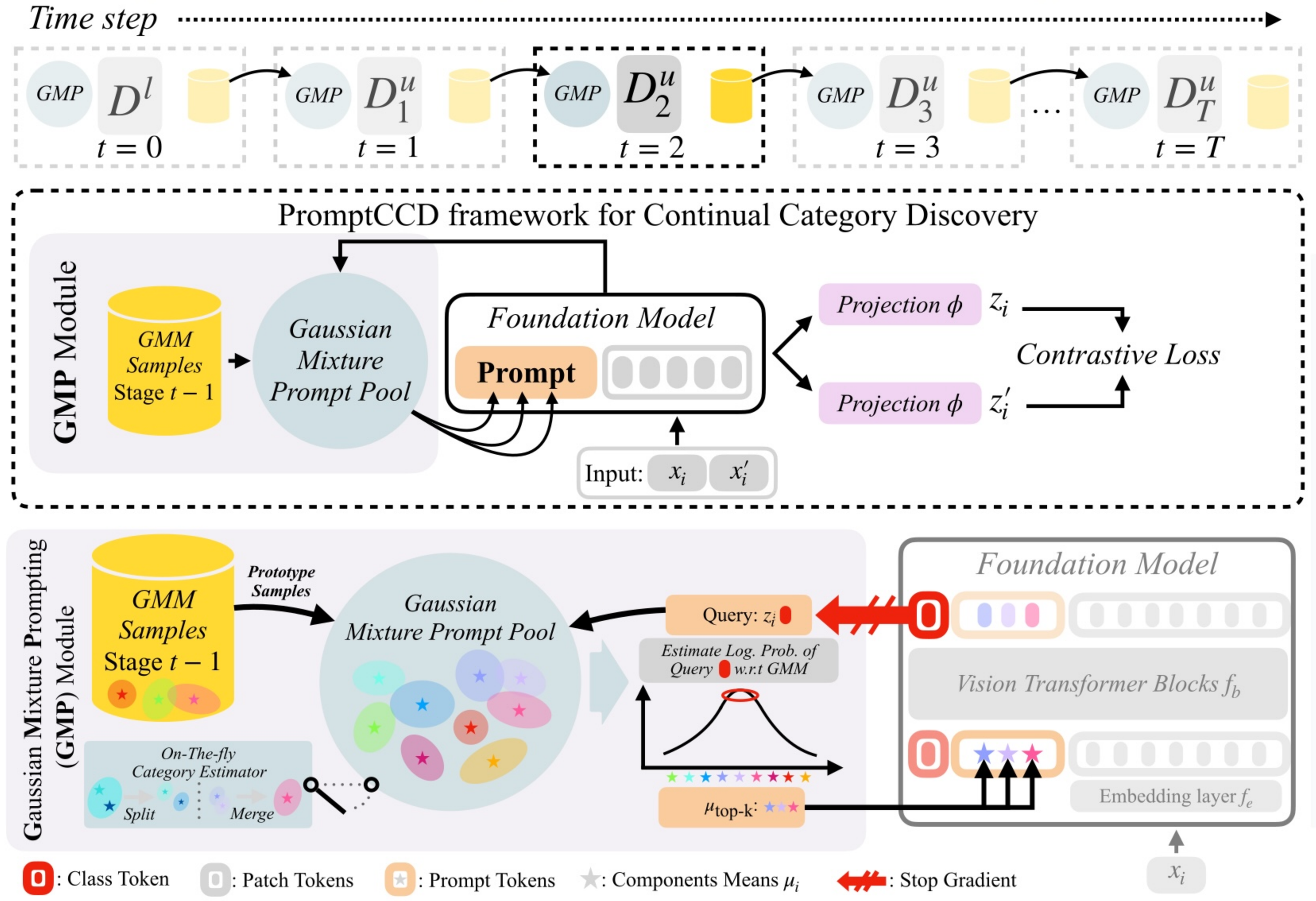


Our baseline adopts **prompt pool module** to adapt vision foundation model for CCD.

Method:
Our baseline extracts query feature which will be used to retrieve top-k most relevant prompts from the prompt pool. These prompts are then used to guide the representation learning process.

Limitation:
(1) Lack of *explicit guidance*, i.e., class labels, which may introduce representation bias,
(2) Fixed-size prompt pool limits scalability,
(3) No mechanism to estimate the number of categories.

PromptCCD: Learning Gaussian Mixture Prompt Pool for CCD



To address the aforementioned limitations in our baseline framework, here, we propose a novel **Gaussian Mixture Prompting (GMP) module**.

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

Key Methods:

- PromptCCD addresses CCD by making use of GMP module to estimate the input **class token**'s feature and retrieves **top-k GMM mean of components** as prompts ($\mu_{\text{top-k}}$) to guide the vision foundation model.

GMP's prompt serves dual roles:
(1) ***task prompt*** (*to instruct model*)
(2) ***class prototypes*** (*explicit guidance*)

- To retain previously learned prompts, it generates prototype samples from the fitted GMM and fit the current GMM with these samples.

PromptCCD-U: Unknown Number of Classes in Unlabelled Data

When the class number in the unlabelled stream of data is unknown, we introduce **PromptCCD-U**, by incorporating a GMM-based category estimation method in our GMP module.
(w/o introducing any extra models or an offline process)

GMP enables **on-the-fly estimation of category numbers** by automatically splitting and merging clusters during learning through assessing the cluster's compactness and seperability.

Comparison with PromptCCD variants:

- PromptCCD *w/*GMP module consistently **outperforms** other PromptCCD variants.
- Our scalable **GMP module** leverages a learnable GMM to construct a flexible pool of prompts which is not restricted to a particular prompt pool size.
- Moreover, unlike other prompt pool designs, GMP offers **on-the-fly category estimation** which is essential for CCD.

Method	CIFAR100 *All*	CIFAR100 *Old*	CIFAR100 *New*	ImageNet-100 *All*	ImageNet-100 *Old*	ImageNet-100 *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

Method	TinyImageNet *All*	TinyImageNet *Old*	TinyImageNet *New*	CUB *All*	CUB *Old*	CUB *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

CCD Benchmark

CCD benchmarks results: Comparison with other methods for CCD leveraging pretrained DINO & DINOv2 models on generic and fine-grained datasets.

Category Number Known

Method	Pretrained Model	CIFAR100 *All*	CIFAR100 *Old*	CIFAR100 *New*	ImageNet-100 *All*	ImageNet-100 *Old*	ImageNet-100 *New*	TinyImageNet *All*	TinyImageNet *Old*	TinyImageNet *New*	Caltech-101 *All*	Caltech-101 *Old*	Caltech-101 *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 *All*	Aircraft *Old*	Aircraft *New*	Stanford Cars *All*	Stanford Cars *Old*	Stanford Cars *New*	CUB *All*	CUB *Old*	CUB *New*	**Avg. results** *All*	**Avg. results** *Old*	**Avg. results** *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

Category Number 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	–	<u>80</u>	<u>90</u>	<u>100</u>	<u>80</u>	<u>90</u>	<u>100</u>	<u>160</u>	<u>180</u>	<u>200</u>	<u>160</u>	<u>180</u>	<u>200</u>
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

More Analysis

PromptCCD Model Component Analysis

top-k Prompts	GMM Samples	C100 Avg. *ACC* *All*	C100 Avg. *ACC* *Old*	C100 Avg. *ACC* *New*	CUB Avg. *ACC* *All*	CUB Avg. *ACC* *Old*	CUB Avg. *ACC* *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

Effectiveness of GMP module on different models

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**