Revisiting Class-Incremental Learning with Pre-Trained Models: Generalizability and Adaptivity are All You Need

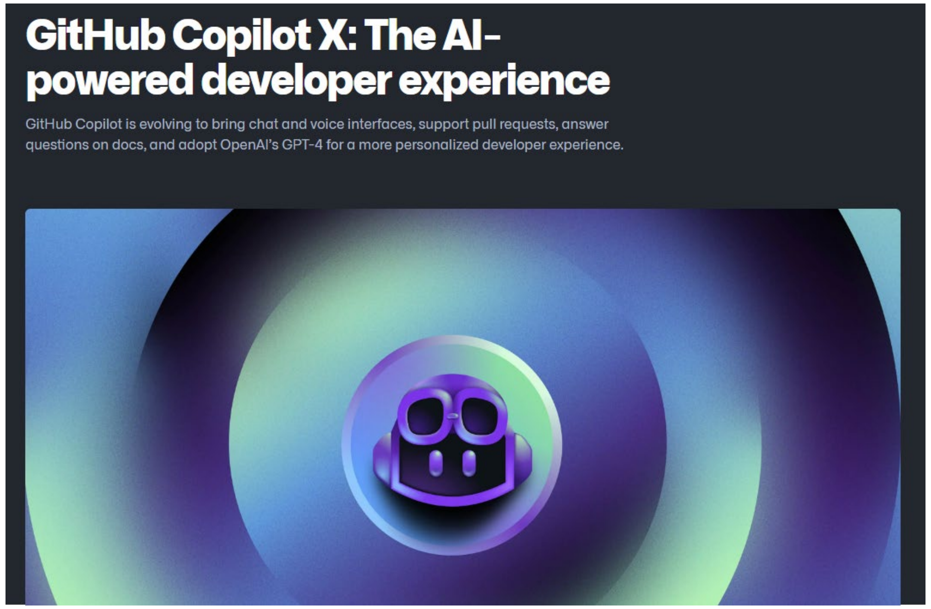
Da-Wei Zhou[1] Zi-Wen Cai[1] Han-Jia Ye[1] De-Chuan Zhan[1] Ziwei Liu[2]

[1]School of Artificial Intelligence, Nanjing University [2]S-Lab, NTU

Outline

- **Background of Continual Learning**
- Related Work
- Aper: Adapt and Merge
- Experiments
- Take Home Message

Recent Advances in AI



GitHub Copilot X: The AI-powered developer experience

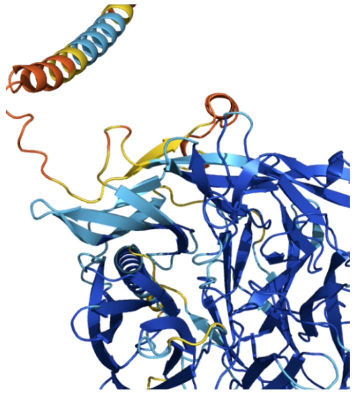
GitHub Copilot is evolving to bring chat and voice interfaces, support pull requests, answer questions on docs, and adopt OpenAI's GPT-4 for a more personalized developer experience.

ChatGPT: New AI chatbot has everyone talking to it

7 December 2022

AlphaFold is an AI system developed by DeepMind that predicts a protein's 3D structure from its amino acid sequence. It regularly achieves accuracy competitive with experiment.

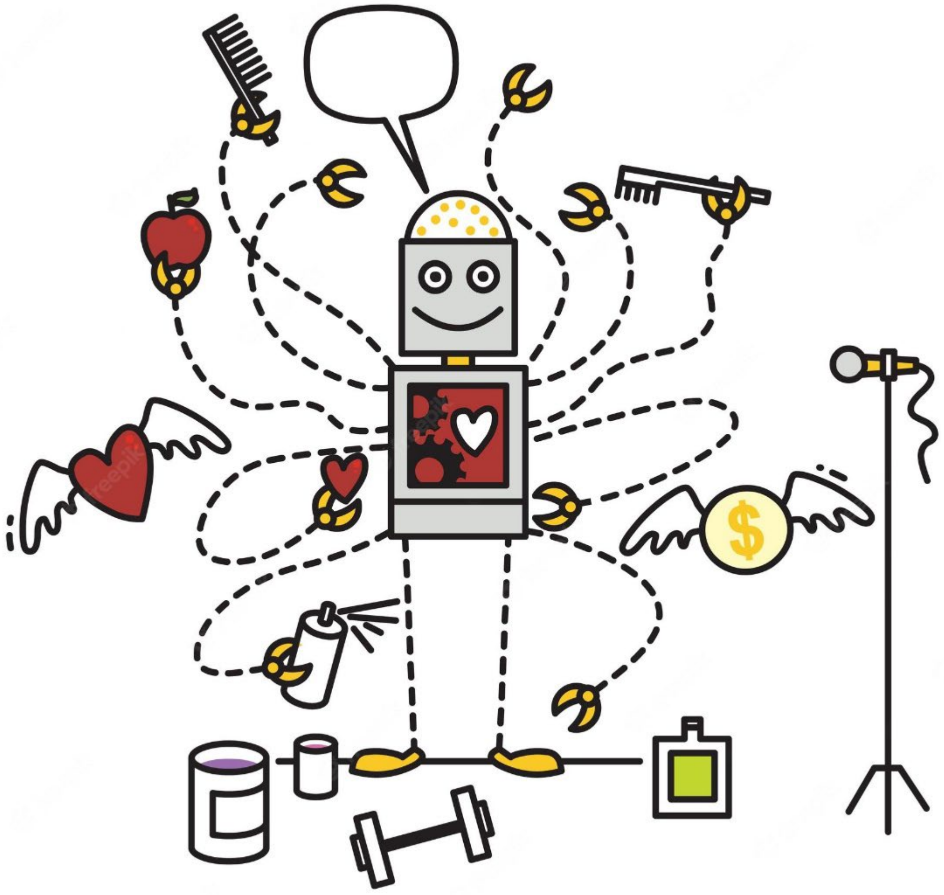
DeepMind and EMBL's European Bioinformatics Institute (EMBL-EBI) have partnered to create AlphaFold DB to make these predictions freely available to the scientific community. The latest database release contains over 200 million entries, providing broad coverage of UniProt (the standard repository of protein sequences and annotations). We provide individual downloads for the human proteome and for the proteomes of 47 other key organisms important in research and global health. We also provide a download for the manually curated subset of UniProt (Swiss-Prot).



Q8I3H7: May protect the malaria parasite against attack by the immune system. Mean pLDDT 85.57.

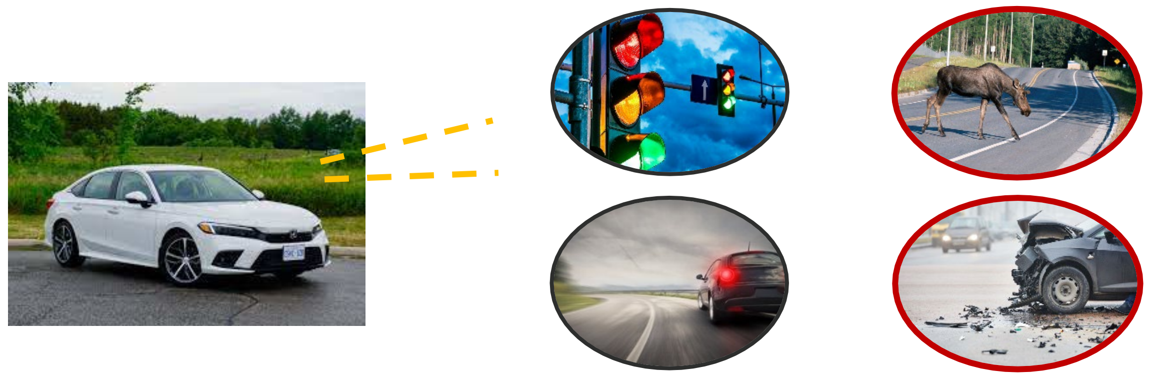
THE NOBEL PRIZE IN CHEMISTRY 2024
David Baker
"for computational protein design"
Demis Hassabis
John M. Jumper
"for protein structure prediction"
THE ROYAL SWEDISH ACADEMY OF SCIENCES

Can AI Manage Multiple Tasks Like Human?



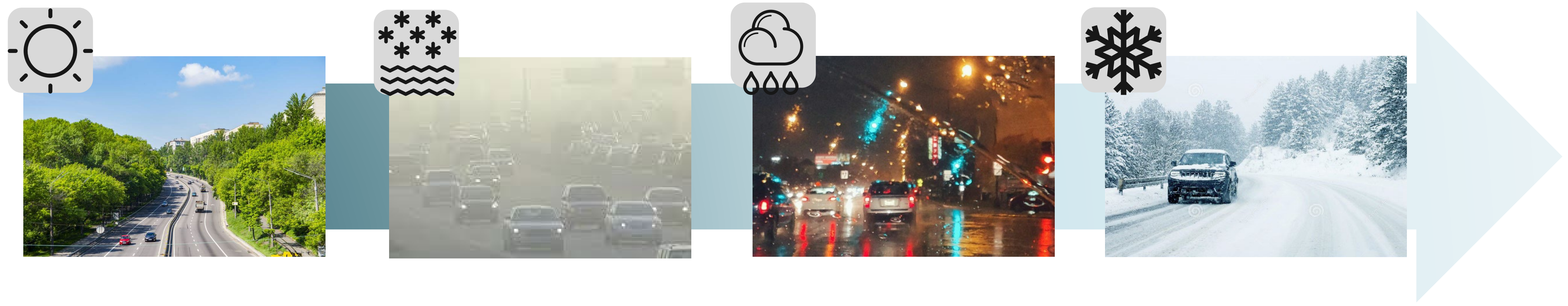
Real World APPs Require Continual Learning

......



Close-word training

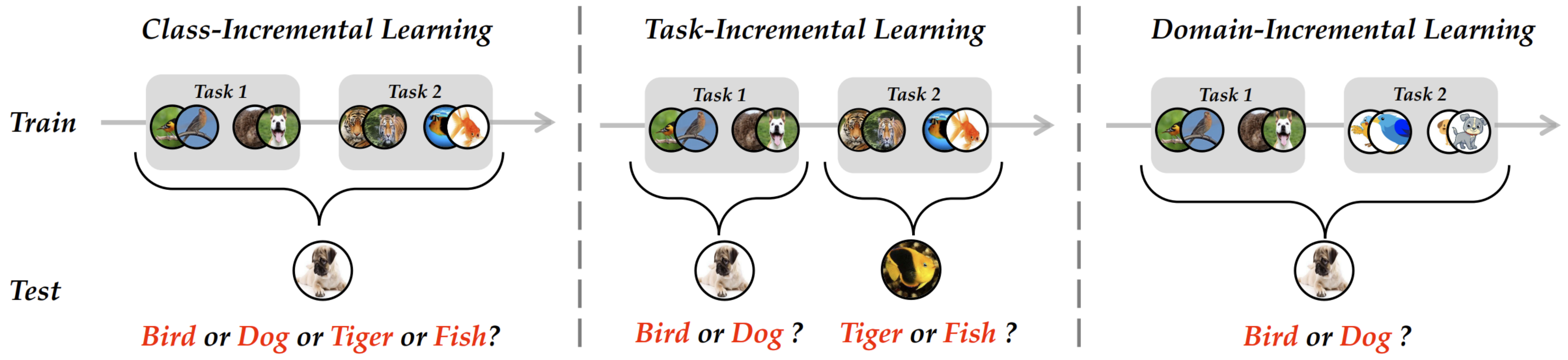
Real-world Application



New scenes will emerge, and self-driving vehicles need updating.

Problem Definition

- *Class-Incremental Learning (CIL)*: classify among all seen classes
- *Task-Incremental Learning (TIL):* classify among each (given) task
- *Domain-Incremental Learning (DIL):* classify among all distributions

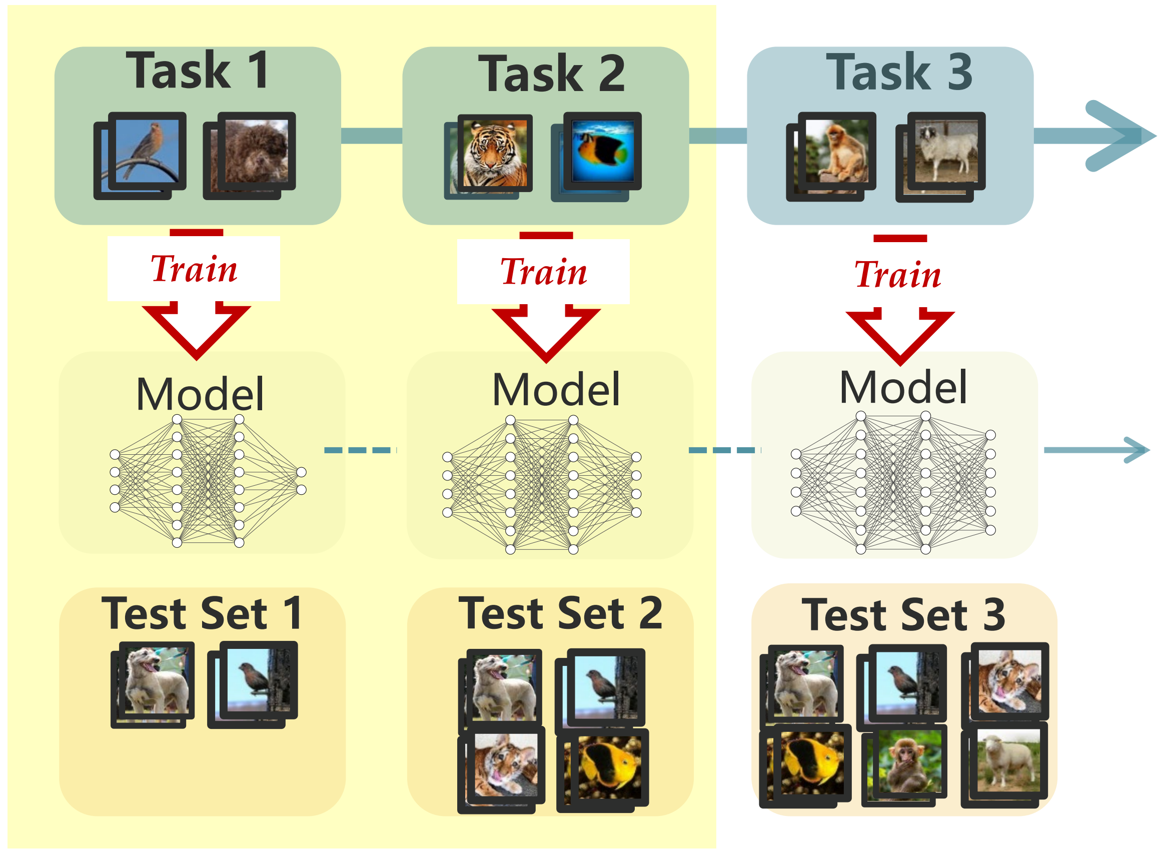


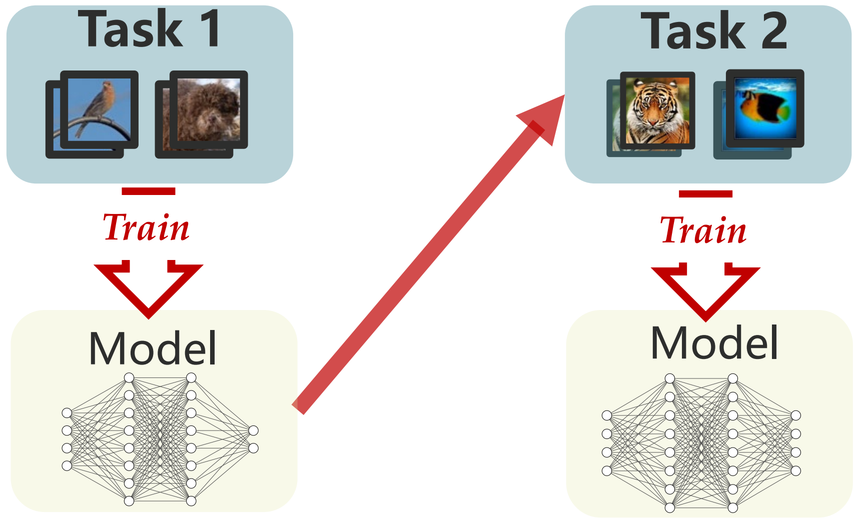
Class-Incremental Learning
Task-Incremental Learning
Domain-Incremental Learning
Train
Task 1
Task 2
Test
Bird or Dog or Tiger or Fish?
Bird or Dog ?
Tiger or Fish ?
Bird or Dog ?

Da-Wei Zhou et al., *Class-Incremental Learning: A Survey.* TPAMI 2024

Expected capabilities of continual learning

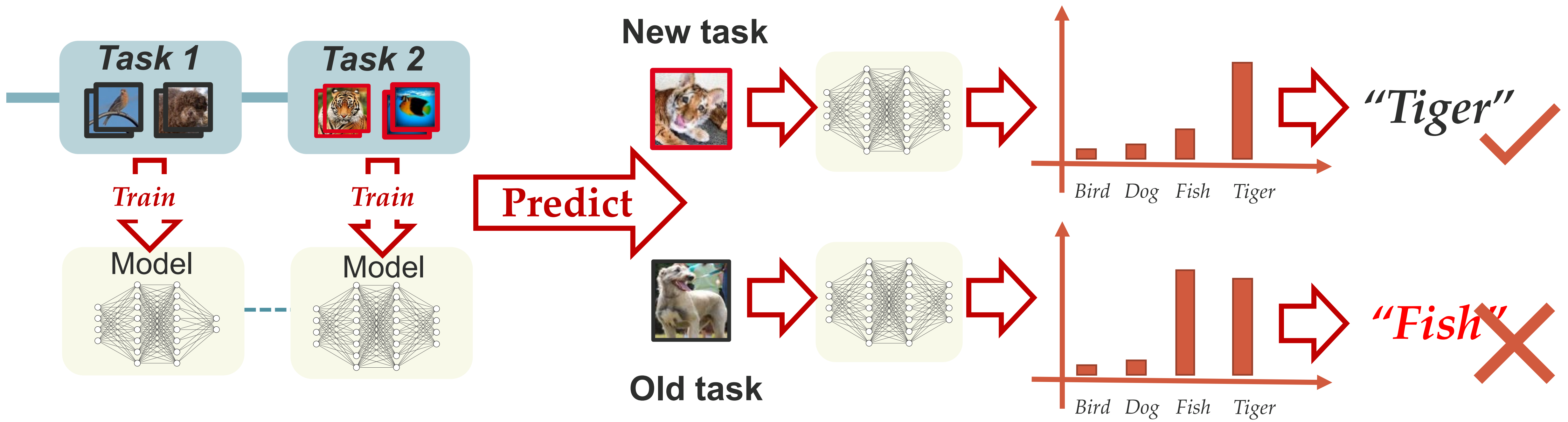
- Target: obtain the knowledge of all tasks seen so far





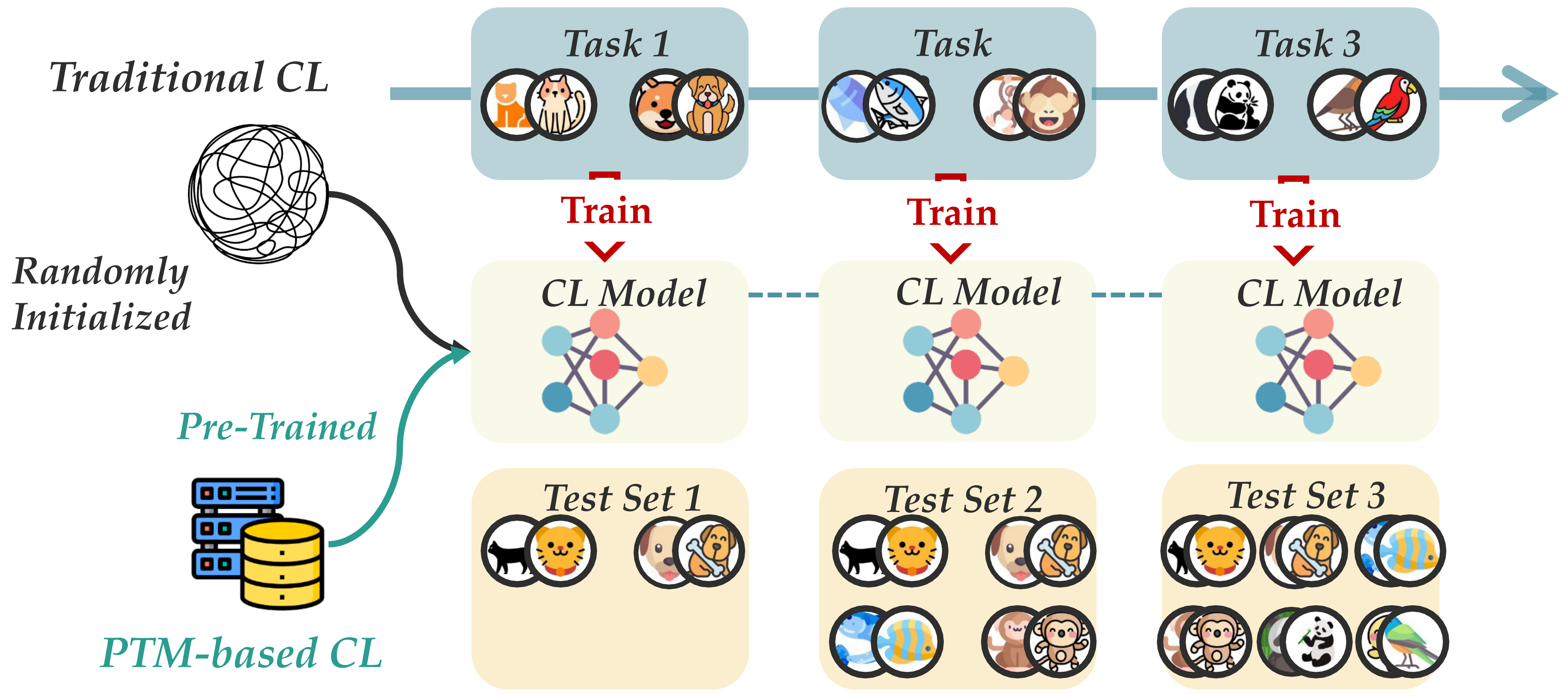
How about *finetuning* the current model with new data?

Catastrophic forgetting



Continual learning of new tasks will *erase* the semantic information of former tasks when learning new tasks.

Continual Learning w/ and w/o PTMs



Recent advances focus on the scenario with strong PTMs

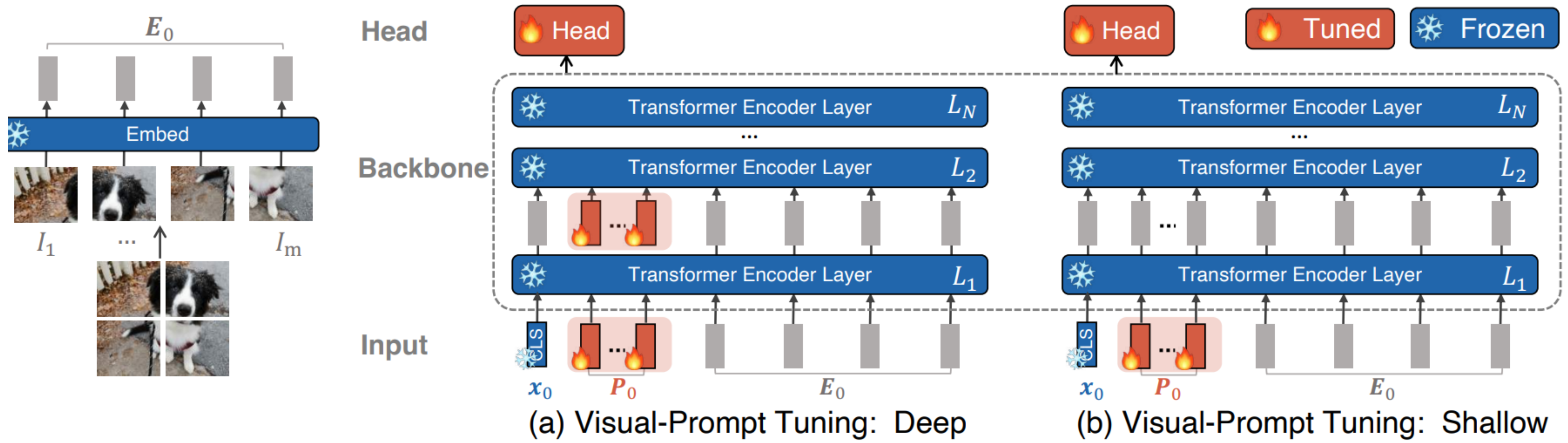
Da-Wei Zhou et al., *Continual Learning with Pre-Trained Models: A Survey.* IJCAI 2024

Outline

- Background of Continual Learning
- **Related Work**
- Aper: Adapt and Merge
- Experiments
- Take Home Message

Related Work: Prompt-based Methods

- **Core Idea**: utilize *lightweight trainable modules, e.g.,* prompts, to adjust the PTM

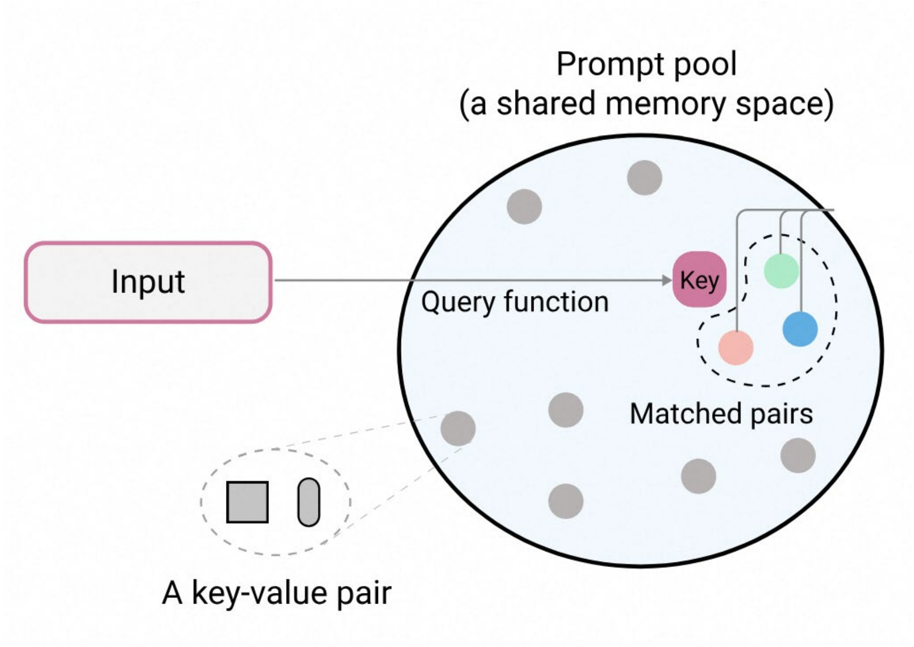


E_0
Embed
I_1
...
I_m
Head
Backbone
Input
Head
Transformer Encoder Layer L_N
Transformer Encoder Layer L_2
Transformer Encoder Layer L_1
CLS
x_0
P_0
E_0
(a) Visual-Prompt Tuning: Deep
Head
Tuned
Frozen
(b) Visual-Prompt Tuning: Shallow

Prompts are learnable tokens in the model, which enables model tuning with pre-trained weights frozen.

Menglin Jia et al., *Visual Prompt Tuning*. ECCV 2022.

Prompt Pool



Key: learnable
Value: Prompt parameter

- Saving a ***prompt pool***, which collects a set of prompts [Wang et al., CVPR'22]

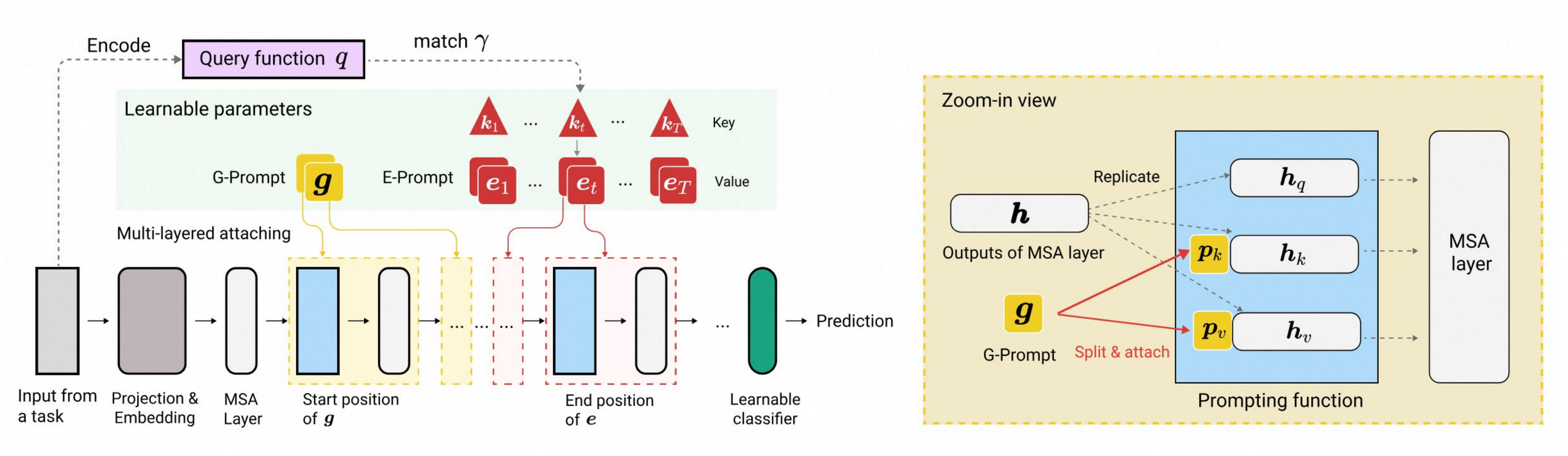
$$\min_{\boldsymbol{P} \cup W} \sum_{(\mathbf{x},y)\in\mathcal{D}^b} \ell(W^\top \phi(\mathbf{x};\boldsymbol{P}), y),$$

- Prompts are organized as ***key-value pairs*** to enable prompt selection
- The prompt pool can be seen as the ***external memory***, enabling instance-specific prompting

Zifeng Wang et al., *Learning to Prompt for Continual Learning*. CVPR 2022.

Dual-Level Prompts

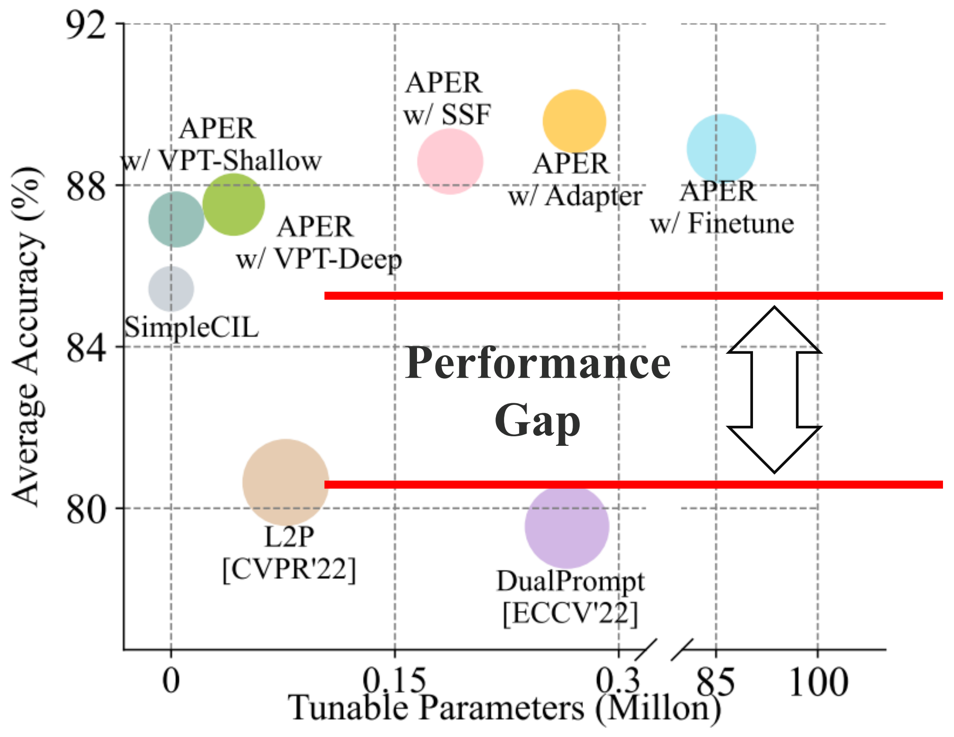
- DualPrompt [Wang et al., ECCV’22] designs two kinds of prompts
  - General (G) Prompts are ***shared for all tasks***;
  - Expert (E) Prompts are optimized ***independently for each task***;



Zifeng Wang et al., *DualPrompt: Complementary Prompting for Rehearsal-free Continual Learning*. ECCV 2022.

Outline

How about directly using the representation?

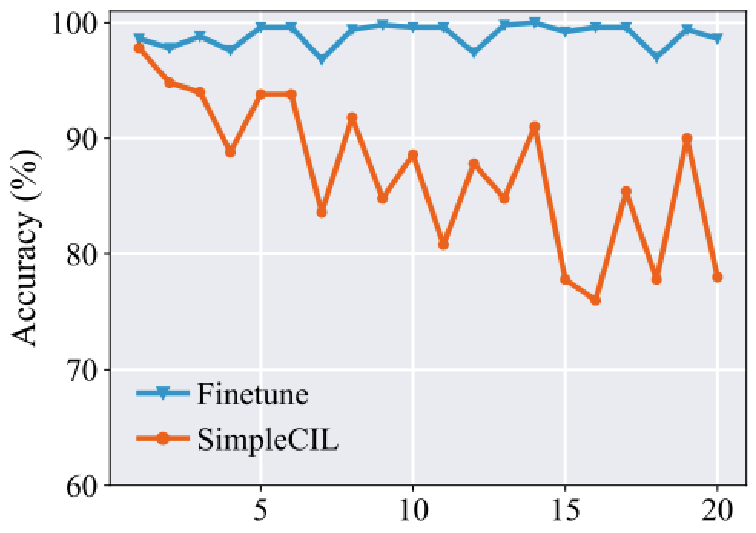


- How to measure the inherent ability of PTMs on these downstream tasks?
- SimpleCIL extracts the center of each class as the classifier:

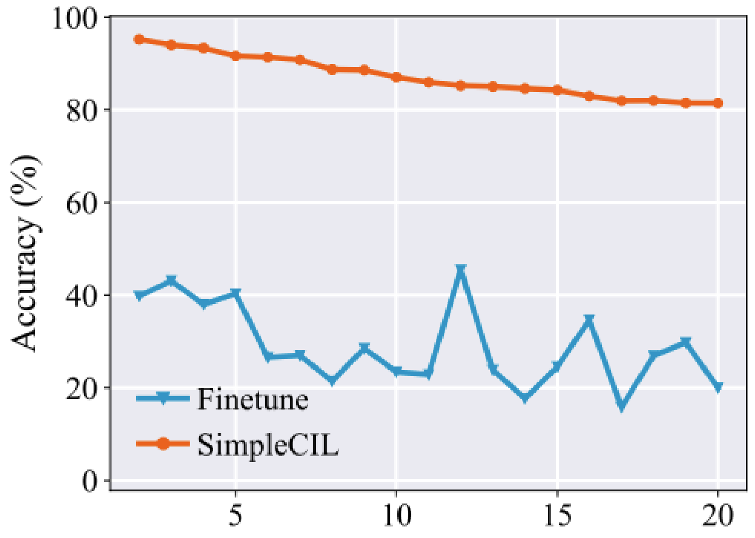
$$\boldsymbol{c}_i = \frac{1}{K}\sum_{j=1}^{|\mathcal{D}^b|}\mathbb{I}(y_j = i)\phi(\mathbf{x}_j)$$

- This ***simple*** solution shows superior performance than many prompt-based methods

Is representation all you need?



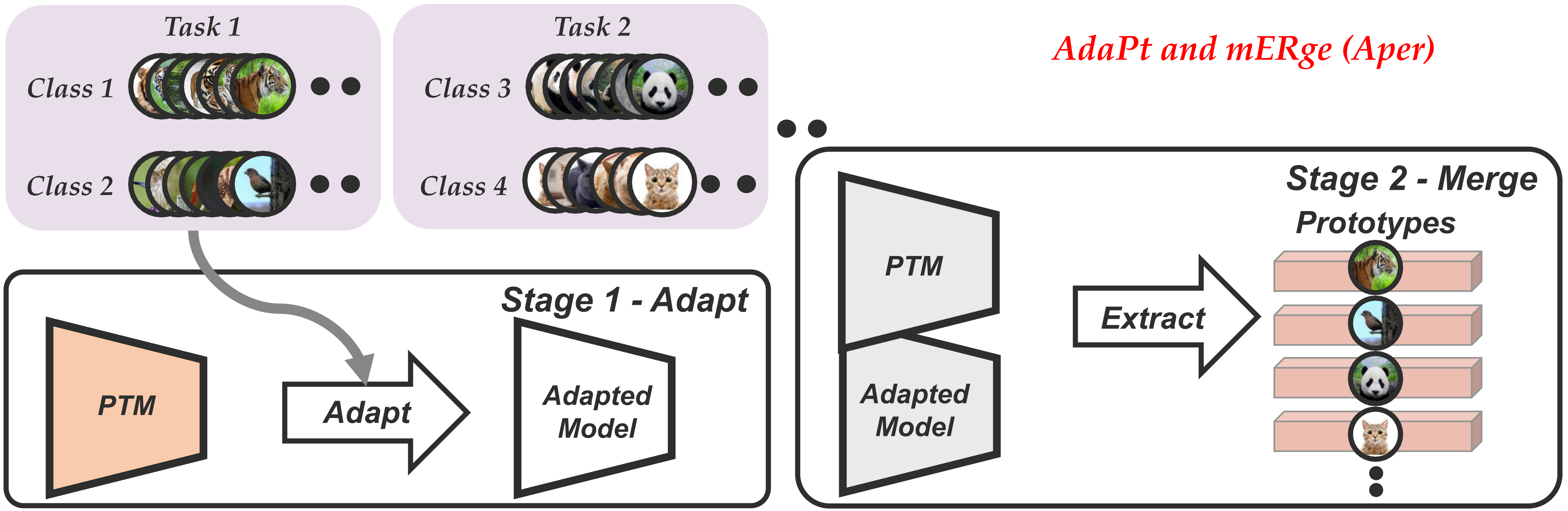
ACC_{new}



ACC_{old}

- Tuning with downstream data can further improve the performance on new classes
- But it hurts the performance of old classes

Concatenating Backbones

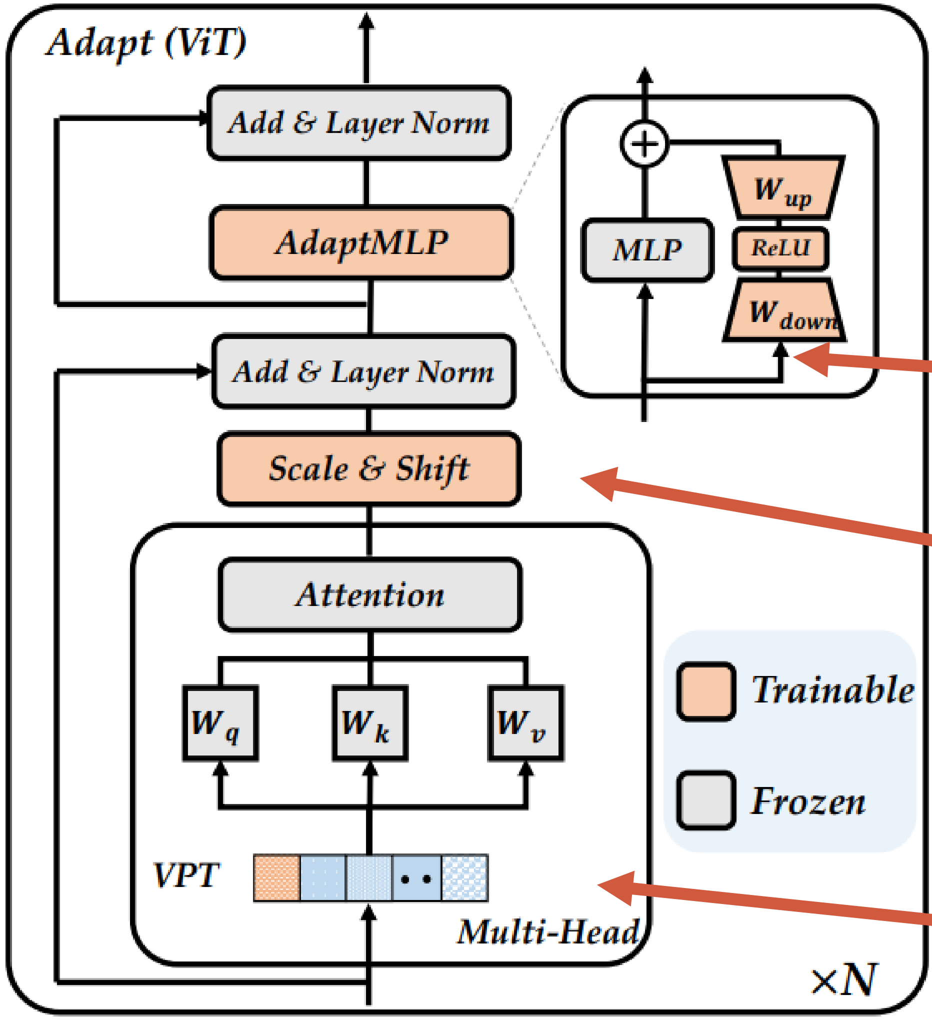


- We unify the advantages of the PTM and adapted model via feature concatenation

$$\boldsymbol{c}_i = \frac{1}{K}\sum_{j=1}^{|\mathcal{D}^b|}\mathbb{I}(y_j = i)[\phi(\mathbf{x}_j), \phi(\mathbf{x}_j; \text{PEFT})]$$

- To resist forgetting, we only adjust the classifier/prototypes after the first stage

How to adapt the model?



Aper is a general framework than can be orthogonally combined with various tuning techniques…

For Vision Transformer, we have:

- Adapter Tuning

$$\mathrm{MLP}(\mathbf{x}_\ell) + \mathrm{ReLU}(\mathbf{x}_\ell W_{\mathrm{down}})W_{\mathrm{up}}$$

- Scale and Shift

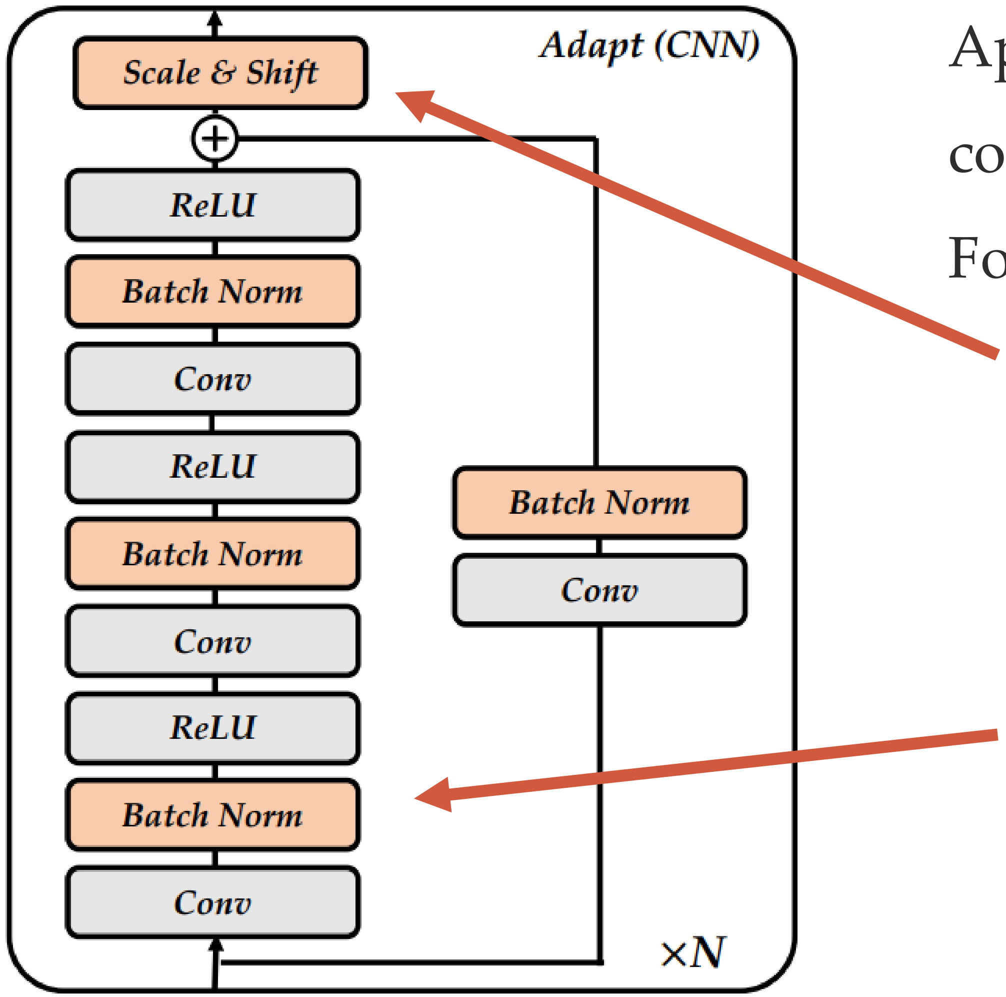
$$\mathbf{x}_o = \gamma \otimes \mathbf{x}_i + \beta$$

- Fully Finetune

$$\min_{\theta_\phi \cup \theta_W} \sum_{(\mathbf{x}_j, y_j) \in \mathcal{D}^1} \ell(f(\mathbf{x}_j), y_j)$$

- Visual Prompt Tuning

How to adapt the model?



Aper is a general framework than can be orthogonally combined with various tuning techniques…

For CNN, we have:

- Scale and Shift

$$\mathbf{x}_o = \gamma \otimes \mathbf{x}_i + \beta$$

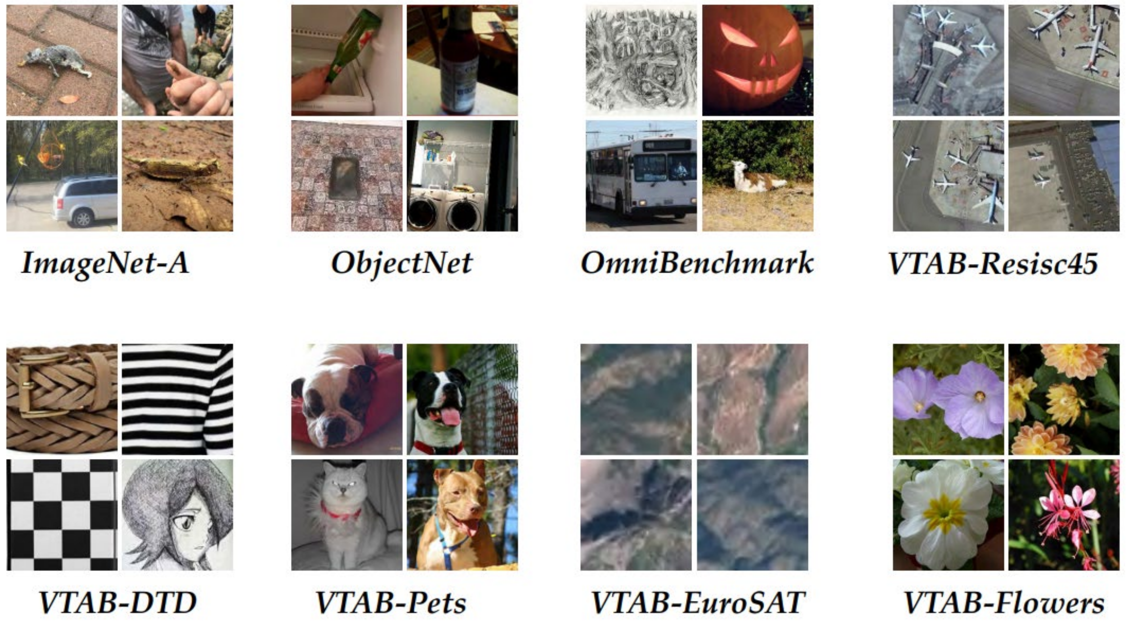
- Fully Finetune

$$\min_{\theta_\phi \cup \theta_W} \sum_{(\mathbf{x}_j, y_j) \in \mathcal{D}^1} \ell(f(\mathbf{x}_j), y_j)$$

- Batch Norm Tuning

Experiments

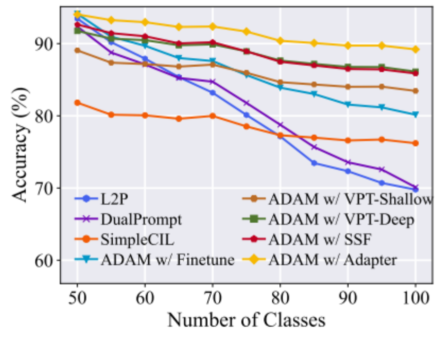
- Establishing new benchmarks in the era of pre-trained model-based CIL



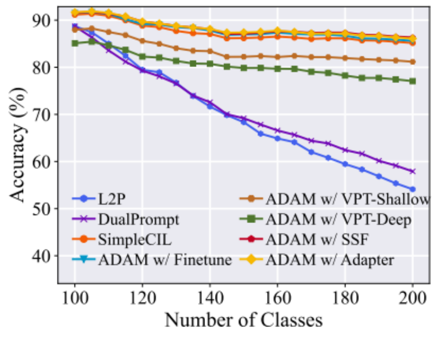
Extensive gap

Experiments

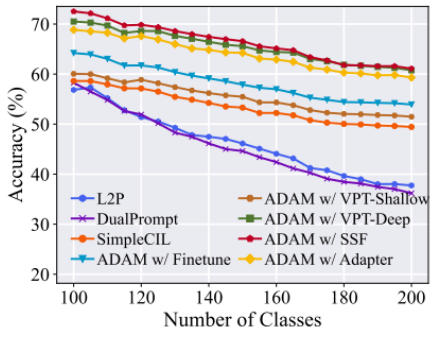
Method	CIFAR B0 Inc5		CUB B0 Inc10		IN-R B0 Inc5		IN-A B0 Inc10		ObjNet B0 Inc10		OmniBench B0 Inc30		VTAB B0 Inc10	
	$\bar{\mathcal{A}}$	$\mathcal{A}_B$	$\bar{\mathcal{A}}$	$\mathcal{A}_B$	$\bar{\mathcal{A}}$	$\mathcal{A}_B$	$\bar{\mathcal{A}}$	$\mathcal{A}_B$	$\bar{\mathcal{A}}$	$\mathcal{A}_B$	$\bar{\mathcal{A}}$	$\mathcal{A}_B$	$\bar{\mathcal{A}}$	$\mathcal{A}_B$
Finetune	38.90	20.17	26.08	13.96	21.61	10.79	21.60	10.96	19.14	8.73	23.61	10.57	34.95	21.25
Finetune Adapter [10]	60.51	49.32	66.84	52.99	47.59	40.28	43.05	37.66	50.22	35.95	62.32	50.53	48.91	45.12
LwF [38]	46.29	41.07	48.97	32.03	39.93	26.47	35.39	23.83	33.01	20.65	47.14	33.95	40.48	27.54
L2P [72]	85.94	79.93	67.05	56.25	66.53	59.22	47.16	38.48	63.78	52.19	73.36	64.69	77.11	77.10
DualPrompt [71]	87.87	81.15	77.47	66.54	63.31	55.22	52.56	42.68	59.27	49.33	73.92	65.52	83.36	81.23
SimpleCIL	87.57	81.26	92.20	**86.73**	62.58	54.55	60.50	49.44	65.45	53.59	79.34	73.15	85.99	84.38
ADAM w/ Finetune	87.67	81.27	91.82	86.39	70.51	62.42	61.57	50.76	61.41	48.34	73.02	65.03	**87.47**	80.44
ADAM w/ VPT-Shallow	90.43	84.57	92.02	86.51	66.63	58.32	57.72	46.15	64.54	52.53	79.63	73.68	87.15	**85.36**
ADAM w/ VPT-Deep	88.46	82.17	91.02	84.99	68.79	60.48	60.59	48.72	67.83	54.65	**81.05**	**74.47**	86.59	83.06
ADAM w/ SSF	87.78	81.98	91.72	86.13	68.94	60.60	**62.81**	**51.48**	**69.15**	**56.64**	80.53	74.00	85.66	81.92
ADAM w/ Adapter	**90.65**	**85.15**	**92.21**	**86.73**	**72.35**	**64.33**	60.53	49.57	67.18	55.24	80.75	74.37	85.95	84.35



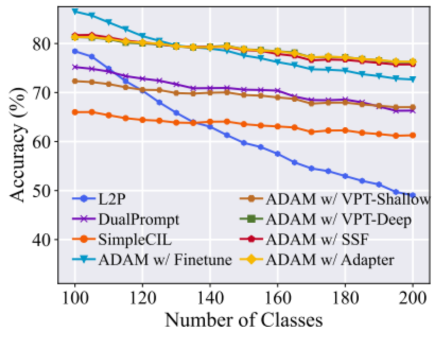
(a) CIFAR B50 Inc5



(b) CUB B100 Inc5



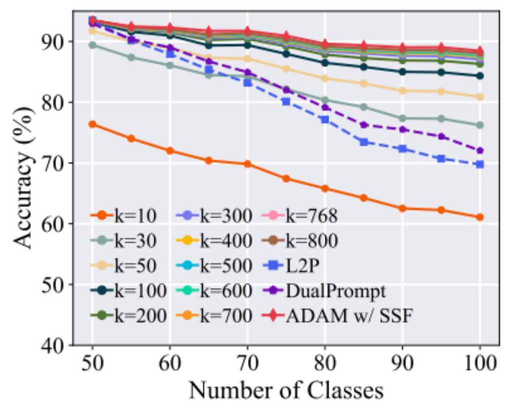
(c) ImageNet-A B100 Inc5



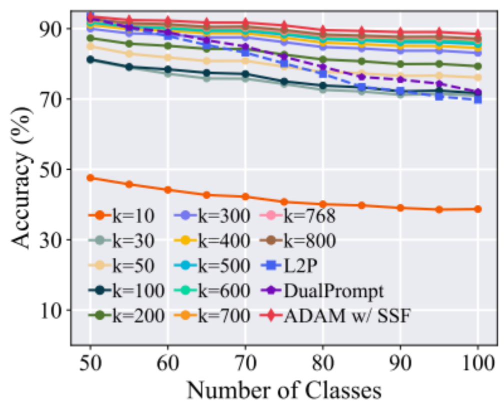
(d) ImageNet-R B100 Inc5

- Obtaining state-of-the-art performance in various benchmarks

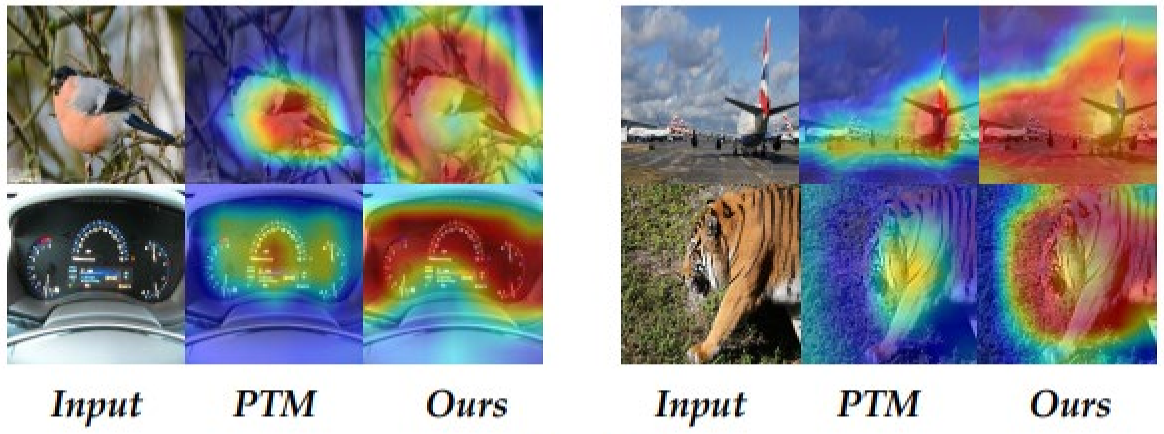
Experiments



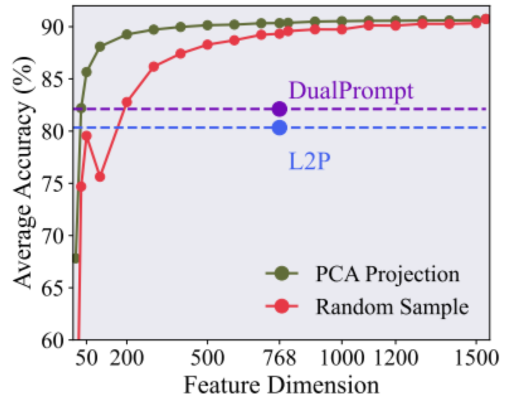
(a) PCA projected features



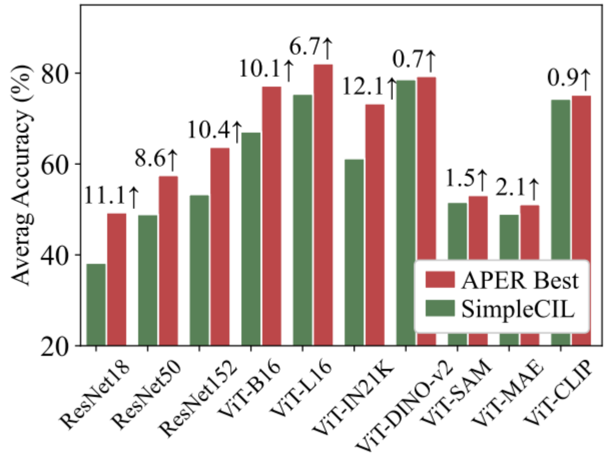
(b) Randomly sampled features



(c) Grad-CAM



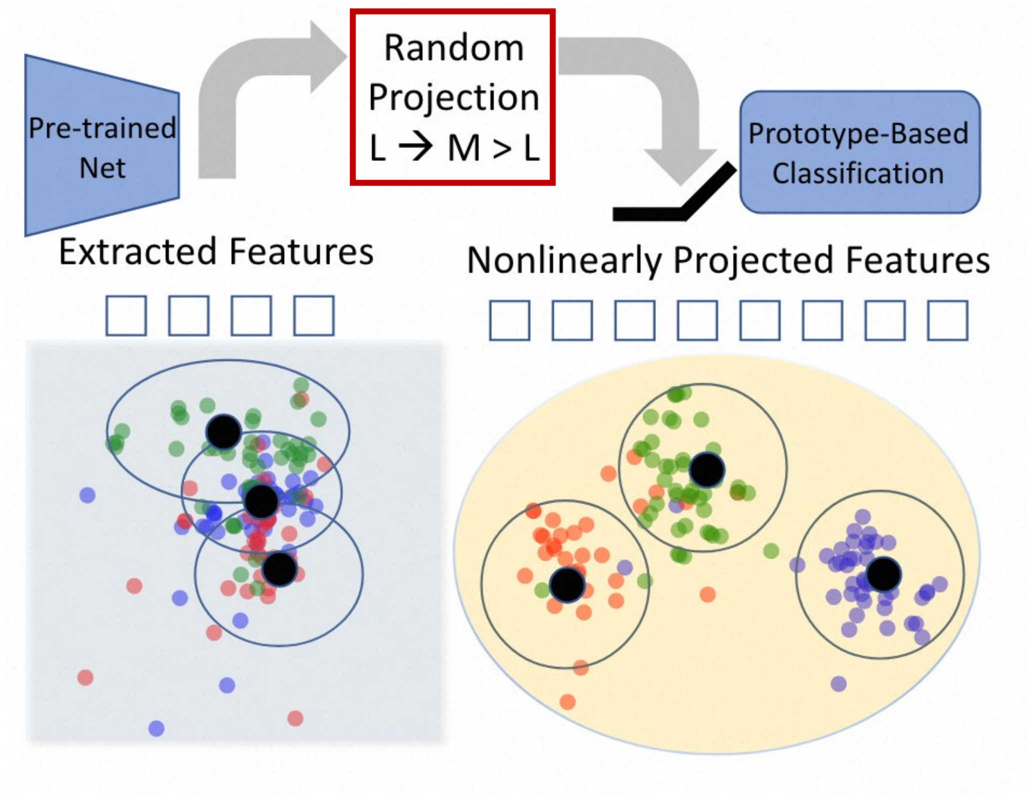
(c) Projected dimension and accuracy



- Beat SOTA even using very limited features
- Improve performance of various backbones
- Enhance the attention map of vanilla PTM

Outline

Random Projection


Pre-trained Net
Random Projection L → M > L
Prototype-Based Classification
Extracted Features
Nonlinearly Projected Features

- RanPAC [Mark et al., NeurIPS'23] designs a ***random projection layer*** to project features into the high dimensional space

$$\mathrm{Proj}\,(\cdot): \mathbb{R}^d \to \mathbb{R}^K, K \gg d$$

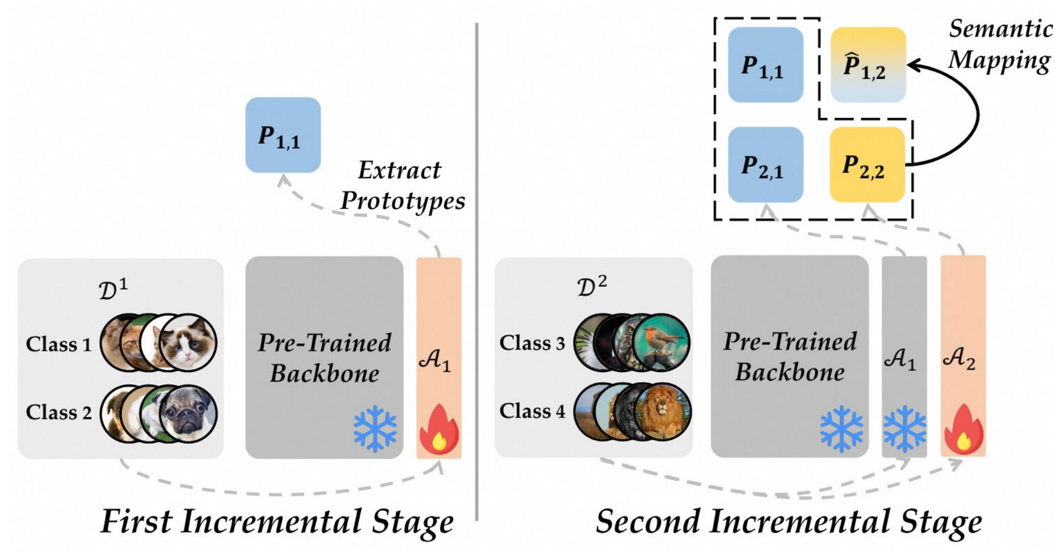
- Calculate the prototypes in the projected space

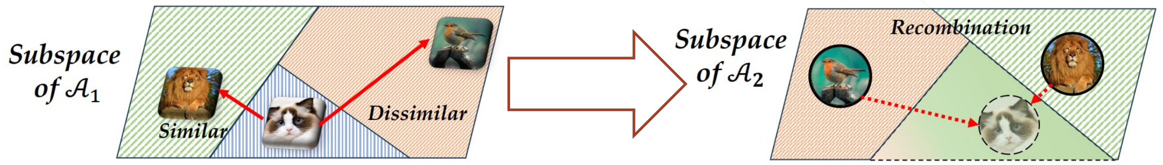
$$\boldsymbol{c}_i = \frac{1}{K} \sum_{j=1}^{|\mathcal{D}^b|} \mathbb{I}(y_j = i)\,\mathrm{Proj}\left(\phi\left(\mathbf{x}_j; \mathrm{PEFT}\right)\right)$$

- Use an online ***LDA classifier*** to remove class-wise correlations for better separability

Mark D. McDonnell et al., *RanPAC: Random Projections and Pre-trained Models for Continual Learning. NeurIPS* 2023

Concatenate Backbones





- What if expanding the feature dimension as data evolves?
- EASE [Zhou et al., CVPR'24] concatenates the feature of multiple backbones

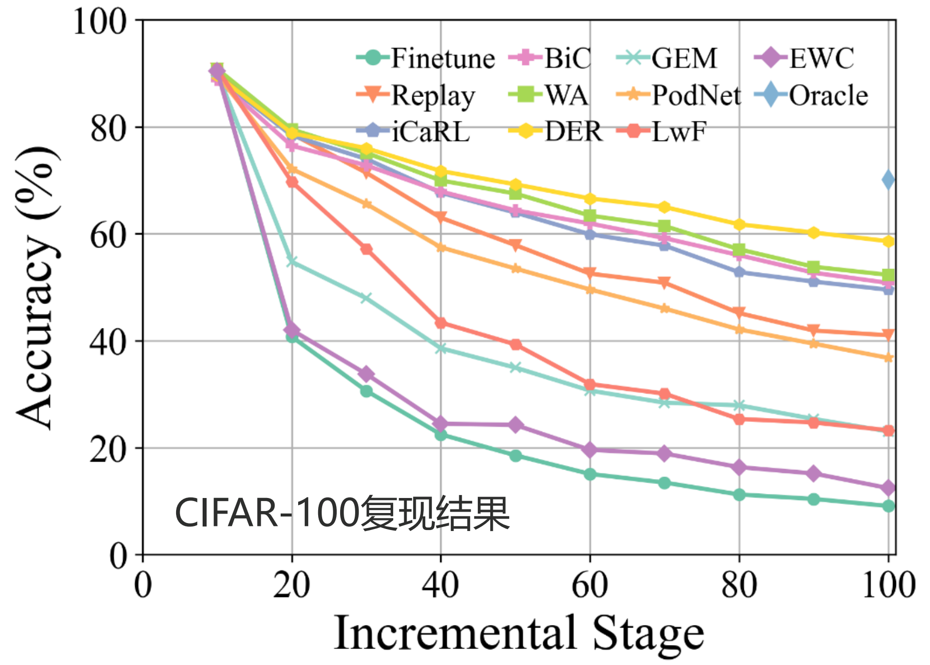
$$\Phi(\mathbf{x}) = [\phi(\mathbf{x}; \mathcal{A}_1), \cdots, \phi(\mathbf{x}; \mathcal{A}_b)] \in \mathbb{R}^{bd}$$

- To complete the missing prototypes of old classes in new spaces, ***semantic information*** is utilized to synthesize prototypes

Da-Wei Zhou et al., *Expandable Subspace Ensemble for Pre-Trained Model-Based Class-Incremental Learning*. CVPR 2024

PyCIL:持续学习工具包

https://github.com/G-U-N/PyCIL



全面 · 基准 · 可扩展 · 长期维护

Da-Wei Zhou, Fu-Yun Wang, Han-Jia Ye, De-Chuan Zhan. PyCIL: A Python Toolbox for Class-Incremental Learning. **SCIENCE CHINA Information Sciences** 2022

Da-Wei Zhou, Qi-Wei Wang, Zhi-Hong Qi, Han-Jia Ye, De-Chuan Zhan, Ziwei Liu Class-incremental learning: A survey. **TPAMI** 2024

周大蔚, 汪福运, 叶翰嘉, 詹德川. 基于深度学习的类别增量学习算法综述. **计算机学报** 2022.

PILOT:（基于预训练模型的）持续学习工具包

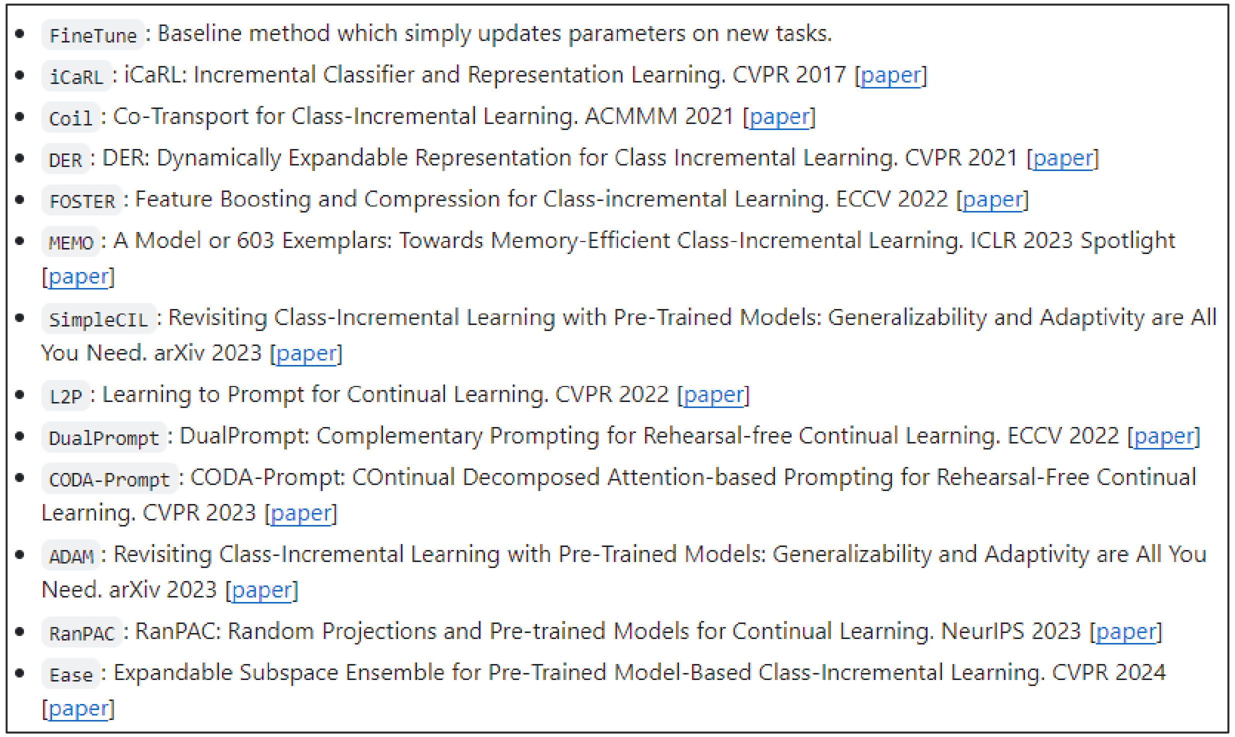
谢谢

LAMDA Learning And Mining from DatA

PILOT

A Pre-trained Model-based Continual Learning Toolbox

https://github.com/sun-hailong/LAMDA-PILOT



- FineTune : Baseline method which simply updates parameters on new tasks.
- iCaRL : iCaRL: Incremental Classifier and Representation Learning. CVPR 2017 [paper]
- Coil : Co-Transport for Class-Incremental Learning. ACMMM 2021 [paper]
- DER : DER: Dynamically Expandable Representation for Class Incremental Learning. CVPR 2021 [paper]
- FOSTER : Feature Boosting and Compression for Class-incremental Learning. ECCV 2022 [paper]
- MEMO : A Model or 603 Exemplars: Towards Memory-Efficient Class-Incremental Learning. ICLR 2023 Spotlight [paper]
- SimpleCIL : Revisiting Class-Incremental Learning with Pre-Trained Models: Generalizability and Adaptivity are All You Need. arXiv 2023 [paper]
- L2P : Learning to Prompt for Continual Learning. CVPR 2022 [paper]
- DualPrompt : DualPrompt: Complementary Prompting for Rehearsal-free Continual Learning. ECCV 2022 [paper]
- CODA-Prompt : CODA-Prompt: COntinual Decomposed Attention-based Prompting for Rehearsal-Free Continual Learning. CVPR 2023 [paper]
- ADAM : Revisiting Class-Incremental Learning with Pre-Trained Models: Generalizability and Adaptivity are All You Need. arXiv 2023 [paper]
- RanPAC : RanPAC: Random Projections and Pre-trained Models for Continual Learning. NeurIPS 2023 [paper]
- Ease : Expandable Subspace Ensemble for Pre-Trained Model-Based Class-Incremental Learning. CVPR 2024 [paper]

涵盖典型CIL算法与基于预训练模型的最优算法

Da-Wei Zhou, Hai-Long Sun, Jingyi Ning, Han-Jia Ye, De-Chuan Zhan. Continual Learning with Pre-Trained Models: A Survey. **IJCAI** 2024

Hai-Long Sun, Da-Wei Zhou, De-Chuan Zhan, Han-Jia Ye. Pilot: A pre-trained model-based continual learning toolbox. **CoRR** 2023