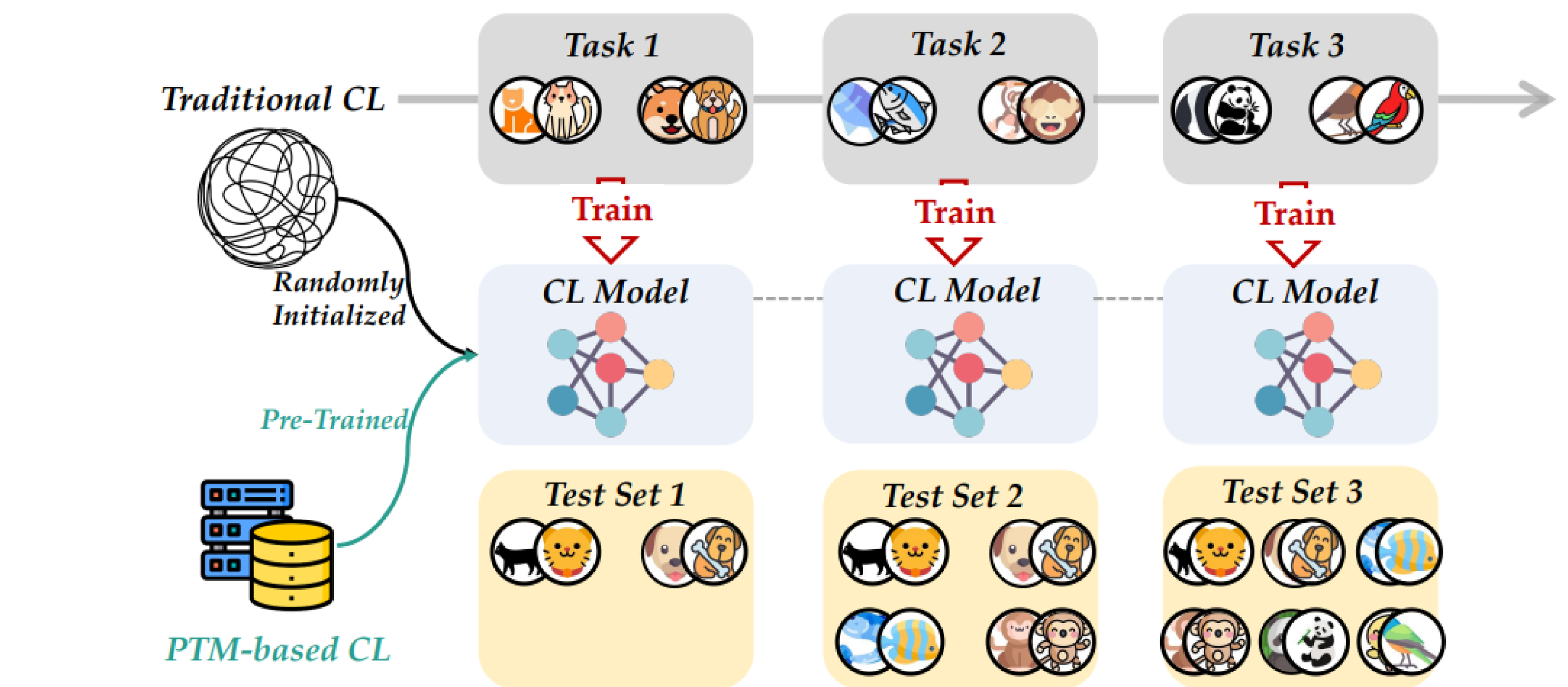
Integrating Task-Specific and Universal Adapters for Pre-Trained Model-based Class-Incremental Learning

Yan Wang, Da-Wei Zhou, Han-Jia Ye
Nanjing University zhoudw@lamda.nju.edu.cn

Highlights

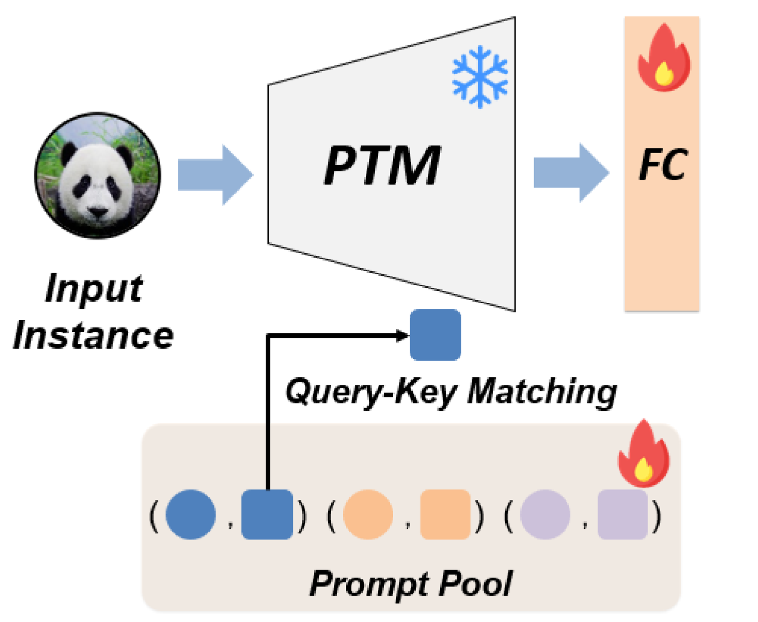
- We train task-specific adapters and introduce an **entropy-based** selection mechanism;
- We propose an **adapter fusion strategy** to construct a universal adapter, which encodes the most discriminative features shared across tasks;
- We combine the most confident task-specific adapter with the universal adapter to enhance prediction accuracy;
- **State-of-the-art** experimental results on benchmarks.

Class-Incremental Learning with Pre-Trained Model



Using PTM as the initialization, we need to sequentially **learn new classes** and **do not forget old ones**.

Current Solution



Trains task keys and prompts; selects prompts via key matching at inference.

Main drawbacks

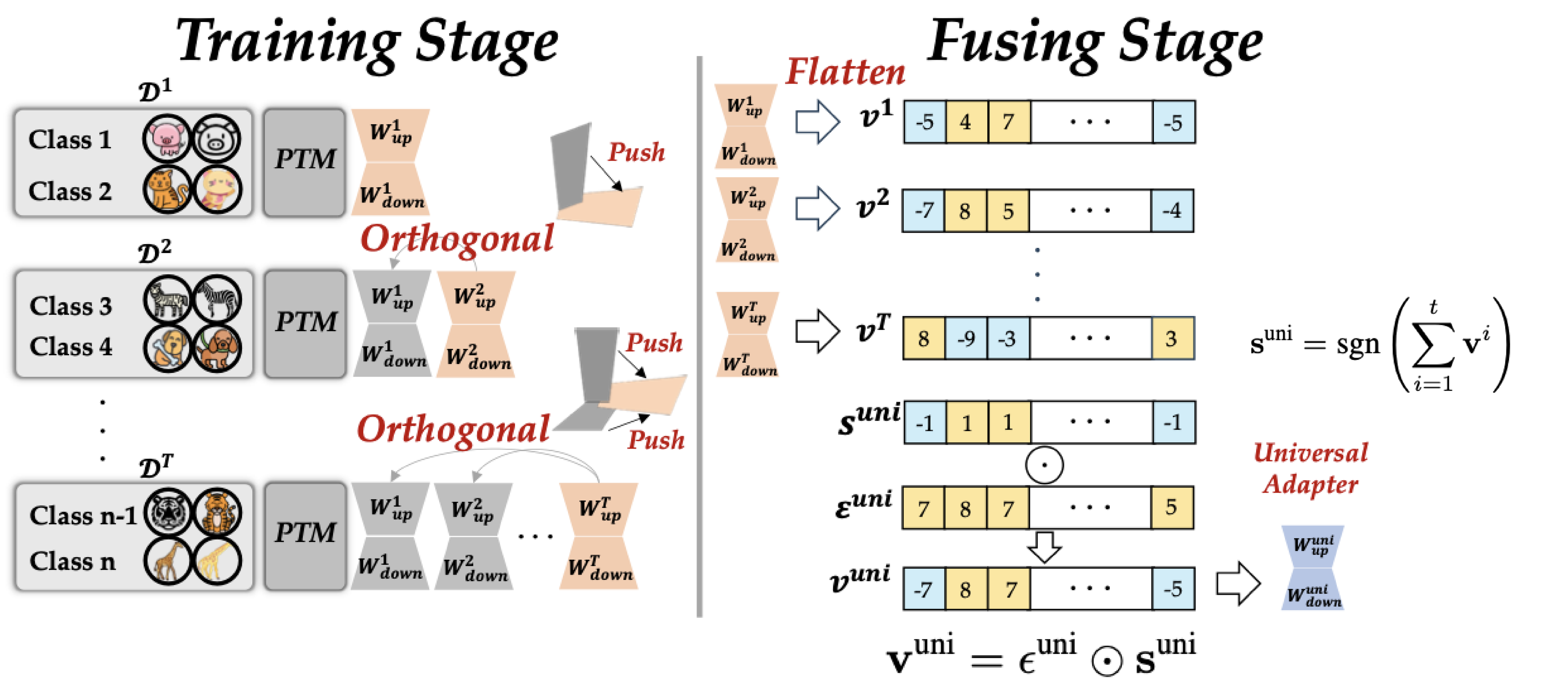
- Performance is highly sensitive to the accuracy of retrieved task-specific prompts during inference.
- The overemphasis on task-specific prompts hinders the capture and leverage of general, transferable knowledge across tasks.

Can we achieve precise module selection while utilizing general knowledge shared across tasks?

Multi-Stage Adapter Fusion

Key idea

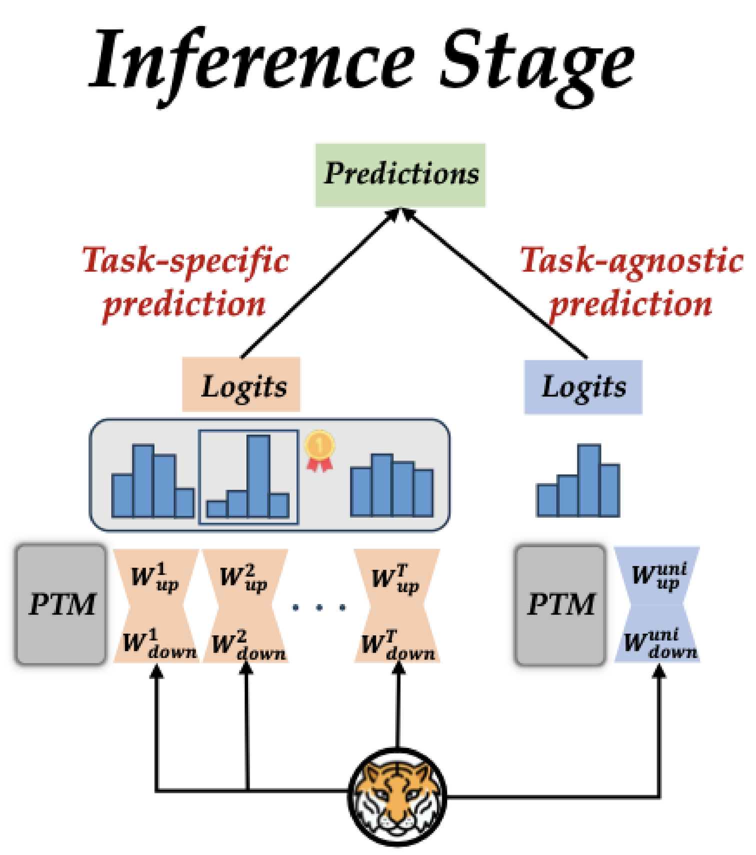
- Train an adapter for each incremental task to encode the ***task-specific information***.
- Fuse these task-specific adapters into ***a universal adapter***, which amalgamates cross-task knowledge while preserving domain-invariant representations.



$$\mathbf{s}^{\text{uni}} = \text{sgn}\left(\sum_{i=1}^{t} \mathbf{v}^i\right)$$

$$\mathbf{v}^{\text{uni}} = \epsilon^{\text{uni}} \odot \mathbf{s}^{\text{uni}}$$

Adapter Selection via Prediction Uncertainty



$$\mathcal{A}^* = \underset{\mathcal{A}_i \in \{\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_t\}}{\arg\min} \left(-\sum_{c=1}^{y_t} f_c(\mathbf{x}; \mathcal{A}_i) \log f_c(\mathbf{x}; \mathcal{A}_i) \right)$$

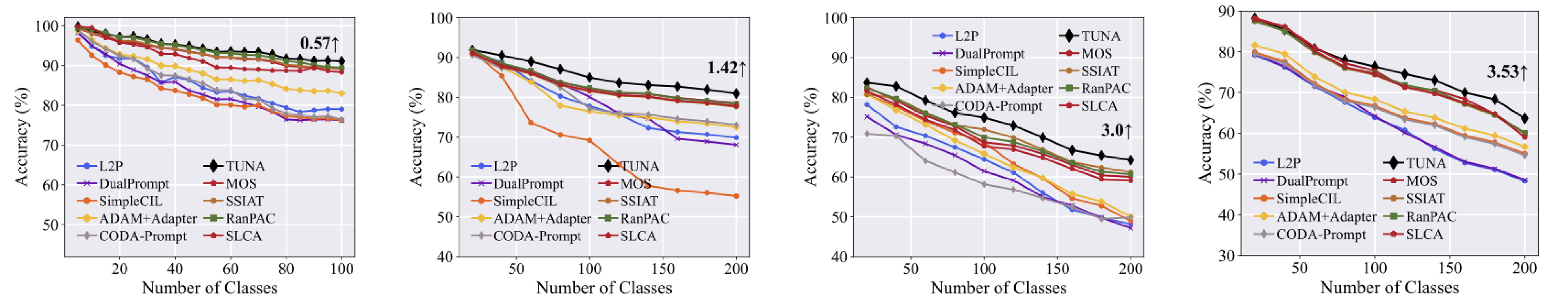
Key idea

- Select the most suitable adapter according to entropy.
- Combine task-specific adapter with the universal adapter during inference.

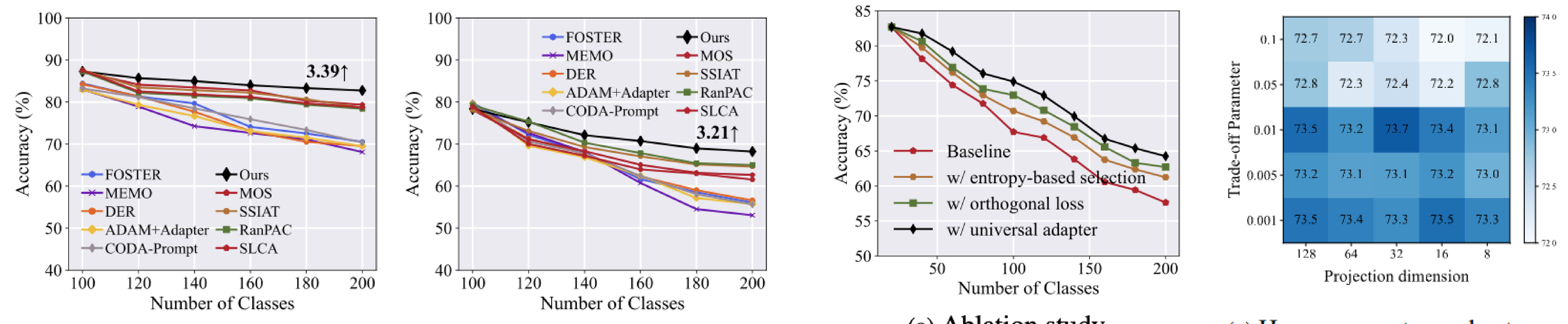
$$y^* = \arg\max_y \left(f_y(\mathbf{x}; \mathcal{A}^*) + f_y(\mathbf{x}; \mathcal{A}_{\text{uni}}) \right)$$

State-Of-The-Art Results

Method	CIFAR B0 Inc5 $\bar{\mathcal{A}}$	CIFAR B0 Inc5 $\mathcal{A}_B$	ImageNet-R B0 Inc20 $\bar{\mathcal{A}}$	ImageNet-R B0 Inc20 $\mathcal{A}_B$	ImageNet-A B0 Inc20 $\bar{\mathcal{A}}$	ImageNet-A B0 Inc20 $\mathcal{A}_B$	ObjectNet B0 Inc20 $\bar{\mathcal{A}}$	ObjectNet B0 Inc20 $\mathcal{A}_B$
L2P [50]	85.94	79.93	75.46	69.77	49.39	41.71	63.78	52.19
DualPrompt [49]	87.87	81.15	73.10	67.18	53.71	41.67	59.27	49.33
CODA-Prompt [43]	89.11	81.96	77.97	72.27	53.54	42.73	66.07	53.29
SLCA [58]	92.49	88.55	81.17	77.00	68.66	58.74	72.55	61.30
SSIAT [45]	93.52	90.07	83.20	78.85	70.83	62.23	73.65	62.45
MOS [44]	93.30	89.25	82.96	77.93	67.08	56.22	74.69	63.62
SimpleCIL [64]	87.57	81.26	61.26	54.55	59.77	48.91	65.45	53.59
APER + Adapter [64]	90.65	85.15	75.82	67.95	60.47	49.37	67.18	55.24
RanPAC [37]	94.00	90.62	82.98	77.94	69.32	61.82	72.76	62.02
EASE [65]	91.51	85.80	81.74	76.17	65.34	55.04	70.84	57.86
TUNA (Ours)	**94.44**	**90.74**	**84.22**	**79.42**	**73.78**	**64.78**	**76.46**	**66.32**



(a) CIFAR B0 Inc5 (b) ImageNet-R B0 Inc20 (c) ImageNet-A B0 Inc20 (d) ObjectNet B0 Inc20



(a) ImageNet-R B100 Inc20 (b) ImageNet-A B100 Inc20 (a) Ablation study (a) Hyperparameters robustness

Conclusion

TUNA can unify task-specific and universal information in a unified framework, achieving a new SOTA in PTM-based CIL.

Homepage **Survey** **Code** **WeChat**