Tired of unrealistic assumptions in continual learning? Try our assumption-free approach—no exemplars, no balance, no limits!

AFCIL: Assumption Free Class Incremental Learning

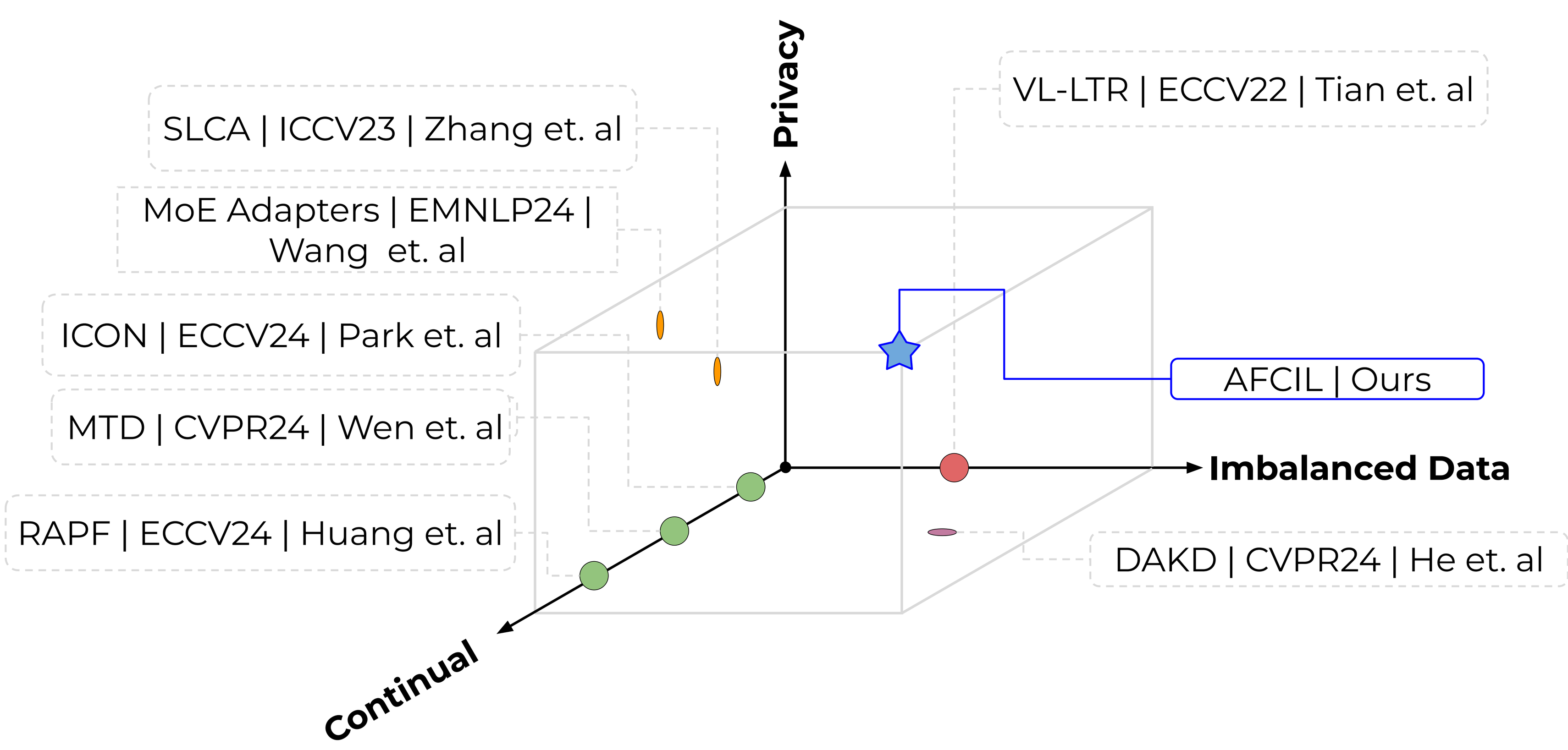
Divya Mehul Rajparia, Rahul Vigneswaran, Hari Chandana Kuchibhotla, Vineeth N Balasubramanian

Indian Institute of Technology Hyderabad, India

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Introduction

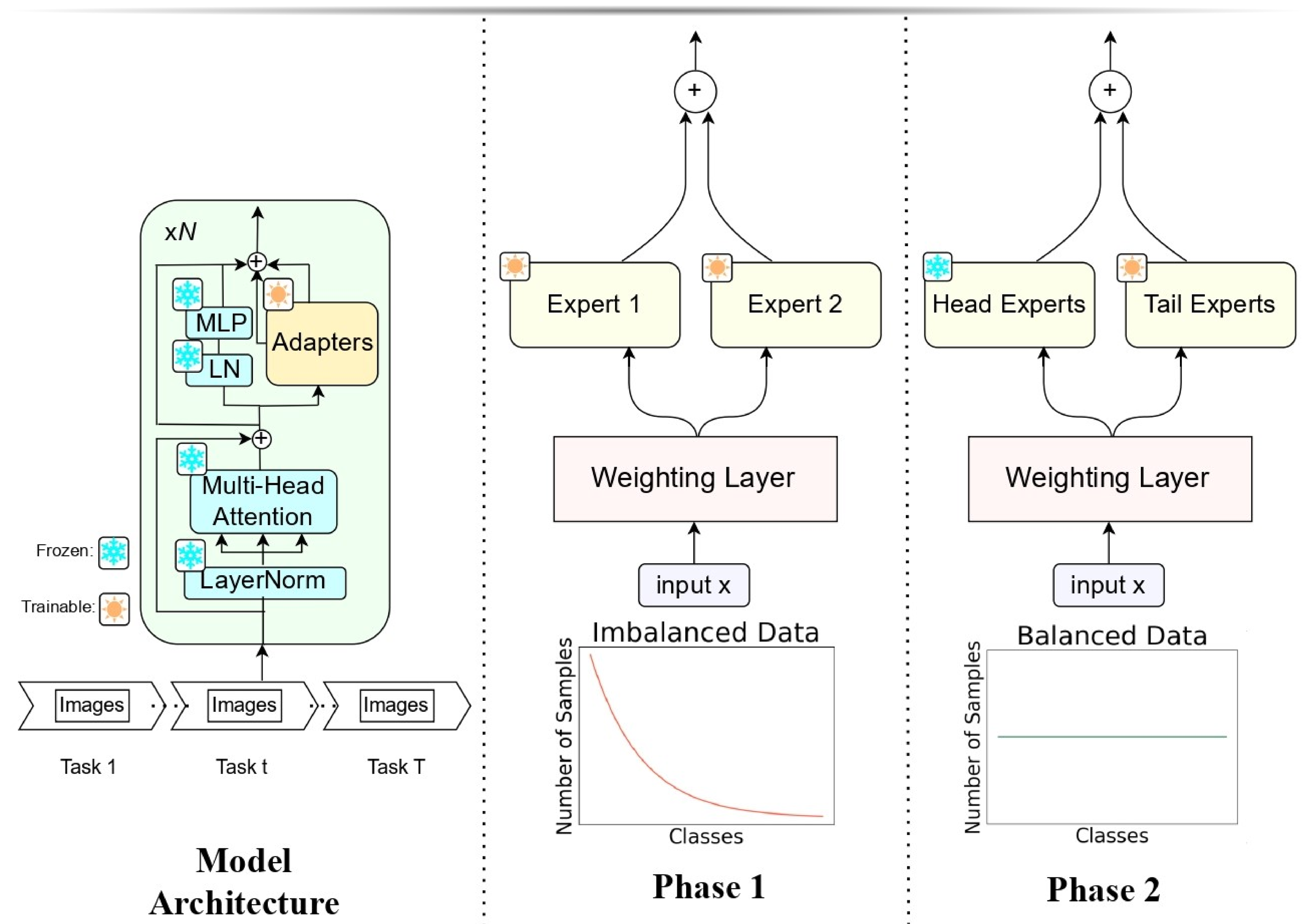
- **Continual Learning (CL)** is a machine learning paradigm where data arrives sequentially as *tasks*, and the model must learn them without forgetting previous ones.
- Most CL methods assume balanced data and mitigate forgetting by storing *exemplars* from past tasks.
- Long-Tailed Class Incremental Learning (LTCIL) relaxes the balance assumption but still depends on exemplars.
- Exemplar-free methods address privacy concerns but struggle with skewed data.
- This exposes a gap: a CL **setup** that is **exemplar-free**, supports **imbalanced data**, uses **pre-trained weights**, and enables a **multimodal** setting.
- We define this as **Assumption-Free Class Incremental Learning (AFCIL)**—free from prior assumptions.



Key Contributions

- We propose a new continual learning setup, **Assumption Free Class Incremental Learning (AFCIL)**, which removes key assumptions made in prior work.
- The AFCIL setup supports long-tailed data, is exemplar-free, and permits the use of pre-trained weights.
- To address the challenges of AFCIL, we introduce **Dynamic Expert Saturation (DES)**, a method that incrementally saturates experts to accommodate both head and tail classes.

DES: Methodology



Methodology

- We propose **Dynamic Expert Saturation (DES)**, which fully utilizes model capacity by first specializing in head classes, then progressively saturating with tail class knowledge.
- To enable efficient training, we insert lightweight trainable adapters (MLPs) into a frozen CLIP backbone, allowing adaptation without modifying the backbone.
- Training proceeds in two phases:
  Phase 1: All experts are trained with an imbalanced sampler to learn head class features.
  Phase 2: Head experts are frozen, and remaining experts are fine-tuned with a balanced sampler to specialize in tail classes, incrementally saturating model capacity.
- Final predictions combine expert outputs through dynamic weighting, leveraging both head and tail expertise.

Results

Accuracies across methods for Assumption Free Class Incremental Learning

	CIFAR100-LT			ImageNet-LT		
ρ	100	50	10	100	50	10
Zero-Shot CLIP	80.36	80.36	80.36	89.71	89.71	89.71
L2P (CVPR '22)	70.70	74.46	79.65	87.04	88.85	90.92
DualPrompt (ECCV '22)	70.53	74.45	79.99	89.03	90.52	92.09
CODA (CVPR '23)	<u>80.79</u>	<u>85.69</u>	**89.72**	89.83	91.00	<u>93.68</u>
SLCA (ICCV '23)	71.90	75.31	79.297	<u>92.08</u>	<u>92.84</u>	93.632
MoE-Adapter (CVPR '24)	79.85	80.05	82.77	89.75	90.12	91.96
DES (Ours)	**87.17** 6.38↑	**89.4** 3.71↑	<u>89.38</u> 0.34↓	**94.51** 2.43↑	**94.8** 1.96↑	**96.09** 2.41↑

Ablation study

Top: Evaluating impact of different components of DES on overall accuracy on CIFAR100LT dataset. **Bottom Left:** Analysis of different freezing strategies for the phase 2 of DES on CIFAR100LT dataset. **Bottom Right:** Analysis of different saturation techniques for DES

Frozen Expert	Balanced Sampler	Accuracy
✗	✗	81.16 (↓6.01)
✓	✗	83.92 (↓3.25)
✗	✓	86.33 (↓0.84)
✓	✓	**87.17**

Freezing Strategy	Accuracy
Top weighted (local)	86.84 (↓0.33)
Top weighted (global)	86.91 (↓0.26)
Round Robin	87.04 (↓0.13)
Random	**87.17**

Saturation Technique	Accuracy
Weighted Loss	85.54 (↓1.63)
Balanced Sampler	**87.17**

Future work

- Implement a more intelligent freezing strategy to freeze experts after phase one of training.
- Store task statistics for classifier alignment, enabling the preservation of knowledge from previous tasks while facilitating adaptation to new tasks and mitigating forgetting.