Meta-Learning Hyperparameters for Parameter Efficient Fine-Tuning

The IEEE/CVF Conference on Computer Vision and Pattern Recognition 2025 (CVPR’25)

Zichen Tian[1], Yaoyao Liu[2], Qianru Sun[3]

[1]Singapore Management University, [2]University of Illinois Urbana-Champaign

The task we are focusing on.

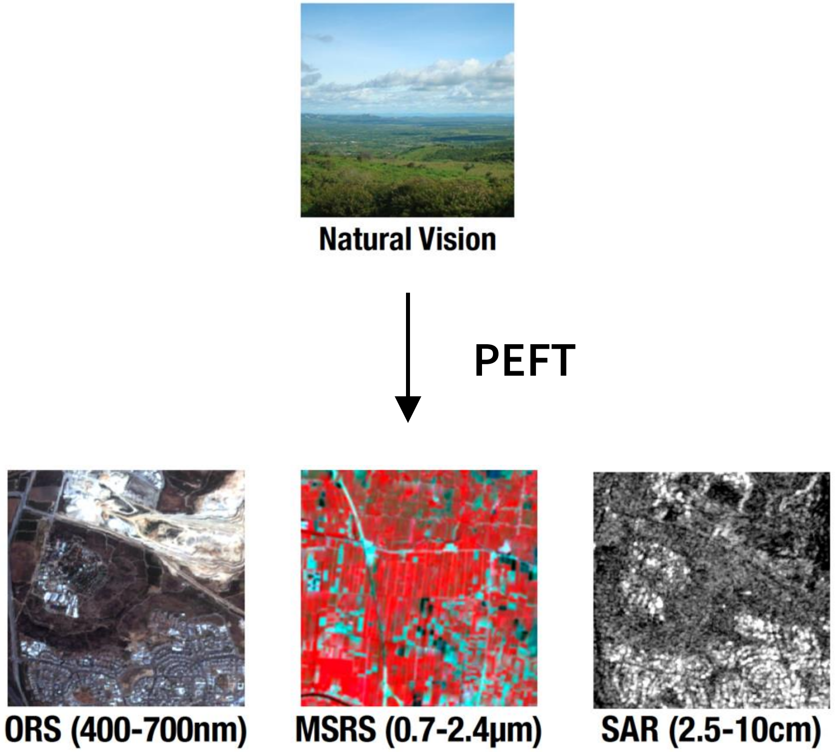
Task & Problems

- **Train-from-Scratch**

  Training from from scratch is impractical for domains with data scarcity and data imbalance (*e.g.*, Remote Sensing).

- **Fine-tuning**
  - Full fine-tuning: computationally expensive, long-tailed issue
  - Parameter Efficient Fine-Tuning (PEFT) reduces computational and data requirements, and mitigates long-tailed issue

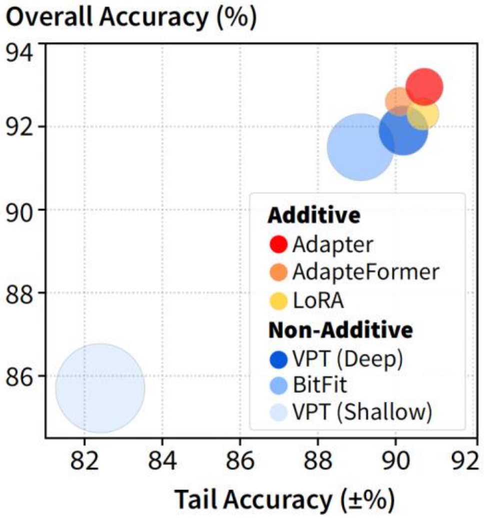
- **Challenge**
  - Hyperparameter sensitivity…



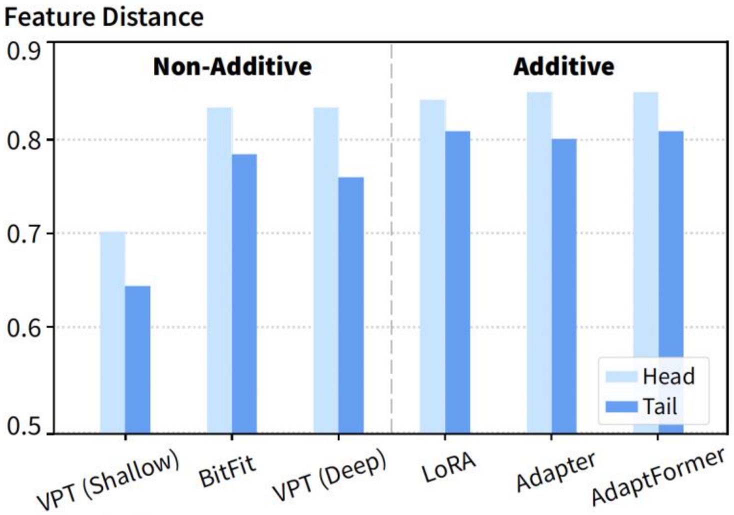
What we identified in this task.

Key Observations

- **Existing PEFT Approaches**
  - Additive PEFT > Non-additive PEFT
- **Three Critical Hyperparameters**
  - Intra-block position (*e.g.*, Q/K/V projection, FFN)
  - Depth (*i.e.*, which layer)
  - Scaling factor (combination weight)
- **Crucial for tail classes**



(a) PEFT Performance

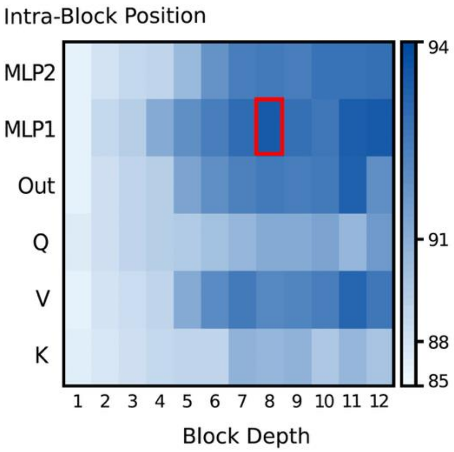


(b) Inter-Class Feature Distance

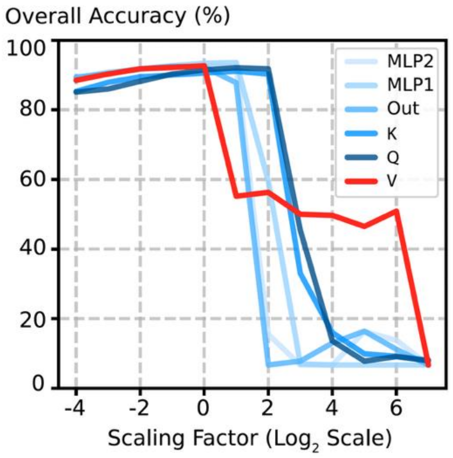
What we identified in this task.

Key Observations & Challenges

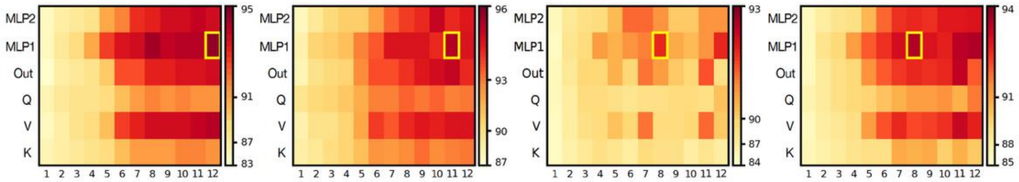
- **Existing PEFT Approaches**
  - Additive PEFT > Non-additive PEFT
- **Three Critical Hyperparameters**
  - Intra-block position (*e.g.*, Q/K/V projection, FFN)
  - Depth (*i.e.*, which layer)
  - Scaling factor (combination weight)
- **Crucial for tail classes**
- **Key Challenges**
  - Manual search: $O(L|S|N_\alpha)$ complexity
  - Mixed discrete-continuous problem
  - Non-monotonic interactions



(c) Position v.s. Block Depth



(d) Position v.s. Scaling Factor



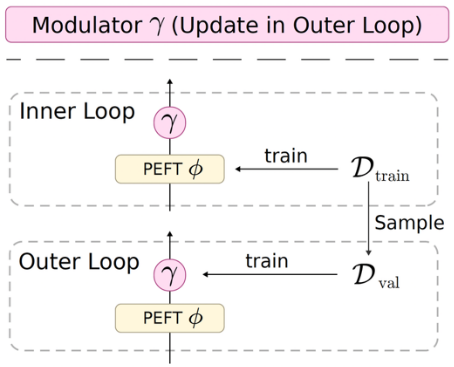
(e.1) Head (e.2) Med (e.3) Tail (e.4) Overall

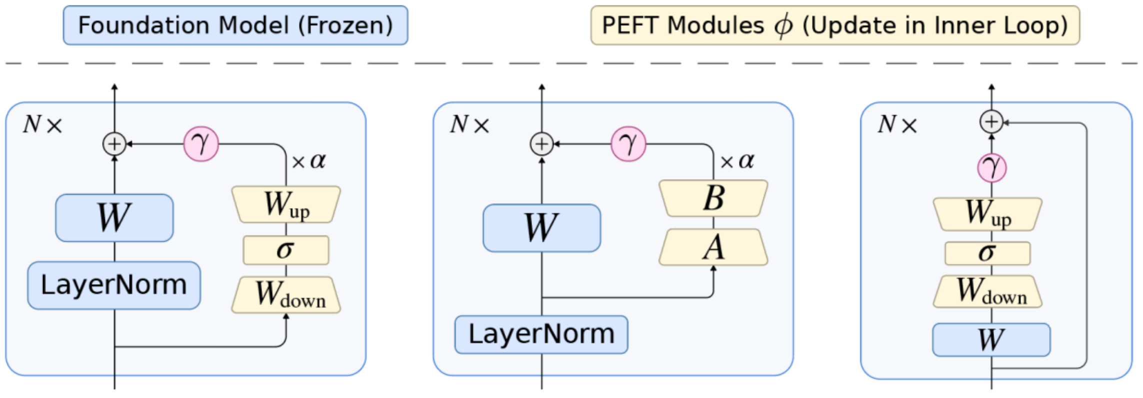
(e) Accuracy heatmap for Head/Med/Tail/Overall

What we propose as a solution.

MetaPEFT Method

- **Unified Modulator:** γ
  - Convert discrete → continuous optimization.
- **Bi-Level Optimization**
  - Inner loop: train PEFT modules
  - Outer loop: optimize modulator





(f.1) AdaptFormer w/ Ours (f.2) LoRA w/ Ours (f.3) Adapter w/ Ours

(f) Architecture of MetaPEFT

Our major results.

Results & Contributions

- **Comparison with SOTA**
  - Three transfer scenarios tested
  - Consistent improvements
  - Tail classes: up to 3.5 times gain
- **Contributions**
  - End-to-end PEFT hyperparameter optimization
  - Converts intractable MINLP → differentiable problem
  - Automatic discovery of optimal configurations
  - Code avail on GitHub

Category	Method	IN21K → iNat2018			IN21K → DOTA			SatMAE → SAR		Avg_{tail}	Avg
		Head	Med	Tail	Head	Med	Tail	Head	Tail		
Non-additive	VPT-Shallow	57.5	65.5	65.9	85.4	89.0	82.4	39.8	68.4	72.23	69.24
	BitFit	57.6	67.3	68.4	92.0	93.7	89.1	32.1	74.7	77.40	71.86
	VPT-Deep	64.9	74.2	75.9	92.1	93.5	90.2	30.6	50.0	72.03	71.43
Additive	Adapter	67.9	76.8	77.7	93.2	94.9	90.6	37.9	75.8	81.37	76.85
	w/ Ours	68.2	76.9	78.1	93.2	**95.1**	90.7	37.7	76.0	81.60	76.99
	AdaptFormer	68.7	76.7	78.0	93.2	94.9	90.1	35.3	**76.7**	81.60	76.70
	w/ Ours	68.7	76.6	78.2	92.7	94.6	90.1	35.5	76.4	81.57	76.60
	LoRA	69.1	77.3	78.5	93.1	93.4	90.7	40.0	72.1	80.43	76.78
	w/ Ours	**70.2**	**78.6**	**79.3**	**93.9**	**95.1**	**91.4**	**40.6**	74.2	**81.63**	**77.91**

Thank you!

Welcome to email me at zichen.tian.2023@phdcs.smu.edu.sg