ECCV 2026 | MALMÖ | SEPT 8–12

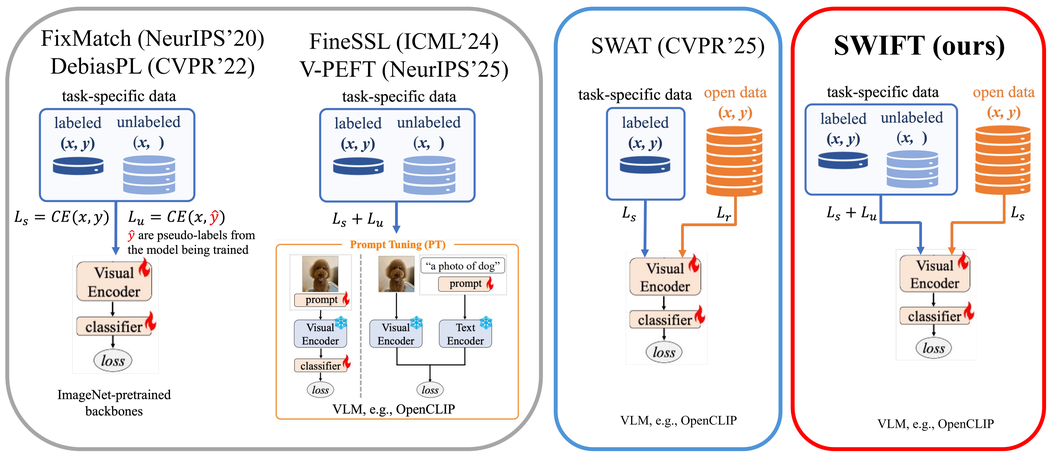
Solving Semi-Supervised Few-Shot Learning from an Auto-Annotation Perspective

Tian Liu[1], Anwesha Basu[1], James Caverlee[1], Shu Kong[2]
[1]Texas A&M University, [2]University of Macau

paper & code

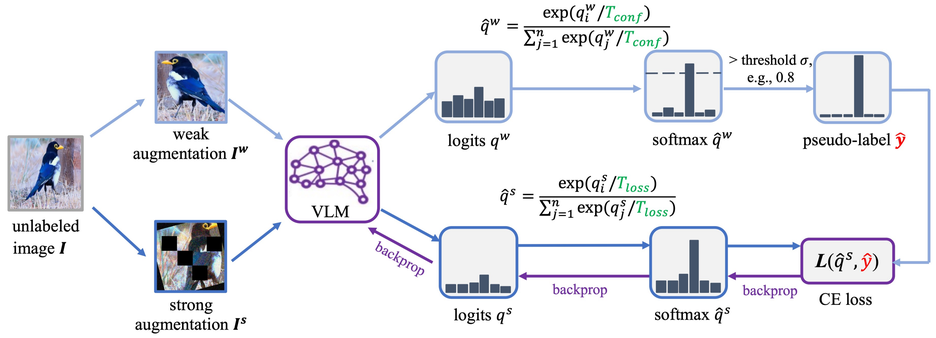
Motivation

- Semi-Supervised Few-Shot Learning (SSFSL) mirrors real-world *auto-annotation,* aiming to learn a model over a few labeled and abundant unlabeled task-specific examples so to automatically annotate the unlabeled ones.
- Existing SSL or SSFSL methods either finetune an ImageNet-pretrained backbone or prompt a frozen VLM, while recent FSL method has achieved significant progress by ***finetuning*** **a VLM and** ***retrieving*** **task-relevant open data**.
- To bridge the gap, we are motivated to explore these techniques for SSFSL.



Challenges

- We start by finetuning the VLM OpenCLIP ViT-B/32 with the representative SSL method *FixMatch* (NeurIPS'20) using the default hyperparameters (e.g., confidence threshold 0.8).
- Yet, it barely rivals with zero-shot methods **REAL** (CVPR'24) and significantly underperforms the few-shot finetuning method **FS-FT** (CVPR'25), which does not exploit unlabeled data.

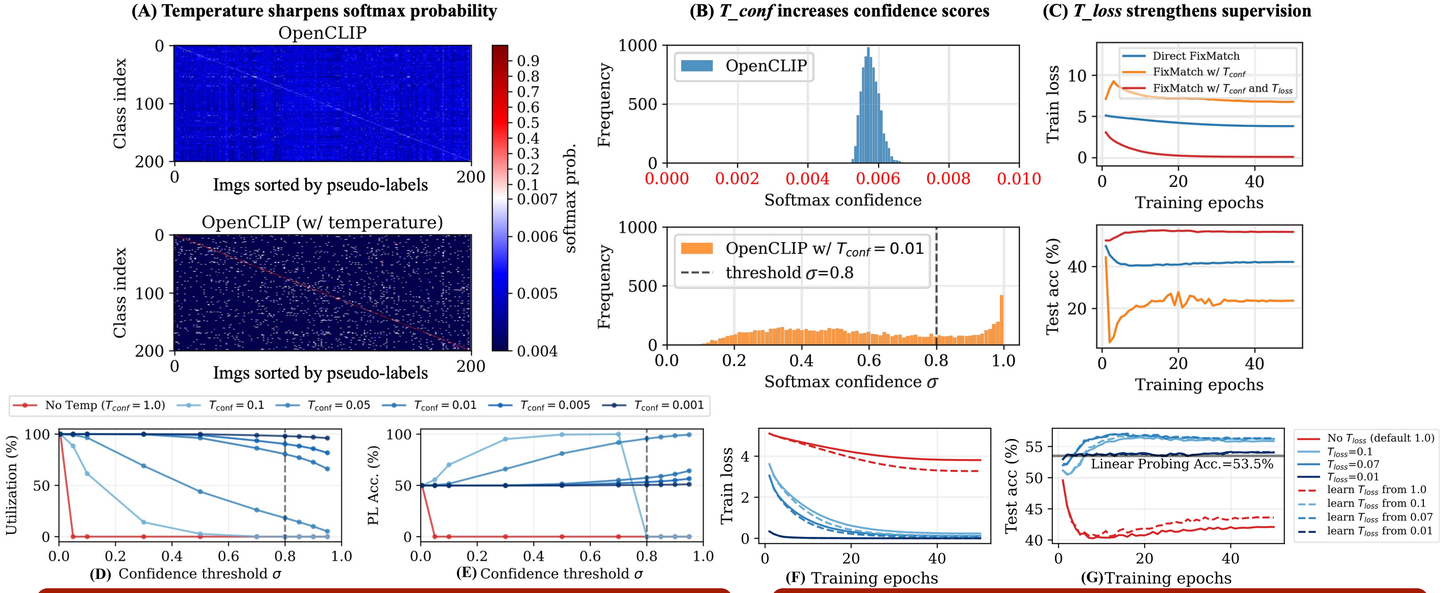


FixMatch workflow (enhanced by our temperatures)

method	training data	backbone	mean acc. over five datasets		
			4-shot	8-shot	16-shot
FS-FT [42]	L	VLM-OpenCLIP	61.0	65.8	69.1
FixMatch [57]	$L+U$	INet-pretrained	30.7	45.3	60.1
FixMatch	$L+U$	VLM-OpenCLIP	39.3 $^{-21.7}$	49.9 $^{-15.9}$	57.2 $^{-11.9}$
DebiasPL [67]	$L+U$	INet-pretrained	34.5	49.0	62.7
DebiasPL	$L+U$	VLM-OpenCLIP	39.6 $^{-21.4}$	49.8 $^{-16.0}$	57.1 $^{-12.0}$

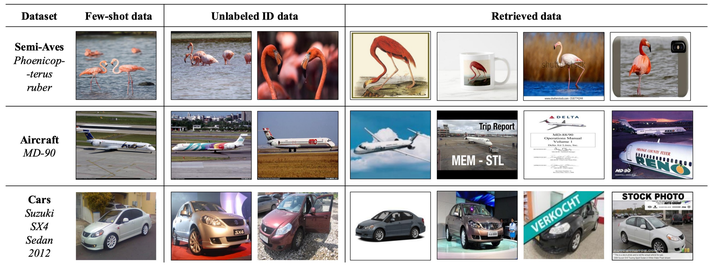
Failure Diagnostics & Remedies by Temperatures

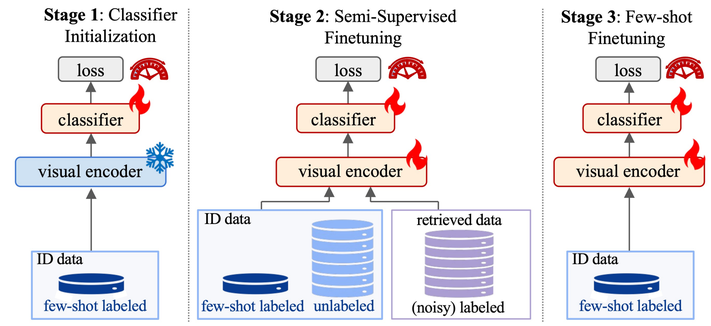
- Root cause: *VLMs produce "flat" distribution of softmax probabilities*, which results in
  - **zero utilization of unlabeled data** (w/ default confidence threshold of 0.8) → T_{conf} sharpens softmax confidences.
  - **weak training supervision** that prevents effective finetuning the VLM → T_{loss} strengthens training supervision.



Method

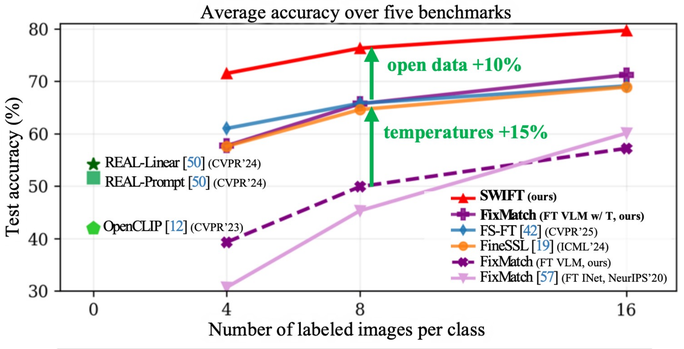
- Our **SWIFT** (stage-wise finetuning with temperatures) exploits retrieved data via a stage-wise training to mitigate domain gaps.





SOTA Performance

- **SWIFT** surpasses prior methods by >5% accuracy.



method	training data	mean accuracy over five datasets					
		OpenCLIP ViT-B/32			DINOv2 ViT-B/14		
		4-shot	8-shot	16-shot	4-shot	8-shot	16-shot
FS-FT [42] CVPR'25	L	61.0	65.8	69.1	56.5	71.1	79.6
FixMatch [57] direct	L+U	39.3	49.9	57.2	50.2	66.3	77.1
+ stage-1: Classifier Init.	L+U	56.3 $^{+17.0}$	59.8 $^{+9.9}$	62.2 $^{+5.0}$	56.7 $^{+6.5}$	71.5 $^{+5.2}$	80.3 $^{+3.1}$
+ stage-2: Semi-Sup. FT	L+U+R	68.8 $^{+11.1}$	73.3 $^{+7.6}$	77.3 $^{+6.1}$	76.3 $^{+19.6}$	82.0 $^{+10.5}$	86.3 $^{+6.0}$
+ stage-3: FS-FT (SWIFT)	L+U+R	**71.5** $^{+2.7}$	**76.3** $^{+3.0}$	**79.7** $^{+2.4}$	**78.2** $^{+1.9}$	**84.4** $^{+2.4}$	**87.8** $^{+1.5}$

References

[1] "The Neglected Tails in Vision Language Model". CVPR, 2024.
[2] "Few-shot Recognition via Stage-wise Retrieval-augmented Finetuning". CVPR 2025.
[3] "FixMatch: Simplifying semi-supervised learning with consistency and confidence". NeurIPS 2020.
[4] "Debiased Learning from naturally imbalanced pseudo-labels". CVPR 2022.
[5] "Erasing the bias: Fine-tuning foundation models for semi-supervised learning". ICML 2024.
[6] "Revisiting semi-supervised learning in the era of foundation models". NeurIPS 2025.