ECCV 2026 | MALMÖ | SEPT 8-12

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




Tian Liu, Anwesha Basu, James Caverlee, Shu Kong

Texas A&M University, University of Macau

Website and code: https://tian1327.github.io/SWIFT

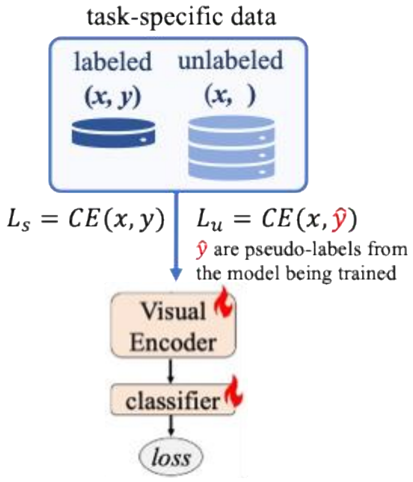
Motivation

- 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.

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**.

FixMatch (NeurIPS'20)
DebiasPL (CVPR'22)

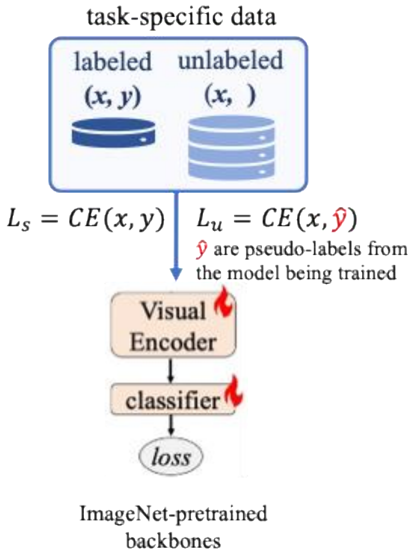


ImageNet-pretrained
backbones

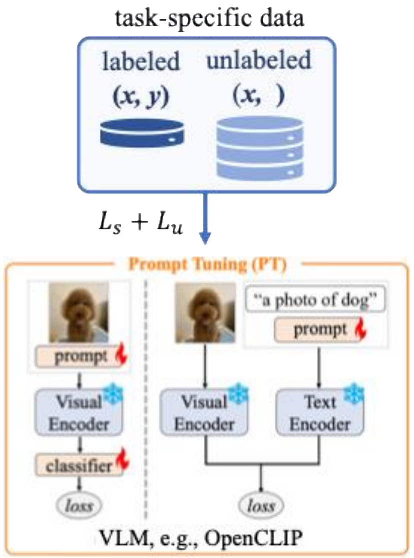
Motivation

- 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**.

FixMatch (NeurIPS'20)
DebiasPL (CVPR'22)



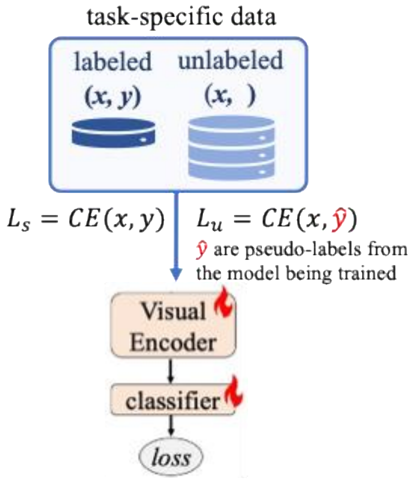
FineSSL (ICML'24)
V-PEFT (NeurIPS'25)



Motivation

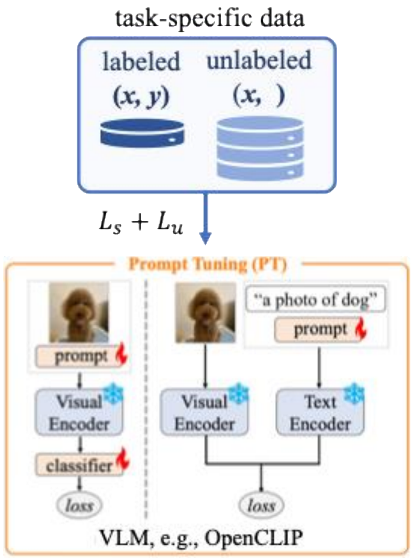
- Semi-Supervised Few-Shot Learning (SSFSL) mirrors real-world *auto-annotation,* aiming to learn a model over a few labeled and abundant unlabeled task-specific examples so to automatically annotate the unlabeled ones.
- Existing SSL or SSFSL methods either finetune an ImageNet-pretrained backbone or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

FixMatch (NeurIPS'20)
DebiasPL (CVPR'22)

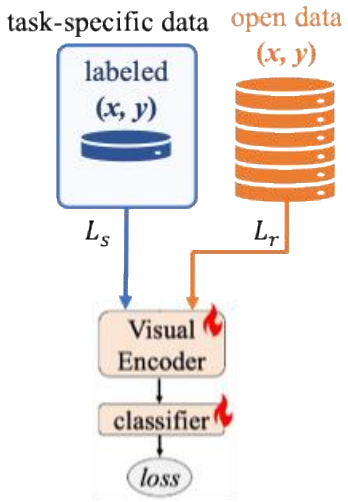


ImageNet-pretrained backbones

FineSSL (ICML'24)
V-PEFT (NeurIPS'25)



SWAT (CVPR'25)

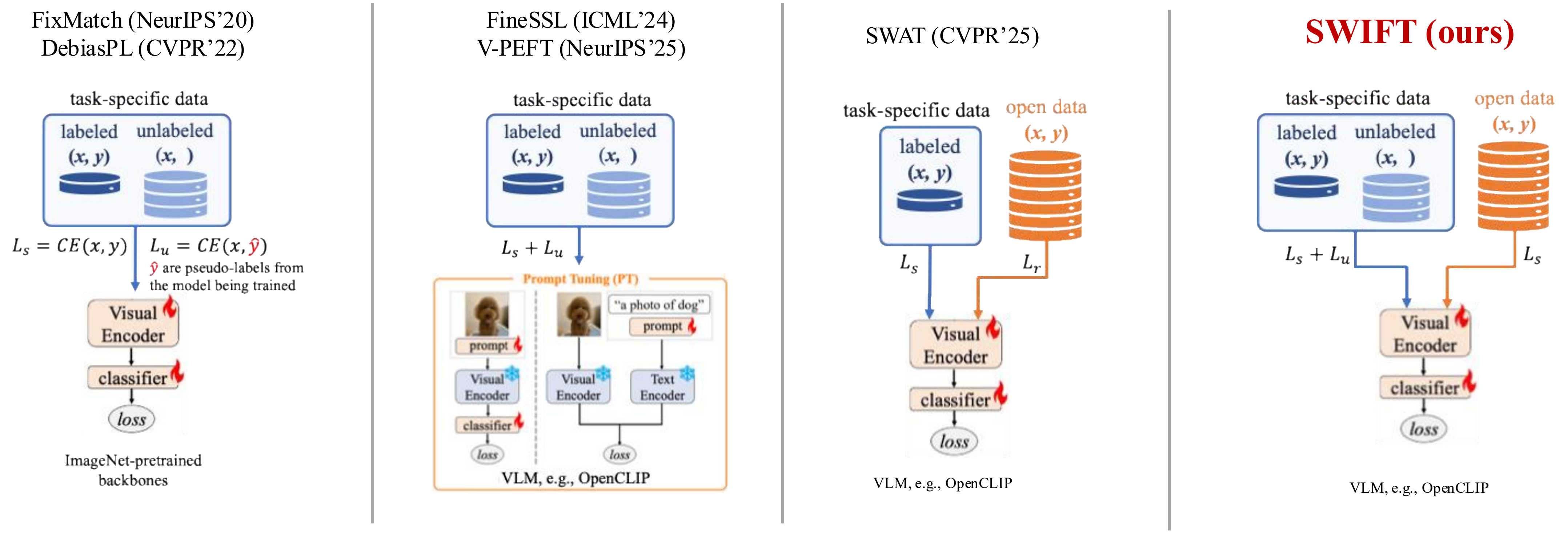


VLM, e.g., OpenCLIP

Liu et al. Few-Shot Recognition via Stage-Wise Retrieval-Augmented Finetuning, CVPR'25

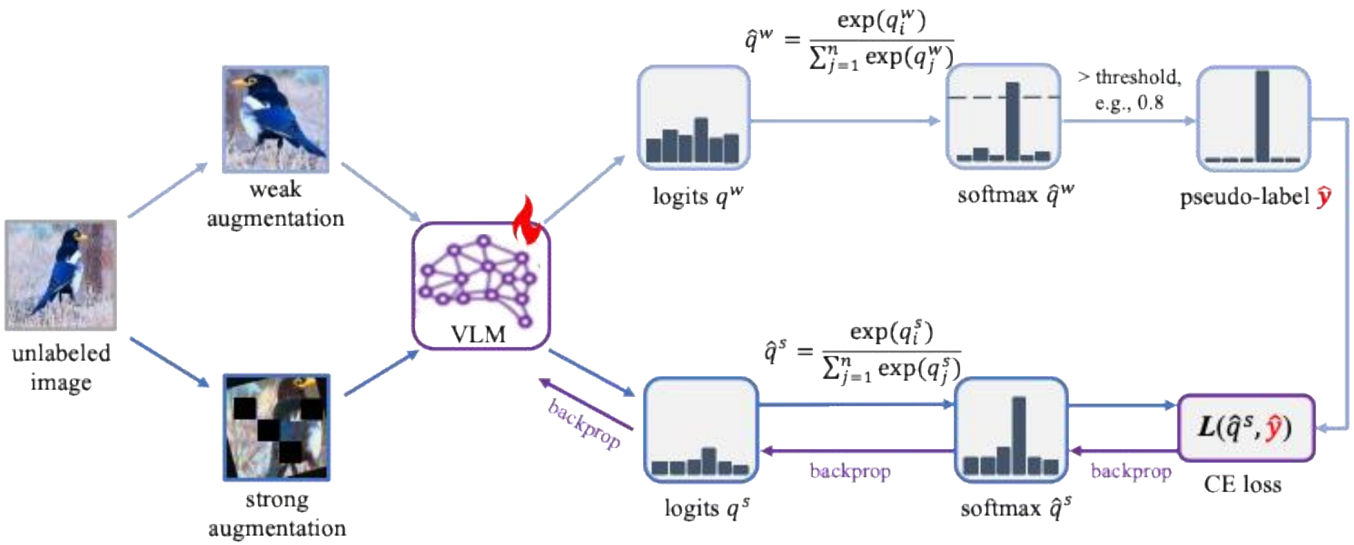
Motivation

- 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.



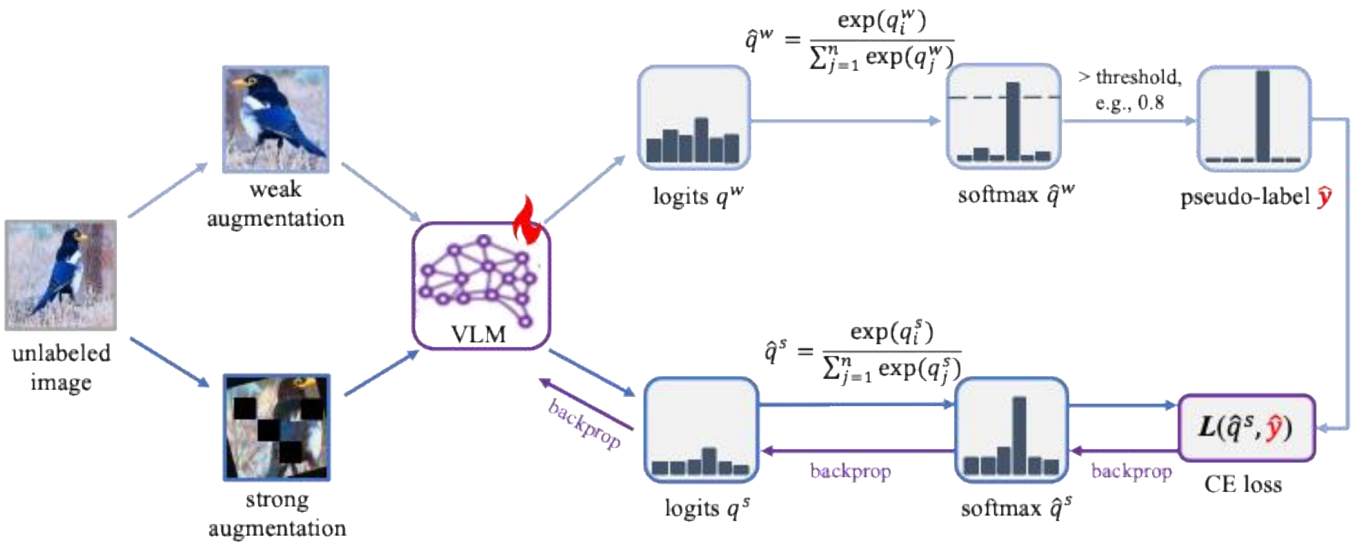
Challenges

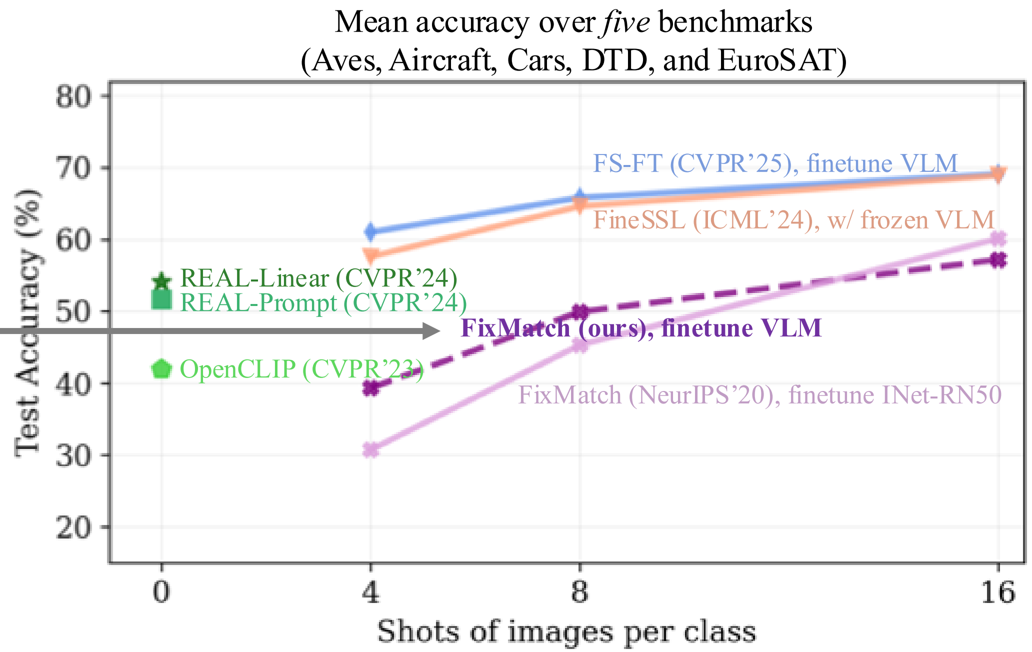
- We start by finetuning the VLM OpenCLIP ViT-B/32 with the representative SSL method FixMatch (NeurIPS'20) using the default hyperparameters (e.g., confidence threshold 0.8).



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.

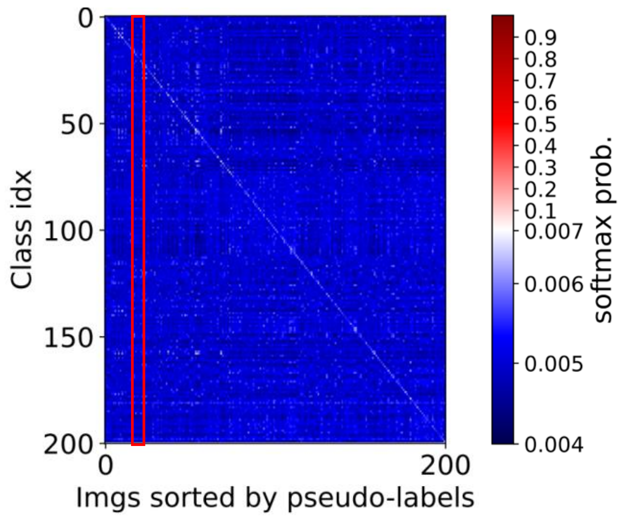




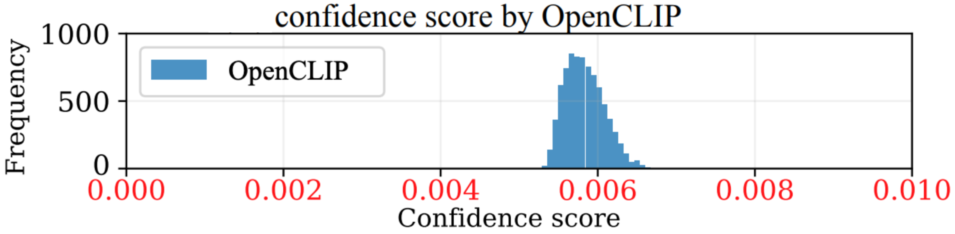
Liu et al. Few-Shot Recognition via Stage-Wise Retrieval-Augmented Finetuning, CVPR'25
Parashar et al. The Neglected Tails in Vision-Language Models. CVPR'24

Failure Diagnostics

Root cause: VLMs produce “flat” distribution of softmax probabilities



semi-Aves dataset (200 classes)

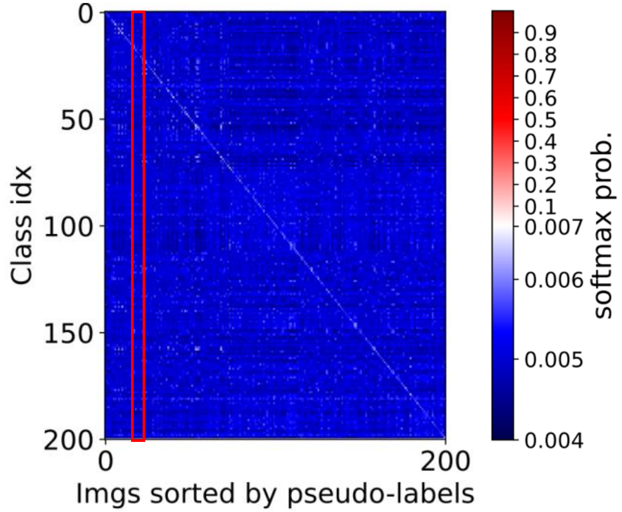


confidence = $\max(\hat{q})$

Failure Diagnostics

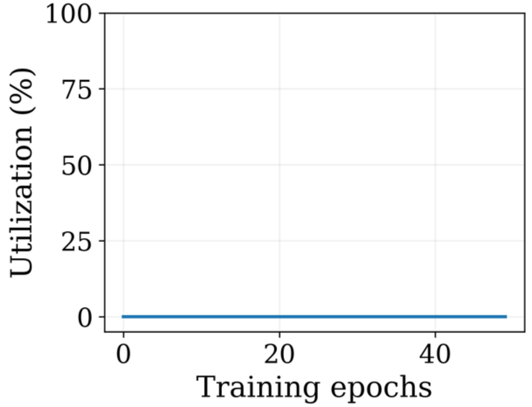
Root cause: VLMs produce “flat” distribution of softmax probabilities, which results in:

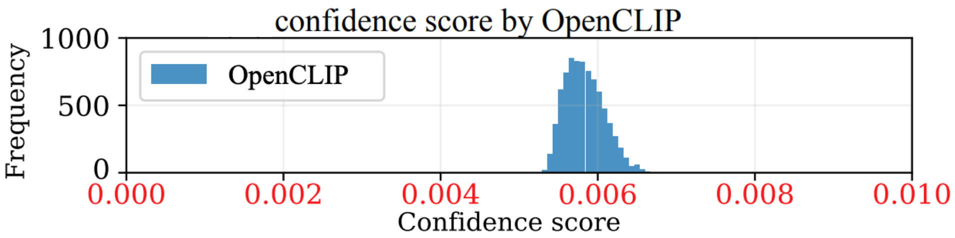
- **zero utilization of unlabeled data** (w/ default confidence threshold of 0.8);



semi-Aves dataset (200 classes)

Conf. > 0.8



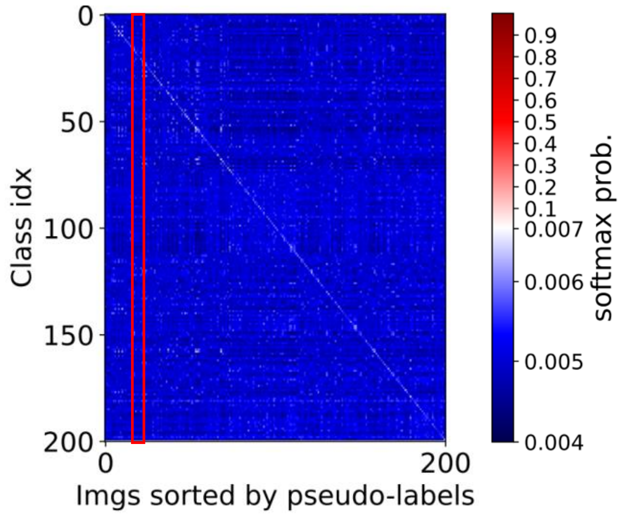


confidence = $\max(\hat{q})$

Failure Diagnostics

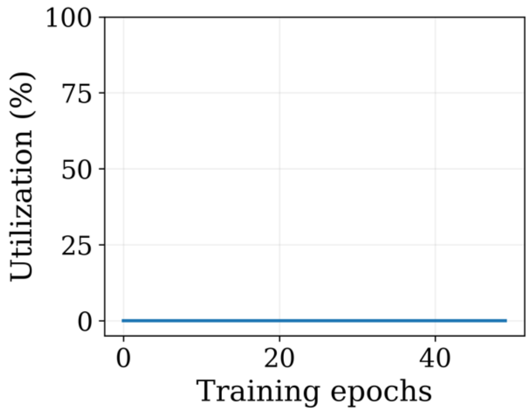
Root cause: VLMs produce “flat” distribution of softmax probabilities, which results in:

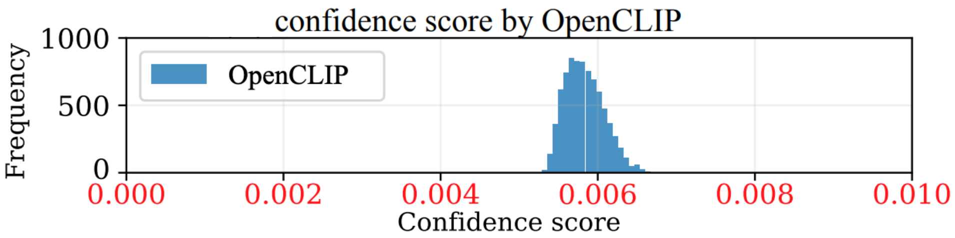
- **zero utilization of unlabeled data** (w/ default confidence threshold of 0.8);
- **weak training supervision** that prevents effective finetuning the VLM.



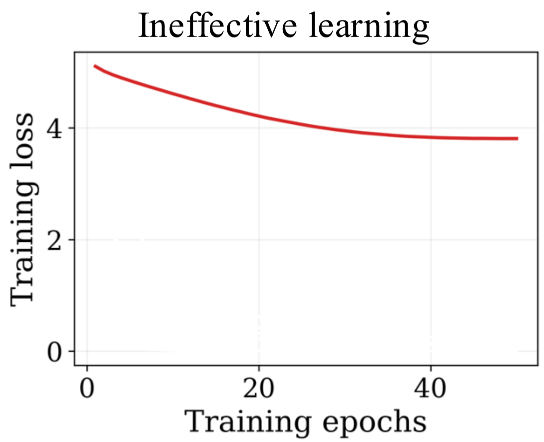
semi-Aves dataset (200 classes)

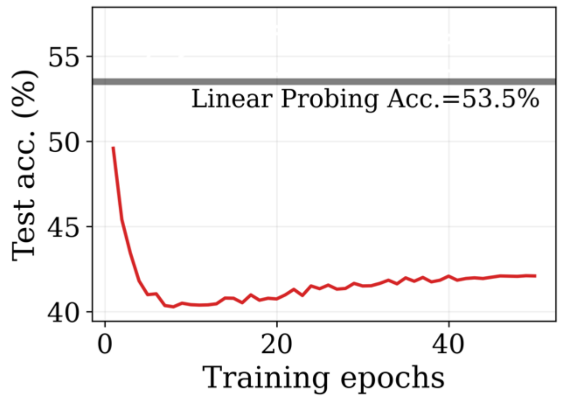
Conf. > 0.8





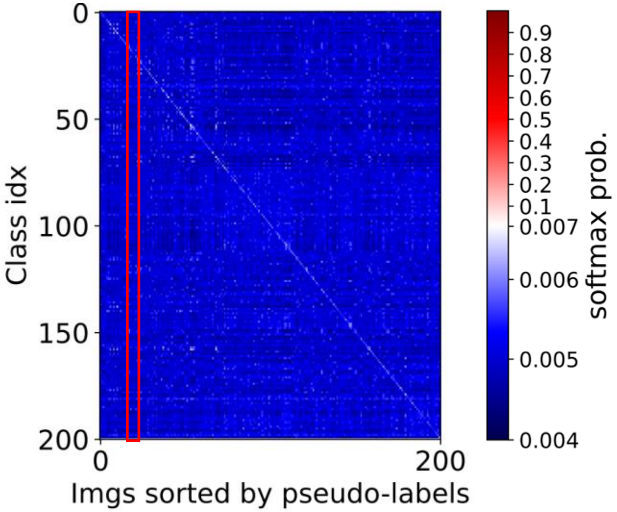
confidence = $\max(\hat{q})$





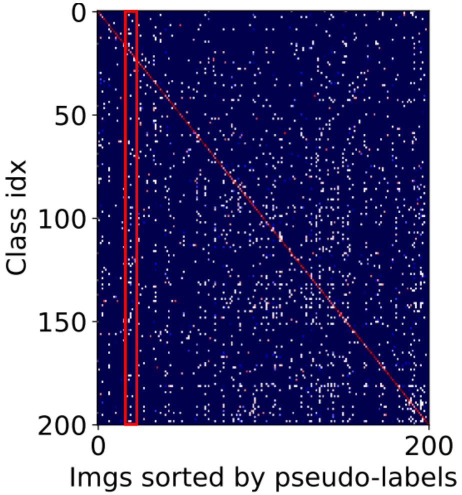
Solution

Temperature-based remedy:



semi-Aves dataset (200 classes)

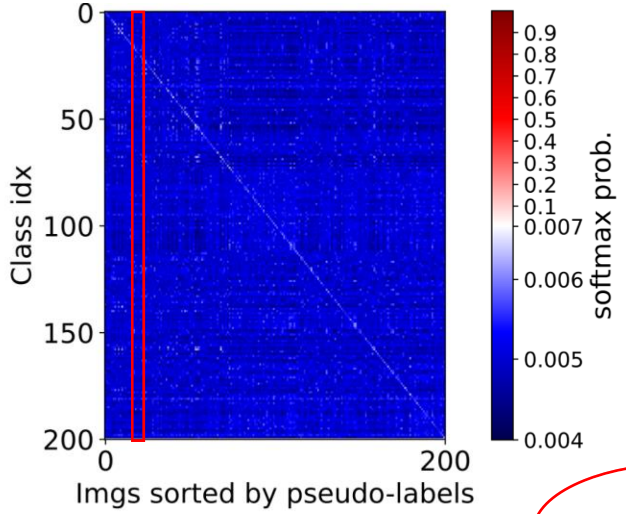
$$\hat{q} = \frac{\exp(q_i/T)}{\sum_{j=1}^{n}\exp(q_j/T)}$$



Solution

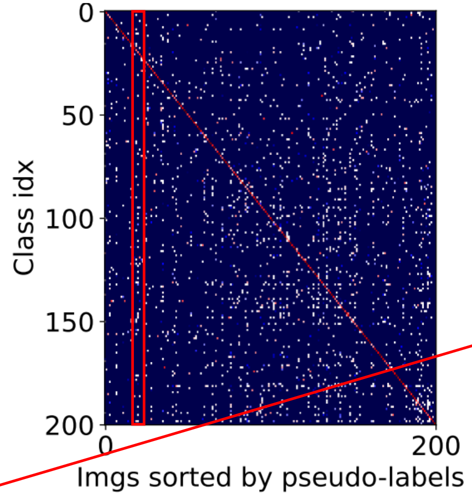
Temperature-based remedy:

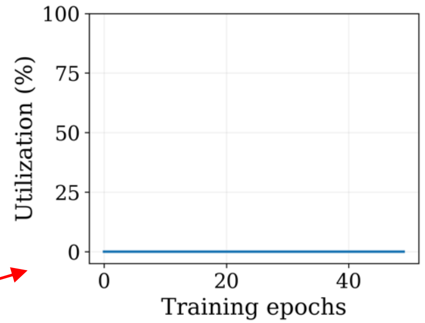
- use a confidence temperature T_{conf} $(e.g., 0.01)$ to sharpen the pseudo-label softmax probability;

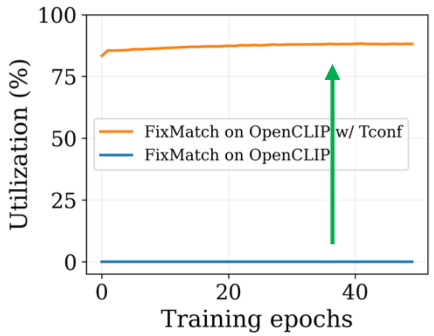


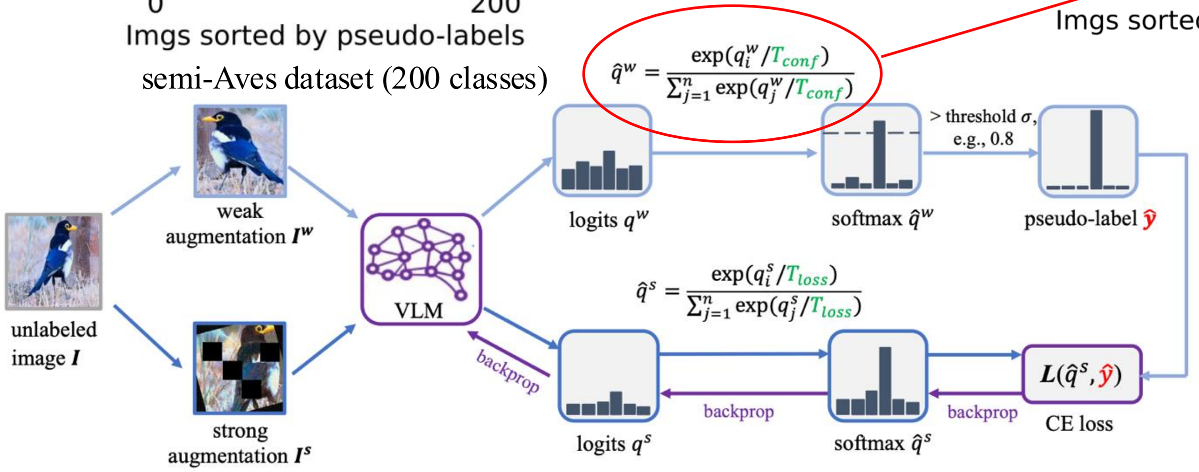
semi-Aves dataset (200 classes)

$$\hat{q} = \frac{\exp(q_i/T)}{\sum_{j=1}^{n} \exp(q_j/T)}$$





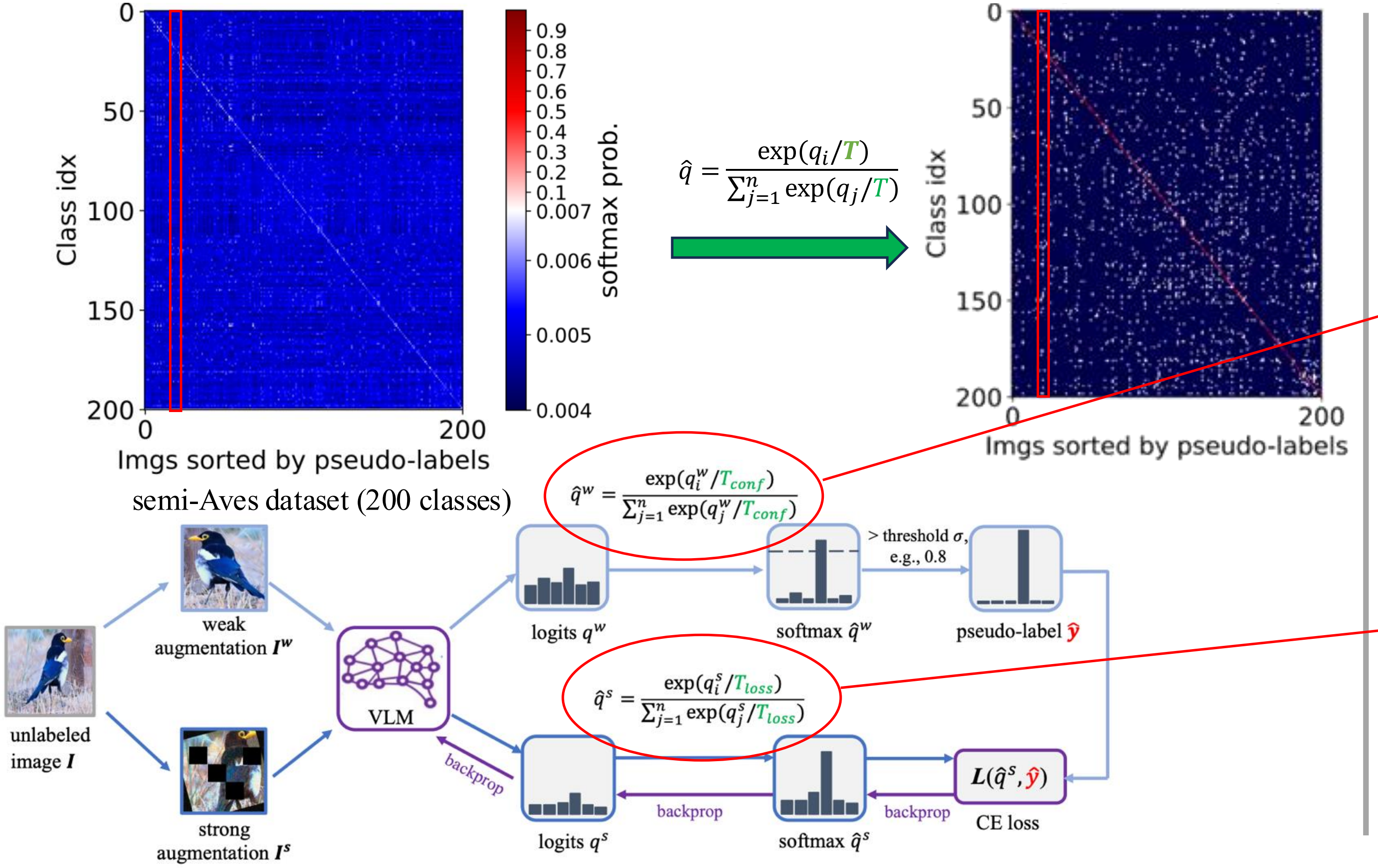


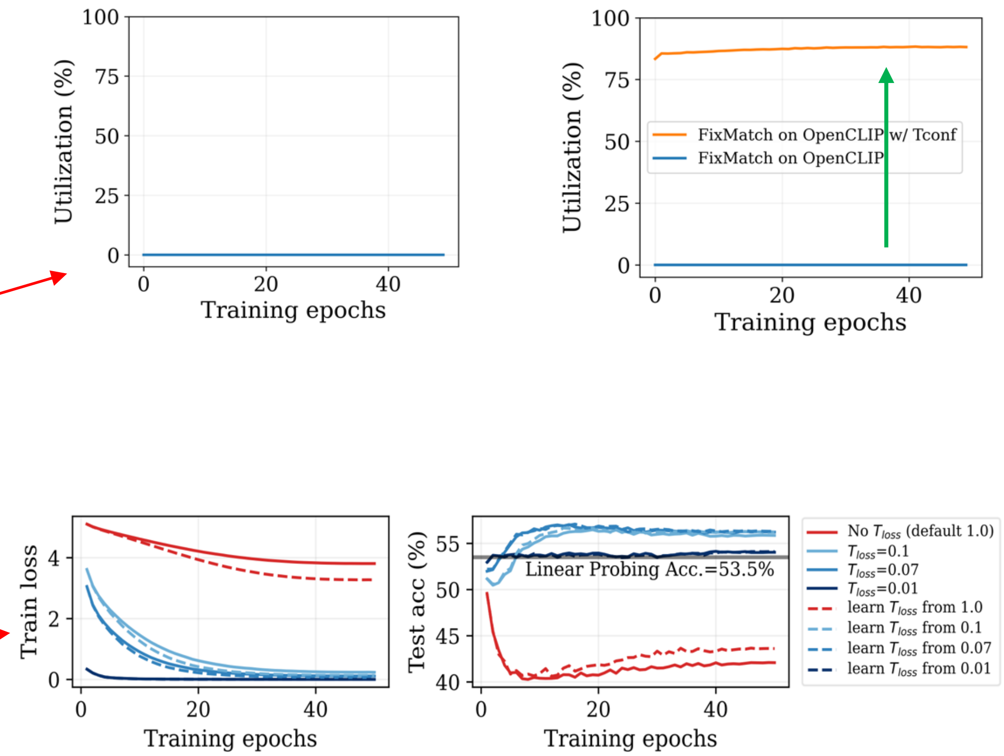


Solution

Temperature-based remedy:

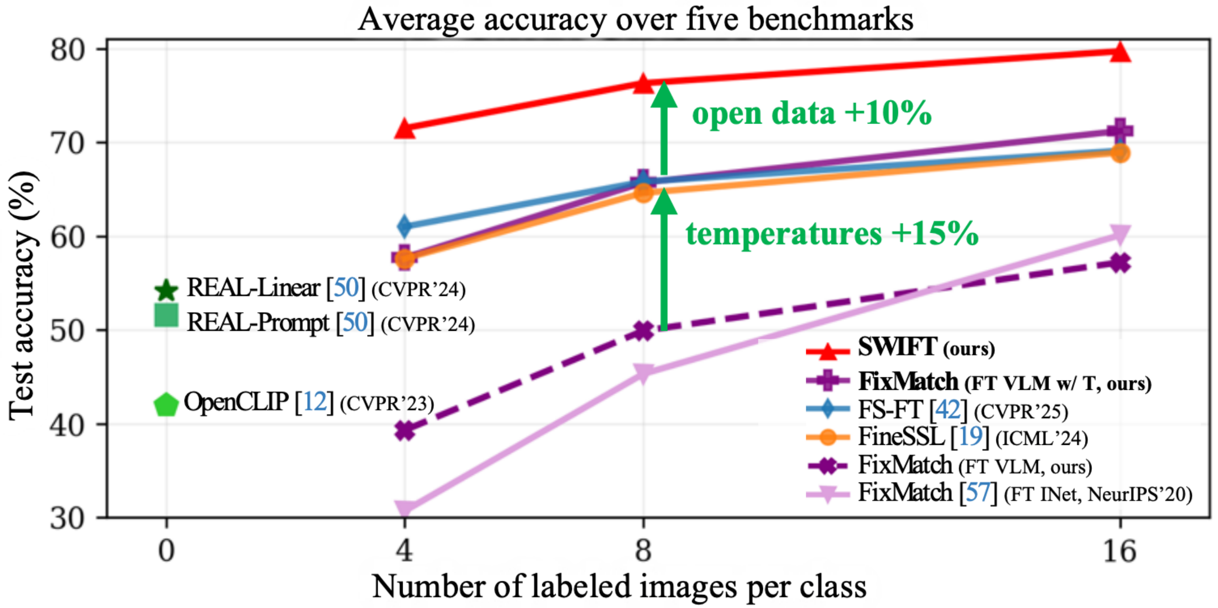
- use a confidence temperature T_{conf} $(e.g., 0.01)$ to sharpen the pseudo-label softmax probability;
- apply a loss temperature T_{loss} (which can be learned during SSL) to strengthen training supervision.

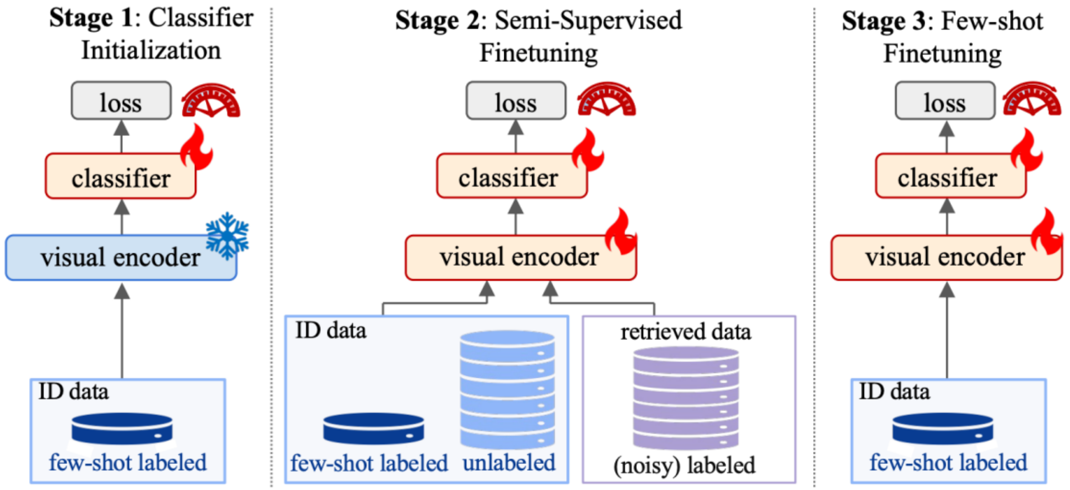




Results

- Our simply temperature-based solution improves FixMatch by over 15% accuracy!
- By further exploiting retrieved open data via stage-wise training, our SSFSL method **SWIFT** achieves another 10% accuracy gains!





Our **SWIFT** method (stage-wise finetuning with temperatures)

Conclusions

- Our work exploits finetuning VLMs and open data for solving SSFSL, advancing real-world *auto-annotation*.
- We reveal the failures of finetuning VLMs for SSL and propose a simple and effective temperature-based solution.
- Our SSFSL method **SWIFT** achieves the state-of-the-art performance.

website and code: https://tian1327.github.io/SWIFT