Co-HSF: Resource-Efficient One-Shot Semi-Supervised Adaptation of Histopathology Foundation Models

Luca Cerny Oliveira, M.S.[1 #], Kartik Patwari, M.S[1], Xiaoguang Zhu, M.S.[1], Sen-Ching Cheung, Ph.D.[2], Brittany N. Dugger, Ph.D.[3] and Chen-Nee Chuah, Ph.D.[1]

[1]Department of Electrical and Computer Engineering, University of California Davis,
[2]Department of Electrical and Computer Engineering, University of Kentucky, [3]Department of Pathology and Laboratory Medicine, University of California Davis.
# Indicates the corresponding author

Histopathology foundation models demonstrate promising zero-shot capabilities and achieve state-of-the-art (SOTA) performance after fine-tuning (Tab. 1). Current fine-tuning methods face key limitations:

- Overfitting on one-shot datasets
- Need for detailed image captions for fine-tuning
- High computational cost of training
- Inability to use unlabeled data effectively

Semi-supervised few-shot learning (SSFSL) offers an alternative, leveraging minimal labeled data alongside unlabeled samples. We study its use for foundation model fine-tuning and propose **Co**-filtered **H**istopathology **S**emi-Supervised **F**ew-Shot (**Co-HSF**):

- Dual-SSFSL training (teacher-student setup)
- Novel co-filtering pseudo-labeling technique
- Effectively exploits unlabeled data while reducing inference times

Co-filtering Pseudo-labeling

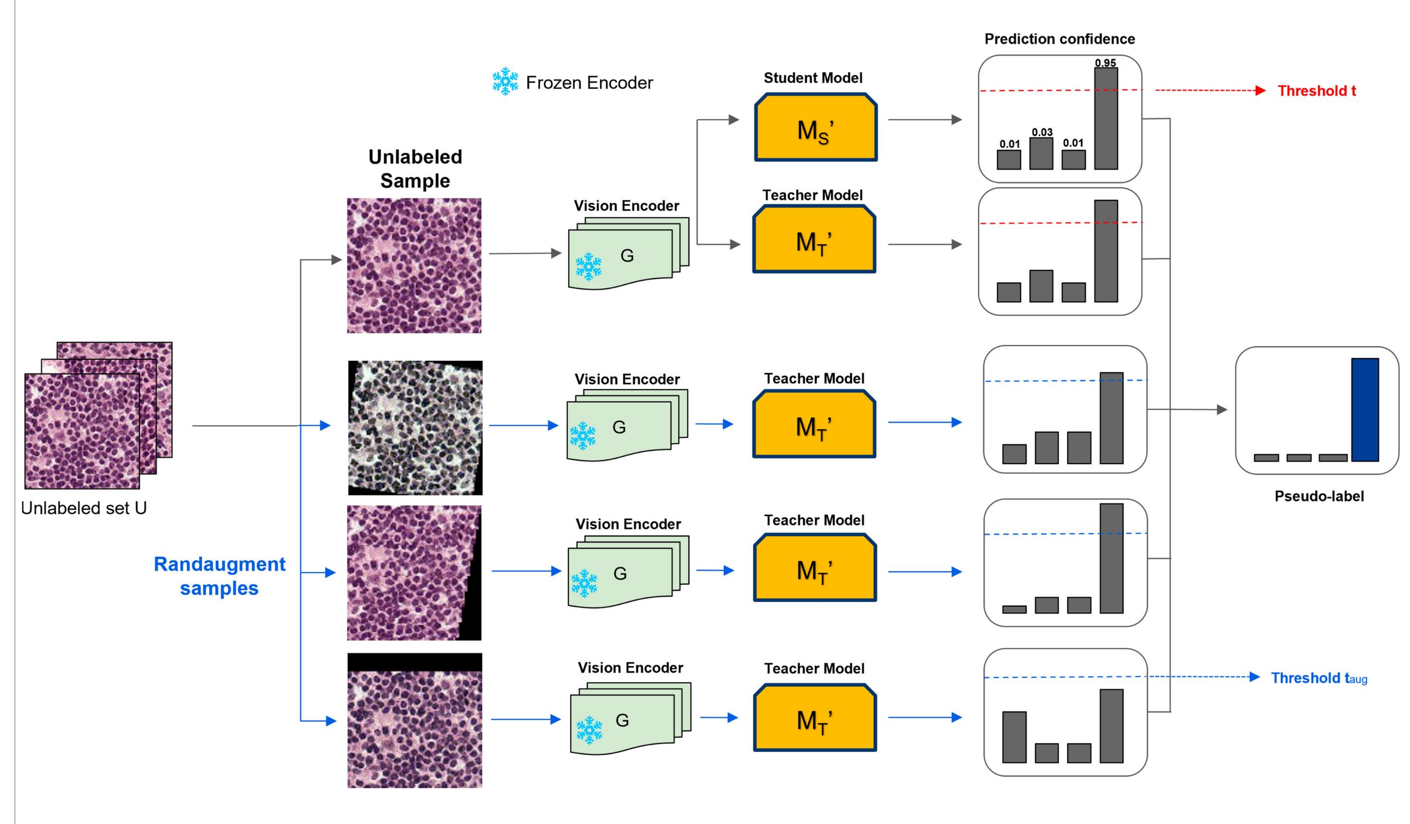

Fig. 2. Overview of the proposed Co-filtering pseudo-labeling strategy: the unlabeled set U is augmented by Randaugment[6] and featurized by CLIP's pre-trained vision encoder G(·). Both student and teacher models evaluate on U and U_{aug} respectively. The prediction confidences from the evaluations are analyzed, if the prediction confidences for the evaluations of a single sample and their HFT-augmented samples are all above the thresholds t and t_{aug} (both calculated from hyperparameter T), we will add the sample to the pseudo-labeled set.

Results

- Tab. 1 shows Co-HSF consistently outperforms both zero-shot and competing fine-tuning approaches in one-shot settings.
- Tab. 3 and Tab. 4 shows Co-filtering outperforms competing SSFSL, generating a more accurate pseudo-labeled set
- Co-HSF demonstrates lowest memory usage and faster inference times than most compared baselines, while leading accuracy performance (see Tab. 4)

Learning Type	Algorithm	PCam		NCT	
		Accuracy (%)	F-1 Score (%)	Accuracy (%)	F-1 Score (%)
Zero-shot	CLIP	56.57	56.56	28.08	23.64
	CLIP (ViT-B-L16)	57.02	53.15	45.46	39.91
	Zhang et al. 2023	53.35	40.39	50.83	46.82
	PLIP (Huang et al. 2023)	68.91	68.76	47.06	46.70
	CONCH (Lu et al. 2024)	82.66	82.57	51.70	47.83
One-shot Linear Probing	SGD (Huang et al. 2023)	80.55±3.61	80.41±4.03	82.26±5.30	81.54±1.46
	Logistic (Lu et al. 2024)	80.75±3.51	80.53±4.25	84.10±1.38	83.49±1.19
One-shot Fine-tuning	CoOp (Zhou et al. 2022)	57.80±0.01	47.04±0.65	24.30±0.02	31.58±6.45
	CLIPath (Lai et al. 2023)	72.18±4.04	70.00±8.44	69.24±1.46	65.65±4.76
One-shot FSL	Hu et al. 2021	80.15±0.78	80.15±0.79	68.22±2.05	68.12±2.10
	Snell et al. 2017	80.75±3.51	80.53±3.66	83.29±5.18	82.19±5.92
	ICI (Wang et al. 2020)	81.69±2.58	81.69±2.59	84.40±4.16	83.20±4.77
	PLCM (Huang et al. 2021)	81.54±1.79	81.49±1.82	87.49±3.62	86.62±4.71
One-shot SSFSL	ICI (Wang et al. 2020)	81.06±2.07	80.99±2.02	87.11±4.07	86.71±4.60
	PLCM (Huang et al. 2021)	82.02±3.43	81.93±3.49	88.76±2.96	88.42±3.35
	Co-HSF (proposed)	**84.41**±1.38	**84.40**±1.38	**90.10**±0.52	**89.82**±1.86

Tab. 1. Quantitative comparison on PCam and NCT-CRC-HE (NCT). CLIP refers to OpenAI CLIP[9]

Methods

- We use CONCH's vision encoder as G(·) (see Fig.1).
- Randaugment[6] augments labeled set X_{aug} for teacher model training and unlabeled set U_{aug} for co-filtering pseudo-labeling (see Fig. 2).
- Hyperparameters: T (for pseudo-label confidence threshold), alpha (for class imbalance mitigation), and step (number of added samples to X each iteration).

Datasets and Evaluation Setup

- Compare average inference time and GPU memory consumption on NVIDIA Tesla T4 and Intel Xeon 4120
- Datasets: (a) PCam[7] is a binary class (control vs. tumor) breast tissue dataset, and (b) NCT-CRC-HE[8] is a nine-class colon tissue dataset containing two different tumor-positive classes.

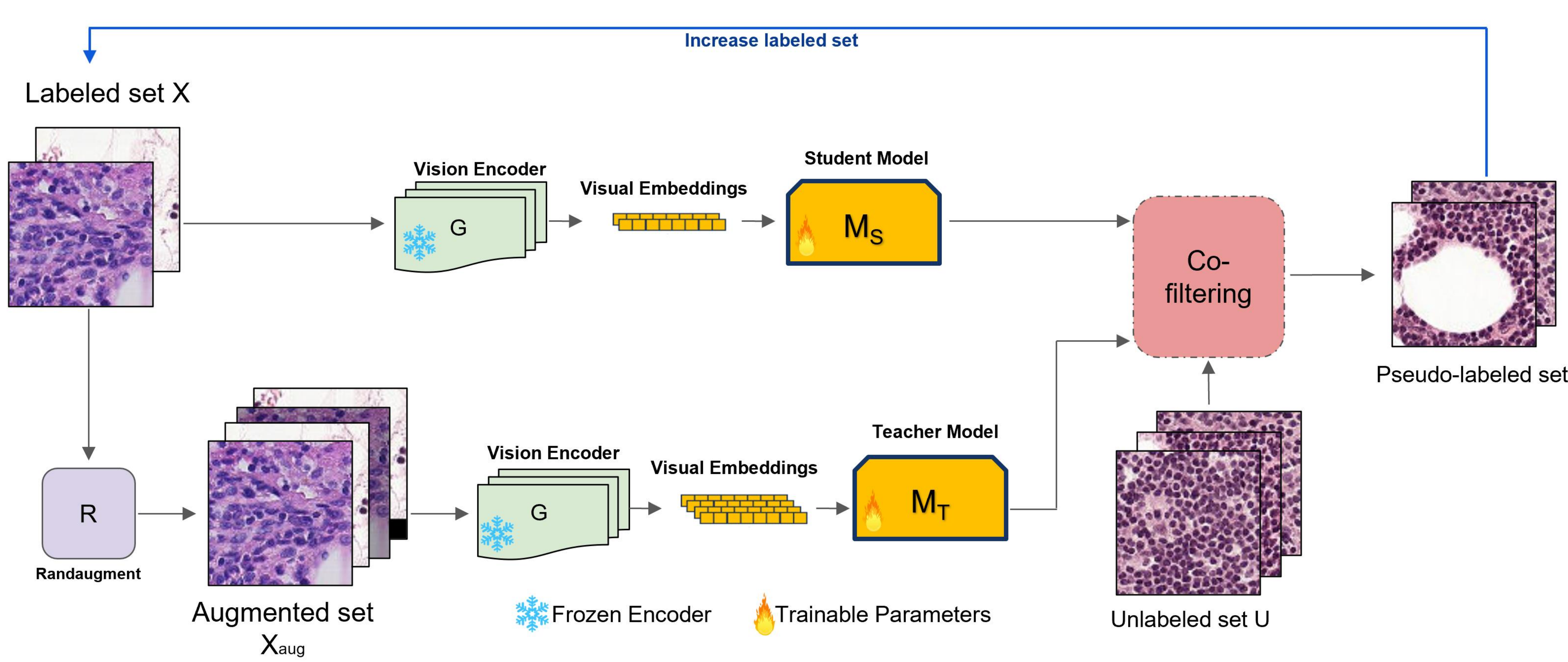

Fig 1. Overview of the proposed framework: the labeled set is augmented by Randaugment[6] and featurized by any CLIP-based pre-trained vision encoder G(·). Both the student and teacher models are trained using the CONCH[5] visual embeddings of X and X_{aug} respectively. The trained models and the unlabeled set are inputted to the Co-filtering algorithm (see Fig. 2), which selects samples and pseudo-labels to be added to X for the next iteration.

Co-filtering Pseudo-labeling Performance

Algorithm	Pseudo-label performance	
	Accuracy (%)	F-1 Score (%)
PLCM	88.04±1.53	84.46±1.74
ICI	89.33±0.89	86.03±1.69
Co-HSF (proposed)	**99.96**±0.10	**99.20**±1.60

Tab. 2. Quantitative comparison on pseudo-label performance for NCT-CRC-HE8

Algorithm	Pseudo-label performance	
	Accuracy (%)	F-1 Score (%)
PLCM	80.17±2.56	80.01±2.58
ICI	96.03±7.94	**96.04**±7.92
Co-HSF (proposed)	**99.05**±0.54	92.60±10.80

Tab. 3. Quantitative comparison on pseudo-label performance for PCam[7]

Resource Utilization

Algorithm	Inference Performance		
	Min	Acc(%)	VRAM(MiB)
CONCH	5.20±0.01	51.70	2222
CLIPath	9.22±0.07	69.24±1.46	2222
Linear Probing	**2.35**±0.01	84.10±1.38	952
ICI	11.86±0.14	87.11±4.07	952
PLCM	2.60±0.01	88.76±2.96	952
Co-HSF	2.44±0.01	**90.10**±0.52	**952**

Tab. 4. Quantitative comparison on test set inference time in minutes (Min), accuracy (Acc), and model VRAM utilization on NCT-CRC-HE dataset

Conclusion

- SSFSL can fine-tune foundation models for histopathology classification with lower GPU utilization and higher accuracy.
- Co-HSF can further improve accuracy and reduce inference time when compared to linear probing, adapter-based, prompt-based, and SSFSL methods.
- Co-HSFis better suited for deployability due to its lower GPU utilization and inference times (see Tab. 4)
- In the future, we aim to test the proposed method over additional, diverse histopathology datasets, different evaluation tasks (e.g. detection), and discuss scenarios such as class-imbalanced unlabeled sets and selection criteria for one-shot labeled dataset.

References

1. Lai et al. 2023. International Conference on Computer Vision Workshops.
2. Zhou et al. 2022. Conference on Computer Vision and Pattern Recognition.
3. Mirza et al. 2024. Advances in Neural Information Processing Systems.
4. Javed et al. 2024. Conference on Computer Vision and Pattern Recognition.
5. Lu et al. 2024. Nature Medicine.
6. Cubuk et al. 2020. Conference on Computer Vision and Pattern Recognition Workshops.
7. Veeling et al. 2018. Medical Image Computing and Computer Assisted Intervention.
8. Kather et al. 2018. Zenodo.
9. Radford et al. 2021. International Conference on Machine Learning.

Acknowledgments: This project was made possible by a grant from the National Institute on Aging (NIA) of the National Institutes of Health (NIH) under Award Number R01AG062517 and U24NS133949, Noyce Initiative UC Partnerships in Computational Transformation Program, NIH-National Institute On Aging awards #2R01-AG0652517, a 2023-2024 UC Davis Chancellor's Fellowship and Child Family Endowed Professorship Funds.