西北工业大学 NORTHWESTERN POLYTECHNICAL UNIVERSITY

空天地海一体化大数据应用技术国家工程实验室
National Engineering Laboratory for
Integrated Aero-Space-Ground-Ocean Big Data Application Technology

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

PEFAT: Boosting Semi-supervised Medical Image Classification via Pseudo-loss Estimation and Feature Adversarial Training

Qingjie Zeng[1*] Yutong Xie[2*] Zilin Lu[1] Yong Xia[1†]

[1] School of Computer Science and Engineering, Northwestern Polytechnical University, China

[2] The University of Adelaide, Australia

Poster Session WED-PM

CVPR 2023 Highlight

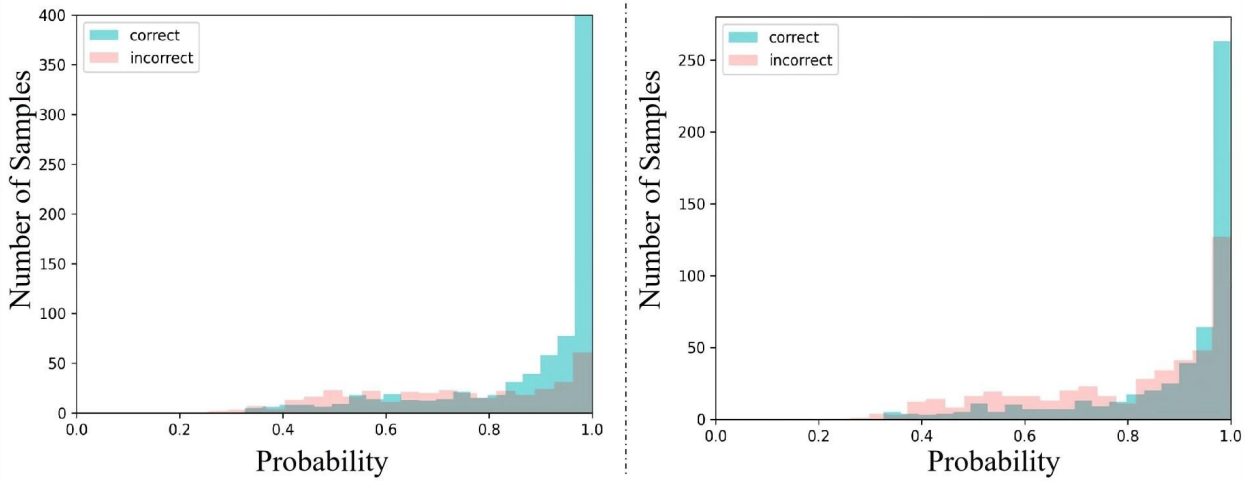
Paper

Code

THE UNIVERSITY of ADELAIDE

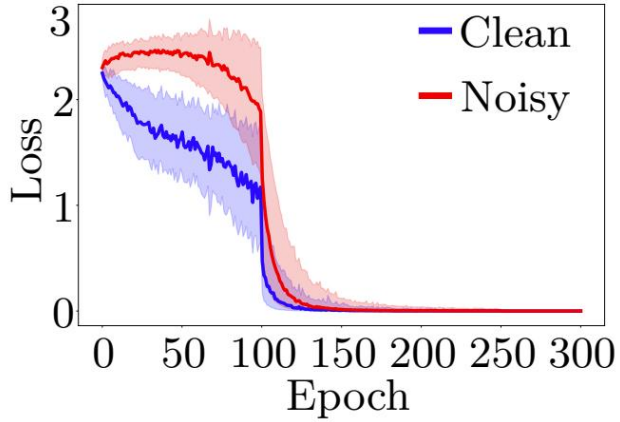
Observation

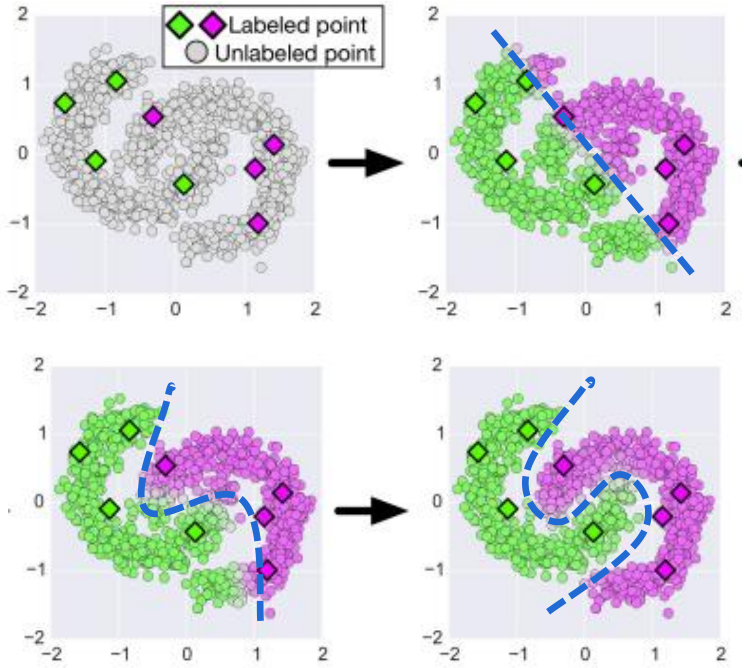
- Correct and incorrect pseudo-labeled data follow the similar probability distribution



Probability distribution on labeled data

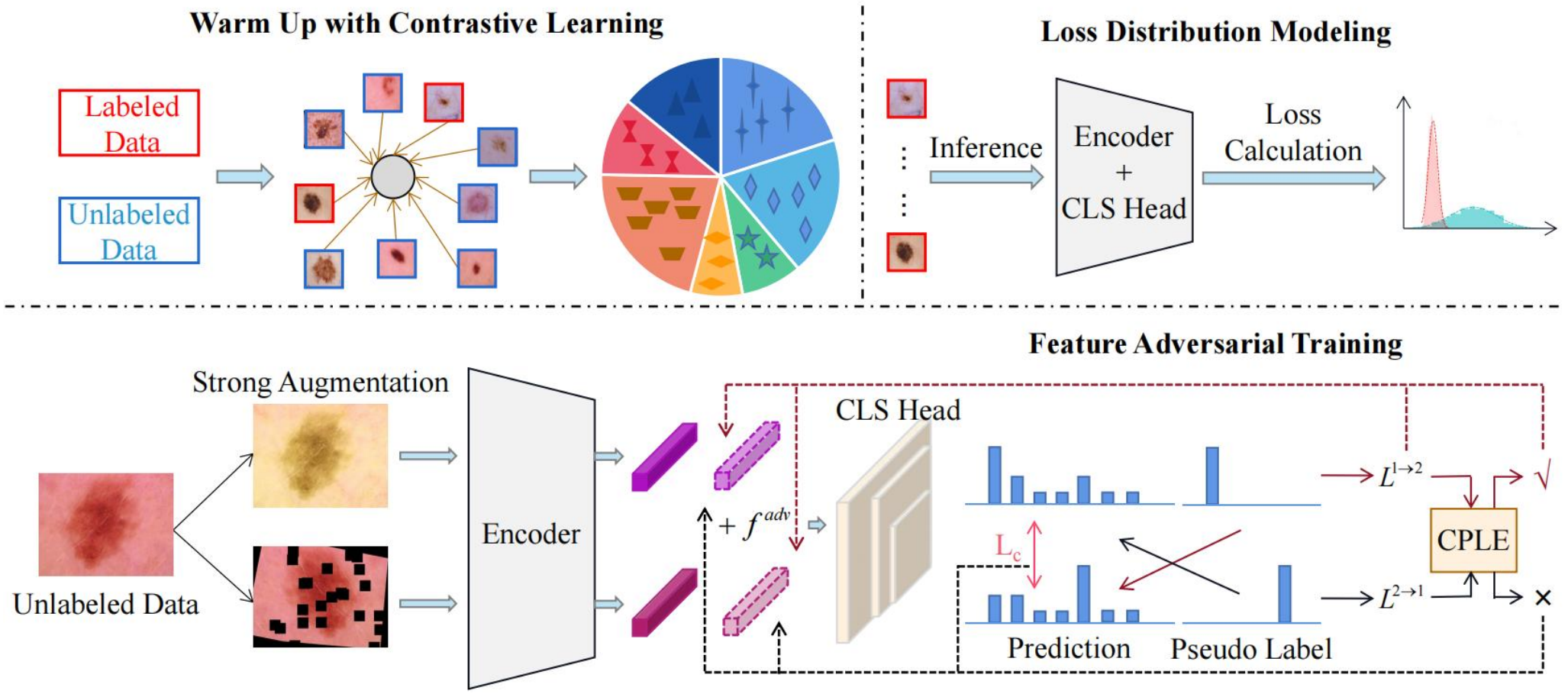
Probability distribution on validation data





Miyato T, Maeda S, Koyama M, et al. Virtual adversarial training: a regularization method for supervised and semi-supervised learning[J]. TPAMI 2018, 41(8): 1979-1993.
Arazo E, Ortego D, Albert P, et al. Unsupervised label noise modeling and loss correction. In: ICML. PMLR, 2019: 312-321.

Method



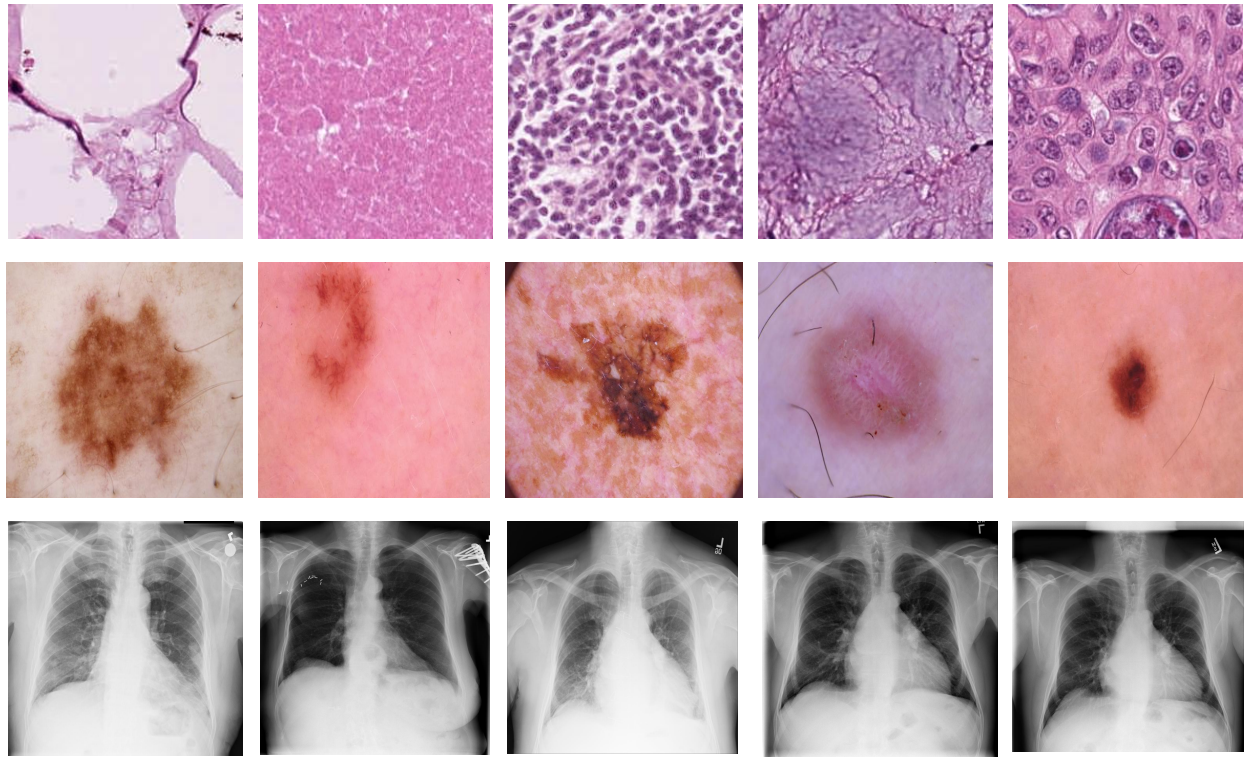
Step one: warm up the model using whole training data.
Step two: build loss prior on labeled data with Gaussian Mixture Model (GMM).
Step three: collect high-quality pseudo-labeled data using fitted GMM and cross pseudo-loss, meanwhile making predictions consistent after injecting feature-level adversarial noise

Datasets

1. **NCT-CRC-HE** (100,000 colorectal cancer histology slides, 9 classes), setting: 100/200 labeled data
2. **ISIC2018** (10,015 skin lesion dermoscopy images, 7 classes), setting: 5%/20% labeled data
3. **Chest X-Ray14** (112,120 chest x-rays from 30805 patients, 14 classes), setting: 2%/5%/10%/15%/20% labeled data

Evaluation Mertics

1. AUC
2. Accuracy
3. Specificity
4. Sensitivity
5. F1-score



Results

Table 1. Performance comparison with other state-of-the-art SSL methods on NCT-CRC-HE dataset. "SENS", "PREC" and "ACC" are Sensitivity, Precision and Accuracy, respectively. We list the evaluation metrics when 100 and 200 labeled data are given. Best and second best results are shown in **bold** and underline, respectively.

Method	NCT-CRC-HE (200 labeled data)					NCT-CRC-HE (100 labeled data)				
	AUC	SENS	PREC	ACC	F1	AUC	SENS	PREC	ACC	F1
Baseline	97.86	78.12	83.06	80.63	76.31	96.48	73.85	76.25	73.29	73.48
MT [33]	98.07	81.89	83.91	81.55	81.19	97.15	77.51	78.81	77.97	77.07
FixMatch [32]	98.43	85.03	84.75	84.81	84.66	97.91	80.59	81.78	80.47	80.28
SimPLE [14]	98.57	85.80	85.56	85.59	85.48	98.01	83.37	83.46	82.72	82.91
CoMatch [20]	98.83	87.94	88.70	86.48	86.24	98.00	84.72	84.58	83.93	84.11
SimMatch [41]	99.02	88.19	88.36	88.31	87.98	98.03	85.07	84.50	84.24	84.43
Ours	**99.08**	**89.68**	**91.18**	**90.29**	**90.12**	**98.25**	**86.82**	**86.78**	**86.01**	**86.33**

Table 2. Performance comparison on ISIC2018 dataset. "SENS", "SPCE" and "ACC" stand for Sensitivity, Specificity and Accuracy, respectively. Evaluation metrics are reported with the percentage of 5% and 20% labeled data. Best and second best results are shown in **bold** and underline, respectively.

Method	ISIC2018 (20% labeled data)					ISIC2018 (5% labeled data)				
	AUC	SENS	SPEC	ACC	F1	AUC	SENS	SPEC	ACC	F1
Baseline	90.90	69.37	91.77	91.42	51.89	84.28	56.32	87.53	85.36	40.96
SRC-MT [24]	93.58	71.47	92.72	92.54	60.68	87.61	62.04	89.36	88.77	46.26
DS^3L [11]	93.85	70.33	92.29	92.53	61.08	85.08	58.82	89.52	89.27	44.19
ACPL [22]	94.36	72.14	-	-	62.23	-	-	-	-	-
RAC-MT [12]	94.42	73.41	92.68	93.27	63.95	87.92	59.34	90.51	91.11	48.54
Ours	**94.87**	**76.72**	**93.45**	**93.68**	**66.15**	**88.64**	**64.10**	**91.25**	**91.81**	**50.96**

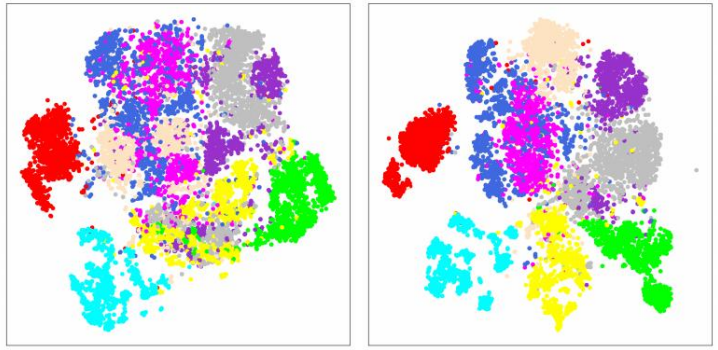
Table 3. Performance of mean AUC on Chest X-Ray14 dataset under the label percentage of 2%, 5%, 10%, 15% and 20%. Note that * denotes the methods employee DenseNet-169 as backbone with 384×384 input size, † means the methods use DenseNet-121 as backbone with 512×512 input size.

Method	Label Percentage				
	2%	5%	10%	15%	20%
Graph XNet* [3]	53.00	58.00	63.00	68.00	78.00
SRC-MT* [24]	66.95	72.29	75.28	77.76	79.23
UPS [29]	65.51	73.18	76.84	78.90	79.92
NoTeacher [34]	72.60	77.04	77.61	-	79.49
$S^2MTS^{2\dagger}$ [23]	74.69	78.96	79.90	80.31	81.06
ACPL† [22]	74.82	79.20	80.40	81.06	81.77
Ours	**75.06**	**79.54**	**80.93**	**81.56**	**82.58**

Ablation and Analysis

Table 4. Ablation study of each module in PEFAT on NCT-CRC-HE dataset. Results are reported in the case of 100 labeled data. $*$ and † denote singly employing FAT on the selected and unselected pseudo-labeled data, respectively.

Method	AUC	SENS	PREC	ACC	F1
Baseline	96.48	73.85	76.25	73.29	73.48
CPLE	98.09	84.57	83.89	84.16	84.65
CPLE+VAT	98.13	85.00	84.14	84.34	84.79
CPLE+FAT*	98.15	85.10	84.66	84.42	84.70
CPLE+FAT†	98.18	85.91	85.76	85.65	85.73
CPLE+FAT	**98.25**	**86.82**	**86.78**	**86.01**	**86.33**



(a) (b)

Figure 4. The t-SNE visualization on NCT-CRC-HE validation set. (a) is the result when using VAT; (b) shows the feature embedding when using FAT.

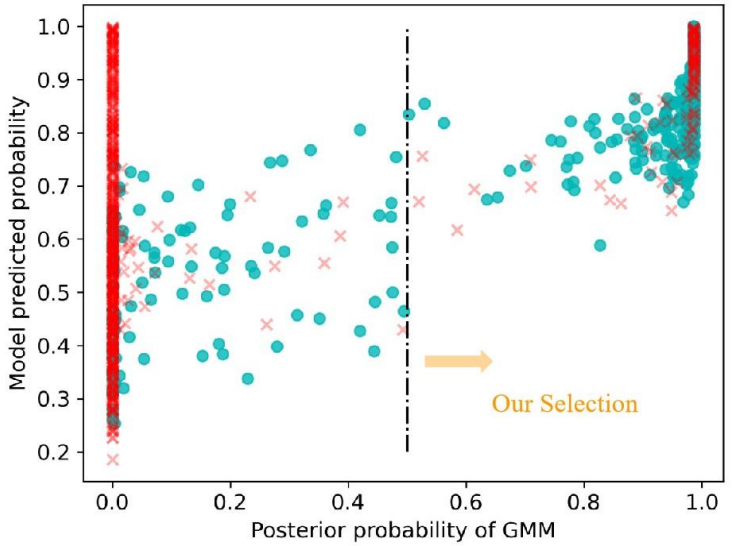


Figure 5. The relation between model predicted probability and posterior probability of GMM. Red × and Turquoise • denote unlabeled data with incorrect and correct pseudo-labels.

Table 5. Experiments conducted on NCT-CRC-HE validation set. $\delta = K$ means using probability threshold K to select pseudo-labeled samples. Ratio = Correct / Selected.

Method	Selected		Unselected ↓	Ratio ↑
	Correct ↑	Incorrect ↓		
δ=0.80	3821	531	5648	87.80
δ=0.85	3172	394	6434	88.95
δ=0.90	2387	257	7356	90.28
δ=0.95	1184	**112**	8704	91.36
CPLE	**6490**	592	**2918**	**91.64**