PSScreen: Partially Supervised Multiple Retinal Disease Screening

Boyi Zheng (boyi.zheng@oulu.fi) and Qing Liu* (qing.liu@oulu.fi)

Center for Machine Vision and Signal Analysis (CMVS), University of Oulu, Finland

Background

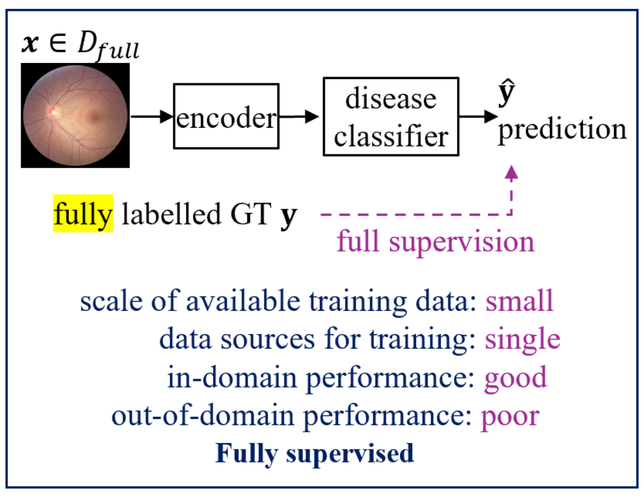
- 🎯 **Aim**: a single model to screen for multi-diseases
- 🚧 **Challenge**: most datasets are partially labelled and follow various distributions

Table1: Public datasets for retinal disease screening.

Open-access datasets		Labels for retinal diseases and the normal state							
names	notations and distributions	normal	diabetic retinopathy	glaucoma	cataract	age-related macular degeneration	hypertensive retinopathy	pathologic myopia	
Kaggle-CAT	$D_1 \sim \mathcal{P}_1$	✓	?	✓	✓	?	?	?	partially labelled
DDR	$D_2 \sim \mathcal{P}_2$	?	✓	?	?	?	?	?	
REFUGE2	$D_3 \sim \mathcal{P}_3$	?	?	✓	?	?	?	?	
ADAM	$D_4 \sim \mathcal{P}_4$	?	?	?	?	✓	?	?	
Kaggle-HR	$D_5 \sim \mathcal{P}_5$	?	?	?	?	?	✓	?	
PALM	$D_6 \sim \mathcal{P}_6$	?	?	?	?	?	?	✓	
ODIR	$D_{full} \sim \mathcal{P}_{full}$	✓	✓	✓	✓	✓	✓	✓	fully labelled

Existing Work

- 🌟 good in-domain performance
- ⚠️ insufficient use of existing public datasets
- ⚠️ poor OOD performance



Overlook the partially labelled data | Overlook the labels | Full use of images and labels

PSScreen

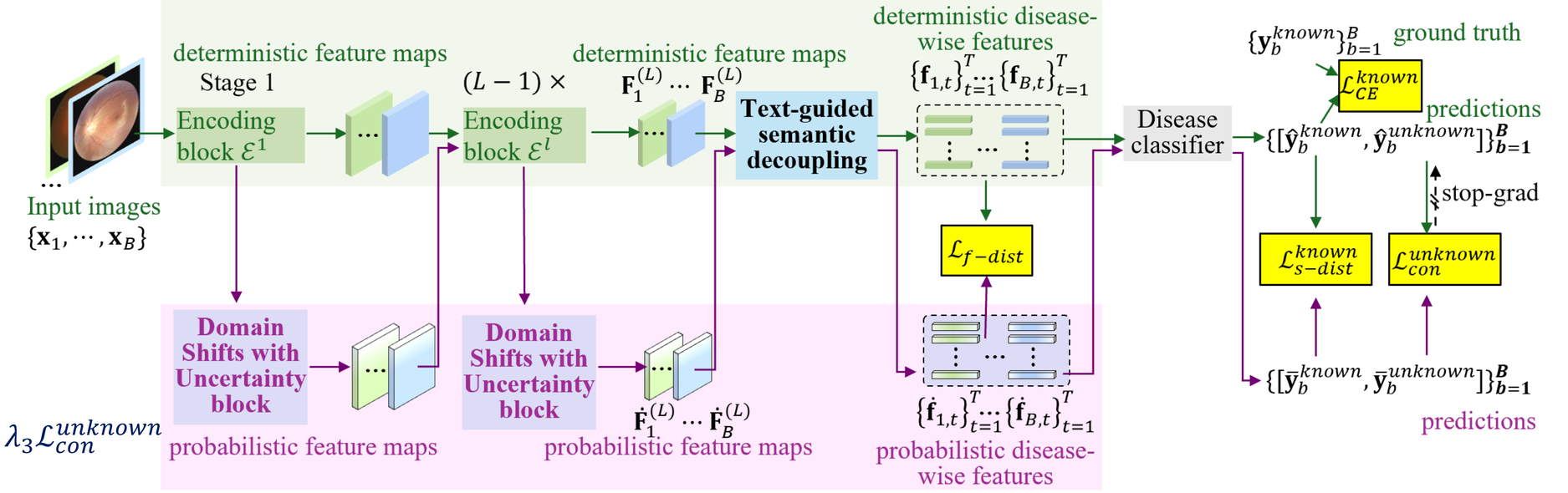
✨ a two-branch self-training framework

Deterministic feature learning branch

Enhance performances ⇄ Pseudo labels for unknown classes

Probabilistic feature learning branch

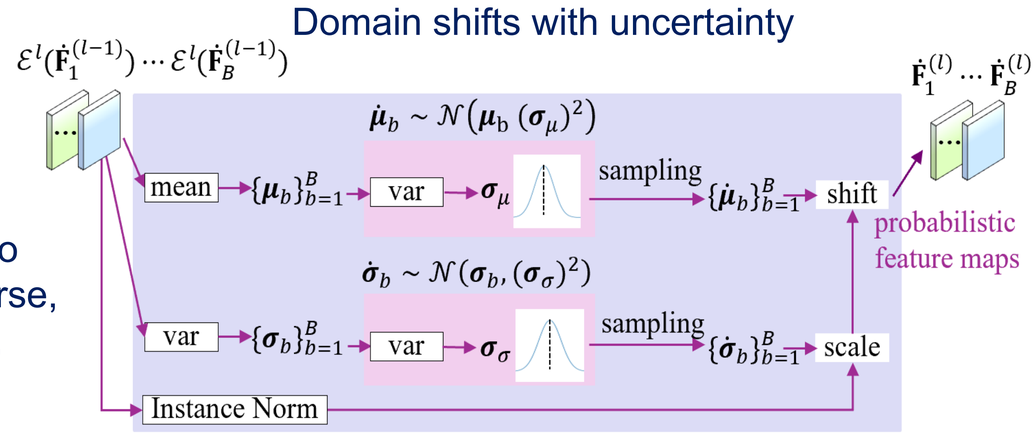
$$\mathcal{L} = \mathcal{L}_{CE}^{known} + \lambda_1 \mathcal{L}_{f-dist} + \lambda_2 \mathcal{L}_{s-dist}^{known} + \lambda_3 \mathcal{L}_{con}^{unknown}$$



Key Components

✨ model deterministic channel-wise statistics as uncertain distributions to simulate diverse, even unseen, domains

Domain shifts with uncertainty



✨ decouple the global feature map into disease-wise features

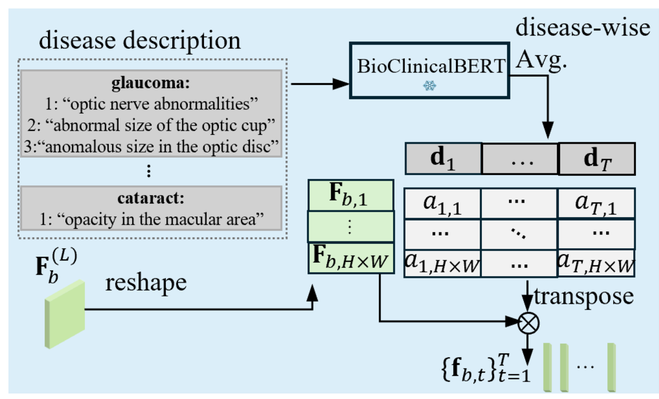
Attention score:

$$\alpha_{t,i} = \frac{\exp(\mathbf{v}^{\mathrm{T}} \tanh(\mathbf{W}_{att}^F \mathbf{F}_{b,i} \odot \mathbf{W}_{att}^d \mathbf{d}_t))}{\sum_{j=1}^{H\times W} \exp(\mathbf{v}^{\mathrm{T}} \tanh(\mathbf{W}_{att}^F \mathbf{F}_{b,j} \odot \mathbf{W}_{att}^d \mathbf{d}_t))}$$

Disease-wise features:

$$\mathbf{f}_{b,t} = \sum_{i=1}^{H\times W} \alpha_{t,i} \cdot \mathbf{F}_{b,i}$$

Text-guided semantic decoupling

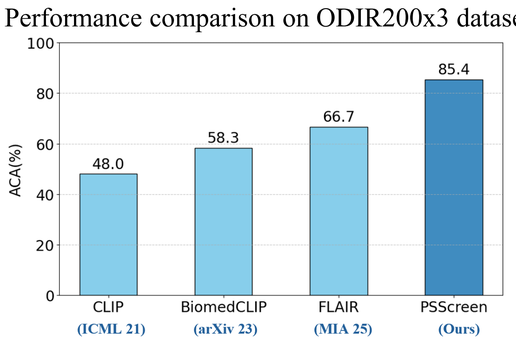


Experimental Validation

Table 2: Performance comparison on Meta (six datasets), Unseen (four datasets) and ODIR.

Methods	Source	Meta		Unseen		ODIR		FPS	GFLOPs	#param (M)
		mF	*mQWK*	*mF*	*mQWK*	*mF*	*mQWK*			
Full supervise	–	–	–	–	–	69.8±0.2	52.6±0.7	–	–	–
MultiNets	–	82.7±0.7	73.0±1.4	62.4±1.2	44.8±2.3	56.4±1.4	26.1±2.1	47.7	5240	837.4
MultiHeads	–	83.6±0.7	75.0±1.3	62.4±1.5	44.2±2.0	56.7±0.6	27.7±0.4	382.6	650	105.3
SST [2]	AAAI22	83.4±1.4	74.7±2.8	63.2±0.9	46.7±1.8	55.6±1.7	26.2±2.5	357.9	670	122.9
SARB [26]	AAAI22	84.0±0.9	75.7±1.8	60.5±1.0	40.3±1.2	61.4±0.9	33.8±1.6	361.6	670	112.7
BoostLU [15]	CVPR23	80.3±0.5	69.5±1.1	44.5±0.8	18.0±0.6	50.6±1.0	17.5±2.1	140.5	650	162.2
HST [3]	IJCV24	83.4±0.4	74.7±0.8	62.3±1.7	44.8±3.8	56.0±0.8	27.5±0.8	357.9	670	123.0
CALDNR [27]	TMM24	82.0±1.1	72.8±2.2	47.6±3.1	19.0±6.7	52.1±0.6	19.1±0.8	138.4	660	120.9
PSScreen (ours)	–	**84.2**±0.3	**76.8**±0.8	**65.9**±0.1	**50.9**±0.1	**64.1**±1.0	**39.8**±1.3	373.1	660	116.8

Performance comparison on ODIR200x3 dataset.



✨ More results

Table 3: Ablation study on key loss terms.

$\mathcal{L}_{f-dist}$	$\mathcal{L}_{s-dist}^{known}$	$\mathcal{L}_{con}^{unknown}$	Meta		Unseen		ODIR	
			mF	*mQWK*	*mF*	*mQWK*	*mF*	*mQWK*
			83.5±0.2	75.0±0.4	63.0±0.8	45.4±1.3	57.0±0.6	28.0±1.0
✓			83.4±0.7	75.3±0.2	63.4±0.1	46.7±0.9	58.0±0.4	29.6±0.5
	✓		83.4±0.2	75.8±0.4	64.0±0.3	47.5±0.7	57.5±1.0	28.7±1.5
		✓	83.3±0.7	75.5±1.1	61.8±1.1	44.1±2.1	61.7±0.9	34.8±1.6
	✓	✓	83.8±0.2	76.4±0.6	65.2±0.3	49.5±0.7	63.5±0.8	38.0±1.5
✓	✓	✓	**84.2**±0.3	**76.8**±0.8	**65.9**±0.1	**50.9**±0.1	**64.1**±1.0	**39.8**±1.3

Table 4: Compatibility with different backbones.

Methods	Meta	Unseen	ODIR
ResNet-101 [11]	75.0±1.3	44.2±2.0	27.7±0.4
+PSScreen	**76.8**±0.8	**50.9**±0.1	**39.8**±1.3
ConvNeXt-T [22]	75.0±0.7	48.2±1.2	25.4±2.3
+PSScreen	**76.6**±0.2	**49.6**±2.1	**37.0**±1.5
ConvNeXt V2-T [33]	76.3±0.7	48.6±1.2	28.9±1.3
+PSScreen	**77.0**±0.2	**49.1**±0.5	**38.3**±1.0
Swin-T [21]	75.8±0.8	49.5±0.7	25.4±0.3
+PSScreen	**79.3**±0.7	**50.8**±0.6	**35.5**±0.9
VMamba-T [20]	76.3±1.0	47.2±0.5	27.0±0.4
+PSScreen	**78.0**±0.7	**49.8**±1.7	**36.9**±1.8

Conclusion

- ✨ two-stream design with deterministic & probabilistic feature learning
- ✨ text-guided semantic decoupling for discriminative feature extraction
- ✨ pseudo label and distillation improve learning from partially labeled data
- ✨ state-of-the-art in- and out-of-domain performance

🌸 Acknowledgments 🌸

Academy Research Fellow project by RCF (Grant 355095), Hunan Provincial NSF of China (Grant 2023JJ30699), CSC–IT Center for Science, Finland

code