Supplementary Materials

Hongyu Hu[1]
mathewcrespo@sjtu.edu.cn
Tiancheng Lin[1]
ltc19940819@sjtu.edu.cn
Yuanfan Guo[1]
gyfastas@sjtu.edu.cn
Chunxiao Li[2]
lcx198910@126.com
Rong Wu[2]
wurong7111@163.com
Yi Xu*[1]
xuyi@sjtu.edu.cn

[1] MoE Key Lab of Artificial Intelligence
Shanghai Jiao Tong University
Shanghai, China

[2] Department of Ultrasound
Shanghai General Hospital
Shanghai Jiao Tong University School of Medicine
Shanghai, China

1 Dataset Details

Breast US-SWE Dataset The dataset collected in [2] consists of 2008 benign pairs and 1446 malignant pairs for diagnosing breast lesions in B-mode ultrasound (US) and shear-wave elastography (SWE) modalities. Main experiments and ablation studies are carried out on this dataset.

Synthesized Retinal Fundus-FFA Dataset The synthesized dataset in [5] is for diagnosis of age-related macular degeneration (AMD) in retina. It orignates from the IChallenge-AMD dataset [3], including 77% non-AMD and 23% AMD patients in fundus modality. Li *et al* [5] applied CycleGAN [6] on IChallenge-AMD dataset to generate corresponding fundus fluorescein aniography (FFA) modality to synthesize a new multi-modality dataset.

BUSI Dataset It contains 487 benign images and 210 malignant images for the diagnosis of breast lesion in ultrasound modality [1]. We transfer models pretrained on Breast US-SWE dataset to BUSI dataset, to validate transfer capacity of our proposed method.

IChallenge-PM Dataset It contains 1200 retinal images in color fundus modality for the classification of pathological myopia (PM) with 50% PM and 50% non-PM [4]. Models prertrained on Synthesized Retinal Fundus-FFA dataset are transferred to IChallge-PM dataset.

2 Experiment Settings in Evaluation

For downstream linear classification and finetuning evaluation, feature vectors from different modalities are concatenated and fed into the classifier. Other multi-modality fusion strategies are not utilized, to avoid mixed performance improvements introduced by our pretraining

method and the fusion method. All the linear classification and finetuning experiments are trained for 50 epochs, with a learning rate of 0.03 and a batchsize of 64.

3 Analysis on Scaling Factor

Fig.1 shows experimental results under different scaling factor. It is observed that pretrained model achieves its best performance with $\alpha = 3$, and the three metrics begin to drop when α is greater than 3. We infer that when scaling factor increases to a rather large value, gradient of global contrastive loss is more significant, and the consistency loss contributes less to overall network optimization. The network might neglect the local anatomical consistency, which is of great help for downstream task.

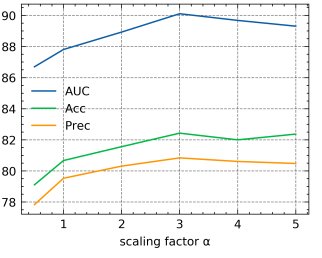

Figure 1: Linear classification results on Breast US-SWE dataset to study the impact of scaling factor in overall loss function.(Unit:%)

References

[1] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in Brief*, 28:104863, 2020. ISSN 2352-3409.

[2] Kun Chen, Yuanfan Guo, Canqian Yang, Yi Xu, Rui Zhang, Chunxiao Li, and Rong Wu. Enhanced breast lesion classification via knowledge guided cross-modal and semantic data augmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 53–63. Springer, 2021.

[3] H Fu, F Li, JI Orlando, H Bogunovic, X Sun, J Liao, Y Xu, S Zhang, and X Zhang. Adam: Automatic detection challenge on age-related macular degeneration. *IEEE Dataport*, 2020.

[4] Huazhu Fu, Jun Cheng, Yanwu Xu, Damon Wing Kee Wong, Jiang Liu, and Xiaochun Cao. Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE transactions on medical imaging*, 37(7):1597–1605, 2018.

[5] Xiaomeng Li, Mengyu Jia, Md Tauhidul Islam, Lequan Yu, and Lei Xing. Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis. *IEEE Transactions on Medical Imaging*, 39(12):4023–4033, 2020.

[6] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.