2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

MEAL: Meta enhanced Entropy-driven Adversarial Learning for Optic Disc and Cup Segmentation

Bingqi Ma[1], Qi Yang[2], Hui Cui[3] and Jiquan Ma[4]

Abstract— **Accurate segmentation of optic disc (OD) and optic cup (OC) can assist the effective and efficient diagnosis of glaucoma. The domain shift caused by cross-domain data, however, affect the performance of a well-trained model on new datasets from different domain. In order to overcome this problem, we propose a domain adaption model based OD and OC segmentation called *Meta enhanced Entropy-driven Adversarial Learning (MEAL)*. Our segmentation network consists of a meta-enhanced block (MEB) to enhance the adaptability of high-level features, and an attention-based multi-feature fusion (AMF) module for attentive integration of multi-level feature representations. For the optimization, an adversarial cost function driven by entropy map is used to improve the adaptability of the framework. Evaluations and ablation studies on two public fundus image datasets demonstrate the effectiveness of our model, and outstanding performance over other domain adaption methods in comparison.**

I. INTRODUCTION

Glaucoma is a chronic retinal disease and one of leading causes of blindness. With recent development of fundus imaging, qualitative and quantitative analysis of morphology of optic disc (OD) and optic cup (OC), and cup-to-disc ratio (CDR) have been used in diagnosis of glaucoma, accurate segmentation of OD and OC in fundus image is an critical step. Well-trained model often work well on the specific dataset, and cannot get satisfactory results on cross-domain data acquired from different scanners or subjects.

To overcome above drawback, unsupervised domain adaption methods have been exploited. Hoffman[1] and Javanmardi[2] have tried adversarial learning on domain adaption to restrict the uncertainty region of medical images. However, the rough optimization of the segmentation resulted in failing to get considerable performance. *p*OSAL[3] proposed an alignment-based approach on basis of adversarial learning, it simply splice the ResNet and adversarial learning and its result is not satisfactory. Recently, BEAL[4] improved the performance of *p*OSAL by introducing entropy-based adversarial learning for segmentation. However, the generalizability of the model is not considered.

In this work, we propose a novel unsupervised domain adaption framework, i.e, *meta enhanced entropy-driven adversarial learning (**MEAL**)* for OD and OC segmentation. Jointly considering accuracy, efficiency and generalization performance. To improve the accuracy, we propose a novel meta enhanced multi-feature attention segmentation network (MAN) to predict mask. Firstly, attention-based multi-feature fusion (AMF) is designed to enhance the perception on boundaries, especially at the soft boundary of OC. Secondly, we innovatively propose a data enhancement method called meta enhanced block (MEB) to improve generalization performance and efficiency of the model. To further obtain better segmented performance on challenging images, we introduce a entropy-based adversarial learning. Experimental results and ablation studies show that our model has achieved improved results on two public datasets RIM-ONE-r3[5] and Drishti-GS[6].

[1] HeiLongJiang University, China mabibgqi1@gmail.com
[2] HeiLongJiang University, China 10197014366@qq.com
[3] La Trobe University, Australia l.cui@latrobe.edu.au
[4] HeiLongJiang University, China majiquan@hlju.edu.cn (corresponding author phone: 18686825590)

II. METHODOLOGY

A. Meta Enhanced Block (MEB)

To improve the generalization and efficiency, we propose a novel meta enhanced block (MEB) inspired by meta-learning methods[7]. MEB has capacity to reduce overfitting to source domain and speed up the model converge to target domain.

MEB is composed of meta dataset x_a and a meta sub-network. Here, meta dataset consists of three different fundus image datasets (ORIGA, MESSIDOR and DRIVE) called 3-way where a way means each fundus image dataset and each way has 5 fundus image samples, called 5-shot. The meta sub-network consists of four convolutional layers for feature extraction, in which the first two layers are followed by pooling operation and the latter two layers have not pooling. In detail, since different fundus image datasets come from different devices, the key to solving the cross-device problem is to solve the low-dimensional pixel differences. The combination of MEB and AMF module can effectively improve the universality of the model. Given fundus images belong to different way of x_a, they were fed in each branch separately and processed by the meta sub-network. Then, these features are concatenated together to generate meta enhanced feature f_m. Compared with traditional artificial noise[8], MEB can get more detail of morphology of OC and OD, which can predict more reliable mask to accelerate model converging.

B. Attention-based Multi-feature Fusion (AMF)

Inspired by Attention U-Net[9], we design a novel AMF to integrate MEB features, low-level and high-level semantic features into the network. Firstly, high-level semantic features f_h and MEB features f_m are reshaped to half size by 1×1 convolution before concatenation operation. Afterwards, we apply ReLU activation and Sigmoid to generate attention weight α, and then multiplied with f_h by α to obtain the fused feature f_α.

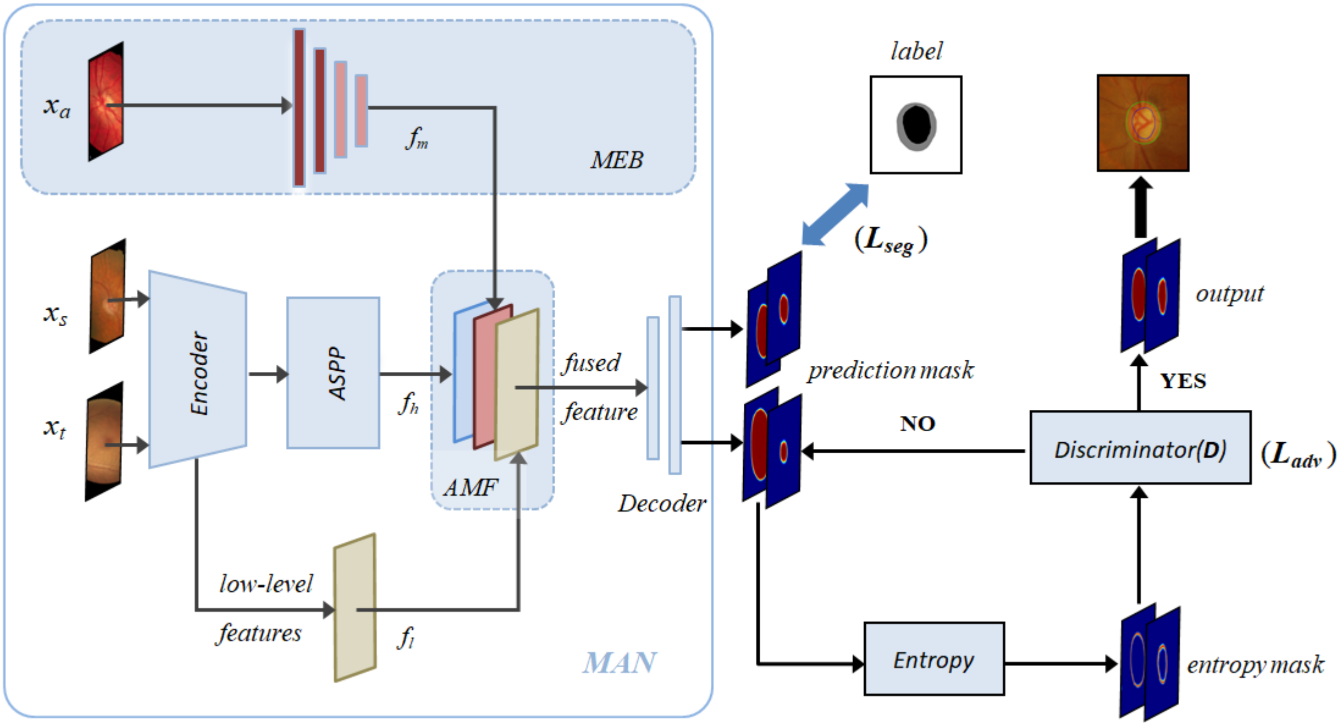

Fig. 1. Illustration of the MEAL domain adaption framework with MAN and entropy-driven adversarial learning. The major components in MAN include DeepLabV3+ decoder, atrous spatial pyramid pooling (ASPP), attention-based multi-feature fusion module (AMF), and meta enhanced block (MEB). We firstly train source domain (x_s) under supervision. Then, target domain (x_t) generates features via trained parameters. Three categories of features are generated, which are meta enhanced features (f_m) from MEB, semantic features enhanced by ASPP (f_h), and low-level features (f_l) from the first convolutional layer of encoder. The features are fused with attentional weights by AMF before feeding to a 2-layer decoder for segmentation mask generation. Entropy-driven discriminator D penalizes the results for the final segmentation.

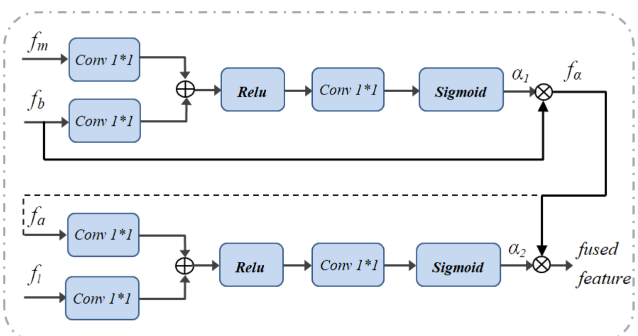

Fig. 2. Architecture of AMF module. Given multiple features including high-level semantic features f_h, meta enhanced features f_m, and low-level features f_l. Input features f_h and f_m are processed first. Attention features are generated by analysing both activations and contextual information provided by the two input features. Then, f_h is scaled with attention coefficients α_1 computed in attention block to generate f_α. f_α and f_l repeat the above operation to get fused feature.

For low-level semantic feature f_l, we do same operation with f_α. The final integrated fused feature is obtained for decoder. The channel attention structure can effectively integrate pixel-level features obtained from MEB into the feature map of the source domain. Besides, the accurate details of the low-level features are extracted and added to the high-level features as supplement. Compared with the conventional concatenation approach[10], the proposed AMF aligns different features with attention weights for effective extraction and identification of soft boundaries.

C. Loss Function

The segmentation predicted by MAN in the target domain may suffer from uncertainties on the boundary of OC and OD. To address this issue, we introduce an entropy-driven adversarial learning to seek a best estimation[4].

Given a source domain image set $\mathcal{I}_S \subset \mathbb{R}^{H\times W\times 3}$ along with ground truth mask $Y_S \subset \mathbb{R}^{H\times W\times 3}$, and a target domain image set $\mathcal{I}_T \subset \mathbb{R}^{H\times W\times 3}$ without ground truth. For each $x_s \in \mathcal{I}_S$, the prediction mask $p^m_{x_s}$ is generated by MAN. Similarly, the target domain prediction mask $p^m_{x_t}$ can be generated from target domain data $x_t \in \mathcal{I}_T$. To realize the entropy-driven adversarial learning, an discriminator D aligns the entropy masks between source domain *E(s)* and target domain *E(t)*. The loss function for discriminator D is defined as Binary Entropy Loss (*BCE*) and formulated as

$$\mathcal{L}_D = \frac{1}{N}\sum_{x_s\in\mathcal{I}_S}\mathcal{L}_{BCE}(E(x_s),1) + \frac{1}{M}\sum_{x_t\in\mathcal{I}_T}\mathcal{L}_{BCE}(E(x_t),0). \quad (1)$$

where *E(x)* is entropy mask of prediction mask encoding by Shannon Entropy as

$$E(x_s) = p^e_{x_s}\cdot\log(p^e_{x_s}), \quad (2)$$

$$E(x_t) = p^e_{x_t}\cdot\log(p^e_{x_t}). \quad (3)$$

Then, the adversarial loss $\mathcal{L}_{adv}^{e}$ is calculated based on $\mathcal{L}_D$ as

$$\mathcal{L}_{adv}^{e} = \frac{1}{M} \sum_{x_t \in \mathcal{I}_t} \mathcal{L}_{BCE}(E(x_t), 1). \tag{4}$$

$\mathcal{L}_{adv}^{e}$ makes target domain prediction converge to manual annotation. Total loss $\mathcal{L}_{tatal}$ is composed of supervised segmentation loss $\mathcal{L}_{seg}$ and adversarial loss $\mathcal{L}_{adv}^{e}$. $\mathcal{L}_{seg}$ is formulated as

$$\mathcal{L}_{seg} = -\frac{1}{N} \sum_{x_s \in \mathcal{I}_S} [y_{x_s}^{m} \cdot \log(p_{x_s}^{m}) + (1 - y_{x_s}^{m}) \cdot \log(1 - p_{x_s}^{m})], \tag{5}$$

where $p_{x_s}^{m}$ is the predicted mask, $y_{x_s}^{m}$ is ground truth. $\mathcal{L}_{tatal}$ is calculated as

$$\mathcal{L}_{tatal} = \mathcal{L}_{seg} + \lambda \mathcal{L}_{adv}^{e}, \tag{6}$$

where λ is an artificial coefficient.

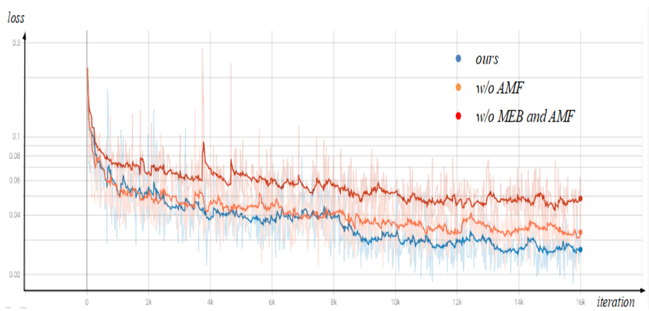

Fig. 3. Visualization of loss functions in our method with different architectures. The AMF improves the efficiency and precise of model and MEB further enhances accuracy of framework after 8000 iterations.

III. EXPERIMENTS AND RESULTS

A. Dataset and Pre-processing

To evaluate the performance of our method, we use public fundus image dataset REFUGE[1] as training and validation dataset of source domain that shot by Zeiss Visucam 500. Drishti-GS[6] and RIM-ONE-r3[5] are used as target domain datasets. Drishti-GS is shot through FOV 30° for diagnosing glaucoma. RIM-ONE-r3 is shot through Nidek AFC210 with a body of a Canon EOS 5D Mark II.

Region of interest (ROI) of size 512×512 centred at OD is cropped as input. We also perform data augmentation including random rotation, flipping, elastic transformation, contrast adjustment and random erasing as previous work[11][12].

TABLE I

COMPARISON WITH OTHER DOMAIN ADAPTION APPROACHES.

Methods	**RIM-ONE-r3**		**Drishti-GS**	
	$DICE_{cup}$	$DICE_{disc}$	$DICE_{cup}$	$DICE_{disc}$
TD-GAN	0.728	0.853	0.747	0.924
Hoffman *et al*	0.755	0.852	0.851	0.959
Javanmardi *et al*	0.779	0.853	0.849	0.961
OSAL-pixel	0.778	0.854	0.851	0.962
*p*OSAL	0.787	0.865	0.858	**0.965**
BEAL	0.810	0.898	0.862	0.961
MEAL(Ours)	**0.821**	**0.912**	**0.870**	0.962

[1]https://refuge.grand-challenge.org/

B. Implementation details

Our module is implemented on PyTorch platform. The discriminator D is optimized with SGD. Segmentation is optimized by Adam. The initial learning rate of Adam is set as 1e-3 and decreased by 0.2 every 100 epochs in total of 200 epochs. The learning rate of discriminator is set as 2.5e-5.

C. Evaluation metrics and comparison methods

Dice coefficients (DICE) score is used to evaluate OD and OC segmentation results. DICE is defined as

$$DICE = \frac{2 \times N_{TP}}{2 \times N_{TP} + N_{FP} + N_{FN}}, \tag{7}$$

where N_{TP}, N_{FP}, and N_{FN} represent the number of true positive, false positive and false negative pixels.

Comparison is done with six domain adaption methods: TD-GAN[13], high-level feature alignment[1], output space-based adaption[2], OSAL-pixel[3], its further improved model *p*OSAL[3], and BEAL[4]. Results is tabulated at Table.I.

D. Segmentation Results

The segmentation results of our method and other methods are shown in Table.I. For RIM-ONE-r3, our model outperformed all other methods. Our model achieved the highest DICE of 0.821 and 0.912 in OC and OD segmentation, followed by the second-best model BEAL. Particularly, we achieved 1.1% and 1.3% higher Dice on OC and OD segmentation than the second-best model on RIM-ONE-r3 dataset. For Drishti-GS dataset, our method achieved the best DICE of 0.870 on OC segmentation. For OD segmentation, our model was the second-best with a DICE of 0.962, which was slightly 0.3% lower than the best model.

Segmentation results on OD and OC by our model and the second-best BEAL are visualized in Fig.4. As shown, BEAL resulted in an inappropriate result on ambiguous border and dark color samples in prediction. Our model achieved better prediction on the boundary.

E. Ablation Study

TABLE II

RESULTS OF ABLATION STUDY ON DIFFERENT COMPONENTS.

Hierarchically fused blocks from:			**RIM-ONE-r3**		**Drishti-GS**	
EAL	MEB	AMF	$DICE_{cup}$	$DICE_{disc}$	$DICE_{cup}$	$DICE_{disc}$
			0.779	0.885	0.841	0.951
✓			0.800	0.898	0.851	0.960
✓	✓		0.806	0.902	0.856	0.958
✓	✓	✓	**0.821**	**0.911**	**0.870**	**0.962**

Our ablation studies are performed to demonstrate the contributions of the major components in our model. The results are given in Table.II. Without using EAL, MEB and AMF modules, the DICE score of OD and OC segmentation are 0.779 and 0.885 on RIM-ONE-r3, and are 0.841 and 0.951 on Drishti-GS. EAL module improved the DICE score on OD and OC of RIM-ONE-r3 to 0.800 and 0.898, and

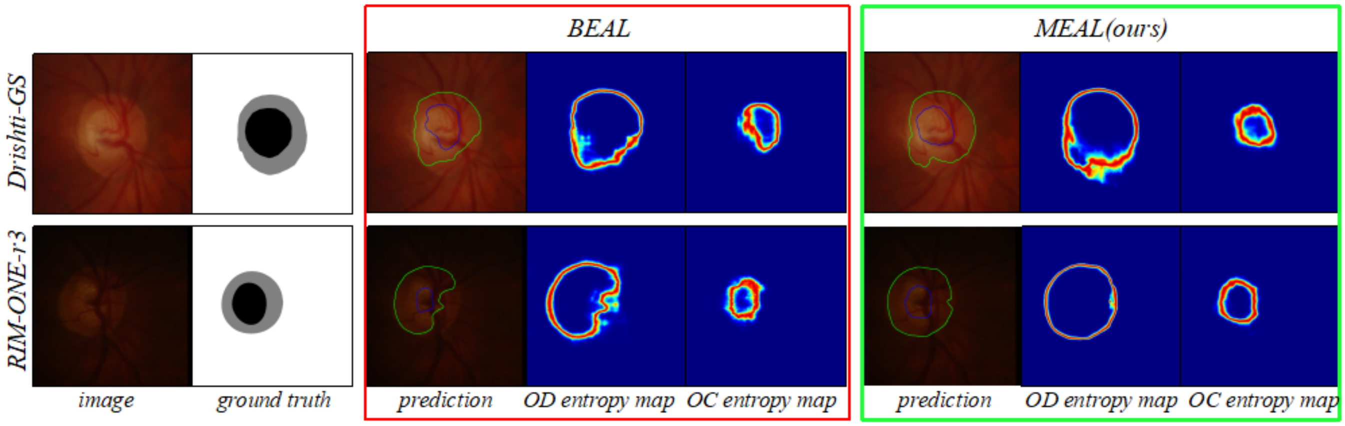

Fig. 4. Qualitative results on Drishti-GS and RIM-ONE-r3 datasets indicate that our method have a better appearance as shown above. Green and blue lines on prediction represent OD and OC borders, respectively. Otherwise, entropy maps of images are also shown, which further clearly appeals that our method is more predominant. The red region in them refers to margin of OD and OC.

those on Drishti-GS to 0.851 and 0.960. The primary reason is that adversarial learning improves the perception and segmentation at uncertain boundaries. MEB module further improved the DICE score of OD and OC to 0.805 and 0.902 on RIM-ONE-r3, the DICE to 0.856 and 0.958 on Drishti-GS. This is benefitted by MEB enhanced the generalizability of our framework on cross-domain segmentation work. The loss functions in Fig.3 further demonstrated the capacity of MEB to reduce overfitting to source domain and speed up the loss function converge on the target domain. AMF module further lifted the DICE score of OD and OC to 0.821 and 0.911 on RIM-ONE-r3, and those to 0.868 and 0.962 on Drishti-GS dataset. The attentional feature fusion mechanism adaptively fused the soft boundaries of OC and OD, which improved the segmentation accuracy. Furthermore, as Fig.3 shown, AMF improved the fitting ability of MEB on the target domain after 8000 iterations when compared with the model without using AMF.

IV. CONCLUSION

We proposed a novel meta enhanced entropy-driven adversarial learning for OD and OC segmentation. The meta-learning based data enhancement module improved the generalizability of the segmentation network and the attention based multi-feature fusion module improved the segmentation accuracy, especially at the soft boundaries. Our model outperformed the other state-of-the-art domain adaption model. However, we have not tested our model on other datasets. In future work, our model will be applied to other tasks in medical image analysis.

REFERENCES

[1] Hoffman J, Wang D, Yu F, et al. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation[J]. arXiv preprint arXiv:1612.02649, 2016.

[2] Javanmardi M, Tasdizen T. Domain adaptation for biomedical image segmentation using adversarial training[C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE, 2018: 554-558.

[3] Wang S, Yu L, Yang X, et al. Patch-based output space adversarial learning for joint optic disc and cup segmentation[J]. IEEE transactions on medical imaging, 2019, 38(11): 2485-2495.

[4] Wang S, Yu L, Li K, et al. Boundary and entropy-driven adversarial learning for fundus image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019: 102-110.

[5] Fumero F, Alayón S, Sanchez J L, et al. RIM-ONE: An open retinal image database for optic nerve evaluation[C]//2011 24th international symposium on computer-based medical systems (CBMS). IEEE, 2011: 1-6.

[6] Jayanthi Sivaswamy, S Krishnadas, Arunava Chakravarty, G Joshi, A Syed Tabish, et al., “A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis,” JSM Biomedical Imaging Data Papers, vol.2, no.1, pp.1004, 2015.

[7] Sung F, Yang Y, Zhang L, et al. Learning to compare: Relation network for few-shot learning[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1199-1208.

[8] Jiahua Dong, Yang Cong, Gan Sun, Bineng Zhong, and Xiaowei Xu, “What can be transferred: Unsupervised domain adaptation for endoscopic lesions segmentation,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp.4023–4032.

[9] Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: Learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.

[10] Hu X, Yu L, Chen H, et al. AGNet: Attention-guided network for surgical tool presence detection[M]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Springer, Cham, 2017: 186-194.

[11] Tsai Y H, Hung W C, Schulter S, et al. Learning to adapt structured output space for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7472-7481.

[12] Kamnitsas K , Baumgartner C , Ledig C , et al. Unsupervised Domain Adaptation in Brain Lesion Segmentation with Adversarial Networks[C]// International Conference on Information Processing in Medical Imaging. Springer, Cham, 2017.

[13] Yue Zhang, Shun Miao, Tommaso Mansi, and Rui Liao, “Task driven generative modeling for unsupervised domain adaptation: Application to x-ray image segmentation,” in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2018, pp. 599–607.