

Supplementary Material: Rethinking Domain Adaptive Optic Disc and Cup Segmentation in Fundus Image through Dynamic Diffusion Flow

Canran Li
cali5184@uni.sydney.edu.au

Dongnan Liu
dongnan.liu@sydney.edu.au

Weidong Cai
tom.cai@sydney.edu.au

School of Computer Science
The University of Sydney
Sydney, Australia

1 Proof of Convergence Conditions

Given an input source image x_i^s and a network f parameterized by θ , we can formulate the prediction result as $f(x_i^s; \theta)$. Moreover, the task gradient of each training process can be expressed as:

$$\mathbf{g}_t = \frac{\partial \mathcal{L}(f(x_i^s; \theta), y_i^s)}{\partial \theta}, \quad (1)$$

where \mathcal{L} denotes the loss function. Similarly, given the updated x_i^s , the task gradient of each training process should be:

$$\mathbf{g}_t' = \frac{\partial \mathcal{L}(f(x_N^s; \theta), y_i^s)}{\partial \theta} \quad (2)$$

$$= \frac{\partial \mathcal{L}(f(\sqrt{1 - \bar{\alpha}_n} - \bar{\gamma}_n x_i^s + \sqrt{\bar{\alpha}_n} I_T^f + \sqrt{\bar{\gamma}_n} I_T^p; \theta), y_i^s)}{\partial \theta}, \quad (3)$$

where $\bar{\alpha}_n$ and $\bar{\gamma}_n$ denote adjustable hyper-parameters, and I_T^f and I_T^p denote feature-level and pixel-level target domain-specific information, respectively. For the domain gradient, we use f_d to represent the network to avoid confusion:

$$\mathbf{g}_d = \frac{\partial \mathcal{L}_{BN}(I_{s,i}^f, I_t^f)}{\partial \theta} + \frac{\partial \mathcal{L}_{BN}(I_{s,i}^p, I_t^p)}{\partial \theta} \quad (4)$$

$$= \frac{\partial \mathcal{L}_{BN}(f_d(\sqrt{1 - \bar{\alpha}_n} - \bar{\gamma}_n x_i^s + \sqrt{\bar{\alpha}_n} I_T^f + \sqrt{\bar{\gamma}_n} I_T^p; \theta), I_T)}{\partial \theta}, \quad (5)$$

where I_T denotes the target domain-specific information from both feature-level and pixel-level. To preserve source domain data's inherent characteristics, each input image x_i^s is a

Dataset	Resolution	Camera	Number of samples	Release year
DRISHTI-GS [9]	2047 × 1760	Unknown	50 Train + 51 Test	2014
RIM-ONE-r3 [10]	1072 × 1424	Canon EOS 5D	99 Train + 60 Test	2015
REFUGE [11]	2124 × 2056	Zeiss Viscucam 50	320 Train + 80 Test	2018

Table 1: Statistics of three fundus datasets used for the proposed method.

fixed value, which can be seen as a constant term. Hence, the variables of Eq 3 are I_T^f and I_T^p , which are decided by the number of target input data stream images N . To preserve the data's inherent characteristics, the following condition should be satisfied:

$$g'_t \rightarrow g_t. \quad (6)$$

Based on Eq 1 and Eq 3, Eq 6 can be also written as:

$$f(\sqrt{1 - \bar{\alpha}_n - \bar{\gamma}_n} x_i^s + \sqrt{\bar{\alpha}_n} I_T^f + \sqrt{\bar{\gamma}_n} I_T^p; \theta) \rightarrow f(x_i^s; \theta). \quad (7)$$

Since $\bar{\alpha}_n$, $\bar{\gamma}_n$ and x_i^s are all constant terms, the convergence condition of Eq 7 is:

$$I_T^f, I_T^p \rightarrow 0. \quad (8)$$

For the domain transformation of source image x_i^s , the domain gradient g_d needs to satisfy $g_d \rightarrow 0$. As shown in the main paper, the \mathcal{L}_{BN} used in our method is cosine similarity loss. Hence, to satisfy $g_d \rightarrow 0$, the following condition should be satisfied:

$$f_d(\sqrt{1 - \bar{\alpha}_n - \bar{\gamma}_n} x_i^s + \sqrt{\bar{\alpha}_n} I_T^f + \sqrt{\bar{\gamma}_n} I_T^p; \theta) \rightarrow I_T. \quad (9)$$

Similar to Eq 7, since $\bar{\alpha}_n$, $\bar{\gamma}_n$ and x_i^s are all constant terms, the convergence condition of Eq 9 is:

$$I_T^f, I_T^p \rightarrow I_T. \quad (10)$$

As mentioned in Sec. 2 (in the main paper), I_T is obtained through the target data stream, which is a non-zero value. Therefore, according to Eq 8 and Eq 10, the convergence condition of domain transformation and the preservation of source domain data's inherent characteristics are opposite, which makes it impossible for the training process to satisfy the convergence conditions of both simultaneously. Hence, we propose a Nash equilibrium strategy to dynamically modify the diffusion steps to achieve a local optimal for our proposed network for the best performance.

2 Datasets

We conduct experiments on three public glaucoma diagnostic datasets: DRISHTI-GS [9], RIM-ONE-r3 [10] and REFUGE [11]. The detailed statistics of the datasets are shown in Table 1.

3 Pseudo-algorithm

For better understanding of our proposed approach, we provide the pseudo-algorithm below (Algorithm 1).

Algorithm 1 DDF-UDA Algorithm

Require: Set of source domain data X^S and set of target domain data X^T .

Student model parameters θ , learning rate η ;

for x_i^s in X^S **do**

1:**Input** image x_i to the student model;

2:**Produce** feature map F ;

3:**Produce** predicted mask y_p ;

4:**Compute** $\mathcal{L}_{ce} = \mathcal{L}_{CE}(y, y_p)$;

5:**Update** $\theta = \theta - \eta \nabla_{\theta} \mathcal{L}_{ce}$;

6:**Sample** N samples from X^T denote as X_N^T ;

for x_n^t in X_N^T **do**

7:**Input** image x_n^t to the teacher model;

8:**Produce** feature-level target domain specific information $I_{t,n}^f$;

9:**Produce** pixel-level target domain specific information $I_{t,n}^p$;

10:**Update** $x_{i,N}' = \sqrt{1 - \hat{\alpha}_n - \hat{\gamma}_n} x_i^s + \sum_{n=0}^N \hat{\alpha}_n I_{t,n}^f + \sum_{n=0}^N \hat{\gamma}_n I_{t,n}^p$;

11:**Input** $x_{i,N}'$ to the student model;

12:**Produce** updated feature map F' ;

13:**Produce** predicted mask y_p' ;

14:**Produce** feature-level domain specific information $I_{t,n}^{f'}$;

15:**Produce** pixel-level domain specific information $I_{t,n}^{p'}$;

16:**Compute** denoised segmentation loss $\mathcal{L}'_{ce} = \mathcal{L}_{CE}(y, y_p')$;

17:**Compute** $\mathcal{L}_{BN} = \mathcal{L}_{cos}(I_{s,i}^{p'}, I_t^p) + \mathcal{L}_{cos}(I_{s,i}^{f'}, I_t^f)$;

18:**Compute** $\mathcal{L}_{con} = \sigma[\sum_{i=1}^N F \cdot (\log F - \log F')]$;

19:**Compute** $\mathcal{L} = \mathcal{L}_{ce} + \mathcal{L}'_{ce} + \mathcal{L}_{con} + \mathcal{L}_{BN}$;

20:**Update** $\theta = \theta - \eta \nabla_{\theta} \mathcal{L}$;

end for

21:**Update** Teacher model with θ using EMA;

end for

Ensure: Model parameters θ .

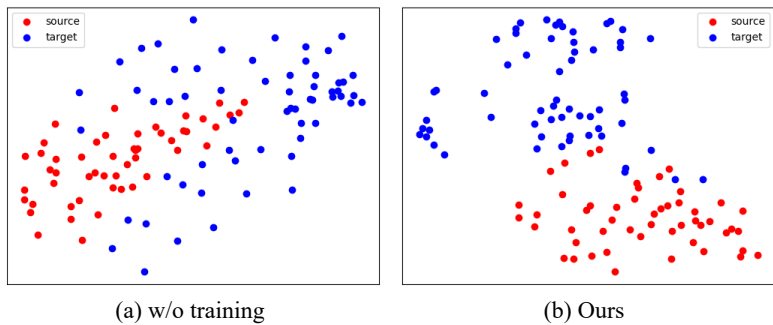


Figure 1: Feature visualization with t-SNE. Red and blue dots denote the source and target data, respectively. (a) A DeepLabv3+ model without any training. (b) A model trained with DDF-UDA.

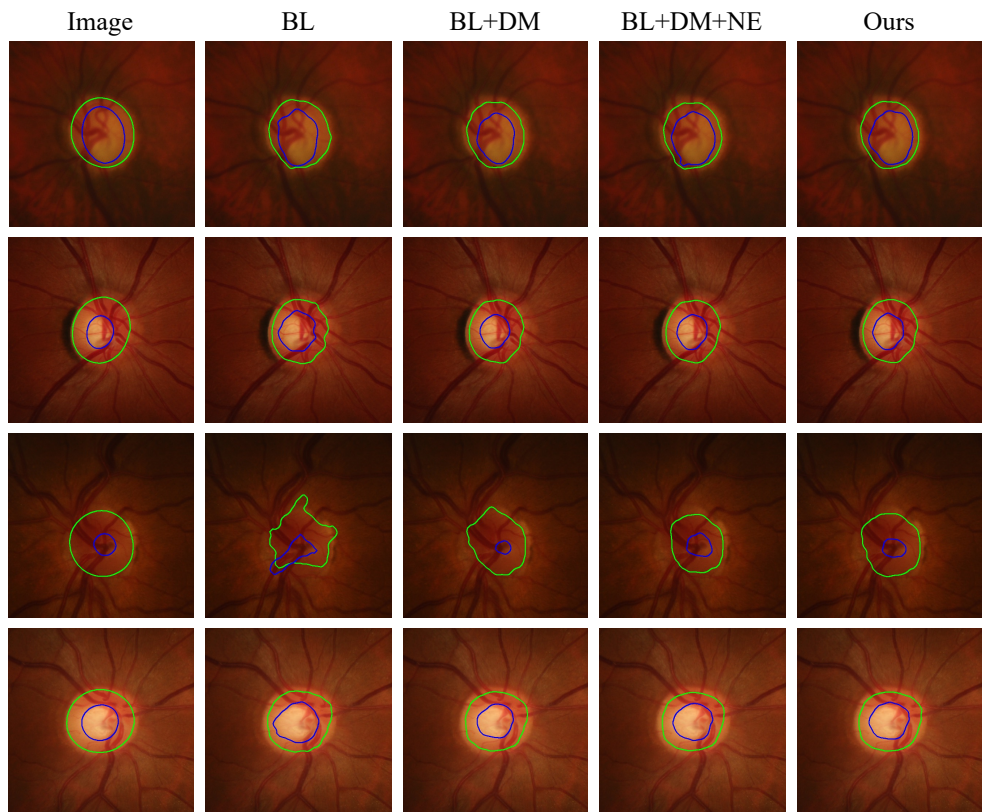


Figure 2: Visualization of OD (green) and OC (blue) segmentation results for the ablation study conducted from Drishti-GS to RIM-ONE-r3. BL: Baseline.

4 Visualization Results

4.1 t-SNE

We show the feature visualization with t-SNE. As shown in Fig. 1, we conduct two different settings on the test set of both Drishti-GS (source) and RIM-ONE-r3 (target). The results demonstrate that our approach achieves promising performance in distinguishing source and target samples.

4.2 Qualitative Results

We provide visualization results for our ablation studies conducted from Drishti-GS to RIM-ONE-r3. As shown in Fig. 2, our proposed method achieves the best performance compared with other network settings.

References

- [1] Francisco Fumero, Silvia Alayón, José L Sanchez, Jose Sigut, and M Gonzalez-Hernandez. Rim-one: An open retinal image database for optic nerve evaluation. In *Proceedings of the 24th IEEE International Symposium on Computer-Based Medical Systems*, pages 1–6, 2011.
- [2] José Ignacio Orlando, Huazhu Fu, João Barbosa Breda, Karel Van Keer, Deepti R Bathula, Andrés Diaz-Pinto, Ruogu Fang, Pheng-Ann Heng, Jeyoung Kim, JoonHo Lee, et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. *Medical Image Analysis*, 59:101570, 2020.
- [3] Jayanthi Sivaswamy, SR Krishnadas, Gopal Datt Joshi, Madhulika Jain, and A Ujjwaft Syed Tabish. Drishti-gs: Retinal image dataset for optic nerve head (onh) segmentation. In *2014 IEEE 11th International Symposium on Biomedical Imaging*, pages 53–56, 2014.