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

Abstract

Impacted by the domain shift issue across varying fundus image datasets collected from different medical centres and devices, the performance of a well-trained optic segmentation network is usually affected when applied to another dataset with different distributions. To handle this issue, the unsupervised domain adaptation (UDA) strategy is widely used to improve the generalization ability of deep learning networks by using unlabeled data. However, existing UDA approaches for optic segmentation tasks are mostly adversarial learning-based, which heavily rely on the balance between the source and target datasets to align the features. In this regard, we propose a diffusion-based framework, named Dynamic Diffusion Flow Unsupervised Domain Adaptation (termed DDF-UDA), for the cross-domain optic disc (OD) and optic cup (OC) segmentation in fundus images. Specifically, we propose an adaption module based on diffusion procedure at both feature and pixel levels to alleviate the cross-domain gaps. In order to modify the domain information of the source image while minimizing changes to its content, we further propose an adjustment strategy based on Nash equilibrium, which could dynamically modify the diffusion steps. Experimental results on public datasets demonstrate that our DDF-UDA can effectively leverage unlabeled data to achieve state-of-the-art performance in OD/OC segmentation.

1 Introduction

Accurate segmentation of the optic disc (OD) and the optic cup (OC) from fundus images is crucial for the clinical diagnosis of glaucoma [14]. In recent years, various OD/OC segmentation methods based on deep learning have been proposed and achieved significant success [27, 32]. However, the performance of these deep learning-based methods is limited when applied to new datasets with similar content due to the domain gaps [1]. To address this challenge, unsupervised domain adaptation (UDA) methods have been introduced [7, 18, 21].

In the OD/OC segmentation UDA tasks, adversarial learning is utilized to minimize the feature gap between different domains and demonstrate impressive results [16, 27, 31].

BEAL [26] uses entropy-driven adversarial training strategies to strengthen the boundary prediction of the target domain image and minimize the feature uncertainty. RDR-Net [6] uses the variational auto-encoder (VAE) to reconstruct images and utilizes dynamic convolution specific to entropy-driven adversarial learning for low-level feature refinement and prediction map alignment, respectively. However, these adversarial learning-based UDA methods heavily rely on the balance between the source and the target datasets. The breakdown of this balance will affect the effectiveness of the UDA methods [17]. If the scale of the source dataset is much larger than the target one, it may cause model bias to the source samples due to the strong inductive bias of the deep learning models. On the other hand, if the source dataset is much smaller than the target domain, it may result in insufficient model training, which further makes it difficult to obtain a reliable representation of the target domain [12, 15].

Hence, instead of using feature alignment or image reconstruction to reduce cross-domain bias, we propose a cross-domain information fusion strategy through the diffusion model. The diffusion model is a latent variable model that has recently been applied to generative models and gets impressive outcomes [9]. Diffusion models consist of a forward diffusion process and a reverse generation process, which can obtain a trajectory containing a sample by repetitively adding Gaussian noise T times to any initial sample and create new data by reverse traversing the Markov chain [8, 22]. Some work based on the diffusion generation model has addressed the UDA problem by utilizing image generation to reduce the domain gap [20, 29, 30]. Unlike these diffusion UDA methods for synthesized images, our approach mainly focuses on the latent feature generation. Specifically, we propose a novel end-to-end UDA network based on the guided diffusion model, which transfers the domain information from the target domain into the source domain through forward diffusion processes.

Previous research [2] has shown that the Batch Normalization (BN) layer can reflect domain-specific information. The output of the BN layer could be seen as a Gaussian distribution with expectation equal to 0 and variance equal to 1. Inspired by recent research [4] which proves the superiority of utilizing the diffusion model from a feature perspective, we conduct diffusion on the feature-level. Accordingly, we define the output of the encoder's final BN layer and the decoder's final BN layer as feature-level and pixel-level domain-specific information, respectively. Then, the target domain's feature-level and pixel-level features are diffused into the source domain dataset as directional noise (which can be seen as guidance). Furthermore, when applying the diffusion strategy to transform the source data, we aim to maintain the content of the source domain images to the greatest extent possible during the transition to the target domain. Therefore, we propose a Nash equilibrium strategy to adjust the diffusion process. This approach ensures that the source domain data is effectively transformed into the target domain while preventing over-transformation (which may lead to the original source domain labels being unusable), thus augmenting the dataset.

The main contributions of this paper are summarized as follows: (1) We propose a novel **D**ynamic **D**iffusion **F**low **U**nsupervised **D**omain **A**daptation framework, named as DDF-UDA, which employs domain diffusion learning to tackle the UDA task associated with joint OD/OC segmentation across diverse fundus image datasets. (2) Instead of using feature alignment to bridge the gap between the source and target domains, DDF-UDA employs a diffusion process at both the feature and pixel levels. This approach enables the transfer of domain information from the target data to the source data, thus establishing domain adaptation. (3) Given that the transformation from the source domain to the target domain, as well as the maintaining of the data content inherent to the source, represent a contradictory game theory problem, we propose a Nash equilibrium strategy to dynamically adjust the

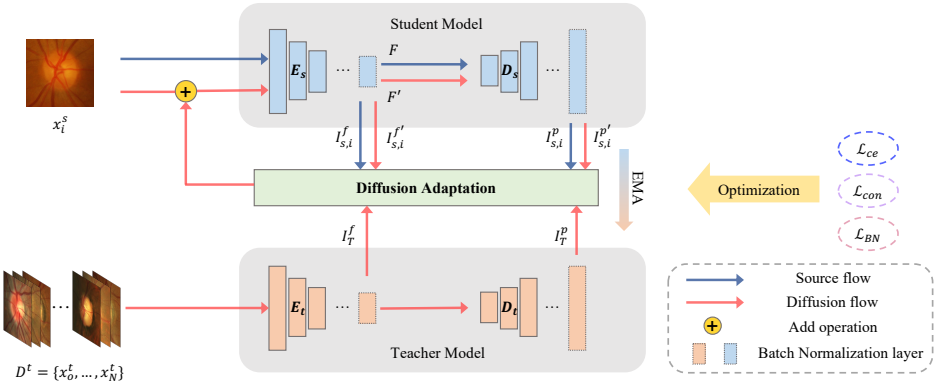

Figure 1: Overview of our proposed DDF-UDA framework. We use different colors to represent student and teacher models (best views in color).

diffusion process to achieve an optimal solution.

We conduct extensive experiments on two kinds of UDA scenarios: imbalance UDA (where a significant difference in data quantity exists between the source and target domains) and traditional UDA. Experimental results demonstrate the effectiveness of the proposed DDF-UDA compared with existing SOTA methods.

2 Method

2.1 Overview of the Framework

The purpose of UDA is to use the source domain data $X^S = \{x_i^s, y_i^s\}_{i=1}^{N_s}$ and the unannotated target domain data $X^T = \{x_i^t\}_{i=1}^{N_t}$ to optimize the segmentation model. Fig. 1 shows an overview of the proposed approach. We propose a novel UDA paradigm based on the diffusion model to improve the model's generalization ability in the target domain dataset. Specifically, our model adapts to different domain distributions by diffusing the domain information from the target dataset into the source domain which has supervised signals.

In detail, our framework employs a dual-stream structure. Given a student network and a teacher network, we use $E_s\&D_s$ and $E_t\&D_t$ to represent the encoder&decoder from different networks, respectively. When x_i^s is input into the student network $E_s\&D_s$, N images from the target dataset are passed to the teacher network $E_t\&D_t$ simultaneously. Firstly, the source domain feature map F of x_i^s could be extracted by E_s and the output from the BN layer of E_s is obtained as the feature-level source domain information $I_{s,i}^f$. Then, with F passed to D_s, the pixel-level source domain information $I_{s,i}^p$ could be obtained through the BN layer of D_s.

As we once send N samples (which can be seen as a subset) from the target dataset to the framework during each source data training step, we could treat this subset of the target dataset as the target data stream $X_N^T = \{x_n^t\}_{n=1}^N$. The target data stream is input into the teacher model to obtain a set of feature-level and pixel-level target domain-specific information, denoted as $I_T^f = \{I_{t,n}^f\}_{n=1}^N$ and $I_T^p = \{I_{t,n}^p\}_{n=1}^N$, respectively. We regard I_T^f and I_T^p as a type of directional noise for conducting domain-level information diffusion on the source input

image to achieve updated $x_{i,N}^{s'}$. $x_{i,N}^{s'}$ is then resent to the student network, and the feature map F' and the feature-level domain-specific information $I_{s,i}^{f'}$ are both obtained through E_s. F' is further passed to D_s to obtain the pixel-level domain-specific information $I_{s,i}^{p'}$. As for the updated $x_{i,N}^{s'}$, the initial label y_i^s is still used as the supervision signal. Hence, one goal of the optimization is to make sure $F \simeq F'$. Furthermore, from the UDA perspective, we strive to ensure that $I_{s,i}^{f'} \simeq I_{t,i}^{f}$, as well as $I_{s,i}^{p'} \simeq I_{t,i}^{p}$, to secure the domain information transformation. Following [25], the student network is updated through optimizing, while the teacher network is updated through exponential moving average (EMA):

$$\theta_t^{(i)} = \omega\theta_t^{(i-1)} + (1-\omega)\theta_s^{(i)}, \tag{1}$$

where $\theta_t^{(i)}$ denotes the teacher network parameters at i_{th}, $\theta_s^{(i)}$ represents the student network parameters at the same step, and ω is the EMA weight hyperparameter ($\omega = 0.999$ in our case).

2.2 Domain Diffusion Processing

Inspired by existing diffusion models [3, 9, 20], we adapt the forward process used for generating diffusion models into a directed process (see Fig. 2). Given a real source domain image sample $x_i^s \sim q(x)$, the domain information transferring process could be formulated as follows [9]:

$$q\left(x_{i,1:N}^{s'} \mid x_i^s\right) = \prod_{n=1}^{N} \mathcal{N}\left(x_{i,n}^{s'}; \sqrt{1-\beta_n} x_{i,n-1}^{s'}, \beta_n I_{t,n}^{f}\right), \tag{2}$$

$$\alpha_n = 1-\beta_n, \bar{\alpha}_n = \prod_{i=1}^{n} \alpha_i, \tag{3}$$

$$x_{i,N}^{s'} = \sqrt{\bar{\alpha}_n} x_i^s + \sum_{n=0}^{N} \sqrt{1-\bar{\alpha}_n} I_{t,n}^{f}, \tag{4}$$

where $x_{i,N}^{s'}$ denotes the updated x_i^s. The product $\prod_{n=1}^{N}$ indicates that the distribution is factorized over n steps, and $\mathcal{N}$ represents a Gaussian distribution. Both α_n and β_n are adjustable hyperparameters.

Additionally, except $I_{t,i}^{f}$, we also introduce pixel-level domain information for the diffusion process. For the pixel-level features, we aim to eliminate semantic information as much as possible while maintaining domain information. Hence, we propose an assumption that the image contains only two types of information: semantic level and domain level. We introduce a Pixel-level Feature Extraction Threshold (PFT). Suppose after passing through the dual-stream model, the segmentation results in a poor agreement between the two models (below a specified threshold) while still maintaining similar performance (which can be seen through the evaluation metrics). We assume that under this scenario, the output of the features by the two decoders mostly consists of domain information [28]. Therefore, we further adapt Eq. 5 as follows:

$$x_{i,N}^{s'} = \sqrt{1-\bar{\alpha}_n-\bar{\gamma}_n} x_i^s + \sum_{n=0}^{N} \sqrt{\bar{\alpha}_n} I_{t,n}^{f} + \sum_{n=0}^{N} \sqrt{\bar{\gamma}_n} I_{t,n}^{p}, \tag{5}$$

where α_n and γ_n both denote adjustable hyperparameters.

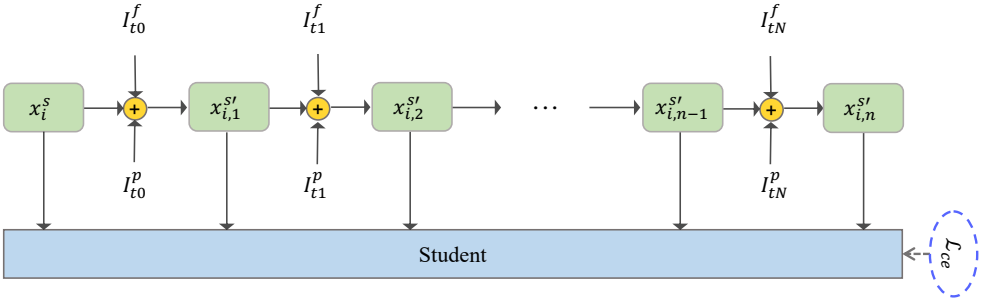

Figure 2: Diffusion adaptation processing. We use both the feature-level target domain information I^f_{tN} and the pixel-level target domain information I^p_{tN} to update image x^s_i during the diffusion process.

2.3 Optimization Processing

For optimization, we use the cross-entropy loss as our supervision loss for segmentation, which can be denoted as $\mathcal{L}_{ce}$:

$$\mathcal{L}_{ce} = -\frac{1}{M}\sum_{i=1}^{M} y^s_i \log(p_{y^s_i}). \tag{6}$$

Noted that $\mathcal{L}_{ce}$ is applied to both the original source input x^s_i and the updated source $x^{s'}_{i,N}$. Similarly, the updated segmentation loss $\mathcal{L}_{ce'}$ are defined as:

$$\mathcal{L}_{ce'} = -\frac{1}{M}\sum_{i=1}^{M} y^s_i \log(p'_{y^s_i}), \tag{7}$$

where M is the total number of pixels, and y^s_i is the ground-truth label. $p_{y^s_i}$ and $p'_{y^s_i}$ are the predicted probability of x^s_i and $x^{s'}_{i,N}$, respectively.

As mentioned in Sec. 2.1, we aim to induce the similarity between domain information from both source and target data for good adaption performance. Additionally, the domain characteristics of $I^{f'}_{s,i}$ and $I^f_{t,i}$, as well as $I^{p'}_{s,i}$ and $I^p_{t,i}$, need to be extremely similar to guarantee that the domain state of the image x^s_i transitions entirely from source to target. Therefore, we propose the BN loss $\mathcal{L}_{BN}$ related to domain information and the consistency loss $\mathcal{L}_{con}$ to ensure the consistency of the source domain supervision model to combine the supervision signal to effectively facilitate model convergence. $\mathcal{L}_{con}$ is calculated by Kullback-Leibler divergence as follows:

$$\mathcal{L}_{con} = \sigma\left[\sum_{i=1}^{N} F \cdot \left(\log F - \log F'\right)\right], \tag{8}$$

where σ denotes the activation function. $\mathcal{L}_{BN}$ is calculated by cosine similarity as follows:

$$\mathcal{L}_{BN} = \mathcal{L}_{cos}(I^{p'}_{s,i}, I^p_t) + \mathcal{L}_{cos}(I^{f'}_{s,i}, I^f_t), \tag{9}$$

$$\mathcal{L}_{cos} = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{I'_{s,i}\cdot I_t}{\left\|I'_{s,i}\right\|\cdot\|I_t\|}\right). \tag{10}$$

Hence, the total loss of the DDF-UDA framework can be formulated as

$$\mathcal{L}_{total} = \mathcal{L}_{ce} + \mathcal{L}'_{ce} + \mathcal{L}_{con} + \mathcal{L}_{BN}. \tag{11}$$

Note that the dual-stream structure and the diffusion process are only employed during the training phase. For inference, the test images are directly sent to the student network to obtain the final prediction masks.

2.4 Nash Equilibrium

As mentioned in Sec. 1, due to the opposite convergence conditions for domain transformation and the preservation of source domain data's inherent content (see details in the supplementary materials), it is difficult to find the optimal solution for model parameters during the training process. Hence, instead of setting a fixed number of target input data stream images N, we set it as a dynamic one based on the training loss to force the training process to reach a local optimum for the best performance of our method.

In detail, when N target samples are proceeded to transfer the domain information, which will facilitate $\mathcal{L}_{BN}$ reduction and convergence. However, this would significantly alter the original information of the source image. Thus, $\mathcal{L}_{con}$ will deteriorate. To this end, for the $\mathcal{L}_{BN}$ and the $\mathcal{L}_{con}$, their convergence conditions are inconsistent, making it challenging to obtain a global optimal solution [5]. This situation presents a game theory problem: the mathematical modeling of rational behaviour in interdependent situations. Therefore, our goal is no longer to attain a globally optimal solution but to achieve local optimal situations for each objective function.

Nash equilibriumis is a game theory strategy, which can dynamically adjust training strategies to achieve local optimal solutions while ensuring the stability of the training process [11, 23]. Drawing inspiration from the Nash equilibrium method, after assigning an initial value to the number of target images, we use the $\mathcal{L}_{BN}$ and $\mathcal{L}_{con}$ values obtained in each diffusion process to correct the number of target images N. During the diffusion process, the $\mathcal{L}_{BN}/\mathcal{L}_{con}$ ratio is noted as $\mathcal{R}_1$. If $\mathcal{R}_1$ is greater than 1, indicating that $\mathcal{L}_{BN}$ is greater than $\mathcal{L}_{con}$, it is necessary to increase N to decrease the $\mathcal{L}_{BN}$. Conversely, we need to reduce the value of N. To this end, the training process will keep establishing a dynamic competition until a state of equilibrium (Nash equilibrium) is achieved by both parties. The following formula shows the objective of the Nash equilibrium strategy:

$$u_1\left(\mathcal{L}^*_{BN}, \mathcal{L}^*_{con}\right) \geq u_0\left(\mathcal{L}_{BN}, \mathcal{L}_{con}\right), \tag{12}$$

$$u_N\left(\mathcal{L}^*_{BN}, \mathcal{L}^*_{con}\right) \geq u_{N-1}\left(\mathcal{L}^*_{BN}, \mathcal{L}^*_{con}\right), \tag{13}$$

where u_0 denotes the first strategy, and u_1 denotes the second strategy. u_N and u_{N-1} denote the strategy for the Nth and $(N-1)$th optimization, respectively. $\mathcal{L}_{BN}$ and $\mathcal{L}_{con}$ are the losses without adjustment, and $\mathcal{L}^*_{BN}$ and $\mathcal{L}^*_{con}$ are the adjusted losses. For better understanding of our proposed approach, we provide the pseudo-algorithm in supplementary materials.

Method	Drishti-GS → RIM-ONE-r3				RIM-ONE-r3 → Drishti-GS			
	DI_{disc}	DI_{cup}	$MIoU_{disc}$	$MIoU_{cup}$	DI_{disc}	DI_{cup}	$MIoU_{disc}$	$MIoU_{cup}$
ρOSAL [27]	0.9153	0.8159	0.8957	0.8383	0.9096	0.8303	0.8614	0.8325
BEAL [26]	0.9208	0.8001	0.9003	0.8345	0.9293	0.8054	0.8927	0.8116
ISFA [16]	0.9252	0.8220	0.9059	0.8471	0.9324	0.8385	0.8961	0.8413
RDR-Net [6]	0.9215	0.8282	0.9028	0.8512	**0.9434**	0.8430	0.9124	0.8455
Ours	**0.9572**	**0.8324**	**0.9551**	**0.8583**	0.9265	**0.8478**	**0.9186**	**0.8655**

Table 1: Comparison OD/OC segmentation results when adapted from Drishti-GS to RIM-ONE-r3 and from RIM-ONE-r3 to Drishti-GS.

3 Experiments and Results

We conduct experiments on three public glaucoma diagnostic datasets: RIM-ONE-r3 [8], DRISHTI-GS [24] and REFUGE [19]. The detailed statistics of the datasets are shown in supplementary materials. We conduct extensive experimental scenarios on OD/OC segmentation for two types of UDA: imbalanced UDA and traditional UDA. For imbalanced UDA, there is a substantial difference in the quantity of data between the source and target domains. We select adapting Drishti-GS to RIM-ONE-r3 and RIM-ONE-r3 to Drishti-GS settings in an imbalanced scenario. In the case of traditional UDA, we adopt the same settings as in the previous works, utilizing the REFUGE dataset as the source domain and the Drishti-GS and RIM-ONE-r3 datasets as the target domains.

3.1 Experiment Details

Evaluation Metrics Various commonly used metrics are used to evaluate the experimental results. These include the Dice Similarity Coefficient (Dice), which offers pixel-level accuracy evaluation. We also employ Mean Intersection over Union (MIoU) to further compare the segmentation performances of different approaches.

Implementation Details Due to the image size variance across different fundus image datasets, we initially crop the original image to yield a 512×512 Region of Interest (ROI) centred on the OD. For computational efficiency, we subsequently resize the cropped image to a 256×256 format as the input of the network. This study employs standard data augmentation techniques to increase the sample size, incorporating strategies such as elastic transformation, Gaussian noise injection, random erasing, as well as random scaling and cropping. For the segmentation network, we apply DeepLabv3+ with a pre-trained MobileNetV2 as the backbone. The model is implemented on a workstation with one NVIDIA GeForce RTX 3090Ti GPU. For optimization, we set the batch size to 8 with the Adam optimizer using "poly" learning rate policy. The initial learning rate is set to 0.001, with a pixel-level feature extraction threshold set to 0.75. For the adjustable hyper-parameters, both $\bar{\alpha}_n$ and $\bar{\gamma}_n$ are set to 0.01.

Method	Drishti-GS		RIM-ONE-r3	
	DI_{disc}	DI_{cup}	DI_{disc}	DI_{cup}
No Adapt	0.930	0.854	0.822	0.703
Upper bound	0.977	0.910	0.969	0.877
Hoffman *et al.* [10]	0.959	0.851	0.852	0.755
Javanmardi *et al.* [13]	0.961	0.849	0.853	0.779
ρOSAL [27]	0.965	0.858	0.865	0.787
BEAL [26]	0.961	0.862	0.898	0.810
ISFA [16]	0.966	0.892	0.908	0.822
RDR-Net [6]	**0.971**	0.893	0.918	0.840
Ours	0.969	**0.893**	**0.962**	**0.850**

Table 2: Quantitative comparison OD/OC segmentation results when adapted from REFUGE to Drishti-GS and adapted from REFUGE to RIM-ONE-r3.

3.2 Performance of the Proposed Method

3.2.1 Results on imbalanced UDA settings

We report the results of our model under imbalanced UDA settings in Table 1. Specifically, we first select DRISHTI-GS and RIM-ONE-r3 as the source and target domains, respectively. We compare the proposed method with four state-of-the-art OD/OC segmentation UDA methods, i.e., ρOSAL [27], BEAL [26], ISFA [16] and RDR-Net [6]. Table 1 reports the segmentation results of adapting Drishti-GS to RIM-ONE-r3 and the segmentation performance when adapting from RIM-ONE-r3 to DRISHTI-GS, with the best results highlighted in bold. For the convenience of description, the dice, and mIoU values of OD/OC are expressed as DI_{disc}, DI_{cup}, $MIoU_{disc}$, and $MIoU_{cup}$.

Table 1 illustrates that in adaptive OD/OC segmentation tasks with imbalanced data, our DDF-UDA consistently yields competitive results across various scenarios. As the results shown in Table 1, which are performed on the Drishti-GS to RIM-ONE-r3 adaptation task, our method achieves state-of-the-art performance on all evaluation metrics. Specifically, DDF-UDA outperforms the prior work in terms of Dice of OD segmentation of 3.57%, improving from 92.15% (by RDR-Net) to 95.72%. Additionally, our proposed method exhibits remarkable performance in MIoU metrics for OD/OC segmentation. In particular, for $MIoU_{disc}$, there were respective increases of 5.23% compared to RDR-Net. In addition, we also report the results adapting from RIM-ONE-r3 to Drishti-GS. As shown in Table 1, under the setting of adapting RIM-ONE-r3 to Drishti-GS, our proposed framework performs better than existing methods on 3 out of 4 evaluation metrics. The above results validate the effectiveness of DDF-UDA, which leverages information diffusion to bridge the domain gap between the source and target domains. Notably, when the target dataset is larger than the source one, our method proves especially effective in adaptation tasks by diffusing target domain information, surpassing the performance of adversarial learning methods which rely on domain alignment. Fig. 3 shows the qualitative results of OD (green) and OC (blue) segmentation adapted from Drishti-GS to RIM-ONE-r3.

3.2.2 Results on traditional UDA settings

To assess the performance of our method on traditional datasets, we employ the REFUGE dataset as the source domain, and use Drishti-GS and RIM-ONE-r3 as the target domains separately. For the results in Table 2, we compare our proposed method with six SOTA UDA

Method	Baseline	DM	NE	PFT	DI_{disc}	DI_{cup}	$MIoU_{disc}$	$MIoU_{cup}$
Drishti-GS → RIM-ONE-r3								
No Adapt	✓				0.8952	0.7425	0.8938	0.8074
	✓	✓			0.9530	0.8203	0.9510	0.8504
	✓	✓	✓		**0.9580**	0.8234	**0.9558**	0.8520
	✓	✓	✓	✓	0.9572	**0.8324**	0.9551	**0.8583**
REFUGE → Drishti-GS								
	✓				0.9298	0.8542	0.9201	0.8700
	✓	✓			0.9543	0.8891	0.9611	0.8890
	✓	✓	✓		0.9628	0.8814	0.9629	0.8865
	✓	✓	✓	✓	**0.9686**	**0.8931**	**0.9640**	**0.9010**

Table 3: Ablation study results of adapting Drishti-GS to RIM-ONE-r3 and adapting REFUGE to Drishti-GS.

methods and two baselines. In addition to the four previously mentioned SOTA OD/OC segmentation UDA approaches, we also compare the globally aligned method proposed by Hoffman et al. [10] and the adversarial learning-based UDA algorithm by Javanmardi et al. [13]. For the baselines, we also train a DeepLabV3+ with source data only ("No Adapt") and with annotated target data only ("Upper bound"), respectively.

The results in Table 2 show that DDF-UDA achieves highly competitive results under the traditional UDA settings. On the Drishti-GS dataset, the Dice scores for OD/OC segmentation are nearly equivalent to the SOTA methods. In the adaptation from REFUGE to RIM-ONE-r3, DDF-UDA achieves superior segmentation performance. Specifically, our method achieves an impressive improvement on Dice score for OD segmentation (91.8% → 96.2%), indicating a significant 4.4% gap and approaching the performance of fully supervised results (96.9%). These reported results demonstrate that in the scenarios where the source domain contains more data than the target domain, our proposed DDF-UDA consistently delivers advanced UDA segmentation performance through dynamically diffusing target domain information. Some additional visualization results are shown in supplementary materials.

3.3 Ablation Study

To assess the impact of various modules within our proposed model, we conduct ablation experiments to evaluate the effectiveness of each module. We conduct these experiments using Drishti-GS as the source domain and RIM-ONE-r3 as the target domain (imbalanced UDA setting), and then using REFUGE as the source domain and Drishti-GS as the target domain (traditional UDA setting) for clarity and consistency. Each module is designated as follows: DM for the Diffusion Module, NE for Nash Equilibrium, and PFT for Pixel-level Feature Extraction Threshold. The outcomes of the different module implementations are visually presented in Table 3. We utilized a model trained only on the source domain as our baseline.

The results illustrate that as the DM, NE, and PFT modules are incrementally integrated into the segmentation model, all evaluation metrics exhibited continuous improvement. It is notable that in the absence of domain adaptation, the baseline displayed relatively poor performance, with the lowest Dice scores for OD and OC in both two tasks. In contrast, the DM module leads to an increase in Dice scores by 5.78% & 7.78% (imbalanced UDA) and 2.45% & 3.49% (traditional UDA) when compared to the baseline. It is important to high-

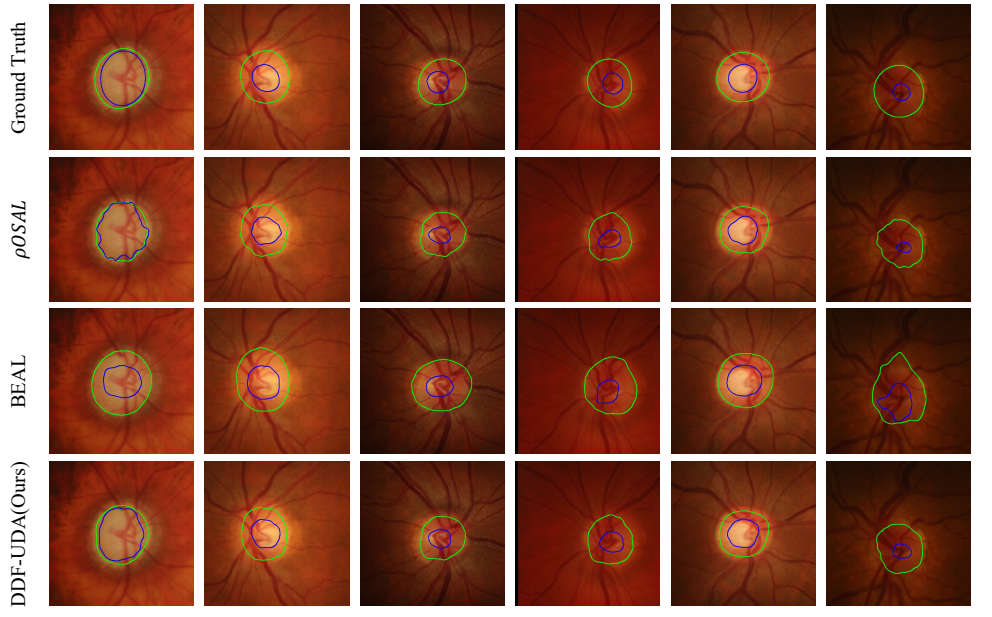

Figure 3: Qualitative results of OD (green) and OC (blue) segmentation, which adapted from Drishti-GS to RIM-ONE-r3.

light that the DM module significantly contributes to improve segmentation accuracy. This is attributed to the DM module dynamically diffusing domain information from the target domain data stream into the source domain training model, thereby improving domain adaptation effects. When the NE module is introduced into the network, segmentation performance also shows varying degrees of improvement. Finally, experimental results demonstrate that the addition of the PFT module can improve both OD/OC segmentation results.

4 Conclusion

In this paper, we propose a novel DDF-UDA method, to tackle the UDA for OD/OC segmentation under both balanced and imbalanced data conditions. Instead of relying on feature alignment to bridge the gap between the source and target domains, our method implements a diffusion process at both feature and pixel levels to gradually transfer domain information from target to source for adaptation. Considering that the transformation from the source domain to the target domain, and the maintenance of the data content of the source domain, is a self-contradictory game theory problem, we introduce a Nash equilibrium to govern the diffusion process and reach an optimal solution. We conduct extensive experiments on different UDA OD/OC segmentation settings, including traditional UDA settings and unbalanced UDA settings, to demonstrate the effectiveness of our method.

References

[1] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79:151–175, 2010.

[2] Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7354–7362, 2019.

[3] Shoufa Chen, Peize Sun, Yibing Song, and Ping Luo. Diffusiondet: Diffusion model for object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19830–19843, 2023.

[4] Xinlei Chen, Zhuang Liu, Saining Xie, and Kaiming He. Deconstructing denoising diffusion models for self-supervised learning. *arXiv preprint arXiv:2401.14404*, 2024.

[5] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *International Conference on Machine Learning*, pages 1081–1090. PMLR, 2019.

[6] Ziyang Chen, Yongsheng Pan, and Yong Xia. Reconstruction-driven dynamic refinement based unsupervised domain adaptation for joint optic disc and cup segmentation. *IEEE Journal of Biomedical and Health Informatics*, 2023.

[7] Jianan Fan, Dongnan Liu, Canran Li, Hang Chang, Heng Huang, Filip Braet, Mei Chen, and Weidong Cai. Revisiting adaptive cellular recognition under domain shifts: A contextual correspondence view. In *Proceedings of the European Conference on Computer Vision*, 2024.

[8] 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.

[9] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[10] Judy Hoffman, Dequan Wang, Fisher Yu, and Trevor Darrell. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. *arXiv preprint arXiv:1612.02649*, 2016.

[11] Junling Hu and Michael P Wellman. Nash q-learning for general-sum stochastic games. *Journal of Machine Learning Research*, 4(Nov):1039–1069, 2003.

[12] Jiaxing Huang, Dayan Guan, Aoran Xiao, Shijian Lu, and Ling Shao. Category contrast for unsupervised domain adaptation in visual tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1203–1214, 2022.

[13] Mehran Javanmardi and Tolga Tasdizen. Domain adaptation for biomedical image segmentation using adversarial training. In *2018 IEEE 15th International Symposium on Biomedical Imaging*, pages 554–558, 2018.

[14] Noor Elaiza Abdul Khalid, Noorhayati Mohamed Noor, and Norharyati Md Ariff. Fuzzy c-means (fcm) for optic cup and disc segmentation with morphological operation. *Procedia Computer Science*, 42:255–262, 2014.

[15] Elyor Kodirov, Tao Xiang, Zhenyong Fu, and Shaogang Gong. Unsupervised domain adaptation for zero-shot learning. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2452–2460, 2015.

[16] Haijun Lei, Weixin Liu, Hai Xie, Benjian Zhao, Guanghui Yue, and Baiying Lei. Unsupervised domain adaptation based image synthesis and feature alignment for joint optic disc and cup segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(1):90–102, 2021.

[17] Canran Li, Dongnan Liu, Haoran Li, Zheng Zhang, Guangming Lu, Xiaojun Chang, and Weidong Cai. Domain adaptive nuclei instance segmentation and classification via category-aware feature alignment and pseudo-labelling. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 715–724. Springer, 2022.

[18] Dongnan Liu, Donghao Zhang, Yang Song, Fan Zhang, Lauren O'Donnell, Heng Huang, Mei Chen, and Weidong Cai. Pdam: A panoptic-level feature alignment framework for unsupervised domain adaptive instance segmentation in microscopy images. *IEEE Transactions on Medical Imaging*, 40(1):154–165, 2020.

[19] 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.

[20] Duo Peng, Ping Hu, Qiuhong Ke, and Jun Liu. Diffusion-based image translation with label guidance for domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 808–820, 2023.

[21] Ringwald, Tobias, and Stiefelhagen Rainer. Unsupervised domain adaptation by uncertain feature alignment. In *British Machine Vision Conference*, 2020.

[22] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

[23] Dale Schuurmans and Martin A Zinkevich. Deep learning games. *Advances in Neural Information Processing Systems*, 29, 2016.

[24] 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.

[25] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems*, 33:596–608, 2020.

[26] Shujun Wang, Lequan Yu, Kang Li, Xin Yang, Chi-Wing Fu, and Pheng-Ann Heng. Boundary and entropy-driven adversarial learning for fundus image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 102–110, 2019.

[27] Shujun Wang, Lequan Yu, Xin Yang, Chi-Wing Fu, and Pheng-Ann Heng. Patch-based output space adversarial learning for joint optic disc and cup segmentation. *IEEE Transactions on Medical Imaging*, 38(11):2485–2495, 2019.

[28] Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, and Xinyi Le. Semi-supervised semantic segmentation using unreliable pseudo-labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4248–4257, 2022.

[29] Zhaohu Xing, Liang Wan, Huazhu Fu, Guang Yang, and Lei Zhu. Diff-unet: A diffusion embedded network for volumetric segmentation. *arXiv preprint arXiv:2303.10326*, 2023.

[30] Yijun Yang, Huazhu Fu, Angelica Aviles-Rivero, Carola-Bibiane Schönlieb, and Lei Zhu. Diffmic: Dual-guidance diffusion network for medical image classification. *arXiv preprint arXiv:2303.10610*, 2023.

[31] Qianbi Yu, Dongnan Liu, Chaoyi Zhang, Xinwen Zhang, and Weidong Cai. Unsupervised domain adaptive fundus image segmentation with few labeled source data. In *British Machine Vision Conference*, 2022.

[32] Rongchang Zhao, Wangmin Liao, Beiji Zou, Zailiang Chen, and Shuo Li. Weakly-supervised simultaneous evidence identification and segmentation for automated glaucoma diagnosis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 809–816, 2019.