BEYOND DOMAIN ALIGNMENT: ANATOMICALLY CONSTRAINED UNSUPERVISED ADAPTATION FOR OPTIC NERVE HEAD SEGMENTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Glaucoma causes irreversible vision loss, and accurate optic disc (OD) and optic cup (OC) segmentation from fundus images is essential for computing the vertical cup-to-disc ratio (vCDR) used in primary screening. Deep learning has automated OD/OC segmentation, but supervised models degrade under domain shift caused by heterogeneous devices, protocols, and populations. Existing unsupervised domain adaptation (UDA) methods align feature or output distributions but neglect OC-OD anatomical relationship, leading to clinically implausible predictions. We propose an anatomically constrained UDA framework for optic nerve head (ONH) segmentation that jointly addresses feature-level domain discrepancy and anatomical-level population variability. The framework combines automated ROI extraction, a segmentation backbone, and decoder adversarial alignment. To preserve valid ONH geometry without target annotations, a differentiable cup-to-disc area ratio is regularized by a two-sided extended log-barrier, with per-sample bounds adapted via class-conditional mechanism driven by a lightweight classifier. Ablation shows that anatomical regularization benefits only when the source-target cup-to-disc distributions differ significantly. We therefore enable it selectively using a two-sample hypothesis test on per-image ratios, evaluated without target labels. Trained on REFUGE as source domain, the gated framework achieves competitive cross-domain performance, with OC $Dice$ of 0.895 and 0.823 and OD $Dice$ of 0.982 and 0.974 on Drishti-GS and on RIM-ONE-r3, respectively.

1 INTRODUCTION

Glaucoma is a chronic progressive optic neuropathy that damages the optic nerve head (ONH), resulting in irreversible vision loss Mary et al. (2016). In 2020, 76 million people aged 40-80 were affected, projected to reach 111.8 million by 2040 Tham et al. (2014). About 50% of the early glaucoma cases are undiagnosed Giangiacomo & Coleman (2009). Therefore, early detection is critical to saving vision. Screening for primary glaucoma often involves ocular fundus imaging of the optic disk (OD) and optic cup (OC) with clinical indicators such as the vertical cup-to-disc ratio (vCDR). Accurate OD and OC segmentation is thus essential for vCDR calculation and early screening.

Manual segmentation of OD and OC by ophthalmologists is laborious, expensive, and prone to variability Naithani et al. (2007), rendering it impractical for large-scale assessment. While deep learning is a promising option Fu et al. (2018); Gu et al. (2019); Wang et al. (2019c), supervised models need a large and uniform number of annotations. Fundus datasets vary in resolution, tone and contrast due to different equipment causing domain shift that hinders generalization of source-trained models.

Unsupervised domain adaptation (UDA) trains on a labeled source domain and adapts it to an unlabeled target domain to address domain shift without target-domain annotations. ONH segmentation UDA is challenging due to cross-domain appearance variation, low contrast and blood vessels. Since vCDR requires accurate OD and OC bounds, poor segmentation affects screening.

To enable large-scale glaucoma screening with precise vCDR measurements and minimal manual labor, we propose an automated and domain-adaptive framework that combines automated region of interest (ROI) extraction, segmentation network with domain-specific batch normalization, decoder-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

tap adversarial alignment and a two-sided extended log-barrier on a differentiable soft cup-to-disc ratio. An auxiliary classifier provides image-level class information that modulates the ratio bounds per sample, and an unsupervised gating rule enables this constraint only when the source–target cup-to-disc distributions differ significantly. The main contributions of this paper are summarized as follows:

- We propose an unsupervised domain adaptation framework for ONH segmentation that jointly addresses feature-level domain discrepancy and anatomical-level population variability, integrating automated ROI extraction, domain specific batch normalization, and decoder-tap adversarial alignment.
- We introduce a differentiable soft-cup-to-disc area ratio, and a class-conditional bound mechanism that leverages an auxiliary classifier to adapt anatomical constraints on a per-sample basis, hence enforcing clinically plausible OD and OC geometry without target-domain annotations.
- We analyze the relationship between anatomical constraint and domain shift, showing that class-conditional bounds yield substantial gains when the source and target cup-to-disc distributions are statistically distinguishable but are detrimental when they are not. We convert this finding into an unsupervised gating rule based on a two-sample hypothesis test that enables the constraint selectively using only source labels and unlabeled target predictions.

2 Related Work

Optic Disc and Optic Cup Segmentation. Traditional OD and OC segmentation includes thresholding, boundary detection and region-based algorithms, such as circular transformation Aquino et al. (2010); Lu (2011), conditional random fields Chakravarty & Sivaswamy (2017), and handcrafted features Abramoff et al. (2007); Joshi et al. (2011); Cheng et al. (2013). However, fundus image quality, lesions, and blood vessels might impair standard approaches, making OD and OC boundary extraction problematic. Convolutional neural networks (CNNs) have shown promise in automatic OD and OC segmentation Fu et al. (2018); Shah et al. (2019); Shankaranarayana et al. (2019); Edupuganti et al. (2018). However, such high performance often requires extensive, consistent-distributed annotated training data. Acquiring extensive annotations is costly and fundus images vary due to imaging equipment and techniques. This causes domain shift, and supervised models often struggle on new datasets Ghafoorian et al. (2017); Dou et al. (2018). Automated OD and OC segmentation remains challenging due to the high density of retinal blood vessels, minor intensity difference between the OC and neuroretinal rim, and soft boundary regions where adjacent structures lack clear boundaries.

Unsupervised Domain Adaptation. Domain shift can be overcome by training the network with source and target data. However, annotating target-domain data during training is impractical in medical image segmentation. UDA provides an alternative by training deep networks on labeled source domains and applying them to unlabeled target domain with distinct visual features Pachade et al. (2021); Kadambi et al. (2020); Liu et al. (2022a); Wang et al. (2019a); Feng et al. (2022); Liu et al. (2022c). Popular UDA approaches include output space adversarial Wang et al. (2019b), which aligns the output space to gain domain-invariant information. However, fundus images show significant brightness, color, and shape changes between the OD and OC. Recent studies have focused on multi-scale UDA methods to address this difficulty Chen & Wang (2021); Liu et al. (2022b).
Previous UDA methods can be broadly grouped into several strategies. Image-level translation transfers target-domain images to the source domain and applies a source-trained network to the images Wang et al. (2019b). High-level feature alignment explores a shared hidden feature space across domain datasets to create structure information of predictions Dou et al. (2018). However, most methods fail on soft border regions due to appearance disparity and poor contrast.

3 Methodology

This section presents the proposed framework, as illustrated in figure 1. The pipeline includes automated ROI extraction for heterogeneous device standardization; a segmentation backbone with

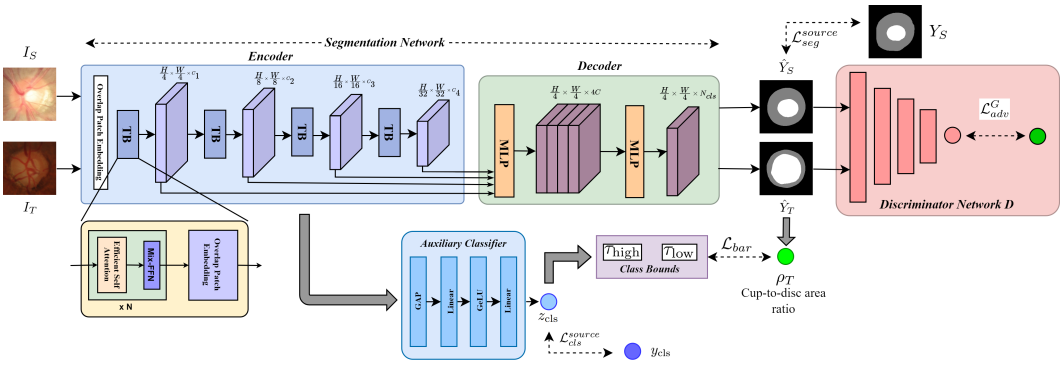


Figure 1: Overview of the proposed anatomically constrained UDA framework for ONH segmentation

domain-specific batch normalization; decoder-tap adversarial alignment to bridge feature-level domain discrepancy; and a two-sided barrier-regularized cup-to-disc ratio. A lightweight classifier provides image-level class information that modulates the ratio-bounds on a per-sample basis to account for inter-class anatomical heterogeneity.

3.1 ROI Extraction

Automated ROI localization around the ONH is performed on all sources and target fundus images to ensure spatial consistency across different acquisition devices. Changune et al. (2026) proposes a traceable image-processing framework for ROI localization. The localized ONH regions are cropped and resized via bilinear interpolation to a uniform spatial resolution of $256 \times 256 \times 3$ pixels. Finally, intensity values across all three RGB channels are normalized to zero mean and unit variance to reduce illumination-related variability across datasets.

3.2 Segmentation Network

The segmentation backbone uses SegFormer Xie et al. (2021), a hierarchical Mix-Transformer encoder (MiT-b5) and a lightweight all-MLP decoder head pretrained on ImageNet Deng et al. (2009). The network receives cropped ONH ROIs as input and performs a 3-class semantic segmentation: background, disc rim, and optic cup.
The MiT-b5 encoder extracts multi-scale feature representations at four spatial resolutions: $\{\frac{H}{4}, \frac{H}{8}, \frac{H}{16}, \frac{H}{32}\}$. The feature map of the final encoder step, $F_{enc} \in \mathbb{R}^{512 \times \frac{H}{32} \times \frac{W}{32}}$, captures high-level contextual semantics. The decoder aggregates these hierarchical representations using a 1×1 convolutions, upsamples to $(\frac{H}{4}, \frac{W}{4})$ and fuses with the decoder representation $F_d \in \mathbb{R}^{768 \times \frac{H}{4} \times \frac{W}{4}}$. Finally, 1×1 projection maps F_d to the 3-class output logits, which are bi-linearly interpolated back to the input resolution to produce soft class probability maps via pixel-wise Softmax.

Domain-Specific Batch Normalization (DSBN). To address covariate shift without adding spatial overhead, DSBN Chang et al. (2019) is applied within the decoder. The decoder's batch normalization layers maintain dedicated running statistics (mean μ and variance σ^2) for each domain, sharing affine learnable parameters (γ, β):

$$\text{DSBN}(x; d) = \gamma . \left(\frac{x - \mu_d}{\sqrt{\sigma_d^2 + \epsilon}} \right) + \beta, d \in \{\text{source}, \text{target}\} \tag{1}$$

This aligns intermediate domain feature distributions before spatial adversarial alignment.

3.3 Adversarial Alignment

Adversarial domain alignment directly bridges the domain shift between source and target distributions via intermediary latent feature representations. Instead of adversarial alignment at the encoder

bottleneck - where features are heavily compressed - the discriminator is at the decoder tap. The fused decoder feature map F_d retain good spatial fidelity under adversarial alignment. A four-layer CNN is used to implement the domain discriminator D. Each layer uses a 3×3 convolution with a LeakyReLU activation, gradually increasing the channel dimension from 64 to 512. A global average pooling (GAP) layer collapses spatial dimensions, and a final linear projection yields a scalar domain logit per unit. Spectral normalization Miyato et al. (2018) is applied to prevent discriminator saturation and generator gradient starvation. No batch normalization is used.
The domain discriminator is optimized using binary cross-entropy loss over source features F_{src} (labeled as 1) and target features F_{tgt} (labeled as 0):

$$\mathcal{L}_{adv}^{D} = -\mathbb{E}_{x_s \sim D_s}[\tilde{y}_s \log D(F_{src})] - \mathbb{E}_{x_t \sim D_t}[(1-\tilde{y}_t)\log(1-D(F_{tgt}))] \tag{2}$$

where $\tilde{y}_s = 0.9$ and $\tilde{y}_t = 0.1$. Simultaneously, the generator is trained with a reverse-label adversarial loss to fool the discriminator into classifying target feature maps as source domain:

$$\mathcal{L}_{adv}^{G} = -\mathbb{E}_{x_t \sim D_t}[\log D(F_{tgt})] \tag{3}$$

The adversarial loss regularizes the decoder's shared latent space, forcing the network to extract domain-invariant OD and OC structural features before passing them to the final classification layer.

3.4 Differentiable Anatomical Constraint

Domain adaptation via unconstrained feature alignment often loses valid structures while transferring representations across domain shifts. While standard adversarial objectives align domain distributions, they lack geometric awareness. This causes excessive expansion of the predicted OC into the neuroretinal rim or contraction of the OD boundary. To mitigate such structural degeneration, an anatomy-aware size constraint based on the relative area of the OC with respect to the optic disc is introduced.
A clinically meaningful measure of this relative geometry is the cup-to-disc area ratio. In retinal imaging, healthy eyes have smaller cup-to-disc area ratios ($\rho_{\text{normal}} \approx 0.351$–$0.377$), while glaucomatous eyes exhibit larger ratios ($\rho_{\text{glaucoma}} \approx 0.618$–$0.626$) due to progressive optic cup enlargement Abdellaoui et al. (2022). Motivated by this anatomical relationship, a differentiable soft cup-to-disc area ratio is computed from the predicted class probabilities:

$$\rho = \frac{A_{\text{cup}}}{A_{\text{disc}}} = \frac{\sum_{p\in\Omega} P_{\text{cup}}(p)}{\sum_{p\in\Omega}\left(P_{\text{rim}}(p) + P_{\text{cup}}(p)\right)}, \tag{4}$$

where Ω denotes the spatial pixel grid and $P_c(p)$ represents the predicted Softmax probability for class c at pixel location p. The area ratio is used as a training-time constraint because it integrated over entire regions, yielding smoother gradients. In contrast, clinical evaluation relies on the vertical cup-to-disc ratio (vCDR), since the vertical axis is the sensitive dimension in glaucoma: horizontal cup-to-disc ratio changes little between normal and glaucomatous eyes Abdellaoui et al. (2022). This formulation is fully differentiable and permits gradients to propagate via both the numerator and denominator. During adaptation, detaching the disc-area term can cause uncontrolled OC expansion.
To constrain the predicted anatomical geometry, the soft ratio ρ is limited to a predefined physiological interval $[\tau_{\text{low}}, \tau_{\text{high}}]$. This anatomical regularizer prevents mathematically valid but clinically implausible segmentation configurations preventing excessive cup expansion ($\rho \to 1$) and near-complete loss of the cup ($\rho \to 0$).

Two-sided extended log-barrier. A two-sided extended log-barrier function $\phi_t(z)$ is incorporated into the loss formulation to limit the predicted ratio ρ within the valid bounds. Conventional barriers become undefined when constraints are violated, which is undesirable early in training. The extended log-barrier penalizes the viable region logarithmically and gradually transitions to a linear penalty when the constraint is violated:

$$\phi_t(z) = \begin{cases} -\frac{1}{t}\log(-z), & z \le -\frac{1}{t^2} \\ tz - \frac{1}{t}\log\left(\frac{1}{t^2}\right) + \frac{1}{t}, & \text{otherwise} \end{cases} \tag{5}$$

The barrier parameter $t > 0$ controls the steepness and linear slope of the penalty. The two-sided barrier loss measures the scalar residuals of lower-bound violation $z_{\text{low}} = \tau_{\text{low}} - \rho$ and upper-bound

violation $z_{\text{high}} = \rho - \tau_{\text{high}}$. The barrier loss $\mathcal{L}_{bar}$ for a mini-batch size of B is formulated as

$$\mathcal{L}_{bar} = \frac{1}{B} \sum_{i=1}^{B} \left[\phi_t \left(\tau_{\text{low}}^{(i)} - \rho^{(i)} \right) + \phi_t \left(\rho^{(i)} - \tau_{\text{high}}^{(i)} \right) \right] \tag{6}$$

3.5 Class-Conditional Anatomical Regularization

A single anatomical fixed interval cannot account for the large cup-to-disc proportion difference between normal and glaucomatous eyes. To address this limitation, the proposed framework introduces a class-conditional sizing mechanism that estimates the glaucoma status of each image and to construct an image-specific anatomical band.

Auxiliary Classifier. An auxiliary glaucoma classification head provides the class information for adaptive constraint construction. A lightweight, layer-normalized MLP (67k parameters) provides class information. The feature map, F_{enc}, is received from the final stage of the MiT-B5 encoder. A GAP layer first aggregates the spatial feature map into 512 dimensions. To prevent the auxiliary classification objective from altering the spatial representations learned for segmentation, the pooled feature is detached from the computational graph using a stop-gradient operator $sg[\cdot]$. The pooled feature then undergoes layer normalization, linear projection to a 128-dimensions, GeLU, dropout ($p = 0.2$), and a linear projection to scalar logit z_{cls}. The classifier therefore operates as an auxiliary prediction pathway without propagating classification gradients into the shared encoder:

$$\mathcal{L}_{\text{cls}}^{\text{source}} = \text{BCEWithLogits}\left(z_{\text{cls}}, y_{\text{cls}}\right). \tag{7}$$

The classification head is supervised using labeled source-domain samples and is weighted by λ_{cls}. This provides class information without letting the auxiliary task dominate segmentation learning.

Confidence-weighted anatomical bounds. The classifier output z_{cls} is converted into a scalar glaucoma probability using a temperature-scaled Sigmoid function:

$$w = \sigma\left(\frac{z_{\text{cls}}}{T}\right), \tag{8}$$

where T denotes the temperature parameter. The classifier prediction may be overconfident when applied to target-domain features due to the source-domain training. A confidence-shrinkage mechanism is hence introduced to reduce the impact of uncertain or miscalibrated predictions. The confidence magnitude is defined as $\kappa = |2w - 1|$, and the shrunken probability is computed as

$$w_s = 0.5 + \kappa(w - 0.5). \tag{9}$$

This preserves direction while reducing uncertain estimates' influence. The w_s is used to interpolate between two clinically motivated anatomical bands: the normal-eye band, $[\tau_{\text{low,N}}, \tau_{\text{high,N}}]$, and the glaucomatous eye band, $[\tau_{\text{low,G}}, \tau_{\text{high,G}}]$. The lower and higher image-specific ratio bounds are parameterized as:

$$\tau_{\text{low}}(w) = w_s \tau_{\text{low,G}} + (1 - w_s)\tau_{\text{low,N}} \tag{10}$$

$$\tau_{\text{high}}(w) = w_s \tau_{\text{high,G}} + (1 - w_s)\tau_{\text{high,N}} \tag{11}$$

Normal-like predictions produce normal bounds, glaucoma-like predictions produce glaucomatous bounds, and uncertain predictions yield intermediate constraints.

Statistical gating of anatomical regularization. The class-conditional bands are enabled only when the source-target cup-to-disc distributions differ significantly. Let $\{\rho_s^{(i)}\}_{i=1}^{n_s}$ denote the per-image cup-to-disc area ratios computed from the labeled source ground-truth masks and $\{\hat{\rho}_t^{(j)}\}_{j=1}^{n_t}$ denote the per-image ratios computed from the unlabeled target predictions after warm-up. A two-sided Kolmogorov–Smirnov (KS) test is applied to these two sets at significance level $\alpha = 0.01$. If the null is rejected, the barrier loss is enabled with class-conditional bounds; otherwise it is disabled. The test requires no target annotations, so gating remains fully unsupervised.

3.6 Objective Function

The framework is trained with a composite objective $\mathcal{L}_{gen}$ that balances supervised source supervision, adversarial alignment, anatomical regularization, and auxiliary classification:

$$\mathcal{L}_{gen} = \mathcal{L}_{seg}^{source} + \lambda_{adv}\mathcal{L}_{adv}^{G} + \lambda_{bar}\mathcal{L}_{bar} + \lambda_{cls}\mathcal{L}_{cls}^{source} \tag{12}$$

The source segmentation loss combines pixel-wise cross-entropy and multi-class Dice loss to address class imbalance:

$$\mathcal{L}_{seg}^{source} = \mathcal{L}_{CE}(\hat{Y}_s, Y_s) + \mathcal{L}_{Dice}(\hat{Y}_s, Y_s) \tag{13}$$

The adversarial generator loss ($\mathcal{L}_{adv}^{G}$) applies reverse-label binary cross-entropy on the decoder-tap features, weighted by $\lambda_{adv} = 10^{-3}$. The barrier loss $\mathcal{L}_{bar}$ enforces the dynamic class-conditional bounds $[\tau_{\text{low}}(w), \tau_{\text{high}}(w)]$ on target predictions, weighted by $\lambda_{bar} = 0.05$ and activated only when the gating test rejects the null of equal source and target cup-to-disc distributions (otherwise $\lambda_{bar} = 0$). The auxiliary classifier loss $\mathcal{L}_{cls}^{source}$ is binary cross-entropy with logits on source samples, weighted by $\lambda_{cls} = 0.5$. The domain discriminator D is updated separately in alternating steps.

4 Experiments and Results

4.1 Datasets and Implementation Details

Datasets. We benchmark the framework on three public clinical fundus image datasets: REFUGE, RIM-ONE-r3, and Drishti-GS. Table 4 in Appendix A summarizes their characteristics. We use five metrics to assess segmentation accuracy and anatomical fidelity: Dice Similarity Coefficient ($Dice_c$), Intersection over Union (IoU_c) (where $c \in \{OD, OC\}$), and the Mean Absolute Error of the vertical cup-to-disc ratio (δ). Definitions are in Appendix B.

Implementation Details. The proposed framework is implemented using PyTorch (v2.1.0) and trained on an NVIDIA A30 Tensor Core GPU. Images are processed in native orientations. Optimization uses the AdamW with weight decay 10^{-4} for all non-batch-normalization parameters. The pretrained MiT-B5 encoder, decoder, auxiliary classifier and domain discriminator use learning rates $\eta_E = 10^{-5}$, $\eta_{Dec} = 10^{-4}$, $\eta_{Cls} = 10^{-4}$, and $\eta_D = 10^{-4}$, respectively. Each iteration uses a batch of $B = 8$ for segmentation and auxiliary classification and an unlabeled target batch for adversarial adaptation. Training runs for 100 epochs. The barrier steepness t is linearly annealed from $t_{start} = 5.0$ to $t_{end} = 50.0$ over training. For the initial 30 epochs, a static warm-up interval $[\tau_{low}, \tau_{high}] = [0.3, 0.6]$ constrains all target samples. Afterwards image-specific class-conditional bounds ($[\tau_{low}(w), \tau_{high}(w)]$) interpolate between the normal-eye band $[\tau_{low,N}, \tau_{high,N}] = [0.25, 0.4]$ and the glaucomatous-eye band $[\tau_{low,G}, \tau_{high,G}] = [0.55, 0.7]$ using the temperature scaled ($T = 2.0$) classifier output.

4.2 Component-wise ablation

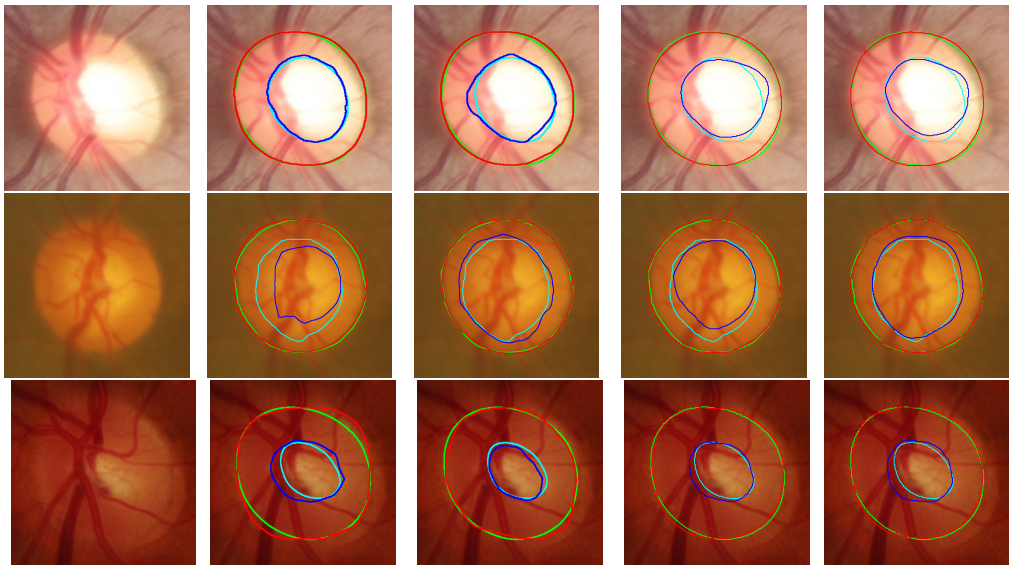
Table 1 presents the component-wise ablation with REFUGE as the source. The source-only base achieves high OD but lower OC performance. Adversarial alignment improves Drishti-GS $Dice_{OC}$, and the classifier further improves $Dice_{OC}$ on RIM-ONE-r3. Fixed size barrier and class-conditional bands yield strongest Drishti-GS performance, but reduce RIM-ONE-r3 $Dice_{OC}$. The benefit of anatomical regularization is not uniform across target domains. Robust cross-domain ONH segmentation must address both **feature-level domain discrepancy** and **anatomical-level population variability**. The domain discriminator targets the former benefiting OC segmentation. Alignment alone does not constrain OC and OD anatomy. The fixed size barrier exposes this limitation: a population-independent prior can be overly restrictive when the target anatomy differs. This motivates the class-conditional bands and gating.

4.3 Qualitative Analysis of Segmentation Boundaries

Figure 2 compares OD and OC boundary predictions across REFUGE (source), Drishti-GS, and RIM-ONE-r3 (target domains). The second column shows the source-only baseline, while the third presents the supervised oracle, providing reference points for assessing domain adaptation. The

Table 1: Component-wise ablation of the proposed framework. Base, Disc., Cls., Size, and Bds. denote the base model, domain discriminator, auxiliary classifier, fixed size barrier, and class-conditional bands, respectively; ✓and — indicate component inclusion and exclusion.

Base	Disc.	Cls.	Size	Bds.	Drishti-GS		RIM-ONE-r3	
					$Dice_{OD}$	$Dice_{OC}$	$Dice_{OD}$	$Dice_{OC}$
✓	—	—	—	—	**0.985**	0.737	0.977	0.810
✓	✓	—	—	—	0.981	0.835	0.976	0.803
✓	—	—	✓	—	0.956	0.874	0.977	0.767
✓	✓	✓	—	—	0.979	0.819	0.974	**0.823**
✓	✓	—	✓	—	0.939	0.881	**0.980**	0.772
✓	✓	✓	✓	—	0.955	0.880	0.966	0.774
✓	—	✓	✓	✓	0.982	0.889	0.978	0.782
✓	✓	✓	✓	✓	0.982	**0.895**	0.968	0.770



(a) Original image (b) Source-only (c) Oracle model (d) Without $\mathcal{L}_{\text{bar}}$ (e) With $\mathcal{L}_{\text{bar}}$

Figure 2: Qualitative effect of the anatomical constraint on cross-domain ONH segmentation. Rows 1–3 correspond to input images from REFUGE, Drishti-GS, and RIM-ONE-r3, respectively. Boundaries are overlaid as follows: OD ground truth in green and OD prediction in red; OC ground truth in cyan and OC prediction in blue.

boundary overlays highlight prediction errors, particularly around the smaller and less clearly delineated OC. The last two columns illustrate the effect of anatomical bounds across target domains with different degrees of anatomical mismatch. Anatomical regularization can improve cup geometry and boundary agreement when source–target mismatch is substantial, but may restrict valid anatomical variation when the distributions are similar, supporting its selective rather than unconditional application.

4.4 Cross-domain generalization and Anatomical Mismatch

To determine when anatomical regularization is beneficial, we evaluate the framework across all six source-target combinations among REFUGE, Drishti-GS and RIM-ONE-r3. Table 2 reports the gating decision along with its effect on the OC segmentation. For the four transfer directions where the KS test rejects the null of equal distributions, class-conditional bounds consistently improve

Table 2: Effect of class-conditional anatomical constraint on the cross domain optic cup segmentation, with significance-gated decisions. ρ_{source} and $\hat{\rho}_{\text{target}}$ denote the source ground-truth and target predicted cup-to-disc area ratios as mean_{std}. The p value is from a two-sided Kolmogorov–Smirnov test between the two distributions; the barrier loss $\mathcal{L}_{bar}$ is enabled when $p < 0.01$. The $-\mathcal{L}_{bar}$ and $+\mathcal{L}_{bar}$ columns use the same trained model. $\Delta\rho_{\text{pred}} = \rho_{\text{pred}} - \rho_{\text{GT}}$ denotes the prediction–ground-truth difference in the mean cup-to-disc area ratio in the target domain.

Source → Target	Statistical significance			$Dice_{OD}$	$Dice_{OC}$		$\Delta\rho_{\text{pred}}$		δ
	ρ_{source}	$\hat{\rho}_{\text{target}}$	p value		$-\mathcal{L}_{bar}$	$+\mathcal{L}_{bar}$	$-\mathcal{L}_{bar}$	$+\mathcal{L}_{bar}$	
REFUGE→Drishti-GS	$0.26_{0.106}$	$0.39_{0.106}$	< 0.001	0.982	0.847	**0.895**	-0.125	+0.005	0.061
RIM-ONE-r3→Drishti-GS	$0.29_{0.189}$	$0.37_{0.124}$	< 0.001	0.954	0.746	**0.889**	-0.185	+0.033	0.069
Drishti-GS→RIM-ONE-r3	$0.55_{0.159}$	$0.41_{0.173}$	< 0.001	0.956	0.727	**0.769**	+0.120	+0.017	0.108
Drishti-GS→REFUGE	$0.55_{0.159}$	$0.36_{0.139}$	< 0.001	0.960	0.786	**0.816**	+0.105	+0.049	0.085
REFUGE→RIM-ONE-r3	$0.26_{0.106}$	$0.29_{0.119}$	0.149	0.974	**0.823**	0.770	+0.004	-0.023	0.056
RIM-ONE-r3→REFUGE	$0.29_{0.189}$	$0.25_{0.086}$	0.112	0.978	**0.867**	0.848	-0.004	-0.008	0.057

$Dice_{OC}$ and reduce $\Delta\rho_{\text{pred}}$. For the two directions where the null hypothesis is not rejected, the bounds reduce $Dice_{OC}$. The effectiveness of the anatomical constraint is governed by whether the source and target cup-to-disc distributions are statistically distinguishable. **Thus, the constraint is an adaptive anatomical safeguard whose benefit scales with source-target morphological mismatch, not a universally beneficial component.**

4.5 Comparison with Existing UDA Methods

Under the gating rule, the full anatomy-aware model is used when source–target cup-to-disc distributions are statistically distinguishable; otherwise, the adversarial/classifier configuration is used. The source-only baseline shows that domain shift mainly harms OC segmentation, while OD delineation remains relatively stable. Existing UDA methods improve OC overlap but often at the cost of OD accuracy or cup-to-disc stability. The gated framework gives a more balanced adaptation, achieving the best OD overlap on both targets and competitive OC overlap, while remaining close to the supervised oracle. It also improves OC segmentation and vCDR error when anatomical mismatch is large, and avoids unnecessary anatomical restriction when it is not.

Table 3: Quantitative comparison with state-of-the-art UDA methods for cross-domain ONH segmentation. REFUGE is used as the labeled source domain. Best UDA results are highlighted in bold.

Method	**Params(M)**	**Drishti-GS**			**RIM-ONE-r3**		
		$Dice_{OD}\uparrow$	$Dice_{OC}\uparrow$	$\delta\downarrow$	$Dice_{OD}\uparrow$	$Dice_{OC}\uparrow$	$\delta\downarrow$
Source-only baseline	82.4	0.985	0.737	0.172	0.977	0.810	0.059
Oracle model	82.4	0.983	0.909	0.049	0.984	0.866	0.047
BEAL Wang et al. (2019a)	12.5	0.961	0.862	-	0.898	0.810	-
pOSAL Wang et al. (2019b)	13	0.965	0.858	0.082	0.865	0.787	0.081
ECSD Liu et al. (2022a)	75	0.965	0.876	0.084	0.865	0.802	0.068
CLR Feng et al. (2022)	13	0.966	0.892	-	0.905	**0.841**	-
TIST Ghamsarian et al. (2023)	135	0.958	0.853	-	0.893	0.823	-
MA-UDA Ji & Chung (2023)	95	0.960	0.861	-	0.902	0.827	-
TriLA Chen et al. (2024)	65	0.962	0.867	-	0.908	0.839	-
HSGMA Qin & Okuwobi (2026)	100	0.968	0.879	0.076	0.966	0.827	**0.055**
DDS-UDA Xiao et al. (2026)	50	0.966	0.875	-	0.906	0.840	-
Proposed (gated)	84.5	**0.982**	**0.895**	**0.061**	**0.974**	0.823	0.056

5 Conclusion

This paper presented an anatomically constrained UDA framework for ONH segmentation that jointly addresses feature-level domain discrepancy and anatomical-level population variability. The method integrates automated ROI extraction, a SegFormer-backbone with DSBN, decoder-tap adversarial alignment and a two-sided log-barrier on a differentiable soft cup-to-disc area ratio. Component-wise ablation showed that anatomical regularization benefits only when the source and

target cup-to-disc distributions differ significantly. This motivated an unsupervised gating rule based on a two-sample hypothesis test that removes the need for a hand-tuned mismatch threshold. Across REFUGE, Drishti-GS and RIM-ONE-r3, the gated framework achieves competitive cross-domain performance and approaches supervised oracle accuracy. Limitations include dependence on KS test power, a single cup-to-disc area ratio prior, and propagation of ROI extraction errors. Future work will address these and extend to multi-source and continual adaptation.

AI USE STATEMENT

In this work, we used generative AI tools solely for editing and proofreading the manuscript, including improvements to grammar, phrasing, clarity, and overall readability. We did not use generative AI tools for research ideation, literature review, experimental design, code generation, data analysis, result interpretation, or the generation of scientific claims. No other tasks with required disclosure are applicable to this work. We have reviewed all AI-assisted edits and verified that they did not alter the scientific meaning, technical content, or conclusions of the paper. We take responsibility for the final content of this work, including all text, claims, and artifacts.

REFERENCES

Taoufik Abdellaoui, Yassine Malek, Houda Brarou, Fouad Elasri, Yassine Mouzari, Karim Reda, and Abdelbarre Oubaaz. Quantitative assessment of optic disc photographs in normal and open-angle glaucoma patients. *Ophthalmology Journal*, 7:12–19, 2022.

Michael D Abramoff, Wallace LM Alward, Emily C Greenlee, Lesya Shuba, Chan Y Kim, John H Fingert, and Young H Kwon. Automated segmentation of the optic disc from stereo color photographs using physiologically plausible features. *Investigative ophthalmology & visual science*, 48(4):1665–1673, 2007.

Arturo Aquino, Manuel Emilio Gegúndez-Arias, and Diego Marín. Detecting the optic disc boundary in digital fundus images using morphological, edge detection, and feature extraction techniques. *IEEE transactions on medical imaging*, 29(11):1860–1869, 2010.

Arunava Chakravarty and Jayanthi Sivaswamy. Joint optic disc and cup boundary extraction from monocular fundus images. *Computer methods and programs in biomedicine*, 147:51–61, 2017.

Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7346–7354. IEEE, 2019.

Shraddha Changune, Vivek Noel Soren, Gautam Das, and Tapan Kumar Gandhi. Bayesian-optimized superpixel-grabcut for traceable optic disc segmentation, 2026. URL `https://arxiv.org/abs/2608.29196`.

Chonglin Chen and Gang Wang. Iosuda: an unsupervised domain adaptation with input and output space alignment for joint optic disc and cup segmentation. *Applied Intelligence*, 51(6):3880–3898, 2021.

Ziyang Chen, Yongsheng Pan, Yiwen Ye, Zhiyong Wang, and Yong Xia. Trila: Triple-level alignment based unsupervised domain adaptation for joint segmentation of optic disc and optic cup. *IEEE Journal of Biomedical and Health Informatics*, 28(9):5497–5508, 2024.

Jun Cheng, Jiang Liu, Yanwu Xu, Fengshou Yin, Damon Wing Kee Wong, Ngan-Meng Tan, Dacheng Tao, Ching-Yu Cheng, Tin Aung, and Tien Yin Wong. Superpixel classification based optic disc and optic cup segmentation for glaucoma screening. *IEEE transactions on medical imaging*, 32(6):1019–1032, 2013.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

Qi Dou, Cheng Ouyang, Cheng Chen, Hao Chen, and Pheng-Ann Heng. Unsupervised cross-modality domain adaptation of convnets for biomedical image segmentations with adversarial loss. *arXiv preprint arXiv:1804.10916*, 2018.

Venkata Gopal Edupuganti, Akshay Chawla, and Amit Kale. Automatic optic disk and cup segmentation of fundus images using deep learning. In *2018 25th IEEE international conference on image processing (ICIP)*, pp. 2227–2231. IEEE, 2018.

Wei Feng, Lin Wang, Lie Ju, Xin Zhao, Xin Wang, Xiaoyu Shi, and Zongyuan Ge. Unsupervised domain adaptive fundus image segmentation with category-level regularization. In *International conference on medical image computing and computer-assisted intervention*, pp. 497–506. Springer, 2022.

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.

Mohsen Ghafoorian, Alireza Mehrtash, Tina Kapur, Nico Karssemeijer, Elena Marchiori, Mehran Pesteie, Charles RG Guttmann, Frank-Erik De Leeuw, Clare M Tempany, Bram Van Ginneken, et al. Transfer learning for domain adaptation in mri: Application in brain lesion segmentation. In *International conference on medical image computing and computer-assisted intervention*, pp. 516–524. Springer, 2017.

Negin Ghamsarian, Javier Gamazo Tejero, Pablo Márquez-Neila, Sebastian Wolf, Martin Zinkernagel, Klaus Schoeffmann, and Raphael Sznitman. Domain adaptation for medical image segmentation using transformation-invariant self-training. In *International conference on medical image computing and computer-assisted intervention*, pp. 331–341. Springer, 2023.

Annette Giangiacomo and Anne Louise Coleman. The epidemiology of glaucoma. In *Glaucoma*, pp. 13–21. Springer, 2009.

Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging*, 38(10):2281–2292, 2019.

Wen Ji and Albert CS Chung. Unsupervised domain adaptation for medical image segmentation using transformer with meta attention. *IEEE Transactions on Medical Imaging*, 43(2):820–831, 2023.

Gopal Datt Joshi, Jayanthi Sivaswamy, and SR Krishnadas. Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment. *IEEE transactions on medical imaging*, 30(6):1192–1205, 2011.

Shreya Kadambi, Zeya Wang, and Eric Xing. Wgan domain adaptation for the joint optic disc-and-cup segmentation in fundus images. *International Journal of Computer Assisted Radiology and Surgery*, 15(7):1205–1213, 2020.

Bingyan Liu, Daru Pan, Zhenbin Shuai, and Hui Song. Ecsd-net: A joint optic disc and cup segmentation and glaucoma classification network based on unsupervised domain adaptation. *Computer Methods and Programs in Biomedicine*, 213:106530, 2022a.

Peng Liu, Charlie T Tran, Bin Kong, and Ruogu Fang. Cada: multi-scale collaborative adversarial domain adaptation for unsupervised optic disc and cup segmentation. *Neurocomputing*, 469:209–220, 2022b.

Zhizhe Liu, Zhenfeng Zhu, Shuai Zheng, Yang Liu, Jiayu Zhou, and Yao Zhao. Margin preserving self-paced contrastive learning towards domain adaptation for medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(2):638–647, 2022c.

Shijian Lu. Accurate and efficient optic disc detection and segmentation by a circular transformation. *IEEE Transactions on medical imaging*, 30(12):2126–2133, 2011.

Viola Stella Mary, Elijah Blessing Rajsingh, and Ganesh R Naik. Retinal fundus image analysis for diagnosis of glaucoma: a comprehensive survey. *IEEE Access*, 2016.

Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*, 2018.

Prashant Naithani, Ramanjit Sihota, Parul Sony, Tanuj Dada, Viney Gupta, Dimple Kondal, and Ravindra M Pandey. Evaluation of optical coherence tomography and heidelberg retinal tomography parameters in detecting early and moderate glaucoma. *Investigative ophthalmology & visual science*, 48(7):3138–3145, 2007.

Samiksha Pachade, Prasanna Porwal, Manesh Kokare, Luca Giancardo, and Fabrice Meriaudeau. Nenet: Nested efficientnet and adversarial learning for joint optic disc and cup segmentation. *Medical Image Analysis*, 74:102253, 2021.

Siyuan Qin and Idowu Paul Okuwobi. Hsgma-net: High-level semantic guidance and multi-scale adaptation for joint optic disc and cup segmentation. In *2026 6th International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA)*, pp. 215–222. IEEE, 2026.

Shivam Shah, Nikhil Kasukurthi, and Harshit Pande. Dynamic region proposal networks for semantic segmentation in automated glaucoma screening. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pp. 578–582. IEEE, 2019.

Sharath M Shankaranarayana, Keerthi Ram, Kaushik Mitra, and Mohanasankar Sivaprakasam. Fully convolutional networks for monocular retinal depth estimation and optic disc-cup segmentation. *IEEE journal of biomedical and health informatics*, 23(4):1417–1426, 2019.

Yih-Chung Tham, Xiang Li, Tien Y Wong, Harry A Quigley, Tin Aung, and Ching-Yu Cheng. Global prevalence of glaucoma and projections of glaucoma burden through 2040: a systematic review and meta-analysis. *Ophthalmology*, 121(11):2081–2090, 2014.

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*, pp. 102–110. Springer, 2019a.

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, 2019b.

Zeya Wang, Nanqing Dong, Sean D Rosario, Min Xu, Pengtao Xie, and Eric P Xing. Ellipse detection of optic disc-and-cup boundary in fundus images. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 601–604. IEEE, 2019c.

Yusong Xiao, Yuxuan Wu, Li Xiao, Gang Qu, Haiye Huo, and Yu-Ping Wang. Dds-uda: Dual-domain synergy for unsupervised domain adaptation in joint segmentation of optic disc and optic cup. *Medical Image Analysis*, pp. 104056, 2026.

Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers. In *Neural Information Processing Systems (NeurIPS)*, 2021.

A DATASETS

REFUGE provides high-quality pixel-level annotations for both the OD and OC, based on independent assessments by seven glaucoma specialists. Moreover, the multi-device acquisition setting introduces image appearance and acquisition variability. In India, **Drishti-GS** was collected in a clinical setting using a distinct imaging protocol and field of view. The OD and OC annotations were obtained by combining manual delineations from four eye-care experts using a majority voting. **RIM-ONE-r3** offers another distinct imaging setting centered on the ONH, using reference masks obtained from two expert ophthalmologists. Collectively, diversity in acquisition conditions, image appearance, anatomical presentation, and annotation procedures allows full assessment of segmentation performance under varied domain shifts.

Table 4: Summary profile of fundus image datasets used in cross-domain evaluation

Dataset	Glaucoma	Healthy	Devices	Resolution
REFUGE	120	1080	Zeiss Visucam 500 / Canon CR-2	2124×2056 / 1634×1634
Drishti-GS	70	31	OPTO Electronic	2047×1759
RIM-ONE-r3	39	85	Nidek AFC-210	2144×1424

B Evaluation Metrics

Overlap-based segmentation accuracy. Spatial overlap between predicted class region masks and expert ground-truth masks is evaluated using both the $Dice$ and IoU. For a target semantic class $c \in$ OD, OC, let $Y_c \in \{0,1\}^{H\times W}$ represent the ground-truth segmentation mask and $\hat{Y}_c \in \{0,1\}^{H\times W}$ represent the threshold binary mask derived from the predicted soft probability map P_c for a target semantic class $c \in$ OD, OC. The $Dice$ measures harmonic spatial overlap, penalizing both false positive and false negative pixel allocations.

$$Dice_c(Y_c, \hat{Y}_c) = \frac{2 \cdot |Y_c \cap \hat{Y}_c|}{|Y_c| + |\hat{Y}_c|} \tag{14}$$

Complementarily, IoU quantifies strict volumetric intersection divided by the union of spatial predictions:

$$\text{IoU}_c(Y_c, \hat{Y}_c) = \frac{|Y_c \cap \hat{Y}_c|}{|Y_c \cup \hat{Y}_c|} \tag{15}$$

Vertical cup-to-disc ratio (vCDR). The vCDR is a structural biomarker for glaucoma assessment, representing the vertical extent of the OC relative to the OD. To evaluate whether the predictions preserve the clinically relevant morphology, we measure the mean absolute error (MAE) of the vCDR value between the predicted and ground-truth binary masks. Let $\text{vCDR}_{gt}^{(i)}$ and $\text{vCDR}_{pred}^{(i)}$ denote the ground-truth and predicted vCDR for the i-th test sample, respectively. The vCDR is computed as the ratio of the vertical diameter of the OC to the vertical diameter of the OD. For a dataset containing N target samples, vCDR MAE (δ), is calculated as:

$$\delta = \frac{1}{N} \sum_{i=1}^{N} |\text{vCDR}_{gt}^{(i)} - \text{vCDR}_{pred}^{(i)}| \tag{16}$$

A lower δ indicates better vertical optic cup-to-disc relationship.

C Cross-Domain Generalization

Table 5: Segmentation performance across all cross-domain combination under the significance gated decision. The gating decision for each transfer direction is reported in table 2.

Source → Target	OD_{Dice}	OC_{Dice}	OD_{IoU}	OC_{IoU}	δ
REFUGE → Drishti-GS	0.982	0.895	0.965	0.816	0.061
REFUGE → RIM-ONE-r3	0.974	0.823	0.949	0.708	0.056
RIM-ONE-r3 → Drishti-GS	0.954	0.889	0.912	0.809	0.069
RIM-ONE-r3 → REFUGE	0.978	0.867	0.957	0.771	0.057
Drishti-GS → REFUGE	0.960	0.816	0.924	0.658	0.085
Drishti-GS → RIM-ONE-r3	0.956	0.769	0.915	0.587	0.108

Table 5 reports the complete cross-domain segmentation performance under the significance-gated decision for all six source–target combinations. The gating decision for each direction is reported in Table 2. Across all transfer directions, OD segmentation remains consistently strong, with OD_{Dice}

between 0.954 and 0.982 and OD_{IoU} between 0.912 and 0.965, confirming that the optic disc boundary is relatively resilient to domain shift. In contrast, OC segmentation exhibits substantially larger variation, with OC_{Dice} ranging from 0.769 to 0.895 and OC_{IoU} from 0.587 to 0.816, reflecting the greater anatomical and appearance variability of the optic cup across datasets. The vCDR error δ ranges from 0.056 to 0.108, with the largest error observed for Drishti-GS $\rightarrow$ RIM-ONE-r3 (0.108), which also yields the lowest OC_{Dice} (0.769). The best overall performance is obtained for REFUGE $\rightarrow$ Drishti-GS ($OC_{Dice} = 0.895$, $\delta = 0.061$) and RIM-ONE-r3 $\rightarrow$ REFUGE ($OD_{Dice} = 0.978$, $OC_{Dice} = 0.867$, $\delta = 0.057$). The results also highlight the asymmetry of domain shift: for example, REFUGE $\rightarrow$ Drishti-GS achieves higher OC_{Dice} than the reverse direction Drishti-GS $\rightarrow$ REFUGE (0.895 vs. 0.816), and RIM-ONE-r3 $\rightarrow$ Drishti-GS outperforms Drishti-GS $\rightarrow$ RIM-ONE-r3 (0.889 vs. 0.769). These findings reinforce that adaptation effectiveness depends on both the source and target populations, and that cross-domain performance should be assessed across multiple transfer directions and anatomical measures.