GrInAdapt: Source-free Multi-Target Domain Adaptation for Retinal Vessel Segmentation

Zixuan Liu[1*], Aaron Honjaya[1*], Yuekai Xu[1*], Yi Zhang[2], Hefu Pan[2], Xin Wang[3], Linda G. Shapiro[1,3], Sheng Wang[1], and Ruikang K. Wang[2,4]

[1] Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, 98195, USA
{zucksliu,shapiro,swang}@cs.washington.edu
[2] Department of Bioengineering, University of Washington
[3] Department of Electrical and Computer Engineering, University of Washington
[4] Department of Ophthalmology, University of Washington
wangrk@uw.edu

Abstract. Retinal vessel segmentation is critical for diagnosing ocular conditions, yet current deep learning methods are limited by modality-specific challenges and significant distribution shifts across imaging devices, resolutions, and anatomical regions. In this paper, we propose GrInAdapt, a novel framework for source-free multi-target domain adaptation that leverages multi-view images to refine segmentation labels and enhance model generalizability for optical coherence tomography angiography (OCTA) of the fundus of the eye. GrInAdapt follows an intuitive three-step approach: (i) grounding images to a common anchor space via registration, (ii) integrating predictions from multiple views to achieve improved label consensus, and (iii) adapting the source model to diverse target domains. Furthermore, GrInAdapt is flexible enough to incorporate auxiliary modalities—such as color fundus photography—to provide complementary cues for robust vessel segmentation. Extensive experiments on a multi-device, multi-site, and multi-modal retinal dataset demonstrate that GrInAdapt significantly outperforms existing domain adaptation methods, achieving higher segmentation accuracy and robustness across multiple domains. These results highlight the potential of GrInAdapt to advance automated retinal vessel analysis and support robust clinical decision-making. Our code is here.

Keywords: Domain adaptation · Retinal imaging · Vessel segmentation.

1 Introduction

Retinal vessel segmentation [10] and quantification [16] are critical for early diagnosis and management of ocular diseases such as retinal vascular changes [3] and diabetic retinopathy [25]. Accurate delineation of arteries and veins [17] facilitates disease detection and monitoring of disease progression [6]. Various

[*] Equal contribution.

imaging modalities—including color fundus photography (CFP) [28] and optical coherence tomography angiography (OCTA) [17]—offer complementary views of the retinal vasculature [5]. However, each modality has inherent trade-offs: CFP provides a wide field of view but often misses fine capillary details, while OCTA delivers high-resolution, three-dimensional information at the cost of being more sensitive to noise and artifacts.

Despite advances in deep learning for segmenting major vessels from CFP [28,23] and extracting fine-grained capillary structures from OCTA [12], current methods remain limited by two factors. First, single-modality approaches are constrained by modality-specific shortcomings—for example, CFP's limited sensitivity to capillaries and OCTA's vulnerability to imaging artifacts [8,13]. Second, models trained on a single source domain struggle to generalize to new devices, fields-of-view, and clinical settings, leading to performance degradation due to distribution shifts [9,15]. While prior work has explored domain adaptation [24,4,14] and multi-modal integration [18,27,19], these approaches typically address non-vessel structure, low-resource, or single-domain scenarios [1,20]. The challenges of adapting vessel structures on OCTA remained less explored.

In this work, we go one step forward to explore a higher-resource but more ambitious setting: how to widely generalize a source model to multiple target domains through a dataset with multi-view images from each target domain. We propose a powerful framework, GrInAdapt, with the key idea to leverage multi-view images to obtain better labels and refine the model across multiple target domains. Specifically, GrInAdapt follows an intuitive three-step approach: first, all images are grounded to an anchor space through registration; second, the labels are refined by integrating predictions from multiple views; and finally, the model is adapted to the target domains. GrInAdapt is simple and robust—even a basic registration applied to imperfect vessel mask predictions can be effective. It demonstrates superiority in achieving label consensus and reducing noise caused by imaging artifacts and low imaging quality. Moreover, GrInAdapt is flexible: it can be easily equipped with ensemble learning and extended to a multi-modal, multi-view setting to further reinforce label robustness via auxiliary modalities.

Extensive experiments on the AI-READI dataset [1] with multi-device, multi-site, and multi-modal retinal images demonstrate that GrInAdapt consistently improves the source model by, on average, a 4% Dice score increase and a 0.42 ASSD reduction. Ablation studies confirm that each component of our three-step approach—registration, label integration, and adaptation—are robust and contribute significantly to the observed performance improvements. Further evaluations reveal robust generalization across sites, unseen locations, and different resolutions, with quantitative gains of 4% and 4.6% in Dice score. These findings underscore GrInAdapt's potential to enhance automated retinal analysis.

2 Preliminaries of Multi-target Domain Adaptation

Although the setting described here applies to various medical imaging modalities, in this work we focus on fundus OCTA and CAVF segmentation - Capil-

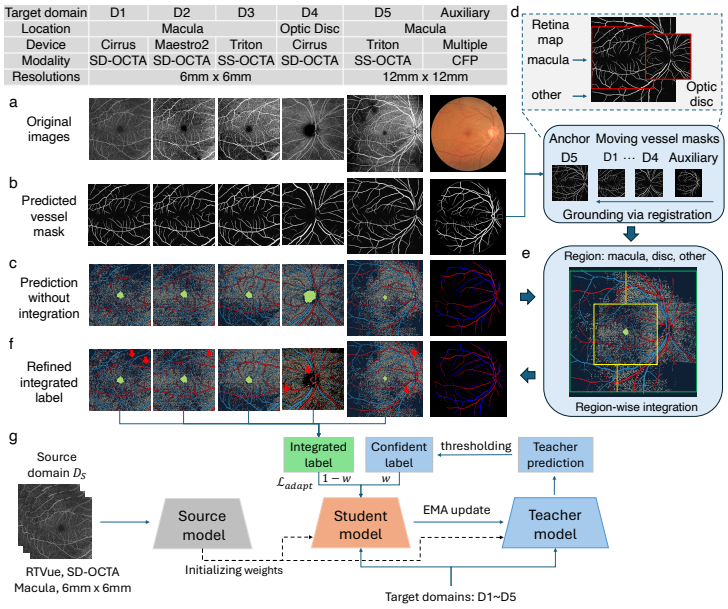

Fig. 1. Overview of GrInAdapt. **a**. Paired *en face* images from different domains. **b**. Predicted vessel masks used for registration. **c**. Prediction from souce models used for integration. **d**. Three regions on retina - macula, optic disc and other. **e**. Different predictions are integrated based on the region. **f**. Refined integrated labels for adapatation. **g**. Adaptation process with student-teacher and adaptive label merging.

laries, Artery, Vein and Foveal avascular zone (FAZ) [17]. Let $\mathcal{S} = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ denote the C-class segmentation labeled source domain, and let there be M unlabeled target domains $\mathbf{T}^M = \{\mathcal{T}_j\}_{j=1}^{M}$, where each $\mathcal{T}_j$ is defined as $\mathcal{T}_j = \{x_i^{t_j}\}_{i=1}^{N_{t_j}}$. In the source-free adaptation setting, only the pre-trained source model f_{θ_S} is available during adaptation. The objective of multi-target domain adaptation (MDA) is to adapt $f_{\theta_S} : \mathcal{X} \to \mathcal{Y}$ so that it performs well across all target domains $\{\mathcal{T}_j\}_{j=1}^{M}$. Adapting to multiple target domains is more challenging because the distribution shift is larger and more diverse. For example, in OCTA imaging, when taking macula-centered 6×6 scans from certain device as the source domain, target domain shifts can arise from several factors (Fig. 1a): imaging quality shift (different devices, D1-D3), location shift (D4), resolution shift (D5), and possible shift caused by the variance of clinicians' skill across different sites.

Subject-level multi-target image pair assumption. Previous DA methods usually focus on a single domain. A naive baseline for MDA is thus to simply treat $\mathbf{T}^M$ as a larger single target domain [9,15]. However, downgrading adaptation performance is expected as the number and diversity of target domains increase. We thus relax data availability, allowing subject-level data pairs across multiple domains, e.g., an eye being imaged from multiple devices, resolutions,

and locations. Specifically, given a subject set $\{s_k\}_{k=1}^{V}$, for each subject s_k, an image is acquired in every target domain j, i.e., $\{x_{s_k}^{t_j}\}_{j=1}^{M}, x_{s_k}^{t_j} \in \mathcal{T}_j$. This cross-target paired information is essential for aligning representations across domains and mitigating the effects of the aforementioned shifts. Optionally, paired data from auxiliary modalities $\mathcal{A}$ such as color fundus photography (CFP) can also be incorporated (D6, Fig. 1a). We assume that $\forall s_k \in V$, paired data $z_{s_k} \in \mathcal{A}$ and a source-free model $f_{\theta_A} : \mathcal{Z} \to \mathcal{W}$ that can generate a C_W-class auxiliary segmentation label $w \in \mathcal{W}$ are available. Here, the primary and auxiliary label space $\mathcal{Y}$ and $\mathcal{W}$ share common semantic classes, i.e., $\mathcal{Y} \cap \mathcal{W} \supseteq \{\text{Artery, Vein}\}$, allowing the auxiliary information to provide complementary cues for predicting the primary segmentation label $y \in \mathcal{Y}$. Under this setting, only training needs subject-level pairs; the adapted model should work on a single image from any domain.

3 GrInAdapt: Grounding, Integrating and Adapting multiple domains

In this section, we introduce GrInAdapt, our framework designed to robustly generalize a pre-trained source model to multiple target domains. GrInAdapt is a three-step process: grounding, integrating, and adapting, each addressing specific challenges arising from domain shifts in retinal vessel segmentation.

3.1 Grounding via Segmented Mask Registration

Inspired by [7,26], we proposed a grounding step that aligns target images of the same subject to a common anchor space to mitigate spatial variability arising from heterogeneous acquisition protocols in multi-target domain adaptation. Given a data bag $\mathcal{I}_{s_k} = \{I_j\}_{j=1}^{M+1} = \{x_{s_k}^{t_j}\}_{j=1}^{M} \cup \{z_{s_k}\}$ of subject s_k, we assumed for every two images (I_{anc}, I_{mov}), there exists a spatial transform $h \in \mathcal{H}$ to map the coordinate system of I_{mov} to that of I_{anc}. In our framework, we first picked an anchor image $I_{anc} \in \mathcal{I}_{s_k}$ and a registration function $R(g) : g(\mathcal{X}_{anc}) \times g(\mathcal{X}_{mov}) \to \mathcal{H}$ to estimate the spatial transformation $h \in \mathcal{H}$ that aligns the coordinates of all other moving images $I_{mov} \in \mathcal{I}_{s_k} \backslash \{I_{anc}\}$ to the coordinates of the anchor image.

We only require an invertible transformation h_j: $\forall h_j \in \mathcal{H}$, there exists an accessible inverse transform h_j^{-1} to map the registered moving image back for the adapting process later. To ensure registration quality, we used the predicted binary vessel probability maps $g(x_{s_k}^{t_j}) = \hat{p}_{s_k}^{t_j,ves}$ as they provide smooth morphological predictions (Fig. 1b). The resulting h_j and h_j^{-1} can be used to register the actual segmentation probability map $\hat{p}_{s_k}^{t_j} \in \mathbb{P}_{\mathcal{X}}^{C}$ (Fig. 1c). A similar process can also be applied to $\hat{p}_{s_k}^{z} \in \mathbb{P}_{\mathcal{X}}^{C_W}$. By doing so, the grounded mask predictions from multiple domains can be efficiently aggregated to share learned information.

3.2 Integrating via Reliable Region-wise Label Merging

Let $\tilde{p}_{s_k}^{t_j} = h(\hat{p}_{s_k}^{t_j})$ and $\tilde{p}_{s_k}^{a_j} = h(\hat{p}_{s_k}^{a_j})$ denote the registered segmentation probability maps for subject s_k from the target and auxiliary modalities, respectively.

With all predictions aligned in a common anchor space, our objective is to refine the segmentation output by fusing these diverse predictions in a manner that is sensitive to regional variations in reliability. The key idea is very intuitive: only reliable domains will be considered for integration. Let $L \subset \mathcal{X}$ denote a defined region within the anchor-space image (e.g., the macular region, optic disc region, or remaining regions), and let $J_L \subset \{1, \ldots, M + M_A\}$ be the index set corresponding to the selected images that have reliable contribution to region L.

For each pixel $u \in L$, the refined soft probability map is computed by averaging the softmax probabilities from the selected predictions (Fig. 1d-f):

$$\hat{p}^{\text{soft}}_{s_k}(u) = h^{-1}\Big(\frac{1}{|J_L|}\sum_{j \in J_L} \tilde{p}^{j}_{s_k}(u)\Big), \quad \hat{y}^{\text{hard}}_{s_k} = \arg\max_{c \in \{1,\ldots,C\}} \hat{p}^{\text{soft},c}_{s_k}(u) \tag{1}$$

where $\tilde{p}^{j}_{s_k}(x) \in \mathbb{P}^C$ is the probability vector (over C classes) predicted by the j-th image for subject s_k. Note that ensemble learning techniques such as bagging [22] multiple model replicas by averaging their predictions can be easily plugged in. The outputs $\hat{p}^{\text{soft}}_{s_k}(x)$ and $\hat{y}^{\text{hard}}_{s_k}(x)$ thereby serve as robust pseudo-labels transformed back to the original space for subsequent adaptation.

Choosing L and corresponding J_L is crucial. We split the retina map into 3 regions: macula, optic disc and other area (Fig. 1d,e). In the macular region, $\mathcal{J}_L$ includes only macula-centered OCTA predictions while excluding CFP images because they lack the fine vessel details for accurate segmentation in the central macula. Conversely, in the optic disc region, we relied on optic disc–centered OCTA predictions in combination with CFP images, which have been demonstrated to yield precise vessel delineations in this area. For all other regions, $\mathcal{J}_L$ comprises both CFP and OCTA predictions to benefit from both modalities.

3.3 Adapting via Teacher-Student Learning with Integrated Label

Unlike conventional domain adaptation methods that primarily focus on denoising the source model's predictions, our approach leveraged the integrated labels to correct and improve the model (Fig. 1g). In the adaptation stage, the integrated label $\hat{y}^{\text{hard}}_{s_k}$ is used for pseudo-supervision. Specifically, initiating f_θ with the source model f_{θ_S}, for each pixel $u \in \mathcal{X}$, we defined a segmentation loss as:

$$\mathcal{L}_{seg} = \frac{1}{|\mathcal{X}|}\sum_{u \in \mathcal{X}} \ell(f_\theta(u), \hat{y}^{\text{hard}}_{s_k,u}), \quad \ell(f(u), y) = \text{Dice}(f(u), y) + \lambda \text{CE}(f(u), y) \tag{2}$$

However, as $\hat{y}^{\text{hard}}_{s_k}$ may still contain noise, we further incorporated a confidence loss exploiting high-confidence model predictions [4] thresholded by τ_c:

$$\mathcal{L}_{conf} = \frac{1}{|\mathcal{X}|}\sum_{x \in \mathcal{X}} \mathbb{I}\Big(\max_c f^c_\theta(u) > \tau_c\Big) \cdot \ell\Big(f_\theta(u), \hat{y}^{\text{hard}}_{s_k,u}\Big), \tag{3}$$

To prevent the model from reinforcing only its own potentially erroneous predictions, we employed a teacher-student framework following [11,24]. The teacher

Table 1. Methods comparison on DSC score across various domains. A, V, F stand for artery, vein and FAZ. Δ indicates improvements compared to the source model. Optic disc F is counted as 0 (fail) / 100 (success) if the output has FAZ / no FAZ prediction.

Methods	C	[24] DSC (%) (mean ± standard deviation) ↑					
		Macula			Optic Disc	Macula	All
		Cirrus(D1)	Maestro2(D2)	Triton(D3)	Cirrus(D4)	Triton(D5)	
		6×6				12×12	
Integrated label (Fig. 1d)	A	74.9±6.9	69.4±4.8	71.2±7.9	63.4±22.0	65.5±6.6	70.0±12.0
	V	77.9±9.9	71.0±5.4	75.1±6.4	64.1±20.3	70.9±4.2	72.6±11.9
	F	82.3±22.6	83.5±22.0	80.3±24.8	100±0.0	75.4±25.4	81.6±23.2
Source model (f_{θ_S})	A	70.0±6.4	64.5±5.0	65.0±9.0	70.3±7.3	58.5±5.9	66.9±7.8
	V	71.3±9.0	64.5±6.0	68.3±8.2	69.6±6.2	62.6±5.0	68.3±7.8
	F	79.4±30.2	89.7±6.3	83.1±23.8	0±0.0	76.5±30.4	83.9±22.6
Ensemble prediction (3 models)	A	71.7±6.2	66.4±4.6	67.0±9.3	63.8±23.4	60.6±7.3	67.2±12.5
	V	73.5±9.3	67.3±5.2	71.1±7.8	63.2±21.0	65.9±5.2	69.2±11.9
	F	87.0±24.5	90.7±6.2	86.3±20.9	0±0.0	83.8±23.1	87.6±19.1
CBMT[24] w/ ensemble prediction	A	70.7±4.4	62.8±5.7	65.1±4.6	71.1±4.4	62.6±3.6	66.9±5.7
	V	71.1±6.4	63.7±5.0	68.0±4.2	69.7±5.0	66.6±3.5	68.0±5.6
	F	0±0.0	83.8±22.2	42.7±45.4	0±0.0	4.2±16.6	42.9±45.0
DPL[4] w/ integrated label	A	73.3±4.9	66.4±5.9	67.8±8.8	75.6±5.1	61.7±6.9	70.1±7.7
	V	74.8±7.1	66.3±6.7	70.1±8.6	73.7±5.0	64.9±6.0	70.9±7.8
	F	88.1±8.7	86.4±9.6	84.2±18.4	100±0.0	78.7±22.4	85.7±14.3
GrInAdapt (Ours)	A	73.1±5.3$_{\Delta 3.1}$	67.2±5.3$_{\Delta 2.7}$	69.0±8.0$_{\Delta 4.0}$	74.4±5.9$_{\Delta 4.1}$	63.3±6.5$_{\Delta 4.8}$	70.5±7.0$_{\Delta 3.6}$
	V	74.6±7.5$_{\Delta 3.3}$	67.7±5.4$_{\Delta 3.2}$	71.6±7.1$_{\Delta 3.3}$	73.5±5.5$_{\Delta 3.9}$	66.9±5.1$_{\Delta 4.3}$	71.7±6.8$_{\Delta 3.4}$
	F	90.9±6.2$_{\Delta 11.5}$	90.3±7.9$_{\Delta 0.6}$	87.3±17.6$_{\Delta 4.2}$	100±0.0	83.5±23.1$_{\Delta 7.0}$	89.0±13.2$_{\Delta 5.1}$

model f_{θ_T} provides more stable targets using weakly-augmented samples and is updated through an exponential moving average (EMA) of the student parameters, which train on strongly-augmented samples: $\theta_T \leftarrow \alpha\theta_T + (1-\alpha)\theta$, with α as the smoothing coefficient. The overall adaptation loss is then formulated as:

$$\mathcal{L}_{adapt} = \lambda(t)\,\mathcal{L}_{conf} + \big(1-\lambda(t)\big)\,\mathcal{L}_{seg}, \tag{4}$$

where $\lambda(t)$ denotes a cosine annealing weight schedule over E epochs. $\lambda(t)$ starts with a low value to prioritize learning from the integrated labels and gradually increases to place greater emphasis on the teacher's high-confidence predictions.

4 Experiments

4.1 Experimental Setup for Multi-target Domain Adaptation

Source and Multiple Target Data Domains. We aimed to segment out a 2D *en face* CAVF mask, while capillaries serve as an auxiliary task. To evaluate GrInAdapt, we picked the OCTA-500 [17] dataset as our source domain and 5 modalities of OCTA scans from the AI-READI dataset [1,20] as our multi-target domains. The OCTA-500 dataset contains 300 6mm × 6mm macula-centered OCTA scans captured by RTVue SD-OCTA device, each paired with a CAVF annotation. The AI-READI dataset has 1,060 patients, 2,112 eyes and in total

10,496 images from 3 sites (UW, UAB, UCSD) with 3 imaging modalities of 6mm × 6mm macula-centered OCTA (imaged by Maestro2 SD-OCTA, Cirrus SD-OCTA, Triton SS-OCTA), one 6mm × 6mm Optic disc centered Cirrus SD-OCTA, and one 12mm × 12mm macula-centered Triton SS-OCTA modality with no CAVF annotations. 3D OCTA flow, 3D OCT structural volume, and 2D superficial *en face* projection map are provided for all domains from both datasets, and the AI-READI dataset has at least one paired 2D *en face* CFP image for each scan. To facilitate the evaluation process, we selected 16 patients from three sites with balanced lateral distribution based on imaging quality, and annotated the CAVF masks by professionally trained ophthalmologists, resulting in a test set of 80 images held-out throughout the training process.

Model Architecture. For the source segmentation model f_{θ_S}, taking both 3D OCTA, 3D OCT and 2D projection map as input, it is a 4-block, 3D Res-UNet equipped with IPN v2 architecture [21,17] and enhanced with a 2D branch that processes superficial retinal features. The 3D branch is designed to capture volumetric contextual information from OCTA and OCT data, while the 2D branch extracts complementary features from en face projections. Nevertheless, GrInAdapt is independent of source model structure. For the auxiliary CFP segmentation model f_{θ_A}, we leveraged the architecture of [28].

Implementation Details. We first trained f_{θ_S} using the OCTA-500 dataset with a 240:10:50 train:val:test split via the standard segmentation loss $l(f(x), y)$ in Eqn. (2), and utilized a trained f_{θ_A} from [28]. f_{θ_S} and f_{θ_A} were used to predict artery-vein masks and generate binary vessel masks for OCTA scans and CFP images. We used a simple key point-based registration with affine transform and AKAZE detector[2,23]. We randomly chose an anchor image and monitored if an image passed registration by thresholding the translation, scaling, shear, and perspective factors extracted from the transform. For each subject, we iterated the anchor image until it successfully registered all images, or failed on all possible anchors. We only picked subjects with all images successfully registered for label integration. For the integration, we used images with a smaller field of view - 6mm×6mm macula-centered and optic disc centered images to split the region. To improve label robustness, we trained 3 replicas of f_{θ_S} and ensembled them by taking the average prediction for each registered image. We only transformed the artery and vein classes back to update the original prediction as we found FAZ labels tended to shrink to the intersection across different domains. In the adaptation stage, we set α=0.995, a learning rate of 8×10^{-5}, and the $\lambda(t)$ schedule to be from 0.1 to 0.9 over 3 epochs. We let λ=1, τ_{artery}=τ_{vein}=0.5, $\tau_{\text{capillaries}}$=0, and a region-wise FAZ thresholding scheme for τ_{faz}. Different levels of Gaussian noise were added to images as weak and strong augmentations.

4.2 Experimental Results

Registration results. GrInAdapt has a high success rate of registration even with a simple, automated, and parameter-free key-point based registration algorithm. Of 2,112 eyes, 1,562 from 868 patients successfully had all of their images registered, resulting in 7,303 registered images with a 74.2% subject-level and

Table 2. Cross-site methods comparison on the DSC score and ASSD distances.

Methods	C	[24] DSC (%) ↑			[24] ASSD (pixel) ↓			
		UW	UAB	UCSD	UW	UAB	UCSD	All
Inegrated	A	70.2±7.8	74.3±6.7	65.7±17.8	1.4±0.7	1.1±0.5	1.9±1.8	1.5±1.2
Label	V	70.7±7.6	78.1±6.1	69.6±17.8	1.3±0.6	1.0±0.5	1.7±1.7	1.3±1.1
(Fig. 1d)	F	73.9±33.8	84.8±14.0	88.4±5.4	1.2±0.7	1.7±2.5	1.3±0.6	1.4±1.5
Source	A	67.0±7.8	70.0±7.2	63.8±7.5	2.0±1.2	1.5±0.7	2.2±1.3	1.9±1.1
Model	V	66.7±7.2	72.3±6.7	66.6±8.5	1.9±1.1	1.6±1.0	2.3±1.8	1.9±1.3
(f_{θ_S})	F	74.1±33.4	90.6±4.9	89.9±7.2	1.5±2.4	0.9±0.7	1.0±0.7	1.2±1.6
GrIn	A	70.3±7.3△3.3	73.3±6.7△3.3	68.1±6.2△4.3	1.6±0.8△.4	1.2±0.5△.3	1.7±0.9△.5	1.5±0.8△.4
Adapt	V	70.3±6.6△3.6	75.1±5.9△2.8	70.2±7.1△3.6	1.4±0.7△.5	1.1±0.6△.5	1.6±0.9△.7	1.4±0.8△.5
(Ours)	F	84.7±20.△10.6	92.2±3.5△1.6	91.5±5.0△1.6	0.9±0.4△.6	0.7±0.3△.2	0.9±0.5△.1	0.8±0.4△.4

81.8% image-level all-success rate. Among the failed image cases, more than 60% of them were found to have too poor imaging quality to generate a valid vessel mask through quality check. GrInAdapt works with adequate registration, evaluated by DSC score between target and registered vessel masks. The median value is 60% for cohorts with successful registration and 7% for failed cohorts. This validated the robustness of the design and provided us a large cohort for label integration and adaptation.

Domain adaptation performance. We compared GrInAdapt with 5 baselines to examine its effectiveness (Table 1). We first validated the good quality of the integrated label (D1-D3, D5) especially on Artery and Vein, since it facilitates adaptation for most domains by serving as the target label. By utilizing it, GrInAdapt enhanced the source model with an average of 4% improvement of Dice score. The improvement scale is generally consistent among the artery and vein classes across the macular region of different domains (D1-D3), suggesting its ability to simultaneously fit multiple domains. Notably, the 12mm×12mm field (D5) had a larger 5.3% average improvement, indicating the power of integrating vessel details from the smaller 6mm×6mm view to the wider but less detailed 12mm×12mm view. For the challenging optic disc domain (D4), GrInAdapt successfully learned to identify the area and not predict FAZ. We further witnessed a substantial improvement in artery and vein performance, especially when integrated label performances are largely downgraded and affected by test set variance. We then ablated GrInAdapt with two modified baselines, CBMT [24] with no integrated label and DPL [4] with integrated label, and found that both the integrated label and designated teach-student scheme contributed to the improvement. We finally demonstrated the similar overall 5-domain artery and vein segmentation performance of GrInAdapt based on one single test image for each domain compared to the integrated label which requires 5 images from all domains in the test set, validating its consistent improvement on all domains and strong generalizability. We further evaluated GrInAdapt across three sites (Table 2) on both Dice and ASSD metrics and again witnessed consistent improvements of Dice and reduced ASSD (on average 0.42) on all sites versus the source model. Compared to the integrated label, GrInAdapt also achieved

comparable results on Artery and Vein segmentation and improved FAZ performance, demonstrating its robust and generalizable performance across sites.

5 Conclusion

We have presented GrInAdapt, a novel framework for multi-target domain adaptation in retinal vessel segmentation. Using multiview and multimodal imaging, GrInAdapt refines pseudo-labels via three steps: grounding via registration, integrating multiple predictions with region-specific fusion, and adapting the source model using a teacher–student framework. Experiments on the large-scale AI-READI dataset demonstrate the superior label consensus and improved segmentation performance across diverse imaging domains achieved by GrInAdapt. GrInAdapt can utilize paired multi-domain data whenever available, e.g., longitudinal analysis where image pairs are collected across time. In future work, we plan to explore advanced registration techniques and incorporate additional modalities, with the goal of reducing annotation burdens and developing a fully automated, clinically deployable retinal vessel analysis system.

Disclosure of Interests. Ruikang Wang receives financial support from Carl Zeiss Meditec, Colgate Palmolive Company, and Estee Lauder Inc; serves as consultant for Carl Zeiss Meditec and Cyberdontics.

References

1. http://orcid. org/0000-0001-6343-2140 Drolet Caroline 4 http://orcid. org/0000-0003-2287-4190 Lucero Abigail 8 Matthies Dawn 7 http://orcid. org/0009 0003-4909-6058 Pittock Hanna 3 Watkins Kate 3 York Brittany 1, P.M., 11, N.P.S.W.X.: Ai-readi: rethinking ai data collection, preparation and sharing in diabetes research and beyond. Nature Metabolism pp. 1–3 (2024)
2. Alcantarilla, P.F., Solutions, T.: Fast explicit diffusion for accelerated features in nonlinear scale spaces. IEEE Trans. Patt. Anal. Mach. Intell **34**(7), 1281–1298 (2011)
3. Cardoso, J.N., Keane, P.A., Sim, D.A., Bradley, P., Agrawal, R., Addison, P.K., Egan, C., Tufail, A.: Systematic evaluation of optical coherence tomography angiography in retinal vein occlusion. American journal of ophthalmology **163**, 93–107 (2016)
4. Chen, C., Liu, Q., Jin, Y., Dou, Q., Heng, P.A.: Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24. pp. 225–235. Springer (2021)
5. Chen, C., Chuah, J.H., Ali, R., Wang, Y.: Retinal vessel segmentation using deep learning: a review. IEEE Access **9**, 111985–112004 (2021)
6. Cunha-Vaz, J.: A central role for ischemia and octa metrics to follow dr progression (2021)

7. Ding, L., Kuriyan, A.E., Ramchandran, R.S., Wykoff, C.C., Sharma, G.: Weakly-supervised vessel detection in ultra-widefield fundus photography via iterative multi-modal registration and learning. IEEE Transactions on Medical Imaging **40**(10), 2748–2758 (2020)
8. Enders, C., Lang, G.E., Dreyhaupt, J., Loidl, M., Lang, G.K., Werner, J.U.: Quantity and quality of image artifacts in optical coherence tomography angiography. PloS one **14**(1), e0210505 (2019)
9. Farahani, A., Voghoei, S., Rasheed, K., Arabnia, H.R.: A brief review of domain adaptation. Advances in data science and information engineering: proceedings from ICDATA 2020 and IKE 2020 pp. 877–894 (2021)
10. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: Blood vessel segmentation methodologies in retinal images–a survey. Computer methods and programs in biomedicine **108**(1), 407–433 (2012)
11. French, G., Mackiewicz, M., Fisher, M.: Self-ensembling for visual domain adaptation. arXiv preprint arXiv:1706.05208 (2017)
12. Gao, M., Guo, Y., Hormel, T.T., Tsuboi, K., Pacheco, G., Poole, D., Bailey, S.T., Flaxel, C.J., Huang, D., Hwang, T.S., et al.: A deep learning network for classifying arteries and veins in montaged widefield oct angiograms. Ophthalmology Science **2**(2), 100149 (2022)
13. Hormel, T.T., Huang, D., Jia, Y.: Artifacts and artifact removal in optical coherence tomographic angiography. Quantitative Imaging in Medicine and Surgery **11**(3), 1120 (2021)
14. Huai, Z., Ding, X., Li, Y., Li, X.: Context-aware pseudo-label refinement for source-free domain adaptive fundus image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 618–628. Springer (2023)
15. Kouw, W.M., Loog, M.: A review of domain adaptation without target labels. IEEE transactions on pattern analysis and machine intelligence **43**(3), 766–785 (2019)
16. Lee, H., Lee, M., Chung, H., Kim, H.C.: Quantification of retinal vessel tortuosity in diabetic retinopathy using optical coherence tomography angiography. Retina **38**(5), 976–985 (2018)
17. Li, M., Huang, K., Xu, Q., Yang, J., Zhang, Y., Ji, Z., Xie, K., Yuan, S., Liu, Q., Chen, Q.: Octa-500: a retinal dataset for optical coherence tomography angiography study. Medical image analysis **93**, 103092 (2024)
18. Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., Wang, T.: A cross-modality learning approach for vessel segmentation in retinal images. IEEE transactions on medical imaging **35**(1), 109–118 (2015)
19. Liu, Z., Xu, H., Woicik, A., Shapiro, L.G., Blazes, M., Wu, Y., Lee, C.S., Lee, A.Y., Wang, S.: Octcube: a 3d foundation model for optical coherence tomography that improves cross-dataset, cross-disease, cross-device and cross-modality analysis. arXiv preprint arXiv:2408.11227 (2024)
20. Owsley, C., Matthies, D.S., McGwin, G., Edberg, J.C., Baxter, S.L., Zangwill, L.M., Owen, J.P., Lee, C.S.: Cross-sectional design and protocol for artificial intelligence ready and equitable atlas for diabetes insights (ai-readi). BMJ Open **15**(2) (2025). https://doi.org/10.1136/bmjopen-2024-097449, https://bmjopen.bmj.com/content/15/2/e097449
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)

22. Sewell, M.: Ensemble learning. RN **11**(02), 1–34 (2008)
23. Shi, D., He, S., Yang, J., Zheng, Y., He, M.: One-shot retinal artery and vein segmentation via cross-modality pretraining. Ophthalmology science **4**(2), 100363 (2024)
24. Tang, L., Li, K., He, C., Zhang, Y., Li, X.: Source-free domain adaptive fundus image segmentation with class-balanced mean teacher. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 684–694. Springer (2023)
25. Tey, K.Y., Teo, K., Tan, A.C., Devarajan, K., Tan, B., Tan, J., Schmetterer, L., Ang, M.: Optical coherence tomography angiography in diabetic retinopathy: a review of current applications. Eye and Vision **6**, 1–10 (2019)
26. Wang, X., Luo, X., Zhuang, X.: Bingo: Bayesian intrinsic groupwise registration via explicit hierarchical disentanglement. In: International Conference on Information Processing in Medical Imaging. pp. 319–331. Springer (2023)
27. Zhao, X., Zhang, J., Li, Q., Zhao, T., Li, Y., Wu, Z.: Global and local multi-modal feature mutual learning for retinal vessel segmentation. Pattern Recognition **151**, 110376 (2024)
28. Zhou, Y., Xu, M., Hu, Y., Lin, H., Jacob, J., Keane, P.A., Alexander, D.C.: Learning to address intra-segment misclassification in retinal imaging. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24. pp. 482–492. Springer (2021)