AMERICAN ACADEMY OF OPHTHALMOLOGY®

One-shot Retinal Artery and Vein Segmentation via Cross-modality Pretraining

Danli Shi, MD, PhD,[1,2,3] *Shuang He, MD,*[3] *Jiancheng Yang, PhD,*[4] *Yingfeng Zheng, PhD,*[3] *Mingguang He, MD, PhD*[1,2,3]

Purpose: To perform one-shot retinal artery and vein segmentation with cross-modality artery-vein (AV) soft-label pretraining.

Design: Cross-sectional study.

Subjects: The study included 6479 color fundus photography (CFP) and arterial-venous fundus fluorescein angiography (FFA) pairs from 1964 participants for pretraining and 6 AV segmentation data sets with various image sources, including RITE, HRF, LES-AV, AV-WIDE, PortableAV, and DRSplusAV for one-shot finetuning and testing.

Methods: We structurally matched the arterial and venous phase of FFA with CFP, the AV soft labels were automatically generated by utilizing the fluorescein intensity difference of the arterial and venous-phase FFA images, and the soft labels were then used to train a generative adversarial network to learn to generate AV soft segmentations using CFP images as input. We then finetuned the pretrained model to perform AV segmentation using only one image from each of the AV segmentation data sets and test on the remainder. To investigate the effect and reliability of one-shot finetuning, we conducted experiments without finetuning and by finetuning the pretrained model on an iteratively different single image for each data set under the same experimental setting and tested the models on the remaining images.

Main Outcome Measures: The AV segmentation was assessed by area under the receiver operating characteristic curve (AUC), accuracy, Dice score, sensitivity, and specificity.

Results: After the FFA-AV soft label pretraining, our method required only one exemplar image from each camera or modality and achieved similar performance with full-data training, with AUC ranging from 0.901 to 0.971, accuracy from 0.959 to 0.980, Dice score from 0.585 to 0.773, sensitivity from 0.574 to 0.763, and specificity from 0.981 to 0.991. Compared with no finetuning, the segmentation performance improved after one-shot finetuning. When finetuned on different images in each data set, the standard deviation of the segmentation results across models ranged from 0.001 to 0.10.

Conclusions: This study presents the first one-shot approach to retinal artery and vein segmentation. The proposed labeling method is time-saving and efficient, demonstrating a promising direction for retinal-vessel segmentation and enabling the potential for widespread application.

Financial Disclosure(s): Proprietary or commercial disclosure may be found in the Footnotes and Disclosures at the end of this article. *Ophthalmology Science 2024;4:100363*

The retina is the only window for monitoring vascular diseases noninvasively. Retinal artery and vein (AV) segmentation are important for downstream applications, such as performing population analyses, diagnosing disease, and planning treatments for the eye, heart, and brain.[1–6] However, segmenting the retinal AV segmentation is a nontrivial task. Although deep learning methods attain state-of-the-art accuracy based on supervised training with large labeled data sets, obtaining manual segmentation labels for the retinal artery and vein requires considerable expertise and high cost. In most retinal AV segmentation data sets, there are only a few manually labeled images.[7] The problem of limited labeled data is exacerbated by eye diseases and differences in image acquisition procedures across cameras and institutions, which can produce wide variations in resolution and image quality.

To overcome these challenges, collective studies have focused on hand-engineered network architectures and data augmentation.[8–11] It is also common to enforce graphical connectivity as vascular prior knowledge.[12,13] However, data augmentation such as random image rotations or random deformations has a limited ability to emulate real variation and provides no clue about the characteristic of the artery and vein structure. Few-shot learning is proposed to address data scarcity on several tasks for natural images and magnetic resonance images,[14,15] where the model learns from a single or a few annotated prototypes. Retinal AV segmentation in a one-shot setting is

https://doi.org/10.1016/j.xops.2023.100363
ISSN 2666-9145/23

challenging yet particularly useful in the clinical scenario, which has never been explored by prior arts.

Although previous work focused on segmenting artery and vein from manual labeling, we aimed at a different way to improve the data efficiency. It is based on our clinical observations in ophthalmology. Fundus fluorescein angiography (FFA) has been the gold standard for assessing vascular disease for decades, which delineates the retinal artery and vein perfusion dynamically in a clear way through contrast dye injection.[16] Furthermore, FFA is an examination frequently performed in the ophthalmic clinic; thus, there are plenty of images available, which provides a great opportunity for incorporating FFA images for automatic labeling.

This study addresses the label deficiency and the limited generalizability of retinal AV segmentation via crossmodality pretraining. To make the labeling automatic, we leveraged several arterial and venous-phase FFA images and registered their vasculature to generate AV soft labels. By pretraining and finetuning, we successively transferred the generic vascular information from FFA to color fundus photography (CFP). We demonstrated the superior generalizability of our method by one-shot segmentation on 6 data sets with different cameras. Our approach is the first method to achieve proven generalizability with automatic soft-label generation and pretraining.

Methods

Data

Figure 1 shows the workflow of the study. For cross-modality pretraining, 2129 CFP images with 12 619 corresponding FFA images in arterial or venous phase (within 10 to 60 seconds after dye injection) from 2143 patients were retrospectively collected from Zhongshan Ophthalmic Center of Sun Yat-Sen University between November 2016 and December 2019. Human subjects from multiple cohorts were included in this study. All patients were anonymized and deidentified, and the study adhered to the tenets of the Declaration of Helsinki. The institutional review board of Zhongshan Ophthalmic Center approved the study. Individual consent for this retrospective analysis was waived. The FFA images were captured by Zeiss FF450 Plus (Carl Zeiss, Inc) and Heidelberg Spectralis (Heidelberg Engineering) with a resolution of 768 × 768. CFP images were captured by Topcon TRC-50XF (Topcon Healthcare) and Zeiss FF450 Plus (Carl Zeiss, Inc), with resolutions ranging from 1110 × 1467 to 2600 × 3200.

We included 6 clinical data sets for generalizability experiments, the RITE (DRIVE-AV),[17] HRF,[18] LES-AV,[19] AV-WIDE,[20,21] and 2 private data sets, PortableAV and DRSplusAV. The private data sets were labeled manually by a retinal specialist using Vessellabel software, as described in the previous study.[11] Data set characteristics are presented in Table 1. In the one-shot setting, we randomly select 1 image from each data set for training and split the remainder 1:1 for validation and testing (Table 2). If the data set had an official train/test split id (RITE), we randomly selected 1 image from the training set and the rest as validation, keeping the testing in line with the official id.

Cross-modality Pretraining

Image registration was based on the retinal vessels extracted from the CFP and FFA. Retinal-vessel maps for CFP images were extracted using the retina-based microvascular health assessment system[11] segmentation module, and vessels in FFA images were extracted by a validated deep learning model.[22] Figure 2 represents the image registration workflow. The arterial phase FFA images were first registered with the venous-phase FFA images, and then both were registered with CFP. We did registration by detecting key points from the corresponding vessel map using AKAZE key point detector[23] and the Nearest-Neighbor-Distance-Ratio for feature-matching and random sample consensus[24] for generating homography matrices and outlier rejection. A validity restriction was added to exclude erroneously registered pairs: the value of the rotation scale was restricted to 0.8 to 1.3, and the absolute value of rotation radian was < 2 before the warping transformation. Furthermore, we filtered out pairs with a poor registration performance (i.e., Dice coefficient < 0.5). Here, the threshold value was empirically set based on the data set in our experiments.

After registration, we generated venous soft labels by subtracting the venous-phase FFA image from its corresponding arterial-phase image and merging the arterial phase and the subtracted venous-phase image to generate an AV soft label.

Table 1. Data Set Characteristics and the Train/Validation/Test Number of Images for the Experiment for One-Shot Artery and Vein Segmentation

Data set	No.	Train/Validation/ Test No.	Location	Field	Size	Eye Disease	Camera
RITE	40	1/19/20	Macula	45°	565×584	DR	CR5 non-mydriatic 3CCD camera (Canon)
HRF	45	1/22/22	Macula	45°	3504×2336	DR, glaucoma	Canon CR-1 nonmydriatic fundus camera (Canon)
AV-WIDE	30	1/14/15	Macula	200°	1300×800 2816×1880 1500×900	DR	Optos 200Tx (Optos PLC)
LES-AV	22	1/10/11	Optic-disc	30°−45°	1620×1444 1958×2196	Glaucoma	Visucam ProNm fundus camera with a ZK-5 (PRO NM/NMFA) 5 MP sensor
PortableAV	30	1/14/15	Macula	50°	4096×3072	Normal, DR, AMD	Mediworks camera (Mediworks)
DRSplusAV	20	1/9/10	Macula Optic-disc	40°×45°	2592×1944	Normal, myopia	iCare DRSplus (Centervue)

AMD = age-related macular degeneration; DR = diabetic retinopathy.

Table 2. Model Performance on the Test Set after Training on One Image from Each Data set

Data set	Vessel	AUC		Accuracy		Dice Score		Sensitivity		Specificity	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
RITE	Artery	0.943	0.019	0.972	0.004	0.724	0.036	0.701	0.045	0.987	0.002
	Vein	0.957	0.01	0.965	0.004	0.709	0.041	0.693	0.029	0.983	0.004
HRF	Artery	0.952	0.015	0.975	0.004	0.685	0.048	0.683	0.044	0.987	0.002
	Vein	0.962	0.007	0.972	0.004	0.696	0.034	0.685	0.026	0.986	0.003
LES-AV	Artery	0.952	0.022	0.980	0.005	0.715	0.054	0.691	0.036	0.991	0.004
	Vein	0.946	0.028	0.974	0.006	0.673	0.069	0.638	0.057	0.989	0.004
AV-WIDE	Artery	0.901	0.026	0.965	0.008	0.585	0.056	0.574	0.052	0.982	0.006
	Vein	0.926	0.015	0.964	0.005	0.586	0.052	0.590	0.043	0.981	0.004
PortableAV	Artery	0.971	0.005	0.976	0.003	0.773	0.022	0.763	0.021	0.988	0.001
	Vein	0.969	0.004	0.965	0.004	0.723	0.022	0.716	0.025	0.982	0.004
DRSplusAV	Artery	0.953	0.012	0.964	0.005	0.729	0.028	0.707	0.032	0.983	0.003
	Vein	0.951	0.006	0.959	0.003	0.674	0.020	0.623	0.013	0.983	0.002

AUC = area under the receiver operating characteristic curve; SD = standard deviation.

We used CFP as input and the corresponding AV soft label as ground truth to train pix2pixHD[25] for cross-modality pretraining. Pix2pixHD is a generative adversarial network (GAN) that aims to model the conditional distribution of real vessels given the input CFP via the following minimax game: the generator G tries to translate CFP to a realistic-looking retinal-vessel map to fool the discriminator D, whereas the discriminator D tries to distinguish the generated vessels from the real label. The multiscale discriminator evaluates the generated images patch by patch, thereby boosting the generator's capability to create a high-resolution, distinct image. The objective function included GAN loss, feature-matching loss, and perceptual loss.

Models were trained with a batch size of 4 and a learning rate of 0.0002. During training, images were resized to 768 × 768 and augmented by random horizontal/vertical flipping, random rotation (0−30 degrees), and random resized crop (ratio from 0.5−2). A total epoch of 20 was preset for each training.

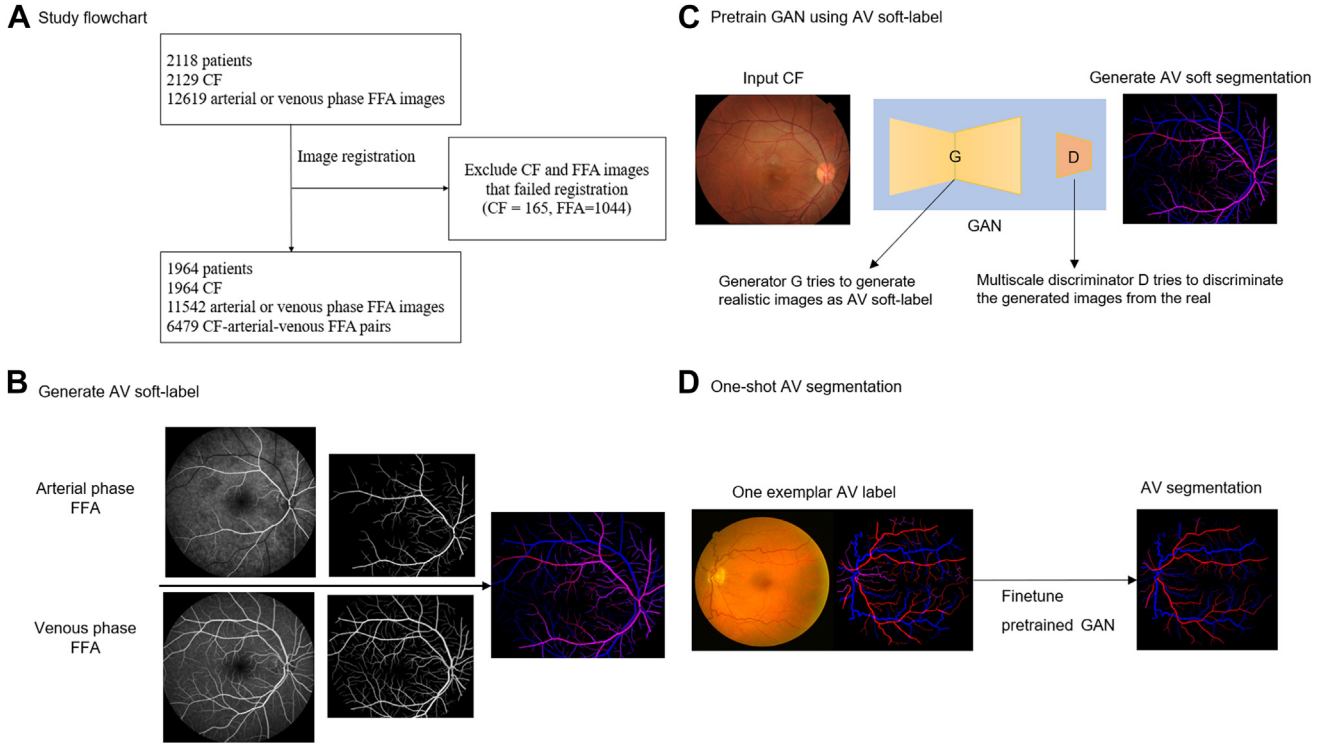

Figure 1. Graphical abstract of the study. A, Color fundus photopgraphy (CFP) and fundus fluorescein angiography (FFA) images were registered. B, Retinal vessels from the arterial-phase FFA and venous-phase FFA were extracted and matched respectively to generate artery-vein (AV) soft labels. C, The soft labels were used to train a generative adversarial network (GAN) to generate realistic AV soft segmentation (pretrain). D, The pretrained GAN was then fine tuned on one exemplar image from each segmentation data set to segment retinal artery and vein (one-shot segmentation). AV = artery-vein; CF = color fundus; FFA = fundus fluorescein angiography; GAN = generative adversarial network.

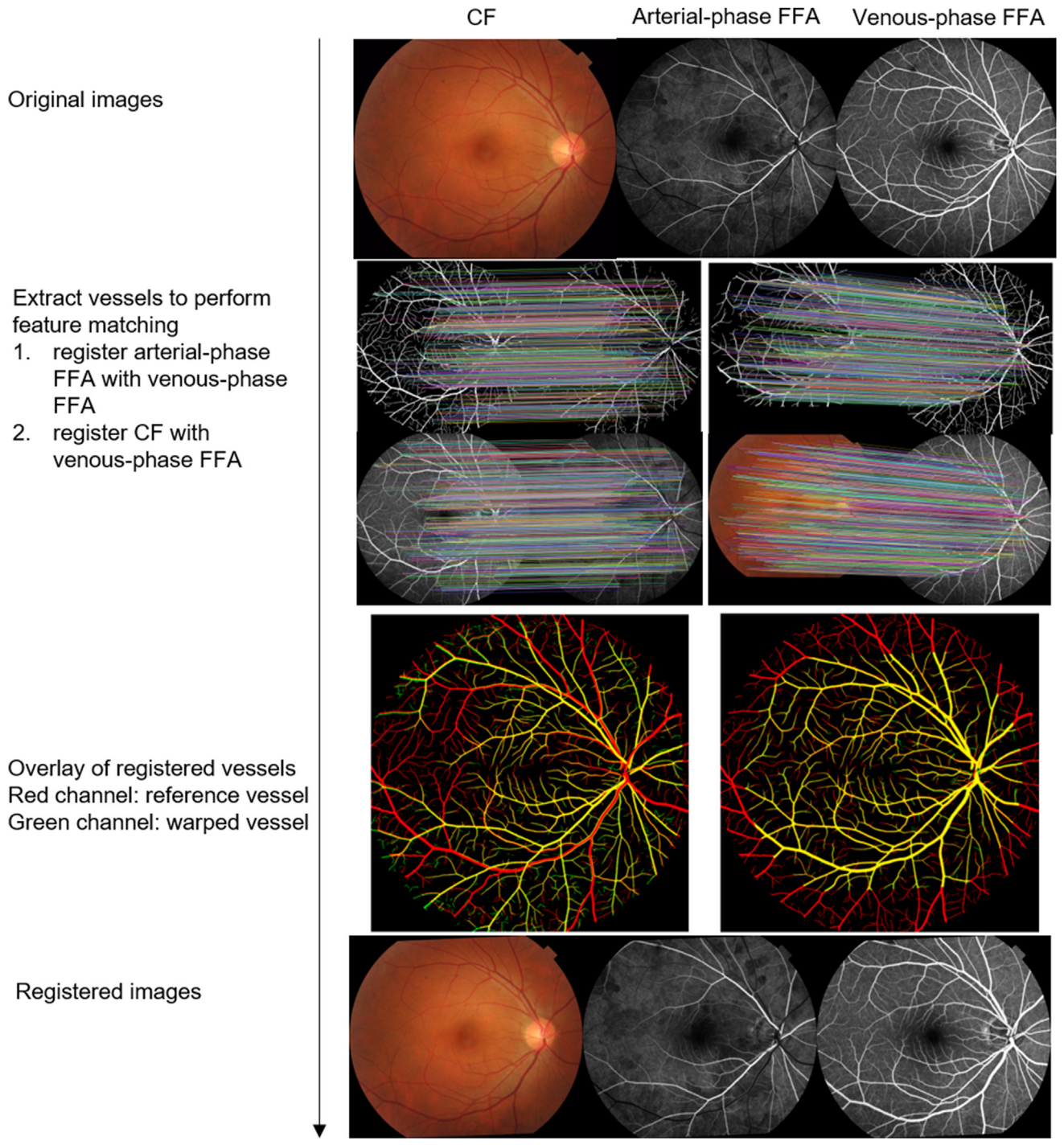

Figure 2. Demonstration of color fundus photography (CFP) and fundus fluorescein angiography (FFA) image registration. First, arterial-phase FFA was registered with venous-phase FFA, then CFP was registered with venous-phase FFA.

We selected the best model from 20 candidate models with the best visual authenticity in terms of the balance of arterial and venous color intensity. This selection was conducted by an ophthalmologist (D.S.) who manually assessed the quality of generation based on 15 images that were randomly chosen from the validation set. The demonstration of the balance of arterial and venous color intensity is shown in Figure 3. Images with good visual authenticity (right, Figure 3) have a balanced red and blue color filling, where arteries and veins were not overwhelmingly filled by adverse color. To provide further information about the quality of image generation,

we reported several quantitative metrics that are commonly used to assess image generation quality. These metrics include the Fréchet inception distance,[26] structural similarity measures,[27] mean absolute error, and the peak signal-to-noise ratio. Higher structural similarity measures, peak signal-to-noise ratio, lower mean absolute error, and Fréchet inception distance values indicate higher quality in generating images.

One-shot AV Segmentation

To demonstrate the generalizability of our pretraining approach, we performed experiments on one-shot segmentation across data sets with apparent different camera modalities, a challenging but practical scenario. We performed comparative experiments for each data set, training on 1 image, and validating and testing on the remaining images.

All models were trained with a batch size of 1 and a learning rate of 0.0002. During training, images were resized to 768 × 768 except for the AV-WIDE data set (resized to 1024 × 1024) and augmented by random horizontal/vertical flipping and random cropping. A total epoch of 15 was preset for each training. The model with the highest area under the receiver operating characteristic curve (AUC) on the validation set was selected. We evaluated the performance of each segmentation model on the test set by AUC, accuracy, Dice coefficient, sensitivity, and specificity.

To investigate the reliability of one-shot finetuning, we conducted experiments by finetuning the pretrained model with a fixed 10 epochs on a different single image for each data set. We then tested the models on the remaining images and calculated the mean and standard deviation of the AUC, accuracy, Dice coefficient, sensitivity, and specificity of multiple models for each data set.

Results

Cross-modality Pretraining

After excluding 165 CFP and 1077 FFA images that failed crossmodal registration, there were 6479 CFP-arterial-venous FFA registered pairs from 1964 participants included in model development. The flowchart is presented in Figure 1. Participants' mean (standard deviation) age was 48.30 (16.60), and 1051 (53.5%) participants were men. The data set was split by 9/1 at patient-level as the training and validation set.

Objective Evaluation

Quantitative comparisons between real and generated AV soft labels were performed on the internal test set, and the mean absolute error, peak signal-to-noise ratio, structural similarity measures, and Fréchet inception distance were 19.38, 20.08, 0.65, and 31.55, respectively.

Subjective Evaluation

We manually chose the eighth epoch, which demostrated the best AV visual authenticity, as the model weights for subsequent finetuning. The visual authenticity on the validation set was demonstrated in 3 scenarios: arterial dominates, venous dominates, and good authenticity (Figure 2). For arterial dominates, the artery (red channel) was fully filled with dye intensity, while the venous channel (blue channel) was also overly filled with red color. For venous dominates, the artery was insufficiently perfused, with low red color intensity. The image with good authenticity was realistic, and the arterial and venous color intensity was balanced.

One-shot Segmentation

Table 2 shows one-shot segmentation performance evaluated on each held-out test set. The proposed method achieved good performance across all cameras, with AUCs ranging from 0.901 to 0.971, accuracy from 0.959 to 0.980,

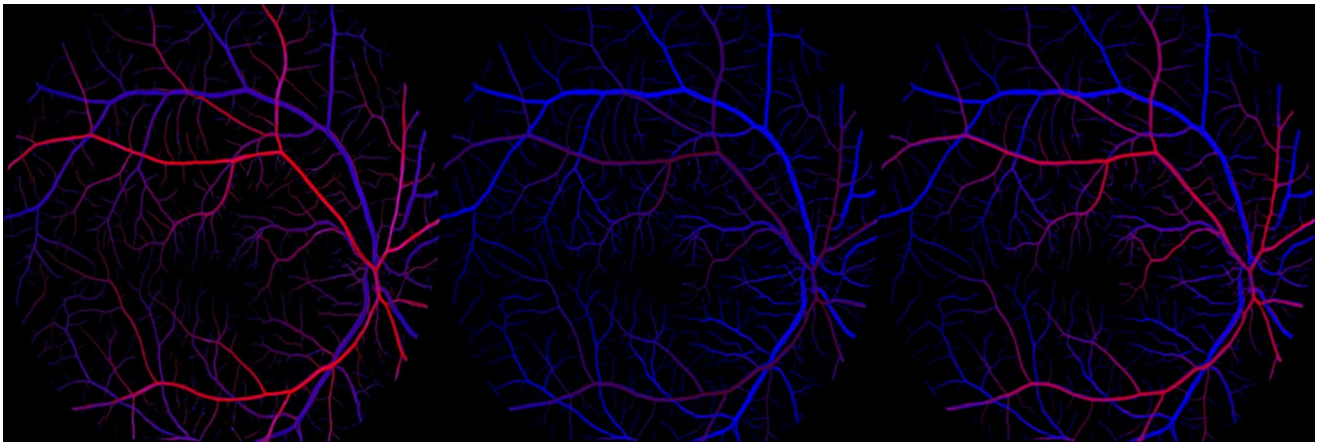

Figure 3. Demonstration of artery and vein visual authenticity. Red denotes arterial fluorescein dye filling, and blue denotes veinous fluorescein dye filling. From left to right, arterial dominates (red color overfills), venous dominates (blue color overfills), and good visual authenticity (most realistic).

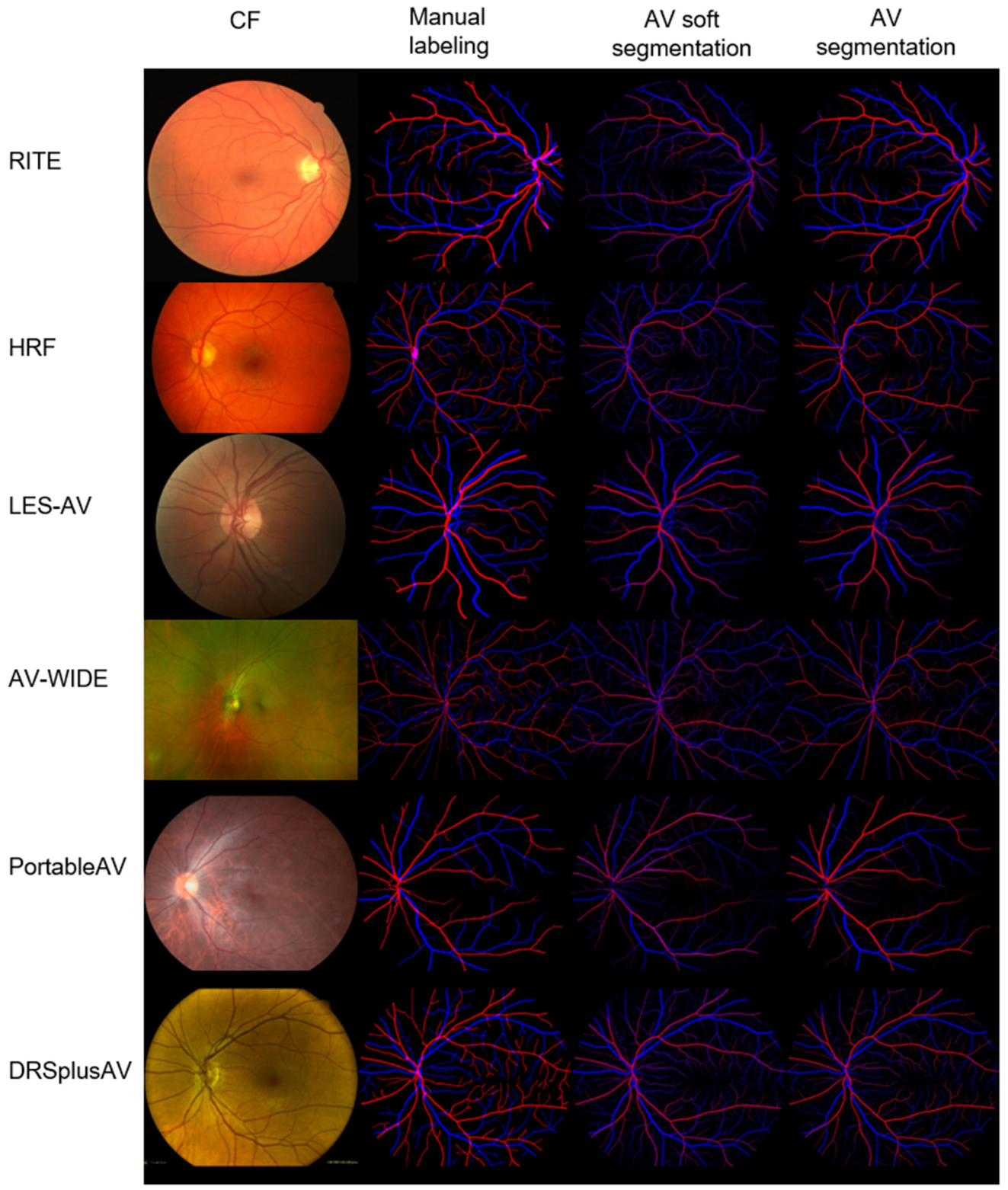

Figure 4. Examples of color fundus photograph (CFP) from each data set (first column), the manual labeling (second column), the soft artery and vein (AV) segmentation predicted by the pretrained model (third column), and the one-shot finetuned AV segmentation results (fourth column).

Dice score from 0.585 to 0.773, sensitivity from 0.574 to 0.763, and specificity from 0.981 to 0.991. Figure 4, from left to right, shows examples of CFP images from different data sets, the manual labeling, the soft AV map predicted by the pretrained weight, and the final AV segmentation on each test set. For comparison, we also

Table 3. Performance Comparison with Full-Data Training Methods

Data set	Method	Vessel	AUC		Accuracy		Dice Score		Sensitivity		Specificity	
			Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
RITE	Morano et al[28]	Artery	0.97	0.002	0.89	0.007	-	-	0.87	0.02	0.91	0.007
		Vein	0.97	0.001			-	-				
	Ma et al[29]	Artery/Vein	-		0.93		-		0.92		0.93	
	Shi et al[11]	Artery	0.95		0.94		0.57		0.86		0.94	
		Vein	0.95		0.94		0.63		0.87		0.95	
	Current method	Artery	0.94	0.02	0.97	0.004	0.72	0.04	0.70	0.05	0.99	0.002
		Vein	0.95	0.01	0.97	0.004	0.71	0.04	0.69	0.03	0.98	0.004
HRF	Karlsson et al[30]	Artery/Vein	-		0.96		0.96		0.97		0.95	
	Shi et al[11]	Artery	0.95		0.93		0.46		0.83		0.93	
		Vein	0.96		0.94		0.50		0.87		0.94	
	Current method	Artery	0.95	0.02	0.98	0.004	0.69	0.05	0.68	0.04	0.99	0.002
		Vein	0.96	0.007	0.97	0.004	0.70	0.03	0.69	0.03	0.99	0.003
LES-AV	Kang et al[31]	Artery/Vein	-		0.92		-		0.94		0.90	
	Shi et al[11]	Artery	0.97		0.95		0.86		0.96		0.58	
		Vein	0.97		0.95		0.85		0.96		0.61	
	Current method	Artery	0.95	0.02	0.98	0.005	0.72	0.05	0.69	0.04	0.99	0.004
		Vein	0.95	0.03	0.97	0.006	0.67	0.07	0.64	0.06	0.99	0.004
AV-WIDE	Khanal et al[32]	Artery/Vein			0.98		0.80	0.003	0.78			
	Shi et al[11]	Artery	0.91		0.95		0.68		0.96		0.45	
		Vein	0.92		0.95		0.73		0.96		0.47	
	Current method	Artery	0.90	0.03	0.97	0.008	0.59	0.06	0.57	0.05	0.98	0.006
		Vein	0.93	0.02	0.96	0.005	0.59	0.05	0.59	0.04	0.98	0.004

AUC = area under the receiver operating characteristic curve; SD = standard deviation.

listed the performance of AV segmentation in previously published papers (Table 3).

The Effect and Reliability of One-shot Finetuning

Table 4 shows the segmentation results of the pretrained model without finetuning when applied to the full images in each data set. Table 5 represents segmentation outcomes after finetuning the model with a fixed 10 epochs using a different single image from each data set iteratively and testing on the remaining images. The segmentation outcomes after one-shot finetuning were better than those without finetuning, as shown by the increased performance on each data set. For each data set, the standard deviations of the segmentation results across models were small, with all of them ranging from 0.001 to 0.102.

Table 4. The Reliability of One-shot Finetuning. Artery and Vein Segmentation Results of the Pretrained Model with No Finetuning

Data set	Vessel	AUC		Accuracy		Dice Score		Sensitivity		Specificity	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
RITE	Artery	0.947	0.012	0.968	0.006	0.418	0.054	0.288	0.046	0.997	0.001
	Vein	0.951	0.010	0.961	0.006	0.384	0.064	0.254	0.047	0.997	0.001
HRF	Artery	0.951	0.010	0.974	0.006	0.480	0.041	0.358	0.051	0.996	0.001
	Vein	0.959	0.006	0.971	0.005	0.463	0.062	0.331	0.054	0.996	0.001
LES-AV	Artery	0.963	0.008	0.978	0.005	0.542	0.055	0.393	0.050	0.998	0.001
	Vein	0.942	0.017	0.967	0.007	0.395	0.073	0.299	0.067	0.993	0.001
AV-WIDE	Artery	0.919	0.014	0.973	0.005	0.378	0.074	0.263	0.059	0.996	0.001
	Vein	0.910	0.010	0.962	0.005	0.107	0.037	0.068	0.024	0.993	0.001
PortableAV	Artery	0.924	0.042	0.968	0.006	0.367	0.059	0.259	0.048	0.995	0.001
	Vein	0.926	0.037	0.964	0.006	0.275	0.061	0.192	0.049	0.993	0.001
DRSplusAV	Artery	0.966	0.007	0.969	0.004	0.451	0.039	0.300	0.034	0.999	0.000
	Vein	0.968	0.005	0.967	0.003	0.434	0.051	0.294	0.040	0.997	0.001

AUC = area under the receiver operating characteristic curve; SD = standard deviation.

Table 5. Segmentation Results after Finetuning the Pretrained Model with a Fixed 10 Epochs on Different Single Image from Each Data set Iteratively, and Test on the Rest of the Images

		AUC		Accuracy		Dice score		Sensitivity		Specificity	
Data Set	Vessel	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
RITE	Artery	0.963	0.014	0.973	0.004	0.674	0.044	0.572	0.058	0.994	0.002
	Vein	0.954	0.016	0.977	0.004	0.661	0.052	0.567	0.067	0.994	0.002
HRF	Artery	0.959	0.014	0.975	0.004	0.667	0.037	0.571	0.075	0.994	0.003
	Vein	0.945	0.021	0.976	0.005	0.625	0.059	0.533	0.102	0.994	0.004
LES-AV	Artery	0.958	0.019	0.980	0.005	0.692	0.058	0.615	0.086	0.994	0.002
	Vein	0.962	0.012	0.983	0.004	0.702	0.049	0.620	0.068	0.995	0.002
AV-WIDE	Artery	0.924	0.013	0.975	0.004	0.524	0.050	0.414	0.055	0.995	0.002
	Vein	0.908	0.019	0.976	0.004	0.518	0.049	0.418	0.052	0.994	0.002
PortableAV	Artery	0.939	0.043	0.971	0.007	0.530	0.075	0.446	0.080	0.991	0.002
	Vein	0.928	0.046	0.970	0.007	0.536	0.074	0.483	0.084	0.989	0.002
DRSplusAV	Artery	0.978	0.006	0.978	0.003	0.718	0.028	0.622	0.043	0.995	0.001
	Vein	0.973	0.010	0.979	0.003	0.716	0.036	0.632	0.055	0.994	0.001

AUC = area under the receiver operating characteristic curve; SD = standard deviation.

Discussion

We presented a cross-modality pretraining method and demonstrated its superiority in an extreme case when only one exemplar is used for training. After finetuning 1 labeled image for each specific image type, the algorithm is versatile across a set of unlabeled images with substantial variation. Our work strives to bridge clinical, algorithm development, and practical generalization compactly.

Previous AV segmentation models were trained using human-labeled ground truth from CFP. There are several drawbacks: first, it is labor-intensive and expertise-demanding. Second, manual labeling is subject to intra-grader and intergrader variation. Third, the artery and vein may be ambiguous to identify in some cases when their color is similar, especially for the small vessels around the optic disc. Fourth, human graders may overestimate the vessel caliber because of the low contrast of the fundus photo, especially for small vessels.

We noted in clinical practice that FFA has been the gold standard for retinal vascular visualization and disease assessment. The injected fluorescein dye circulates within the blood flow, highlighting the vascular wall and the true trajectory of blood flow coming into arteries and then veins. By combing FFA and CFP, our approach is also a type of transfer learning where knowledge from FFA images helps with AV segmentation from CFP. After pretrained by automatically generated AV soft labels, the model captures the robust anatomical nature of the vessel, despite image type variance. The data sets we used to demonstrate the generalizability of the model were composed of different fields of view (ranging from 30 to 200), with different eye diseases and different pixel characteristics (clear and blurry). Images from the PortableAV were taken by a low-cost portable camera (Mediworks, FC162); the affordability of such cameras shows promise for wide deployment. However, image pixel characteristics from the portable camera vary with custom high-quality hospital-based fundus cameras (such as Topcon). Such domain variance may set a bottleneck for deep learning models trained on other common cameras. Here, we demonstrate that, by pretraining and finetuning, the camera variance could be addressed with minimal characteristic data.

Notably, when compared with previously published methods for AV segmentation on public data sets (Table 3),[11,28,30,31,33] the model's performance was slightly lower (for AUC, Dice score, and sensitivity); however, the specificity of our model was nearly perfect (all > 0.98), indicating there were few false-positives by our method. It is also encouraging to know that we trained the model on 1 image from each data set, whereas other methods used full-data training.

We also conducted sensitivity experiments to prove the necessity and reliability of the one-shot finetuning. Although the pretrained model demonstrated the ability to segment artery and vein on new data sets with AUCs > 0.9, the one-shot finetuning can further increase its segmentation performance, as evidenced by the improved AUC, Dice score, and sensitivity on each data set. Moreover, using 1 image to iteratively finetune the model yielded robust generalization.

The study has several advantages. It is the first to utilize FFA information for retinal AV segmentation and is fully automatic. The FFA delineates the artery and vein clearly by dynamic dye filling, which is more objective and accurate than human labeling. The FFA sample includes a series of eye diseases, which contains substantial variation while maintaining stable vascular information. Second, we did comprehensive experiments on 6 data sets with different image cameras and modalities to test the generalizability of the method. Third, even though one-shot AV segmentation is hard and has not been explored before, the performance of our model was reasonably good across data sets. This study, with successful implementation in retinal AV segmentation,

implies a promising perspective to solve the data scarcity issue in supervised medical image analysis by leveraging large-scale weak supervision generated in clinical practice.

There are also limitations to our study. First, models do not achieve supreme segmentation performance, mainly due to the loss of small vessels; future studies are needed to enhance small vessel segmentation performance. Secondly, to create AV soft labels, our framework registered arterial and venous-phase FFA images in a pixel-wise manner, which lends itself to several plausible future extensions (e.g., through deep feature fusion to better unify cross-modality features in an end-to-end manner). It is also worth exploring its extension to other imaging modalities with vessel-like structures. Third, the evaluation method we used to select the pretrained model was subjective, which may limit its applicability. To mitigate this issue, we provided the results of common image generation metrics for reference. Further research is still required to develop evaluation metrics that can quantitatively assess AV visual authenticity.

Conclusion

We presented a cross-modality pretraining method to address the label deficiency problem in retinal AV segmentation and demonstrated the method's generalizability in the one-shot scenario. The method enables generalizable AV segmentation, which is plausible for many downstream applications (e.g., vessel quantification, biomarker identification, and disease prediction). Our approach is likely to have an impact in the field of retinal AV segmentation.

Acknowledgments

The authors acknowledge and thank InnoHK and the HKSAR Government for their assistance and facility.

Footnotes and Disclosures

Originally received: January 7, 2023.
Final revision: June 29, 2023.
Accepted: June 30, 2023.
Available online: July 6, 2023. Manuscript no. XOPS-D-23-00004R3.

[1] Centre for Eye and Vision Research (CEVR), Hong Kong SAR, China.

[2] The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China.

[3] State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangzhou, China.

[4] Swiss Federal Institute of Technology in Lausanne (EPFL), Lausanne, Switzerland.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s): D.S., M.H.: Patent — "A method for cross-labeling retinal artery and vein."

Supported by Global STEM Professorship Scheme (P0046113). The sponsor or funding organization had no role in the design or conduct of this research.

HUMAN SUBJECTS: Human subjects from multiple cohorts were included in this study. All patients were anonymized and deidentified, and the study adhered to the tenets of the Declaration of Helsinki. The institutional review board of Zhongshan Ophthalmic Center approved the study. Individual consent for this retrospective analysis was waived.

No animals were used in this study.

Author Contributions:

Conception and design: Shi and M. He

Data collection: Shi, S. He, Zheng, and M. He

Analysis and interpretation: Shi and Yang

Obtained funding: M. He

Overall responsibility: M. He

Abbreviations and Acronyms:
AUC = area under the receiver operating characteristic curve; **AV** = artery-vein; **CFP** = color fundus photography; **FFA** = fundus fluorescein angiography; **GAN** = generative adversarial network.

Keywords:
Cross-modality pretraining, Domain generalization, One-shot, Retinal artery and vein segmentation.

Correspondence:
Mingguang He, MD, PhD, FRANZCO, Chair Professor of Experimental Ophthalmology, School of Optometry, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China. E-mail: mingguang.he@polyu.edu.hk.

References

1. Cano J, Farzad S, Khansari MM, et al. Relating retinal blood flow and vessel morphology in sickle cell retinopathy. *Eye (Lond)*. 2020;34:886–891.
2. Wang SB, Mitchell P, Liew G, et al. A spectrum of retinal vasculature measures and coronary artery disease. *Atherosclerosis*. 2018;268:215–224.
3. Farrah TE, Webb DJ, Dhaun N. Retinal fingerprints for precision profiling of cardiovascular risk. *Nat Rev Cardiol*. 2019;16:379–381.
4. Cheung CY, Xu D, Cheng CY, et al. A deep-learning system for the assessment of cardiovascular disease risk via the measurement of retinal-vessel calibre. *Nat Biomed Eng*. 2021;5:498–508.
5. Ballerini L, Fetit AE, Wunderlich S, et al. Retinal biomarkers discovery for cerebral small vessel disease in an older population. In: Papież BW, Namburete AIL, Yaqub M, Noble JA, eds. *Medical Image Understanding and Analysis*. Springer International Publishing; 2020:400–409.
6. Czakó C, Kovács T, Ungvari Z, et al. Retinal biomarkers for Alzheimer's disease and vascular cognitive impairment and dementia (VCID): implication for early diagnosis and prognosis. *GeroScience*. 2020;42:1499–1525.

7. Khan SM, Liu X, Nath S, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *The Lancet Digital Health*. 2021;3:e51−e66. https://doi.org/10.1016/S2589-7500(20)30240-5.
8. Mookiah MRK, Hogg S, MacGillivray TJ, et al. A review of machine learning methods for retinal blood vessel segmentation and artery/vein classification. *Med Image Anal*. 2021;68:101905.
9. Zhou Y, Yu H, Shi H. Study group learning: improving retinal vessel segmentation trained with noisy labels. In: de Bruijne M, Cattin PC, Cotin S, et al., eds. *Medical Image Computing and Computer Assisted Intervention − MICCAI 2021*. Springer International Publishing; 2021:57−67. https://doi.org/10.1007/978-3-030-87193-2_6.
10. Zhou Y, Xu M, Hu Y, et al. Learning to address intra-segment misclassification in retinal imaging. In: de Bruijne M, Cattin PC, Cotin S, et al., eds. *Medical Image Computing and Computer Assisted Intervention − MICCAI 2021*. Springer International Publishing; 2021:482−492. https://doi.org/10.1007/978-3-030-87193-2_46.
11. Shi D, Lin Z, Wang W, et al. A deep learning system for fully automated retinal vessel measurement in high throughput image analysis. *Front Cardiovasc Med*. 2022;9:823436.
12. Chen W, Yu S, Ma K, et al. TW-GAN: topology and width aware GAN for retinal artery/vein classification. *Med Image Anal*. 2022;77:102340. https://doi.org/10.1016/j.media.2021.102340.
13. Khanal A, Motevali S, Estrada R. Fully automated tree topology estimation and artery-vein classification. *arXiv*. Published online February 1, 2022. Preprint. https://doi.org/10.48550/arXiv.2202.02382
14. Kang D, Cho M. Integrative few-shot learning for classification and segmentation. *arXiv*. Published online March 29, 2022. Preprint. https://doi.org/10.48550/arXiv.2203.15712
15. Zhao A, Balakrishnan G, Durand F, et al. Data augmentation using learned transformations for one-shot medical image segmentation. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Publications; 2019:8535−8545. https://doi.org/10.1109/CVPR.2019.00874.
16. Burns SA, Elsner AE, Gast TJ. Imaging the retinal vasculature. *Annu Rev Vis Sci*. 2021;7:129−153.
17. Hu Q, Abràmoff MD, Garvin MK. Automated separation of binary overlapping trees in low-contrast color retinal images. In: Mori K, Sakuma I, Sato Y, et al., eds. *Medical Image Computing and Computer Assisted Intervention − MICCAI 2013*. Springer International Publishing; 2013:436−443.
18. Budai A, Bock R, Maier A, et al. Robust vessel segmentation in fundus images. *Int J Biomed Imaging*. 2013;2013:154860. https://doi.org/10.1155/2013/154860.
19. Orlando JI, Barbosa Breda J, van Keer K, et al. Towards a glaucoma risk index based on simulated hemodynamics from fundus images. In: *Medical Image Computing and Computer Assisted Intervention − MICCAI 2018*. Springer International Publishing; 2018:65−73. https://doi.org/10.1007/978-3-030-00934-2_8.
20. Estrada R, Tomasi C, Schmidler SC, Farsiu S. Tree topology estimation. *IEEE Trans Pattern Anal Mach Intell*. 2015;37:1688−1701.
21. Estrada R, Allingham MJ, Mettu PS, et al. Retinal artery-vein classification via topology estimation. *IEEE Trans Med Imaging*. 2015;34:2518−2534.
22. Ding L, Bawany MH, Kuriyan AE, et al. A novel deep learning pipeline for retinal vessel detection in fluorescein angiography. *IEEE Trans Image Process*. 2020;29:6561−6573.
23. Alcantarilla P, Nuevo J, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces, 13.1−13.11. In: *Proceedings of the British Machine Vision Conference 2013*. British Machine Vision Association; 2013. https://doi.org/10.5244/C.27.13.
24. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun ACM*. 1981;24:381−395.
25. Wang TC, Liu M, Zhu J, et al. High-resolution image synthesis and semantic manipulation with conditional GANs. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE Publications; 2018:8798−8807. https://doi.org/10.1109/CVPR.2018.00917.
26. Heusel M, Ramsauer H, Unterthiner T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: Guyon I, Von Luxburg U, Bengio S, et al., eds. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. NeurIPS Proceedings; 2017.
27. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600−612.
28. Morano J, Hervella ÁS, Novo J, Rouco J. Simultaneous segmentation and classification of the retinal arteries and veins from color fundus images. *Artif Intell Med*. 2021;118:102116.
29. Ma W, Yu S, Ma K, et al. Multi-task neural networks with spatial activation for retinal vessel segmentation and artery/vein classification. *arXiv*. Published online July 18, 2020. Preprint. https://doi.org/10.48550/arXiv.2007.09337
30. Karlsson RA, Hardarson SH. Artery vein classification in fundus images using serially connected U-Nets. *Comput Methods Programs Biomed*. 2022;216:106650.
31. Kang H, Gao Y, Guo S, et al. AVNet: a retinal artery/vein classification network with category-attention weighted fusion. *Comput Methods Programs Biomed*. 2020;195:105629.
32. Khanal A, Estrada R. Dynamic deep networks for retinal vessel segmentation. *Front Comp Sci*. 2020;2. https://doi.org/10.3389/fcomp.2020.00035.
33. Ma W, Yu S, Ma K, et al. Multi-task neural networks with spatial activation for retinal vessel segmentation and artery/vein classification. In: Shen D, Liu T, Peters TM, et al., eds. *Medical Image Computing and Computer Assisted Intervention − MICCAI 2019*. 11764. Springer International Publishing; 2019:769−778.