Cross-domain Controllable Generation Enables Zero-shot Fundus Image Analysis

Kaiwen Li[1,2,3], Hangzhou He[1,2,3], Jiakui Hu[1,2,3], Zhengjian Yao[1,2,3], Shuang Zeng[1,2,3], Ourui Fu[1,2,3], Lei Zhu[1,2,3], and Yanye Lu[1,2,3,4]

[1] Institute of Medical Technology, Peking University Health Science Center, Peking University, Beijing, China
[2] Department of Biomedical Engineering, Peking University, Beijing, China
[3] National Biomedical Imaging Center, Peking University, Beijing, China
[4] Corresponding Author
kaiwenli@stu.pku.edu.cn, yanye.lu@pku.edu.cn

Abstract. Precise analysis of ophthalmic diseases necessitates accurate vessel segmentation, image registration, and bifurcation detection across multi-modal images. However, automating these tasks is hindered by expensive manual annotation, which is further exacerbated by various imaging modalities and diseases. Existing generative methods for reducing labeling burden primarily struggle with two challenges: (1) reliance on target-domain image-label pairs and (2) difficulties in achieving a proper balance between controllability and appearance realism. In addition, there is a lack of effective methods for generating modality-dependent labels that can support multiple tasks. In this paper, we propose a unified framework that first generates modality-specific vessel masks and bifurcation labels, followed by the synthesis of realistic target-modality images that are structurally consistent with the generated labels, for zero-shot multi-modal and multi-task fundus image analysis. Specifically, we first rely on vascular dynamics simulations to design a flexible pipeline to synthesize labels. Then, by disentangling structure and appearance via phase maps during the generation, we are able to generate realistic images without requiring image-label pairs from target modalities. Our framework is evaluated under four paradigms across three tasks, including segmentation, registration, and keypoint detection, where our method achieves strong zero-shot performance. Our code is available at: https://github.com/kaiwenli325/UniVessel.

Keywords: Controllable generation · Vessel label generation · Fundus image analysis.

1 Introduction

Advancements in vessel segmentation [17], image registration [18], and vessel bifurcation detection [27] have facilitated accurate analysis of ophthalmic diseases. However, most of these improvements rely on vessel-centric annotations. Given the diversity of modalities and tasks in ophthalmology, together with

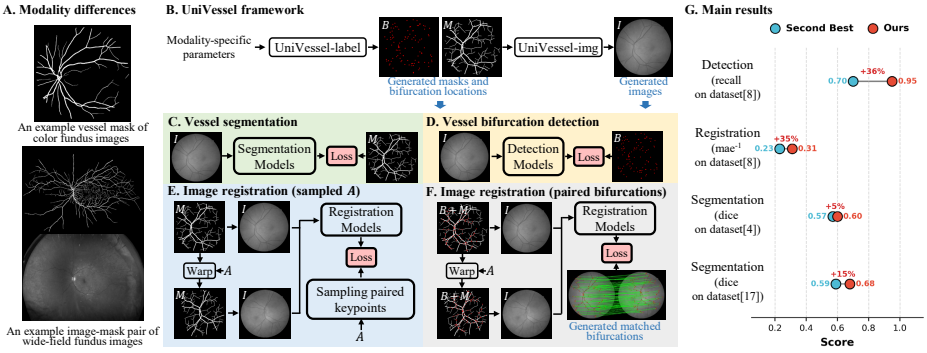

Fig. 1. (A) Illustration of modality differences. (B) The framework of UniVessel. (C-F) Four representative tasks supported by UniVessel and (G) their main results.

disease-induced appearance variations, large-scale and fine-grained vessel annotation is prohibitively expensive. Consequently, research on label-efficient methods [17,10,3,2] has gained widespread attention in recent years.

Typically, some methods rely on target image-label pairs to generate pseudo-labels for task-specific supervision [17,10], or learn their mapping to generate new pairs [3,6,7,31], both of which require target manual labels. In contrast, other methods [2,5] utilize cycle-consistency [32] and borrow vessel masks from other modalities as structure control to generate target-modality images. However, jointly modeling structure and style often leads to a trade-off between controllability and realism. Moreover, since vessel patterns differ substantially across modalities, the modality-related nature of vessel masks exacerbates discrepancies between generated and real images. For example, vessel masks of color fundus images are not suitable for guiding the generation of wide-field fundus images characterized by dense vessel networks, as shown in Fig. 1 A. Alternatively, a potential approach is to synthesize vessel masks from scratch, allowing for the flexible integration of modality-dependent characteristics. Therefore, it is imperative to jointly generate modality-specific labels and corresponding images that are both structurally controllable and stylistically realistic, to reduce labeling burden and enable zero-shot tasks.

To address these limitations, we propose UniVessel, a unified framework for generating modality-specific vessel labels and corresponding realistic images, enabling multiple zero-shot tasks. As shown in Fig. 1 B, in UniVessel, based on initial vessel networks [1], we first establish a topology-aware label generation pipeline, UniVessel-label, to produce vessel masks and bifurcation coordinates that reflect target-modality characteristics. Unlike methods based on tissue metabolism and dynamics [26,21,15,1], which can simulate anatomically plausible networks but fail to accommodate modality-specific fields of view (FOV) and vessel densities, UniVessel-label generates modality-aware labels providing reliable control signals for image synthesis and supervision for downstream tasks. Then, we design a structure and style decoupled generation

paradigm, UniVessel-img, where two diffusion models are employed to separately handle mask-controlled generation and modality-specific style synthesis, with image phase maps serving as an intermediate bridge. By learning the control relationship solely from paired data in public datasets, UniVessel-img can synthesize images across diverse modalities without requiring target-domain image-label pairs. UniVessel is evaluated on three tasks (Fig. 1 C-F), including vessel segmentation, image registration, and bifurcation detection, where UniVessel achieves outstanding zero-shot performance without supervision from real target image-label pairs, as shown in Fig. 1 G.

2 Related work

2.1 Vessel network generation

To simulate vessel structures, based on the space colonization algorithm, [26] further added physiological constraints to control the branching of vessels. By computing oxygen concentration and vascular endothelial growth factor, [21] simulated retinal vascular plexuses to construct optical coherence tomography angiography labels. Subsequently, [15] improved small capillary generation by incorporating space colonization theory. [1] employed a biophysical principle, Murray's law, to decide parameters such as vessel diameters and branching distances, thereby synthesizing vessel networks and deriving vessel masks. However, identical vessel networks can appear substantially different across modalities due to different sources of image contrast. The above methods lack a unified and flexible pipeline for modality adaptation and provide limited control over visual attributes such as vessel tortuosity and thickness distribution.

2.2 Fundus image generation

To generate paired images and vessel masks, [3] used the decoder of an autoencoder for mask generation along with a mask-controlled image generator, trained by target image–label pairs and an adversarial loss. [7] trained separate mask and image generators using manually annotated data, and leveraged the synthesized samples to augment the original dataset. Following a similar architecture, [6] further introduced a super-resolution network to improve image quality. Using an unpaired CycleGAN-based framework, [2,5] transformed vessel masks from the DRIVE dataset [28] into images of the target modality. These methods either require a certain amount of paired target-domain data, or employ vessel masks that fail to reflect the structure characteristics of the target modality.

3 Methods

3.1 Overview

Our UniVessel framework consists of two main modules: UniVessel-img generates **style-realistic images that structurally follow the input masks**;

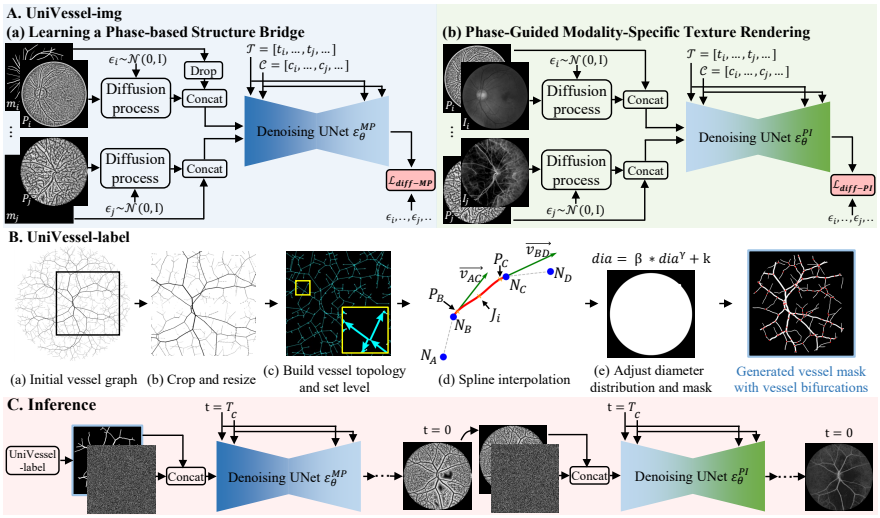

Fig. 2. (A) The training of UniVessel-img. (B) The generation pipeline of vessel masks and bifurcations in UniVessel-label. (C) The inference process of UniVessel.

UniVessel-label generates vessel masks and bifurcation coordinates that **conform to the characteristics of different modalities**. To train the controllable generation model, we leverage paired public data to learn the structure correspondence between vessel masks and images. Moreover, instead of directly using images as generation targets, we introduce local phase maps [14] as a novel intermediate modality. This effectively resolves the style interference caused by domain shifts between public and target datasets, which otherwise hinders the learning of structure control. These maps explicitly disentangle structure from style by encoding structure features (e.g., edges and lines) while suppressing spectral/device-dependent intensity and contrast variations. Therefore, adopting a DDPM-based generation paradigm [9], UniVessel-img employs denoising UNets ϵ_θ^{MP} and ϵ_θ^{PI} to learn mask-controlled phase map generation and transform phase maps into realistic target images, respectively. During inference, masks generated from UniVessel-label serve as the original control signals for UniVessel-img, where the final generated image-label pairs can support multiple downstream tasks.

3.2 Structure-to-Style Generation via Modality Bridging

Learning a Phase-based Structure Bridge To generate the phase maps as a modality bridge, at each spatial location, an image can be represented by even- and odd-symmetric components based on concepts of the analytic signal. The local phase is then defined as the arctangent of their ratio [14].

Formally, as shown in Fig. 2 A(a), taking image I_j sampled from the target domain as an example, its local phase map is computed and denoted as P_j. Since

I_j does not have a corresponding vessel mask, its control mask m_j is set all-zero. Meanwhile, I_i is sampled from the public datasets, along with its control mask m_i and computed local phase map P_i. To prevent the network from incorrectly transferring non-zero control masks exclusively to phase maps of public datasets, we randomly set m_i to an all-zero image with probability P_{drop}:

$$m_i = \alpha \cdot m_i, \qquad \alpha \sim \text{Bernoulli}(P_{drop}) \tag{1}$$

For each phase map in the input minibatch, the diffusion process samples noise $\epsilon \sim \mathcal{N}(0, \mathbf{I})$ and a timestep $t \in \{1, ..., T\}$ to perturb P into the noisy $\tilde{P}$. $\tilde{P}$ is concatenated with its control signal m along the channel dimension and fed into the denoising UNet ϵ_θ^{MP}, along with c representing its modality and the timestep t. Then, ϵ_θ^{MP}, which aims to generate phase maps of different modalities, is trained by:

$$\mathcal{L}_{diff-MP} = \frac{1}{K} \sum_{k=1}^{K} \|\epsilon_k - \epsilon_\theta^{MP}(\tilde{P}_k, m_k, c_k, t_k)\|_2^2 \tag{2}$$

Phase-Guided Modality-Specific Texture Rendering As shown in Fig. 2 A(b), using phase maps as control, UniVessel-img further employs another denoising UNet, ϵ_θ^{PI}, to synthesize target-modality images. At this stage, all images I_k are associated with phase maps P_k computed from themselves. Following a similar training scheme as ϵ_θ^{MP}, with I_k and P_k as the generated target and control signal, respectively, the loss of ϵ_θ^{PI} is calculated as follows:

$$\mathcal{L}_{diff-PI} = \frac{1}{K} \sum_{k=1}^{K} \|\epsilon_k - \epsilon_\theta^{PI}(\tilde{I}_k, P_k, c_k, t_k)\|_2^2 \tag{3}$$

3.3 Modality-aware Automated Label Generation

Based on the trained UniVessel-img, UniVessel-label generates the vessel mask serving as the control mask for ϵ_θ^{MP} to generate the corresponding phase map which then guides ϵ_θ^{PI} to synthesize the paired image. Specifically, regarding the generation process of UniVessel-label, as shown in Fig. 2 B(a), an initial vessel graph is first simulated following L-system and Murray's law [1,23], and then cropped according to a modality-specific FOV. As Murray's law governs vessel diameter during vessel branching and growing, we accordingly determine the inheritance relationships between nodes, enabling flexible control over branching hierarchy. To produce naturally curved vessels, we determine spline control points P_B and P_C along the directions N_A to N_C and N_D to N_B, respectively, ensuring that large curvature does not occur near the nodes. The remaining interpolation points, such as J_i, are randomly distributed around the line connecting N_B and N_C. Finally, the diameter distributions of large and small vessels are controlled using a power function with parameters γ, β and k, and vessels are extracted within device-specific regions. Small random perturbations are added

to these parameters to account for intra-modality variations. Overall, UniVessel-label flexibly generates modality-specific labels by adapting branching levels to imaging depths and tuning the FOV and tortuosity alongside parameters (γ, β and k) according to the target modality.

3.4 Supporting fundus image analysis tasks

Based on the image–label pairs generated by UniVessel following the inference process in Fig. 2 C, a wide range of fundus image analysis tasks can be performed in a zero-shot manner without supervision from real target image-label pairs. Here, we demonstrate four representative paradigms, as shown in Fig. 1 C-F:

1. Vessel segmentation: networks are trained on generated images, using the corresponding generated masks as ground truth.

2. Bifurcation detection: networks are trained on generated images, supervised by the corresponding generated bifurcation coordinates.

3. Registration (sampled A): networks take pairs of generated images and their A-warped counterparts as inputs, supervised by numerous point correspondences sampled via the affine matrix A.

4. Registration (paired bifurcations): networks are trained on the same warped image pairs, guided by anatomically meaningful bifurcations as supervision.

4 Experiments

Implementing details and Datasets: UniVessel is implemented in PyTorch on NVIDIA L40s and trained for 1000 epochs using the Adam optimizer with an initial learning rate of 0.0001, which is gradually decayed. The minibatch size is set to 6, and the input image resolution is $1 \times 512 \times 512$. UniVessel employs 4000 denoising steps with a linear noise scheduler. P_{drop} is set to 0.7.

For testing the zero-shot ability of UniVessel, we employ images of Laser speckle contrast imaging (LSCI) [17] and ultra-widefield fundus photography (FP) [4] for segmentation. FIRE dataset [8] is used for registration and bifurcation detection. For learning the control relationship, 80 images from FIVES [11] and 80 images from RVD [12] are employed.

Results: we first present images generated by different methods based on UniVessel-label's masks. For DDPM [9] and Rectified Flow [19], images are synthesized under the guidance of public paired data, without using phase maps. As shown in Fig. 3, unpaired generation approaches (CUT[25], CycleGAN[32]) tend to preserve either stylistic or structure consistency. Meanwhile, DDPM and Rectified Flow prioritize appearance and do not always adhere to the control masks. In contrast, benefiting from the decoupled design, UniVessel-img can follow the control masks to synthesize realistic images.

For **segmentation**, we compare UniVessel with 1-shot segmentation methods (IFA[24], HSNet[22], PATNet[16]) and generative methods which synthesize datasets to train a UNet for segmentation, and 150 vessel masks generated by

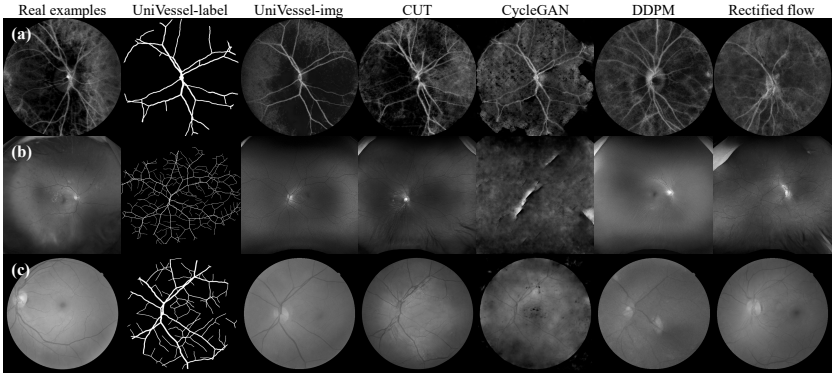

Fig. 3. Images generated from masks of UniVessel-label:(a)LSCI, (b)FP, (c)FIRE.

Table 1. Comparisons of segmentation on LSCI and FP with values in percentage (%).

	LSCI				FP			
Methods	Dice↑	Jaccard↑	Kappa↑	HD95↓	Dice↑	Jaccard↑	Kappa↑	HD95↓
IFA (1-shot)	8.15	4.26	0.49	41.18	52.34	34.07	48.16	36.10
HSNet (1-shot)	7.09	3.70	0.37	38.22	57.04	40.23	55.36	32.21
PATNet (1-shot)	3.99	2.05	1.52	41.65	55.34	38.57	53.72	36.89
CUT	48.45	32.55	45.63	44.40	15.15	8.24	9.27	74.13
CycleGAN	59.11	42.64	55.66	33.79	33.62	20.31	30.72	29.13
Rectified Flow	19.65	10.94	7.32	55.37	34.81	21.61	33.39	59.19
DDPM	22.66	12.85	12.05	46.44	37.86	23.55	35.01	24.43
UniVessel(ours)	**68.49**	**52.30**	**65.91**	**26.66**	**59.89**	**42.99**	**58.36**	**17.97**

UniVessel-label are used as control signals and supervision. As shown in Table. 1, UniVessel achieves the best performance across all metrics on both datasets, even outperforming the 1-shot methods. Compared with other generative approaches, UniVessel produces structurally consistent and realistic images even under large domain gaps, which in turn leads to improved segmentation performance.

For **registration**, we compare UniVessel with unsupervised (GDP-ICP [30], DASC [13]), cross-domain zero-shot (MIFNet [20]), and generation methods which synthesize datasets to train XoFTR [29](Fig. 1 E). We further evaluate UniVessel using SuperRetina [18](Fig. 1 F). MAE and RMSE of the distances between paired annotated points after registration are computed, at the resolution of 1024×1024. MedianE and MaxE denote the median and maximum distances, respectively, while SR indicates the rate of successfully estimated transformation matrices. As shown in Table. 2, UniVessel achieves the best performance, demonstrating its superiority and multi-task adaptability.

For **bifurcation detection**, the generated images are used to train a UNet for detection. Since not all bifurcations in FIRE dataset are labeled, recall on the annotated ones is used as the primary metric. A ground-truth point is considered detected if a predicted point lies within a predefined distance which is set to 10

Table 2. Comparisons of registration on FIRE

Methods	MAE↓	RMSE↓	MedianE↓	MaxE↓	SR↑
GDP-ICP	14.08	48.66	2.15	289.37	97%
DASC	113.02	181.69	23.23	641.52	100%
MIFNet	4.31	6.84	2.38	63.71	100%
$\text{CUT}_{\text{XoFTR}}$	93.17	169.32	9.09	642.00	100%
$\text{CycleGAN}_{\text{XoFTR}}$	91.77	202.30	9.86	1754.48	100%
$\text{Rectified Flow}_{\text{XoFTR}}$	27.23	114.71	5.19	1053.29	100%
$\text{DDPM}_{\text{XoFTR}}$	112.96	181.66	23.37	640.00	100%
$\text{UniVessel}_{\text{SuperRetina}}$(ours)	9.96	20.79	3.70	174.79	100%
$\text{UniVessel}_{\text{XoFTR}}$(ours)	**3.18**	**4.99**	**1.77**	**60.18**	**100%**

Table 3. Comparisons of bifurcation detection on FIRE.

Methods	Recall↑ (10pix)	MAE↓ (10pix, R0.5)	RMSE↓ (10pix, R0.5)	Recall↑ (5pix)	MAE↓ (5pix, R0.5)	RMSE↓ (5pix, R0.5)
CUT	60.60	7.12	7.73	19.85	-	-
CycleGAN	65.22	5.87	6.58	32.76	-	-
Rectified Flow	70.90	6.23	6.93	33.73	-	-
DDPM	61.71	6.11	6.91	28.06	-	-
UniVessel(ours)	**94.93**	**5.60**	**6.29**	**52.84**	**3.49**	**3.73**

Table 4. Ablations on segmentation(LSCI), registration(FIRE) and detection(FIRE)

Settings			Segmentation		Registration			Detection(10pix)		
Phase	Canny	Drop	Dice↑	HD95↓	MAE↓	RMSE↓	SR↑	Recall↑	MAE↓	RMSE↓
	✓		21.94	156.36	106.86	180.32	39%	86.72	6.46	7.16
	✓	✓	27.33	76.80	104.99	179.56	44%	85.45	6.07	6.77
✓			59.52	33.12	3.29	5.41	100%	90.89	6.50	7.28
✓		✓	**68.49**	**26.66**	**3.18**	**4.99**	**100%**	**94.93**	**5.60**	**6.29**

and 5 pixels. For fair comparisons, we select the predicted points with the highest confidence such that the recall reaches 0.5, and compute their MAE and RMSE. Again, UniVessel achieves the best performance in Table. 3, and under the more stringent 5-pixel threshold, it is the only method whose recall exceeds 0.5.

As shown in Table 4, **ablation** studies on segmentation, registration, and detection demonstrate the effectiveness of the key designs in our method. By using phase maps as a modality bridge to disentangle structure from style, UniVessel-img generates realistic and structurally controllable images. Furthermore, with the dropping mechanism, UniVessel-img prevents the generated results from biasing towards the public data. Collectively, these designs enable UniVessel to achieve the best performance.

5 Conclusion

In this work, we propose a novel generation framework that simultaneously produces multiple types of annotations along with corresponding realistic images. Our method requires only guidance from a subset of public datasets to generate images with large inter-modality variations. Experimental results on segmentation, registration, and detection demonstrate the effectiveness of our approach.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (82371112, 62501020); in part by the Science Foundation of Peking University Cancer Hospital (JC202505); in part by the China National Postdoctoral Program for Innovative Talents (BX20250368); in part by the Peking University Medicine plus X Pilot Program-Artificial Intelligence and Medical Development Initiative.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Brown, E.E., Guy, A.A., Holroyd, N.A., Sweeney, P.W., Gourmet, L., Coleman, H., Walsh, C., Markaki, A.E., Shipley, R., Rajendram, R., et al.: Physics-informed deep generative learning for quantitative assessment of the retina. Nature Communications **15**(1), 6859 (2024)
2. Chen, H., Shi, Y., Bo, B., Zhao, D., Miao, P., Tong, S., Wang, C.: Real-time cerebral vessel segmentation in laser speckle contrast image based on unsupervised domain adaptation. Frontiers in Neuroscience **15**, 755198 (2021)
3. Costa, P., Galdran, A., Meyer, M.I., Niemeijer, M., Abràmoff, M., Mendonça, A.M., Campilho, A.: End-to-end adversarial retinal image synthesis. IEEE transactions on medical imaging **37**(3), 781–791 (2017)
4. Ding, L., Kuriyan, A.E., Ramchandran, R.S., Wykoff, C.C., Sharma, G.: Prime-fp20: Ultra-widefield fundus photography vessel segmentation dataset (2020)
5. Fu, S., Xu, J., Chang, S., Yang, L., Ling, S., Cai, J., Chen, J., Yuan, J., Cai, Y., Zhang, B., et al.: Robust vascular segmentation for raw complex images of laser speckle contrast based on weakly supervised learning. IEEE Transactions on Medical Imaging **43**(1), 39–50 (2023)
6. Go, S., Ji, Y., Park, S.J., Lee, S.: Generation of structurally realistic retinal fundus images with diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2335–2344 (2024)
7. He, Y., Ge, R., Tang, H., Liu, Y., Su, M., Coatrieux, J.L., Shu, H., Chen, Y., He, Y.: Conditional virtual imaging for few-shot vascular image segmentation. IEEE Transactions on Medical Imaging (2025)
8. Hernandez-Matas, C., Zabulis, X., Triantafyllou, A., Anyfanti, P., Douma, S., Argyros, A.A.: Fire: fundus image registration dataset. Artificial Intelligence in Vision and Ophthalmology **1**(4), 16–28 (2017)
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)

10. Hou, J., Ding, X., Deng, J.D.: Semi-supervised semantic segmentation of vessel images using leaking perturbations. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 2625–2634 (2022)
11. Jin, K., Huang, X., Zhou, J., Li, Y., Yan, Y., Sun, Y., Zhang, Q., Wang, Y., Ye, J.: Fives: A fundus image dataset for artificial intelligence based vessel segmentation. Scientific data **9**(1), 475 (2022)
12. Khan, M.W., Sheng, H., Zhang, H., Du, H., Wang, S., Coroneo, M., Hajati, F., Shariflou, S., Kalloniatis, M., Phu, J., et al.: Rvd: a handheld device-based fundus video dataset for retinal vessel segmentation. Advances in Neural Information Processing Systems **36**, 18203–18224 (2023)
13. Kim, S., Min, D., Ham, B., Do, M.N., Sohn, K.: Dasc: Robust dense descriptor for multi-modal and multi-spectral correspondence estimation. IEEE transactions on pattern analysis and machine intelligence **39**(9), 1712–1729 (2016)
14. Kovesi, P., et al.: Image features from phase congruency. Videre: Journal of computer vision research **1**(3), 1–26 (1999)
15. Kreitner, L., Paetzold, J.C., Rauch, N., Chen, C., Hagag, A.M., Fayed, A.E., Sivaprasad, S., Rausch, S., Weichsel, J., Menze, B.H., et al.: Synthetic optical coherence tomography angiographs for detailed retinal vessel segmentation without human annotations. IEEE Transactions on Medical Imaging **43**(6), 2061–2073 (2024)
16. Lei, S., Zhang, X., He, J., Chen, F., Du, B., Lu, C.T.: Cross-domain few-shot semantic segmentation. In: European conference on computer vision. pp. 73–90. Springer (2022)
17. Li, K., He, H., Zeng, S., Zhang, X., Li, Y., Zhu, L., Lu, Y.: Points-supervised fundus vessel segmentation via shape priors and contrastive learning. IEEE Transactions on Medical Imaging (2025)
18. Liu, J., Li, X., Wei, Q., Xu, J., Ding, D.: Semi-supervised keypoint detector and descriptor for retinal image matching. In: European conference on computer vision. pp. 593–609. Springer (2022)
19. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. arXiv preprint arXiv:2209.03003 (2022)
20. Liu, Y., Sun, Z., Yu, B., Zhao, Y., Du, B., Xu, Y., Cheng, J.: Mifnet: Learning modality-invariant features for generalizable multimodal image matching. IEEE Transactions on Image Processing (2025)
21. Menten, M.J., Paetzold, J.C., Dima, A., Menze, B.H., Knier, B., Rueckert, D.: Physiology-based simulation of the retinal vasculature enables annotation-free segmentation of oct angiographs. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 330–340. Springer (2022)
22. Min, J., Kang, D., Cho, M.: Hypercorrelation squeeze for few-shot segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 6941–6952 (2021)
23. Murray, C.D.: The physiological principle of minimum work: I. the vascular system and the cost of blood volume. Proceedings of the National Academy of Sciences **12**(3), 207–214 (1926)
24. Nie, J., Xing, Y., Zhang, G., Yan, P., Xiao, A., Tan, Y.P., Kot, A.C., Lu, S.: Cross-domain few-shot segmentation via iterative support-query correspondence mining. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3380–3390 (2024)
25. Park, T., Efros, A.A., Zhang, R., Zhu, J.Y.: Contrastive learning for unpaired image-to-image translation. In: European conference on computer vision. pp. 319–345. Springer (2020)

26. Rauch, N., Harders, M.: Interactive synthesis of 3d geometries of blood vessels. In: Eurographics (Short Papers). pp. 13–16 (2021)
27. Roy, N.D., Biswas, A.: Detection of bifurcation angles in a retinal fundus image. In: 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR). pp. 1–6. IEEE (2015)
28. Staal, J., Abràmoff, M.D., Niemeijer, M., Viergever, M.A., Van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE transactions on medical imaging **23**(4), 501–509 (2004)
29. Tuzcuoğlu, Ö., Köksal, A., Sofu, B., Kalkan, S., Alatan, A.A.: Xoftr: Cross-modal feature matching transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4275–4286 (2024)
30. Yang, G., Stewart, C.V., Sofka, M., Tsai, C.L.: Alignment of challenging image pairs: Refinement and region growing starting from a single keypoint correspondence. IEEE Trans. Pattern Anal. Machine Intell **23**(11), 1973–1989 (2007)
31. Zhang, L., Jindal, B., Alaa, A., Weinreb, R., Wilson, D., Segal, E., Zou, J., Xie, P.: Generative ai enables medical image segmentation in ultra low-data regimes. Nature Communications **16**(1), 6486 (2025)
32. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)