

Diffusion-Enhanced Transformation Consistency Learning for Retinal Image Segmentation

Xiang Li¹, Huihui Fang^{3,4}, Mingsi Liu⁵, Yanwu Xu^{2,4}, and Lixin Duan¹

¹ Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen, China

lxduan@uestc.edu.cn

² School of Future Technology, South China University of Technology, Guangzhou, China

³ College of Computing and Data Science, Nanyang Technological University, Singapore.

⁴ Pazhou Laboratory, Guangzhou, China

fanghuihuibit@163.com

⁵ Assumption University of Thailand, Bangkok, Thailand

Abstract. Retinal image segmentation plays a critical role in rapid disease detection and early detection, such as assisting in the observation of abnormal structures and structural quantification. However, acquiring semantic segmentation labels is both expensive and time-consuming. To improve label utilization efficiency in semantic segmentation models, we propose Diffusion-Enhanced Transformation Consistency Learning (termed as DiffTCL), a semi-supervised segmentation approach. Initially, the model undergoes self-supervised diffusion pre-training, establishing a reasonable initial model to improve the accuracy of early pseudo-labels in the subsequent consistency training, thereby preventing error accumulation. Furthermore, we developed a Transformation Consistency Learning (TCL) method for retinal images, effectively utilizing unlabeled data. In TCL, the prediction of image affine transformations acts as supervision for both image elastic transformations and pixel-level transformations. We carry out evaluations on the REFUGE2 and MS datasets, involving the segmentation of two modalities: optic disc/cup segmentation in color fundus photography, and layer segmentation in optical coherence tomography. The results for both tasks demonstrate that DiffTCL achieves relative improvements of 5.0% and 2.3%, respectively, over other state-of-the-art semi-supervised methods. The code is available at: <https://github.com/lixiang007666/DiffTCL>.

Keywords: Retinal Image Segmentation · Diffusion · Transformation Consistency Learning · Semi-Supervised.

1 Introduction

Deep learning has made significant progress in image segmentation [19,22,6], but traditional deep learning segmentation models still require a large number

of segmentation labels for training. This challenge is particularly acute in the processing of retinal images, such as color fundus photography (CFP) [5] and optical coherence tomography (OCT) images [12]. Manually annotating these complex-structured retinal images requires highly specialized medical knowledge and a significant amount of time. Therefore, it is crucial to improve the efficiency of label utilization.

Semi-supervised learning methods can achieve segmentation using a small number of labels. Mean Teacher (MT) [20] enhances temporal ensembling by averaging model weights, enhancing accuracy, and enabling efficient training with limited labeled data. Uncertainty-Aware Mean Teacher (UA-MT) [24] improves MT by incorporating uncertainty awareness, specifically targeting left atrium segmentation. Li et al. [11] introduced SASSNet, which tackles medical image segmentation by enforcing a geometric shape constraint, leading to improved shape estimation. Chen et al. [4] presented CPS, implementing consistency regularization and achieving state-of-the-art (SOTA) performance in semantic segmentation. Yang et al. [23] proposed UniMatch, redefining consistency frameworks and expanding perturbation strategies, thereby outperforming existing methods on multiple benchmarks. These methods provide innovative approaches to semi-supervised learning and medical image segmentation, emphasizing consistency and uncertainty while introducing novel perturbation techniques.

Semi-supervised learning has been widely applied in medical image segmentation tasks, effectively addressing the challenge of limited labeling in retinal image data [18,12]. Notably, pre-training the model can enhance the representation of image features, thereby enhancing the model’s performance on datasets with limited labels [16]. Pre-training not only helps the model avoid overfitting, particularly in scenarios with limited labels, but also significantly accelerates the model’s convergence [1]. However, a frequently overlooked issue is that many semi-supervised learning models lack an excellent pre-training step.

To further enhance label utilization efficiency and address insufficient pre-training, we propose a Diffusion-Enhanced Transformation Consistency Learning (DiffTCL) for retinal images. The main contributions are as follows: 1) For the semi-supervised setting of retinal image segmentation, we propose for the first time Transformation Consistency Learning (TCL). The prediction of image affine transformations serves as supervision for both image elastic transformations and pixel-level transformations, efficiently utilizing unlabeled data; 2) We design self-supervised diffusion pre-training, which provides a reasonable initial model, improves the accuracy of early pseudo-labels in TCL, and avoids error accumulation; 3) DiffTCL has been applied to datasets of two modalities, CFP and OCT, showing relative performance improvements of 5.0% and 2.3%, respectively, compared to other SOTA methods.

2 Proposed method

The implementation of DiffTCL is illustrated in **Fig. 1**, consisting of two main steps. The first step involves self-supervised diffusion pre-training of the entire

network. The second step employs the proposed Transformation Consistency Learning (TCL) method, using the prediction of image affine transformation (AT) as supervision for both image elastic transformation (ET) and pixel-level transformation (PT) [21], thereby effectively utilizing unlabeled data. Additionally, it conducts supervised training in parallel with a limited amount of labeled data. In TCL, to ensure consistency with PT, inverse transformations are applied to the outputs of the AT and ET branches. Furthermore, we refine the pseudo-labels generated by the AT branch to improve semi-supervised learning performance. DiffTCL enhances the model’s understanding of the intrinsic properties of images, such as shape, texture, and edge information.

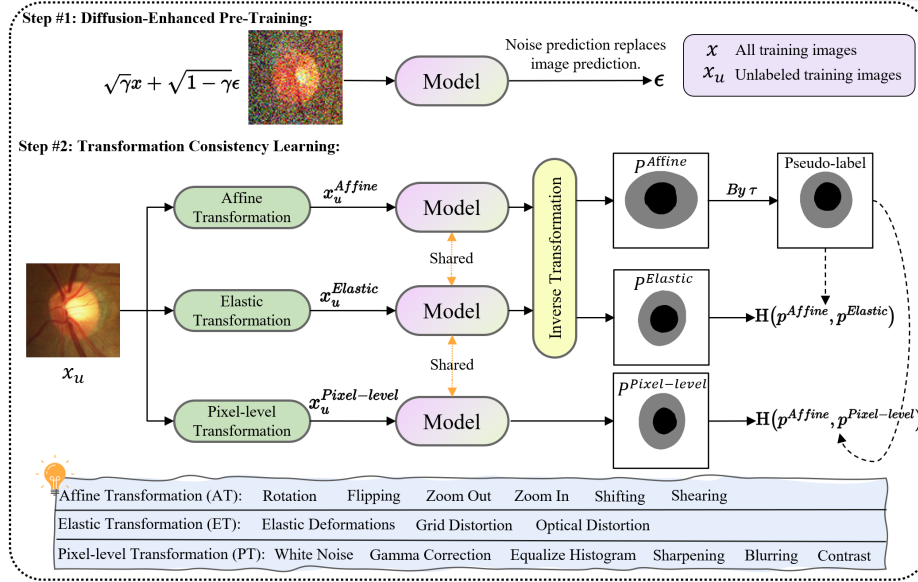


Fig. 1. Overview of our DiffTCL.

2.1 Network architecture

DiffTCL is not limited to a specific architecture. We use DeepLabV3+ as the base network [3], with the widely used ResNet50 serving as its encoder. The initial weights of ResNet50 are trained on ImageNet-1k [17]. Compared to DeepLabV3, DeepLabV3+ incorporates an additional decoder module that enhances the integration of low-level and high-level features, thereby improving the accuracy of segmentation boundaries. Furthermore, the included Atrous Spatial Pyramid Pooling (ASPP) module [3] solves the challenge of multi-scale representation.

2.2 Diffusion-Enhanced pre-training

Recently, there has been a resurgence of interest in denoising autoencoders [9], particularly with Denoising Diffusion Probabilistic Models (DDPMs) [8] gaining prominence. The primary distinction between DDPMs and traditional denoising autoencoders is that DDPMs are specifically trained to remove Gaussian noise introduced into images, with this noise originating from a Gaussian distribution of varying variances. Conversely, traditional denoising autoencoders are generally designed to eliminate Gaussian noise of constant variance. Additionally, DDPMs focus on training an autoencoder to predict noise patterns instead of directly reconstructing the clean image.

Inspired by DDPMs, we investigated the effectiveness of representations learned through diffusion pre-training for retinal image segmentation tasks. Given an input image $\mathbf{x}$ and a scalar noise level γ , a noisy image $\tilde{\mathbf{x}}$ is generated by adding Gaussian noise ϵ :

$$\tilde{\mathbf{x}} = \sqrt{\gamma}\mathbf{x} + \sqrt{1-\gamma}\epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (1)$$

where $\mathbf{x}$ is attenuated by $\sqrt{\gamma}$, and ϵ is attenuated by $\sqrt{1-\gamma}$ to ensure that the random variable variance of $\tilde{\mathbf{x}}$ is 1 if the variance of $\mathbf{x}$ is 1. DDPMs simulate the entire diffusion process from a clear image to pure noise (and its reverse) by uniformly and randomly sampling the noise intensity γ from the range $[0, 1]$. Based on this, we have developed a diffusion pre-training objective function to learn high-quality representations:

$$\mathbb{E}_{\mathbf{x}} \mathbb{E}_{\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \mathbb{E}_{\gamma \sim p(\gamma)} \left\| f_{\theta}(\sqrt{\gamma}\mathbf{x} + \sqrt{1-\gamma}\epsilon) - \epsilon \right\|_2^2, \quad (2)$$

where f_{θ} denotes an image-to-image translation architecture such as DeepLabV3+, and $p(\gamma)$ defines the noise schedule for DDPMs. The loss function represents a single iteration of the diffusion process as modeled by DDPMs. Optimizing the diffusion pre-training’s objective function yields an initial model that improves the accuracy of early pseudo-labels in subsequent consistency training, thereby preventing error accumulation.

2.3 Transformation Consistency Learning (TCL)

TCL thoroughly investigates the transformations of original images, enhancing the model’s understanding of the intrinsic properties of images and ensuring the complete utilization of unlabeled data. This section corresponds to Step 2 in **Fig. 1**. Algorithms in semi-supervised segmentation aim to fully explore unlabeled images $\mathcal{D}_u = \{x_u^i\}$ with a limited amount of annotations from labeled images $\mathcal{D}_l = \{(x_l^i, y_l^i)\}$. For the implementation of TCL, each unlabeled image x_u undergoes three types of transformations simultaneously (AT, ET, and PT) [21], as shown at the bottom of **Fig. 1**. Then, the objective function is a combination of the supervised loss $\mathcal{L}_s$ and the unsupervised consistency loss $\mathcal{L}_u$:

$$\mathcal{L} = \frac{1}{2} (\mathcal{L}_s + \mathcal{L}_u), \quad (3)$$

where the supervised term $\mathcal{L}_s$ is the cross-entropy loss between the model predictions and the ground truths. The consistency loss $\mathcal{L}_u$ ensures that the model predictions under the ET and PT branches are consistent with those of the AT branch, and it can filter out low-confidence samples, as follows:

$$\mathcal{L}_u = \frac{1}{2B_u} \left(\sum \mathbb{1} \left(\max(p^{\text{Affine}}) \geq \tau \right) H(p^{\text{Affine}}, p^{\text{Elastic}}) + \sum \mathbb{1} \left(\max(p^{\text{Affine}}) \geq \tau \right) H(p^{\text{Affine}}, p^{\text{Pixel-level}}) \right), \quad (4)$$

where B_u represents the batch size for unlabeled data, τ is a pre-defined confidence threshold. $\max(p^{\text{Affine}})$ is the highest probability value in this probability distribution, representing the model's confidence in its prediction for a certain class. The indicator function $\mathbb{1}$ outputs 1 if $\max(p^{\text{Affine}}) \geq \tau$, indicating high model confidence, and 0 otherwise. H is the cross-entropy loss, used to minimize the entropy between two probability distributions.

3 Experiments

3.1 Dataset and implementation

REFUGE2 dataset [5]. The REFUGE2 dataset is a publicly available CFP dataset dedicated to glaucoma, featuring segmentation annotations for the optic disc (OD) and optic cup (OC). It contains 2000 CFP images (1200 training images, 400 validation images, and 400 test images.) and was provided by Zhongshan Ophthalmic Center (ZOC) in MICCAI. To balance foreground and background classes, we utilize a U-Net for the coarse segmentation of the OD and OC regions, yielding cropped images of 500×500 pixels. Ultimately, in the training set, the number of labeled images $|\mathcal{D}_l| = 60$, and the number of unlabeled images $|\mathcal{D}_u| = 1140$.

MS dataset [12]. The MS dataset comprises OCT data from 35 subjects, including 14 healthy controls (HC) and 21 multiple sclerosis (MS) patients. Each retinal OCT encompasses 49 B-scans with a resolution of 496×1024 pixels, annotated for 8 retinal layers: RNFL, GCIP, INL, OPL, ONL, IS, OS, and RPE [12]. Each B-scan is divided into two nonoverlapping samples by width, and then center-cropped to a height of 384. This process results in 3430 retinal OCT scans, each 384×512 pixels in size, divided into 2940 training, 196 validation, and 294 test images. For the training, only 5 labeled images from diverse subjects were used, that is, $|\mathcal{D}_l| = 5$ and $|\mathcal{D}_u| = 2935$.

Implementation details. DiffTCL was implemented using torch-2.1.0 and tested on a system equipped with 24Gi GeForce RTX 4090 GPUs (4 cards). To make a fair comparison with prior works, we primarily adopted DeepLabV3+ based on ResNet50 as our segmentation model. In Step 1, the model's output channel size is 3. In Step 2, the output channel size of the model varies depending on the task, being 3 for OD and OC segmentation, and 9 for layer segmentation. During training, each mini-batch consists of 4 labeled and 4 unlabeled images. The initial learning rate was set to 0.001, using an SGD optimizer. Training proceeds for 300 epochs, utilizing a polynomial learning rate scheduler. Regarding the hyperparameter τ , we set it to 0.90.

3.2 Results

The REFUGE2 results. In **Table 1**, we conducted experiments on OC and OD segmentation for CFP, comparing them with other supervised and semi-supervised SOTA methods, respectively. U-Net served as the baseline for comparison. DiffTCL achieved Dice scores of 86.5 for OC segmentation and 95.4 for OD segmentation. Compared to UniMatch, it showed an average Dice improvement of 5.0%. Additionally, visual comparisons for OD and OC segmentation are provided in **Fig. 2**, illustrating that DiffTCL’s segmentation results closely align with the ground truth. This success is attributed to DiffTCL’s effective initial model and full utilization of unlabeled image data.

Table 1. On the REFUGE2 test dataset, OC and OD segmentation comparisons with SOTA methods showed Dice score (%) improvements. Note that for the semi-supervised setting, only 5% labeled samples were used on the training set.

	Method	OC	OD	Mean
Supervised	Baseline	82.3	92.1	87.2
	M-Net [7]	84.9	94.0	89.5
	RBA-Net [14]	85.2	94.5	89.9
Semi-supervised	MT [20]	81.5	90.9	86.2
	UA-MT [24]	82.8	93.6	88.2
	CCT [15]	80.7	92.1	86.4
	SASSNet [11]	82.4	93.7	88.1
	DTC [13]	83.0	93.3	88.2
	CPS [4]	83.3	93.2	88.3
	UniMatch [23]	81.8	91.6	86.7
	DiffTCL (proposed)	86.5	95.4	91.0

The MS results. In **Table 2**, we carried out experiments on OCT layer segmentation, comparing them with other supervised and semi-supervised SOTA methods. DiffTCL achieved the best results in the segmentation of each layer, with an average Dice score of 88.32. Compared to UniMatch, which also employs consistency learning, our method achieved a 2.3% improvement in Dice scores. To further demonstrate the performance of DiffTCL, we provide qualitative comparison results. As shown in **Fig. 3**, there are fewer over-segmented and under-segmented pixels.

Ablation study. We performed an in-depth analysis to evaluate the contribution of each element in our DiffTCL method, as depicted in **Table 3**, through ablation studies on REFUGE2. Performance dipped when any transformation branch within TCL, like ET or PT, was removed. A complete implementation of

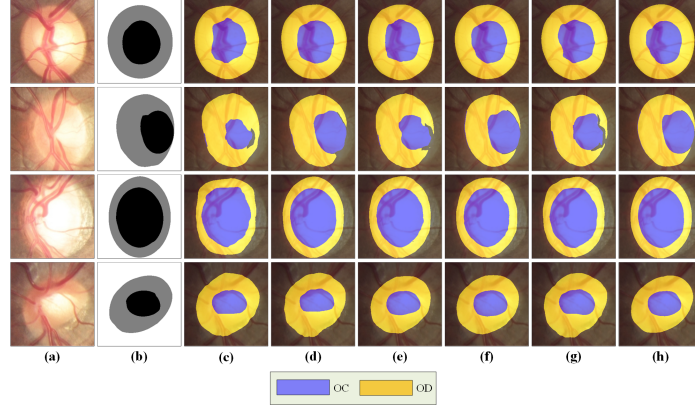


Fig. 2. Conducting a qualitative comparison between DiffTCL and SOTA methods for both OC and OD segmentation on the REFUGE2 test dataset, while employing a training set with only 5% of labeled samples. (a) Original CFP, (b) Ground truth, (c) MT, (d) UA-MT, (e) SASSNet, (f) CPS, (g) UniMatch, (h) DiffTCL.

Table 2. Based on the MS test dataset, the retinal layer segmentation results were compared with SOTA methods, showing an improvement in Dice score (%). Note that only 5 labeled samples were used in the training set.

Method		Retinal layers								
		RNFL	GCIP	INL	OPL	ONL	IS	OS	RPE	Mean
Supervised	Baseline	88.07	88.09	77.65	84.57	90.64	79.21	80.60	82.84	83.96
	MGU-Net [10]	89.50	90.19	79.14	81.80	89.86	76.54	79.83	80.44	83.41
	MTF [2]	89.51	88.10	82.65	85.08	92.15	82.71	82.85	83.83	85.86
Semi-supervised	MT [20]	89.00	89.92	81.22	84.79	90.93	81.26	81.73	84.48	85.42
	UA-MT [24]	89.35	90.88	82.01	86.28	91.42	78.74	80.59	85.22	85.56
	CCT [15]	89.08	89.86	82.12	84.55	91.11	80.54	79.19	80.59	84.63
	SASSNet [11]	89.11	89.94	82.07	84.53	90.95	78.92	80.71	82.93	84.90
	DTC [13]	89.26	90.44	82.87	84.99	90.24	77.18	81.31	85.03	85.16
	CPS [4]	89.25	90.48	81.68	86.16	92.05	82.21	84.67	84.00	86.31
	UniMatch [23]	89.53	90.97	81.19	86.54	92.29	81.95	84.23	83.68	86.30
	DiffTCL (proposed)	90.69	91.14	84.98	88.63	92.98	84.60	86.09	87.42	88.32

TCL reached an mean Dice score of 88.6, thanks to exploiting a variety of original image transformations that improved the model’s understanding of the intrinsic properties of images. Incorporating self-supervised diffusion pre-training into TCL (DiffTCL) not only boosts performance but also speeds up convergence, lowering training costs. Additionally, analyses in **Table 4** and **Table 5** on visible label quantities in REFUGE2 and MS datasets’ training sets demonstrated that, even with a minimal number of visible labels, we achieved Dice improvements of 2.58 and 3.94 times over the supervised method.

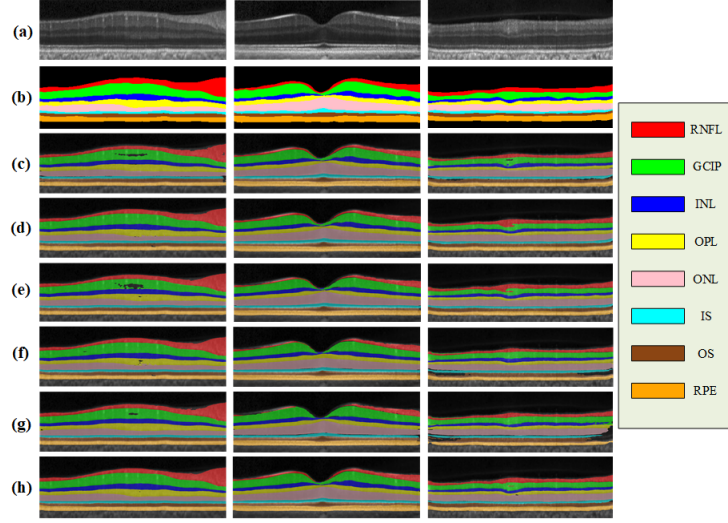


Fig. 3. Qualitative comparison of layer segmentation on the MS test dataset with other methods, with only 5 labeled samples in the training set. (a) Original OCT images, (b) Ground truth, (c) MT, (d) UA-MT, (e) SASSNet, (f) CPS, (g) UniMatch, (h) DiffTCL.

Table 3. The ablation experiment is on the REFUGE2 test dataset.

Method	OC Dice (%)	OD Dice (%)	Mean (%)
TCL w/o ET	67.2	83.4	75.3
TCL w/o PT	50.4	61.5	56.0
TCL	83.3	93.9	88.6
DiffTCL (proposed)	86.5	95.4	91.0

Table 4. Average Dice score (%) comparison with supervised method (DeepLabV3+) on REFUGE2 test dataset, where the percentage indicates the proportion of labeled training images from REFUGE2. Δ measured against DeepLabV3+.

Method	5%	10%	20%	25%	50%
DeepLabV3+ w/ ResNet50	35.3	56.7	62.2	65.5	68.2
DiffTCL (proposed)	91.0	91.2	91.5	91.5	92.0
Gain (Δ)	$2.58\times$	$1.61\times$	$1.47\times$	$1.40\times$	$1.35\times$

4 Conclusions

This study details DiffTCL, a semi-supervised segmentation method for retinal images. DiffTCL’s core components are Diffusion-Enhanced pre-training and Transformation Consistency Learning (TCL). Diffusion-Enhanced pre-training

Table 5. Comparison of Dice scores (%) with supervised methods on the MS test dataset under different numbers of labeled images. The values provided in the table are the average Dice across 8 segmentation layers.

Method	5	10	50	100	1470
DeepLabV3+ w/ ResNet50	22.44	35.33	64.18	71.75	80.08
DiffTCL (proposed)	88.32	89.25	89.77	90.19	91.10
Gain (Δ)	$3.94\times$	$2.53\times$	$1.40\times$	$1.26\times$	$1.14\times$

ensures an optimal initial model, preventing error accumulation. TCL utilizes predictions from image affine transformations as supervision for elastic and pixel-level transformations, efficiently utilizing unlabeled data and enhancing the model’s understanding of the intrinsic properties of images. Experiments on images from two modalities (CFP and OCT) demonstrate DiffTCL’s superior performance over existing SOTA methods, particularly in scenarios with limited labeled data. DiffTCL provides a powerful solution for enhancing the accuracy and efficiency of retinal image segmentation with a small number of labels.

Acknowledgments. This research is supported by Pazhou Laboratory’s basic cloud computing platform.

Disclosure of Interests. The authors declare no competing interests.

References

1. Brempong, E.A., Kornblith, S., Chen, T., Parmar, N., Minderer, M., Norouzi, M.: Denoising pretraining for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 4175–4186 (June 2022)
2. Cao, J., Liu, X., Zhang, Y., Wang, M.: A multi-task framework for topology-guaranteed retinal layer segmentation in oct images. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 3142–3147 (2020). <https://doi.org/10.1109/SMC42975.2020.9283408>
3. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV) (September 2018)
4. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2613–2622 (June 2021)
5. Fang, H., Li, F., Fu, H., Sun, X., Cao, X., Son, J., Yu, S., Zhang, M., Yuan, C., Bian, C., et al.: Refuge2 challenge: Treasure for multi-domain learning in glaucoma assessment. arXiv **abs/2202.08994** (2021)
6. Fang, H., Li, F., Fu, H., Wu, J., Zhang, X., Xu, Y.: Dataset and evaluation algorithm design for goals challenge. In: Antony, B., Fu, H., Lee, C.S., MacGillivray, T., Xu, Y., Zheng, Y. (eds.) Ophthalmic Medical Image Analysis. pp. 135–142. Springer International Publishing, Cham (2022)

7. Fu, H., Cheng, J., Xu, Y., Wong, D.W.K., Liu, J., Cao, X.: Joint optic disc and cup segmentation based on multi-label deep network and polar transformation. *IEEE Transactions on Medical Imaging* **37**(7), 1597–1605 (2018). <https://doi.org/10.1109/TMI.2018.2791488>
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) *Advances in Neural Information Processing Systems*. vol. 33, pp. 6840–6851. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf
9. Larrazabal, A.J., Martínez, C., Glocker, B., Ferrante, E.: Post-dae: Anatomically plausible segmentation via post-processing with denoising autoencoders. *IEEE Transactions on Medical Imaging* **39**(12), 3813–3820 (2020). <https://doi.org/10.1109/TMI.2020.3005297>
10. Li, J., Jin, P., Zhu, J., Zou, H., Xu, X., Tang, M., Zhou, M., Gan, Y., He, J., Ling, Y., Su, Y.: Multi-scale gcn-assisted two-stage network for joint segmentation of retinal layers and discs in peripapillary oct images. *Biomed. Opt. Express* **12**(4), 2204–2220 (Apr 2021). <https://doi.org/10.1364/BOE.417212>, <https://opg.optica.org/boe/abstract.cfm?URI=boe-12-4-2204>
11. Li, S., Zhang, C., He, X.: Shape-aware semi-supervised 3d semantic segmentation for medical images. In: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racocanu, D., Joskowicz, L. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020*. pp. 552–561. Springer International Publishing, Cham (2020)
12. Lu, Y., Shen, Y., Xing, X., Ye, C., Meng, M.Q.H.: Boundary-enhanced semi-supervised retinal layer segmentation in optical coherence tomography images using fewer labels. *Computerized Medical Imaging and Graphics* **105**, 102199 (2023). <https://doi.org/https://doi.org/10.1016/j.compmedimag.2023.102199>, <https://www.sciencedirect.com/science/article/pii/S0895611123000174>
13. Luo, X., Chen, J., Song, T., Wang, G.: Semi-supervised medical image segmentation through dual-task consistency. *Proceedings of the AAAI Conference on Artificial Intelligence* **35**(10), 8801–8809 (May 2021). <https://doi.org/10.1609/aaai.v35i10.17066>, <https://ojs.aaai.org/index.php/AAAI/article/view/17066>
14. Meng, Y., Meng, W., Gao, D., Zhao, Y., Yang, X., Huang, X., Zheng, Y.: Regression of instance boundary by aggregated cnn and gcn. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M. (eds.) *Computer Vision – ECCV 2020*. pp. 190–207. Springer International Publishing, Cham (2020)
15. Ouali, Y., Hudelot, C., Tami, M.: Semi-supervised semantic segmentation with cross-consistency training. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2020)
16. Rodrigues, R., Couto, P.: Semi-supervised learning for ecg classification. In: *2021 Computing in Cardiology (CinC)*. vol. 48, pp. 1–4 (2021). <https://doi.org/10.23919/CinC53138.2021.9662693>
17. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)* **115**(3), 211–252 (2015). <https://doi.org/10.1007/s11263-015-0816-y>
18. Sedai, S., Antony, B., Rai, R., Jones, K., Ishikawa, H., Schuman, J., Gadi, W., Garnavi, R.: Uncertainty guided semi-supervised segmentation of retinal layers in oct images. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. pp. 282–290. Springer International Publishing, Cham (2019)

19. Shen, Y., Li, J., Zhu, W., Yu, K., Wang, M., Peng, Y., Zhou, Y., Guan, L., Chen, X.: Graph attention u-net for retinal layer surface detection and choroid neovascularization segmentation in oct images. *IEEE Transactions on Medical Imaging* pp. 1–1 (2023). <https://doi.org/10.1109/TMI.2023.3240757>
20. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*. vol. 30. Curran Associates, Inc. (2017), https://proceedings.neurips.cc/paper_files/paper/2017/file/68053af2923e00204c3ca7c6a3150cf7-Paper.pdf
21. Uysal, E.S., Bilici, M., Zaza, B.S., Ozgenc, M.Y., Boyar, O.: Exploring the limits of data augmentation for retinal vessel segmentation. *arXiv abs/2105.09365* (2021)
22. Wang, J., Li, X., Cheng, Y.: Towards an extended efficientnet-based u-net framework for joint optic disc and cup segmentation in the fundus image. *Biomedical Signal Processing and Control* **85**, 104906 (2023). <https://doi.org/10.1016/j.bspc.2023.104906>, <https://www.sciencedirect.com/science/article/pii/S1746809423003397>
23. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y.: Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 7236–7246 (June 2023)
24. Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A.: Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation. In: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. pp. 605–613. Springer International Publishing, Cham (2019)