DISTRIBUTION ALIGNMENT INFORMED THRESHOLDING FOR SEMI-SUPERVISED CURVILINEAR STRUCTURE SEGMENTATION

Yuhao Mo[1,2], *Po Peng*[2], *Bihan Wen*[3], *Xulei Yang*[4], *Ce Zhu*[5], *Xun Xu*[4,2†]

1. China Telecom Sichuan Branch, China 2. Southwest Jiaotong University, China 3. Nanyang Technological University, Singapore
4. Institute for Infocomm Research (I2R), A*STAR, Singapore 5. University of Electronic Science and Technology of China, China

ABSTRACT

Curvilinear structure segmentation using deep neural networks is often limited by the high cost of annotation. Semi-supervised learning (SSL) helps mitigate this dependency on extensive annotated data. State-of-the-art SSL approaches generate pseudo-labels for unlabeled data, which are then used for further model training. These methods primarily focus on calibrating thresholds to binarize the predictions. In this work, we assume that when labeled and unlabeled data are similar, the foreground-to-background ratio should be consistent between them. To leverage this assumption, we calibrate the threshold by minimizing the distribution gap between labeled ground truth and pseudo-labels on unlabeled data. Our proposed threshold calibration can be integrated with existing SSL methods. We evaluate its effectiveness on four datasets, demonstrating that our method outperforms current state-of-the-art SSL techniques, especially in scenarios with very low labeled data.

Index Terms— Test-Time Adaptation, Missing Feature Channel, Self-Training, Data Imputation

Segmenting curvilinear structures has wide applications in areas such as defect identification and medical diagnosis. State-of-the-art approaches often employ deep learning paradigms, which require large amounts of labeled data for training. However, collecting labeled data can be expensive and may require specialized expertise, such as in the case of medical images. Semi-supervised learning (SSL) aims to address the challenge of high annotation costs by leveraging inexpensive, unlabeled data. Prevailing SSL methods often adopt a self-training paradigm, generating pseudo-labels for unlabeled data and using these pseudo-labels to supervise model updates [1] [2] [3]. Since incorrect pseudo-labels can significantly harm the model's generalization [4] [5], careful calibration of pseudo-labels is crucial for the success of SSL. Built upon the assumption that pseudo-label accuracy correlates with model confidence (e.g., the highest probability of the posterior), a fixed threshold is often employed to filter out less confident unlabeled data [6]. More recent approaches focus on dynamically adjusting both local and global thresholds to account for the increasing model confidence during training [7] [8].

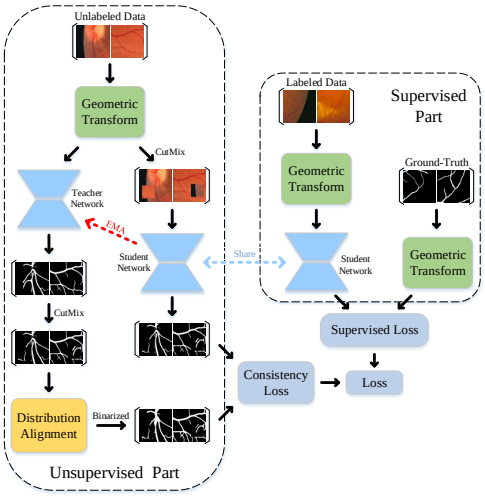

Fig. 1. Overview of the proposed distribution alignment informed thresholding for semi-supervised segmentation.

Contrary to the SSL methods developed for classification tasks, where class labels are often considered balanced, curvilinear segmentation experiences a severe class imbalance between foreground and background pixels. Directly employing existing SSL methods [9] yields suboptimal performance. Motivated by the common class imbalance in SSL, recent attempts have proposed resampling the unlabeled samples to achieve a balanced labeled/pseudo-labeled set for training [10]. However, the resampling approach is less suitable for segmentation tasks as repeating pixels is impractical. Instead of resampling, class-imbalanced SSL via reweighting offers more flexibility and has demonstrated success [11] [12] [13]. However, deriving class-specific weights can be unintuitive and may introduce additional hyper-parameters for tuning.

In this work, we build on the assumption that the ratios between foreground and background pixels are mostly consistent between the labeled and unlabeled data. Under this assumption, we propose refining the pseudo-label predictions on the unlabeled data to minimize the foreground-background distribution mismatch between labeled and unlabeled pseudo-labels. Specifically, we introduce a KL-Divergence measure to quantify the distribution mismatch, and the optimal

† Correspondence to: Xun Xu <alex.xun.xu@gmail.com>. This work was supported in part by Sichuan Science and Technology Program under Project 2023NSFSC1421, the Agency for Science, Technology and Research, Singapore (A*STAR) under Grant M23L7b0021 and in part by the National Natural Science Foundation of China (NSFC) under Grant 62106078.

threshold is derived by minimizing this KL-Divergence. We demonstrate the effectiveness of the proposed dynamic thresholding approach across multiple semi-supervised curvilinear structure segmentation tasks, achieving state-of-the-art performance. In summary, we propose an adaptive thresholding method for effective semi-supervised curvilinear structure segmentation, as illustrated in Fig. 1. The threshold is determined by aligning the distributions between labeled and unlabeled data. Extensive results on multiple curvilinear structure segmentation tasks demonstrate the effectiveness of the proposed method.

1. METHODOLOGY

1.1. Overview of Semi-Supervised Segmentation

We begin with an overview of the semi-supervised image segmentation framework. We define the labeled training samples as $\mathcal{D}_l = \{(x_i, y_i) \mid x_i \in \mathbb{R}^{H\times W\times 3}, y_i \in \{0,1\}^{H\times W}\}_{i=1\ldots N_l}$, the unlabeled training samples as $\mathcal{D}_u = \{x_j\}_{j=1\ldots N_u}$, and denote the segmentation model as $f(x;\theta) \in [0,1]^{H\times W}$. The objective is to train the segmentation model on the combined labeled and unlabeled data $\mathcal{D}_l \cup \mathcal{D}_u$. We adopt a highly effective semi-supervised practice known as self-training [6]. This approach involves two learning paths: for labeled data $\mathcal{D}_l$, we define a supervised segmentation loss $\mathcal{L}_l(x_i, y_i)$, typically instantiated as cross-entropy or dice loss [14]. For unlabeled data $\mathcal{D}_u$, we apply a geometric transformation $g(\cdot)$, followed by a cutmix augmentation $c(x_i, x_j)$ [15]. Cutmix randomly crops a patch from image x_j and pastes it onto x_i. This augmentation can be applied to both input raw images and segmentation masks. We maintain two networks: a teacher network $\hat{\theta}$, which is an exponential moving average of the student network parameters θ. We introduce a threshold τ to binarize the teacher model's predictions $\hat{y}_i$ which are used to supervise the student network as $\mathcal{L}_u = \frac{1}{|\mathcal{D}_u|}\sum_{x_i\in\mathcal{D}_u}\mathcal{L}_l(x_i, \mathbb{1}(\hat{y}_i \geq \tau))$. The threshold τ is typically treated as a hyper-parameter and may require additional labeled data for rigorous tuning.

1.2. Distribution Alignment for Thresholding

Determining the optimal threshold without referring to a separate validation set is non-trivial. In this work, we assume that the label distribution between labeled and unlabeled data is roughly identical, which provides a principled way to estimate the threshold. This identical distribution assumption leads us to explore the optimal threshold as the one that minimizes the discrepancy between the prior distribution on labeled data and the pseudo-labels on the unlabeled data. We denote the labeled prior distribution Q and pseudo-label distribution P subject to threshold τ in Eq. 1.

$$\begin{aligned} Q &= [\frac{\sum_{ihw}\mathbb{1}(y_{ihw}=0)}{C}, \frac{\sum_{ihw}\mathbb{1}(y_{ihw}=1)}{C}], \\ s.t.\ C &= \sum\mathbb{1}(y_{ihw}=0) + \sum\mathbb{1}(y_{ihw}=1) \end{aligned} \tag{1}$$

$$\begin{aligned} P &= [\frac{\sum_{ihw}\mathbb{1}(\hat{y}_{ihw}<\tau)}{\hat{C}}, \frac{\sum_{ihw}\mathbb{1}(\hat{y}_{ihw}\geq\tau)}{\hat{C}}], \\ s.t.\ \hat{C} &= \sum\mathbb{1}(\hat{y}_{ihw}<\tau) + \sum\mathbb{1}(y_{ihw}\geq\tau) \end{aligned} \tag{2}$$

The optimal threshold τ^* is thus obtained by minimizing the KL-Divergence between P and Q as follows.

$$\tau^* = \arg\min_{\tau\in[0,1]} KL(Q||P(\tau)) \tag{3}$$

The KL-Divergence in Eq. 3 does not have a gradient w.r.t. τ, prohibiting us from optimizing τ in a gradient based method. However, since Q is fixed and P can be efficiently calculated given a τ, we choose to discretize the search space of τ and implement an exhaustive search for τ with a increment of 0.05.

1.3. Incremental Distribution Estimation

The labeled prior distribution Q can be easily estimated from the labeled dataset $\mathcal{D}_l$ in an offline manner. The pseudo label distribution P depends on the teacher model's predictions and is subject to weight updates. A naive approach to estimate P would involve running forward passes for all unlabeled data before each training iteration. However, this would introduce extremely high computational overhead for each update step. Therefore, we adopt an incremental distribution estimation following the update rule below, where $\mathcal{B}$ refers to the current minibatch.

$$P(\tau) = \begin{bmatrix} \beta P_0(\tau*) + (1-\beta)\frac{\sum_{x_i\in\mathcal{B}}\sum_{hw}\mathbb{1}(\hat{y}_{ihw}<\tau)}{\hat{C}}, \\ \beta P_1(\tau*) + (1-\beta)\frac{\sum_{x_i\in\mathcal{B}}\sum_{hw}\mathbb{1}(\hat{y}_{ihw}\geq\tau)}{\hat{C}} \end{bmatrix} \tag{4}$$

1.4. Final Training Procedure

The final training procedure follows a two-step alternative update between the thresholding hyper-parameter and semi-supervised learning. In the first step, we estimate the optimal threshold τ^* by Eq. 3. We then update model weights by gradient descent with the combined labelled and unlabeled losses as follows.

$$\mathcal{L}_{ssl} = \frac{1}{|\mathcal{D}_l|}\sum_{x_i,y_i\in\mathcal{D}_l}\mathcal{L}_l(x_i,y_i) + \frac{\lambda}{|\mathcal{D}_u|}\sum_{x_j\in\mathcal{D}_u}\mathcal{L}_l(x_j, \mathbb{1}(\hat{y}_j\geq\tau*)) \tag{5}$$

2. EXPERIMENTS

2.1. Datasets and Experimental Setup

Datasets: **CrackForest** [16] consists of 118 annotated images of road cracks. **EM128** [17] is designed for cell membrane segmentation, containing 30 labeled images divided into 16 regions of 128 × 128 pixels. **DRIVE128** [18] originates from a diabetic retinopathy screening program, featuring images at 584 × 565 pixels divided into 16 regions of 128 × 128 pixels. **STARE128** [19] is a fundus image database for retinal blood vessel segmentation, including 20 images with resolutions of 605 × 700 pixels, each divided into 25 non-overlapping regions of 128 × 128 pixels. The dataset is split into 393/92 images

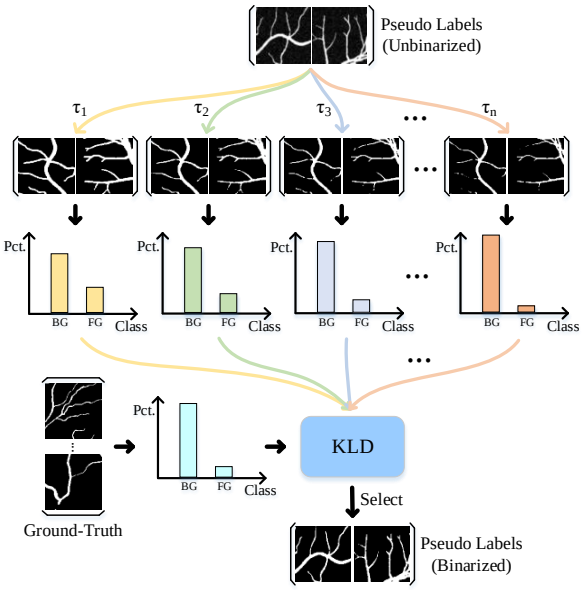

Fig. 2. Illustration of binarizing pseudo predictions $\hat{y}_i$ with different thresholds $\tau_1, \cdots \tau_n$.

for training and testing, resulting in STARE128. For semi-supervised segmentation evaluation, we adopt the data split proposed in [9].

Hyperparameters: Both fully supervised and semi-supervised training use the Adam optimizer with a learning rate of 0.001 and an EMA hyperparameter of 0.999 for teacher model update. We apply a batch size of 32 and 100 training epochs. For incremental distribution estimation, we adopt $\beta = 0.9$. The consistency loss weight is chosen as $\lambda = exp(-10(1 - \frac{e}{100})^2)$ where e refers to current training epoch.

Evaluation Metric: We employ a threshold of 0.5 to binarize the predictions. Subsequently, we calculate the Intersection over Union (IoU) and Dice coefficient by comparing these binarized predictions with the ground-truth. Both metrics formulate segmentation as a binary classification task, with background pixels set to 0 and foreground pixels set to 1. We report the performance as the mean and standard deviation over 5 random runs.

2.2. Competing Methods

Fully Supervised Methods: **FCN** [20] utilizes fully convolutional layers for segmentation model. **SegNet** [21], inspired by autoencoders, achieves accurate pixel-level segmentation using deep convolutional encoders and decoders. **U-Net** [22] introduces a symmetrical U-shaped structure with skip connections to retain global context and local detail information. **ResU-Net** [23] integrates residual blocks from ResNet into U-Net. **BCDU-Net** [24] is a UNet variant with a 3-level encoder-decoder structure that utilizes batch normalization without densely connected convolutions.

Semi-Supervised Methods: **MeanTeacher** [25] incorporates a teacher network that provides smooth supervision signals through exponential moving average (EMA) of student network parameters. **VAT** [26] introduces virtual adversarial perturbations to enhance robustness by maximizing instability near input data. **CPS** [27] enforces consistency between two segmentation networks with different initializations, using one-hot labels as supervision signals for each other. **CutMix** [28] implements data augmentation by cutting and pasting regions of input images, maintaining consistency between teacher and student predictions. **SemiDial** (Ours) proposed method incorporating distribution alignment for semi-supervised segmentation. All semi-supervised methods use BCDU-Net as the backbone.

2.3. Results Analysis

Quantitative Results: We present quantitative comparisons across the five datasets in Tab. 1. From the results, several observations can be made. Firstly, SemiDial consistently outperforms existing semi-supervised segmentation methods at both 1% and 5% labeling budgets. Specifically, the performance gap between SemiDial and the second best method (CutMix) is more pronounced at the 1% labeling budget, highlighting the importance of selecting reliable thresholds for stable SSL performance at extremely low labeling budgets. Furthermore, SemiDial's performance at the 5% labeling budget approaches the upper bound achieved by fully supervised methods, suggesting that SSL offers a promising approach to balancing segmentation performance and labeling efforts.

Qualitative Results: The qualitative experimental results are shown in Fig. 3. As depicted, with 5% labeled data, SemiDial demonstrates cleaner overall prediction results. In more detail, the method accurately identifies more fine cracks and blood vessels without misclassifying them as background. However, like other methods, SemiDial still exhibits limitations in predicting these features consistently across samples. This analysis suggests that the challenge lies in the low contrast between these features and the background, posing difficulty for the model in accurate identification.

2.4. Ablation Study

We investigate the effectiveness of semi-supervised learning (SSL), data augmentation (Data Aug), and distribution alignment (Dist Align) on two datasets at a 1% labeling budget. The ablation study results are presented in Tab. 2. We make the following observations from the results. Firstly, applying the original CutMix model significantly improves upon the fully supervised baseline, demonstrating the effectiveness of consistency-based semi-supervised learning in leveraging unlabeled data to enhance segmentation performance. Secondly, employing strong geometric transformations further enhances segmentation performance. Additionally, when distribution alignment is used to estimate the optimal threshold τ^*, we explore two approaches. Initially, employing the mean square error (MSE) loss for threshold estimation shows improved results over using a fixed threshold of 0.5. However, utilizing KL-Divergence yields superior performance compared to MSE loss for threshold estimation.

Table 1. Evaluation of competing methods for semi-supervised segmentation on curvilinear structure datasets. We report IoU and Dice coefficient as evaluation metrics. All numbers are in %.

		CrackForest		EM128		DRIVE128		STARE128	
		IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice
Fully Supervised	Labeled Pct.	100%		100%		100%		100%	
	FCN [20]	60.50 ± 0.59	75.39 ± 0.46	60.90 ± 0.46	75.70 ± 0.36	42.83 ± 0.17	59.97 ± 0.17	46.46 ± 0.32	63.44 ± 0.30
	SegNet [21]	66.26 ± 1.87	79.70 ± 1.36	57.64 ± 0.52	73.13 ± 0.42	47.33 ± 1.24	64.24 ± 1.14	44.57 ± 4.28	61.58 ± 4.14
	UNet [22]	68.09 ± 0.68	81.01 ± 0.48	66.34 ± 0.17	79.76 ± 0.13	59.87 ± 0.21	74.90 ± 0.16	62.42 ± 0.83	76.87 ± 0.63
	ResU-Net [23]	67.08 ± 0.12	80.30 ± 0.09	67.48 ± 0.07	80.58 ± 0.05	58.76 ± 0.17	74.02 ± 0.13	60.34 ± 0.62	75.26 ± 0.49
	BCDU-Net [24]	70.43 ± 0.25	82.65 ± 0.17	67.51 ± 0.01	80.60 ± 0.01	61.57 ± 0.04	76.21 ± 0.03	64.16 ± 0.03	78.17 ± 0.03
Semi-Supervised	Labeled Pct.	5%		5%		5%		5%	
	MT [25]	64.37 ± 0.77	78.33 ± 0.57	63.92 ± 0.09	77.99 ± 0.06	51.50 ± 0.10	67.99 ± 0.09	57.58 ± 0.08	73.08 ± 0.06
	VAT [26]	56.34 ± 0.43	72.07 ± 0.35	54.89 ± 0.75	70.87 ± 0.63	37.71 ± 0.24	54.77 ± 0.25	40.45 ± 0.54	57.60 ± 0.55
	CPS [27]	59.77 ± 0.01	74.82 ± 0.01	62.44 ± 0.15	76.88 ± 0.11	50.71 ± 0.01	67.29 ± 0.01	47.04 ± 0.07	63.98 ± 0.06
	CutMix [28]	66.51 ± 0.08	79.89 ± 0.06	65.09 ± 0.02	78.85 ± 0.01	53.12 ± 0.03	69.39 ± 0.03	59.15 ± 0.76	74.33 ± 0.59
	SemiDial (Ours)	**66.81 ± 0.11**	**80.10 ± 0.08**	**65.13 ± 0.03**	**78.89 ± 0.03**	**54.20 ± 0.23**	**70.30 ± 0.19**	**59.71 ± 0.41**	**74.77 ± 0.33**
	Labeled Pct.	1%		1%		1%		1%	
	MT [25]	46.27 ± 0.39	63.27 ± 0.36	54.11 ± 0.17	70.22 ± 0.15	40.70 ± 0.98	57.85 ± 0.99	33.26 ± 0.23	49.92 ± 0.26
	VAT [26]	41.95 ± 1.20	59.10 ± 1.20	47.76 ± 0.01	64.65 ± 0.01	32.08 ± 0.51	48.58 ± 0.58	26.58 ± 0.15	42.00 ± 0.19
	CPS [27]	47.99 ± 0.01	64.85 ± 0.01	55.41 ± 0.09	71.31 ± 0.07	41.89 ± 0.08	59.05 ± 0.08	32.12 ± 0.08	48.62 ± 0.09
	CutMix [28]	61.66 ± 0.18	76.28 ± 0.14	56.81 ± 0.12	72.46 ± 0.10	46.34 ± 1.74	63.33 ± 1.63	36.46 ± 0.40	53.44 ± 0.43
	SemiDial (Ours)	**62.69 ± 0.15**	**77.06 ± 0.11**	**57.95 ± 0.15**	**73.38 ± 0.12**	**50.36 ± 0.47**	**66.98 ± 0.41**	**40.49 ± 0.41**	**57.64 ± 0.42**

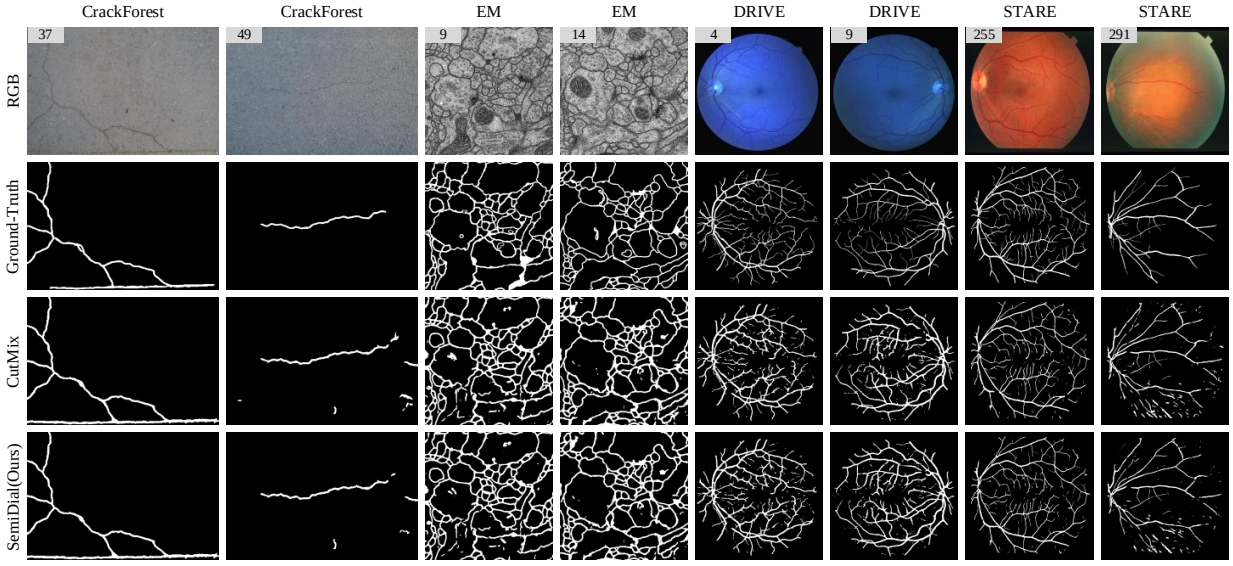

Fig. 3. Qualitative results comparing different methods on selected samples. The number on the top left indicate the image index within each dataset.

Table 2. Ablation study on DRIVE128 and STARE128 datasets with 1% labeling budget.

SSL	DistAlign	DataAug	DRIVE128	STARE128
-	-	-	33.31 ± 0.14	29.80 ± 0.48
CutMix	-	-	43.46 ± 0.73	33.81 ± 2.70
CutMix	-	GeoTForm	46.34 ± 1.74	36.46 ± 0.40
CutMix	MSE	GeoTForm	48.62 ± 1.45	37.36 ± 0.78
CutMix	KLD	GeoTForm	50.36 ± 0.47	40.49 ± 0.41

3. CONCLUSION

In this paper, we propose a novel semi-supervised method for the semantic segmentation of curvilinear structures. Our method mitigates biased and collapsed predictions caused by data imbalance through the distribution alignment strategy. Experimental results across various datasets demonstrate that our approach predicts more detailed features within curvilinear structures and maintains robust performance even with minimal labeled data. Additionally, the distribution alignment strategy is versatile and can be easily integrated into existing semi-supervised image segmentation framework.

4. REFERENCES

[1] D Berthelot, N Carlini, I Goodfellow, N Papernot, A Oliver, and C A Raffel, "Mixmatch: A holistic approach to semi-supervised learning," in *Adv. Neural Inf. Process. Syst.*, 2019.

[2] Q Xie, Z Dai, E Hovy, T Luong, and Q Le, "Unsupervised data augmentation for consistency training," 2020.

[3] Q Xie, M-T Luong, E Hovy, and Q V Le, "Self-training with noisy student improves imagenet classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020.

[4] H Li, Z Wu, A Shrivastava, and L S Davis, "Rethinking pseudo labels for semi-supervised object detection," in *Proc. AAAI Conf. Artif. Intell.*, 2022.

[5] L Yang, W Zhuo, L Qi, Y Shi, and Y Gao, "St++: Make self-training work better for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022.

[6] K Sohn, D Berthelot, N Carlini, Z Zhang, H Zhang, C A Raffel, E D Cubuk, A Kurakin, and C-L Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," in *Adv. Neural Inf. Process. Syst.*, 2020.

[7] Y Wang, H Chen, Q Heng, W Hou, Y Fan, Z Wu, J Wang, M Savvides, T Shinozaki, B Raj, et al., "Freematch: Self-adaptive thresholding for semi-supervised learning," in *Int. Conf. on Learn. Represent.*, 2023.

[8] B Zhang, Y Wang, W Hou, H Wu, J Wang, M Okumura, and T Shinozaki, "Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling," in *Adv. Neural Inf. Process. Syst.*, 2021.

[9] X Xu, M C Nguyen, Y Yazici, K Lu, H Min, and C-S Foo, "Semicurv: Semi-supervised curvilinear structure segmentation," *IEEE Trans. Image Process.*, 2022.

[10] C Wei, K Sohn, C Mellina, A Yuille, and F Yang, "Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021.

[11] L-Z Guo and Y-F Li, "Class-imbalanced semi-supervised learning with adaptive thresholding," in *Int. Conf. Mach. Learn.*, 2022.

[12] Z Lai, C Wang, H Gunawan, S-C S Cheung, and C-N Chuah, "Smoothed adaptive weighting for imbalanced semi-supervised learning: Improve reliability against unknown distribution data," in *Int. Conf. Mach. Learn.*, 2022.

[13] D Guan, J Huang, A Xiao, and S Lu, "Unbiased subclass regularization for semi-supervised semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022.

[14] C H Sudre, W Li, T Vercauteren, S Ourselin, and M J Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *DL-MIA ML-CDS: Third Int. Workshop*, 2017.

[15] G French, S Laine, T Aila, M Mackiewicz, and G Finlayson, "Semi-supervised semantic segmentation needs strong, varied perturbations," in *BMVC*, 2019.

[16] Y Shi, L Cui, Z Qi, F Meng, and Z Chen, "Automatic road crack detection using random structured forests," *IEEE Trans. Intell. Transp. Syst.*, 2016.

[17] I Arganda-Carreras, S C Turaga, D R Berger, D Cireşan, A Giusti, L M Gambardella, J Schmidhuber, D Laptev, S Dwivedi, J M Buhmann, et al., "Crowdsourcing the creation of image segmentation algorithms for connectomics," *Frontiers in Neuroanatomy*, 2015.

[18] J Staal, M D Abràmoff, M Niemeijer, M A Viergever, and B Van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imaging*, 2004.

[19] A D Hoover, V Kouznetsova, and M Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imaging*, 2000.

[20] J Long, E Shelhamer, and T Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2015.

[21] V Badrinarayanan, A Kendall, and R Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.

[22] O Ronneberger, P Fischer, and T Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput.-Assist. Interv.*, 2015.

[23] F I Diakogiannis, F Waldner, P Caccetta, and C Wu, "Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, 2020.

[24] R Azad, M Asadi-Aghbolaghi, M Fathy, and S Escalera, "Bi-directional convlstm u-net with densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops.*, 2019.

[25] A Tarvainen and H Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," in *Adv. Neural Inf. Process. Syst.*, 2017.

[26] T Miyato, S-i Maeda, M Koyama, and S Ishii, "Virtual adversarial training: a regularization method for supervised and semi-supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2018.

[27] X Chen, Y Yuan, G Zeng, and J Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021.

[28] S Yun, D Han, S J Oh, S Chun, J Choe, and Y Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019.