OCTAve: 2D *en face* Optical Coherence Tomography Angiography Vessel Segmentation in Weakly-Supervised Learning with Locality Augmentation

Amrest Chinkamol *Member, IEEE*, Vetit Kanjaras, Phattarapong Sawangjai, Yitian Zhao, Thapanun Sudhawiyangkul, Chantana Chantrapornchai*, Cuntai Guan *Fellow, IEEE* and Theerawit Wilaiprasitporn* *Senior Member, IEEE*

***Abstract*— Objective: While the microvasculature annotation within Optical Coherence Tomography Angiography (OCTA) can be leveraged using deep-learning techniques, expensive annotation processes are required to create sufficient training data. One way to avoid the expensive annotation is to use a type of weak annotation in which only the center of the vessel is annotated. However, retaining the final segmentation quality with roughly annotated data remains a challenge. Methods: Our proposed methods called OCTAve provide a new way of using weak-annotation on microvasculature segmentation. Since the centerline labels are similar to scribble annotations, we attempted to solve this problem by using the scribble-based weakly-supervised learning method. Even though the initial results look promising, we found that the method could be significantly improved by adding our novel self-supervised deep supervision method based on Kullback-Liebler divergence. Results: The study on large public datasets with different annotation styles (i.e., ROSE, OCTA-500) demonstrates that our proposed method gives better quantitative and qualitative results than the baseline methods and a naive approach, with a p-value less than 0.001 on dice-coefficients and a lot fewer artifacts visually seen. Conclusion: The segmentation results are both qualitatively and quantitatively superior to baseline weakly-supervised methods when using scribble-based weakly-supervised learning augmented with self-supervised deep supervision, with an average drop in segmentation performance of less than 10%. Significance: This work gives a new perspective on how weakly-supervised learning can be used to reduce the cost of annotating microvasculature, which can make the annotating process easier and reduce the amount of work for domain experts.**

***Index Terms*— Optical coherence tomography angiography, vessel segmentation, deep neural network, self-supervised learning, weakly-supervised learning.**

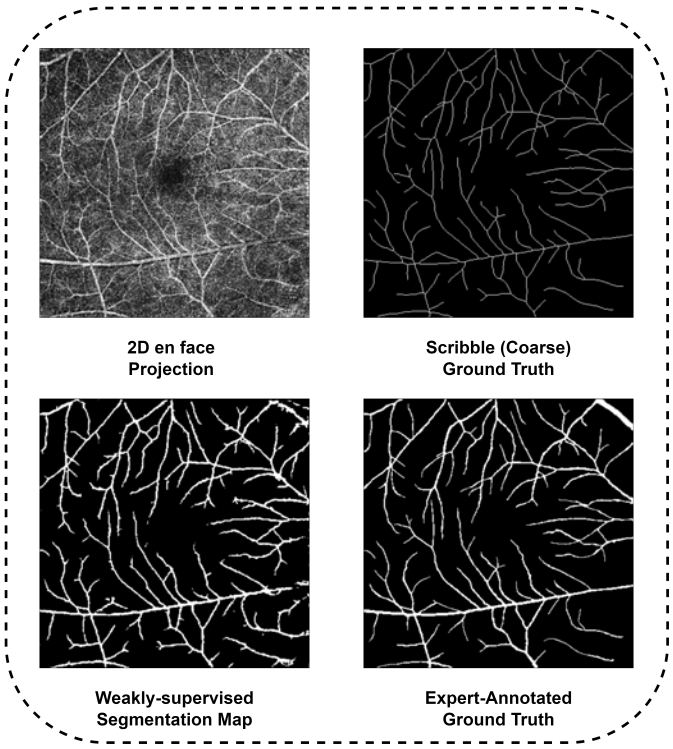

Fig. 1: Alternative vessel segmentation output learned from the coarse annotation while achieving a limited loss in the segmentation performance.

This work was supported by PTT Public Company Limited, The SCB Public Company Limited and National Research Council of Thailand (N41A640131) *(* corresponding author: Chantana Chantrapornchai, Theerawit Wilaiprasitporn).*

A. Chinkamol, P. Sawangjai, T. Sudhawiyangkul and T. Wilaiprasitporn are with Bio-inspired Robotics and Neural Engineering (BRAIN) Lab, School of Information Science and Technology (IST), Vidyasirimedhi Institute of Science & Technology (VISTEC), Rayong, Thailand (e-mail: amrest.c@vistec.ac.th, theerawit.w@vistec.ac.th).

V. Kanjaras is with Kamnoetvidya Science Academy (KVIS), Rayong, Thailand (e-mail: vetit_k@kvis.ac.th).

Y. Zhao is with the Cixi Institute of Biomedical Engineering, Ningbo Institute of Industrial Technology, Chinese Academy of Sciences, Ningbo 315201, China

C. Chantrapornchai is with High-Performance Computing and Networking Center (HPCNC), Department of Computer Engineering, Faculty of Engineering, Kasetsart University, Bangkhen Campus, Bangkok, Thailand (email: fengcnc@ku.ac.th)

C. Guan is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore

I. Introduction

THE visualization of retinal microvasculatures has been an important ongoing research problem, as effective

visualization techniques can help experts reveal details that are useful for diagnosis and analyzing pathological markers. Many current researches focus on techniques that can be used to improve the effectiveness of retinal vessel visualization. One of the techniques, called Optical Coherence Tomography Angiography (OCTA), a non-invasive imaging technique, has been shown to be helpful in the diagnosis of Diabetic Retinopathy (DR), providing a good alternative to the invasive Fluorescein Angiography [1]–[4]. Additionally, in the recent work, Ma *et al*. discovered that the fractal dimension obtained from the 2D *en face* projection of the microvascular structure within the parafovea area from the OCTA imaging technique has the potential to be used as an auxiliary biomarker for the Alzheimer's Disease (AD) [5]. However, OCTA technique is prone to various noises and artifacts, such as shadow and projection artifacts. Thus, vessel segmentation maps obtained from the OCTA imaging may result in the false detection of vascular structures, which can lead to misdiagnosis of the diseases.

Recent researches attempted to extract the vessel morphology from either 2D or 3D OCTA images in order to provide an accurate representation of the retinal vasculature with both conventional [6]–[8] and deep learning methods [5], [9]–[13]. The conventional methods i.e., filters, and thresholding. These methods require the manual calibration to be perform due to various capturing devices, and may also fail under certain conditions. The processed outputs are also susceptible to noises and artifacts presented in the OCTA images. For the deep learning methods, the manual calibration is not required and the resulting model can be generalized when there are enough data for training. While not requiring the manual calibration, a large amount of data need to be annotated by experts in order to perform the fully-supervised training.

Although, there exist many deep learning approaches on the task of vessel segmentation in OCTA images, to the best of our knowledge, only the work by Xu *et al*. [14] attempted to mitigate the tenuous work in the labeling process of retinal vasculature in OCTA and Fundus images. The authors proposed the semi-supervised learning scheme that only requires some patches of label on the meaningful and informative area of the images. For this approach, the need to fully-annotate the vessel for each patches and the need to determine which part of the image is actually contained the most useful information is still required.

In an attempt to rectify the expensive annotation process of the retinal vasculature for the 2D *en face* OCTA images, we investigate the possibility of applying the scribble-like annotation – the approach recently been used on the various other image segmentation tasks on both medical and non-medical images [15]–[20] – which can reduce the workload of experts dramatically since it does not require an expert-level precision in the annotation process. In particular, we propose a novel weakly-supervised learning framework for vessel segmentation tasks on the en face OCTA images, called *OCTAve*. OCTAve enabled the usage of the scribble-like annotation as a supervised label to be used in tandem with experts' fully-annotated unpaired ground truths, unrelated to the training inputs, via the adversarial game between the Segmentor and Discriminator networks to augment the model learning [19]. The Segmentor training is constrained by our novel deep supervision mechanism called **Self-Supervised Deep Supervision (SSDS)**, which can significantly increases the model segmentation performance without requiring additional data (Fig. 1). Our contributions to the field of study can be listed as follows:

- We propose a method that can rectify the expensive labelling cost in 2D *en face* OCTA images by incorporating the weakly-supervised learning method which reduces the need of fully-annotated labels.
- We propose a novel training mechanism called *Self-Supervised Deep Supervision* used in the training of weakly-supervised model along with the adversarial deep supervision mechanism. The results in Table I and II shows that our method significantly increased segmentation performance in both the weakly-supervised and the fully-supervised settings.
- To the best of our knowledge, our work is the first in the field of retinal vessel analysis which applies the weakly-supervised learning together with the adversarial and self-supervised training on the deep neural nets.

II. Related Works

In this section, we provide brief reviews of prior work on segmentation tasks using OCTA images and its projection and the prior work on the applications of several weakly supervised learning algorithms in medical imaging. In order to show the progress made in the tasks explored with OCTA image data, and the possible application of existing weakly-supervised learning technique to rectify our aforementioned problem introduced in the previous section.

A. Segmentation Tasks in OCTA

Numerous works investigated the vessel segmentation tasks and the area segmentation tasks such as foveal avascular zone segmentation [11], [12], [21], [22], by using deep neural networks on *en face* OCTA images due to their performance being significantly higher and more robust than that of hand-crafted filters [4]. L. Mou *et al*. addressed the problem of curvilinear structure segmentation by proposing CS^2-Net [13], which incorporates a Channel and Spatial Attention Module into a UNet-like autoencoder, resulted in the superior accuracy compared to the commonly used medical image segmentation algorithms such as U-net [23], Unet++ [24], UNet3+ [25] Attention UNet [26], R2U-net [27] and CE-net [9]. Li *et al*. published the OCTA-500 dataset, which included pixel-level labeling for large retinal vessel and foveal avascular zone segmentation by proposed the novel 3D to 2D segmentation model IPNv2, which takes 3D volumetric OCTA as an input and generated a 2D segmentation map [12].

In the recent study, Ma *et al*. published the ROSE dataset which contains *en face* OCTA images with fully annotated vessel labels obtained from multiple vasculature layers: superficial vascular complex (SVC), deep vascular complex (DVC), and inner retina vascular plexus (SVC+DVC), as well as their proposed state-of-the-art architecture, OCTA-Net [5], a

novel coarse-to-fine network that employs detailed pixel-level annotation from a consensus of several experts and vessel centerline annotation to train a dual U-Net-like architecture with shared weights on the encoder stack in an attempt to aid the model learning the vascular structure in a joint-learning manner.

However, in these studies on the segmentation tasks, none had considered a way to mitigate the work that needs to be done by the experts in the labeling process despite the laborious work of the pixel-level annotation.

B. Application of Weakly-supervise Learning in Medical Imaging

Weakly-supervised learning has been in a focus of research in the deep learning field, with the goal of reducing the workload for the expert labeling while concurrently achieving an outstanding performance.

Jia *et al*. proposed a weakly-supervised method for cancerous zone segmentation from histopathology images with the image-level annotation and constrained learning, which outperformed state-of-the-art architectures on a large-scale dataset illustrates the reducing amount of annotation work required [28]. Fries *et al*. tackled the task of classification of the aortic valve malformations from cardiac MRI images by using weakly supervision to generate noisy MRI labels and illustrated higher scores in all metrics compares to fully-supervised model on manual-labelled data [29]. Xing *et al*. approached the task of central serous chorioretinopathy (CSC) segmentation by two-stages learning architectures for weakly-supervision with image-level-only annotation, which greatly reduced the amount of labeling task [30].

Vepa *et al.* proposed an automated cerebral vascular segmentation by using active contour as a weak annotation generator. Their results illustrated slight lower scores but significantly reducing annotation time comparing to manual labeling for weak label [31]. Gondal *et al*. achieved high accuracy, low false positives with high sensitivity in detecting lesion region in retinal images by using lesion-level and image-level annotation for weakly-supervised boundary localization, achieving commensurate or even better performance than fully-supervise method [32]. Liu *et al*. used scribble annotation enhanced with uncertainty-aware self-ensembling and transformation-consistent approaches for weakly-supervised COVID19 infectious area segmentation task from CT images; the result on several data sets presents higher efficiency than other weakly-supervised method while obtaining similar performance to fully-supervised method [20].

The above weakly-supervised medical imaging approaches all show the performance improvement and manual labelling workload decrement. Thus, it is worthwhile to investigate the use of weakly supervised learning for medical image-related tasks, primarily to reduce the laborious work required of the human experts and to increase data availability. In this paper, we found that the scribble-like ground truth can be used to mitigate the previously described issues with the weakly-supervised technique for 2D *en face* OCTA vessel segmentation.

Fig. 2: Example of the training data used to train the Segmentor (x, y_s) and the Discriminator (y_u) from OCTA-500 6M.

III. Methods

Our proposed method is illustrated in Fig. 3. The network architecture adopts the work of Valvano *et al* [19] and significantly augmented by our novel self-supervised deep supervision mechanism as presented in the Section III-B.

The approach assumes the existence of a freely accessible, expert-labeled dataset. The dataset may be indirectly-related to our dataset. For example, images may be retrieved from the different devices, but the underlying ground truth is similar to our target dataset; we called this dataset an unpaired dataset. Examples of the data are shown in Fig. 2. To utilize the available expert-made annotations to help in the training of weakly labeled datasets, the concept of domain transfers using a generative adversarial network (GAN) [33] is deployed. The Segmentor network attempts to fool the Discriminator network which attempts to judge the difference between the expert-made segmentation map of the unpaired dataset and the model-predicted segmentation map of the targeted dataset. The goal is to make the Segmentor learn about the shape prior from the unpair expert-annotated ground truth. Thus, the adversarial game played between the Segmentor and the Discriminator enabling the Segmentor to produce a segmentation map resembling to the expert-annotated ground truth.

Let us denote Σ and Δ as the Segmentor and the Discriminator network respectively. We assign x, y_s and y_u as an input, a scribble-like ground truth, and a unpair fully-annotated ground truth respectively, where $\hat{y}, \hat{a}^i|_{i=0}^d$ and c are a predicted segmentation map, a set of attention maps and the discriminator output. The functional of each part can be described into equations as follow:

$$\Sigma(x) = \hat{y}, \hat{a}^i|_{i=0}^d \tag{1}$$

$$\Delta(\hat{a}^i|_{i=0}^d) = c; c \in \mathbb{R} \tag{2}$$

The network is trained to fool the Discriminator at multiple resolutions, by exploiting the stack of decoder layers. The adversarial game is played at the discriminative outputs from each decoder layer. Then, the outputs are fed to the adversarial attention gates as in Fig. 4 before passing through the discriminators. The adversarial game optimization at multi-scale outputs of the UNet are named as *adversarial deep supervision*.

In the above equations, $\hat{a}^i|_i^d = \{\hat{a}^0, \hat{a}^1, ..., \hat{a}^d\}$ is a set of attention map outputs from the adversarial attention gates, where d is the depth of the segmentor's attention gates.

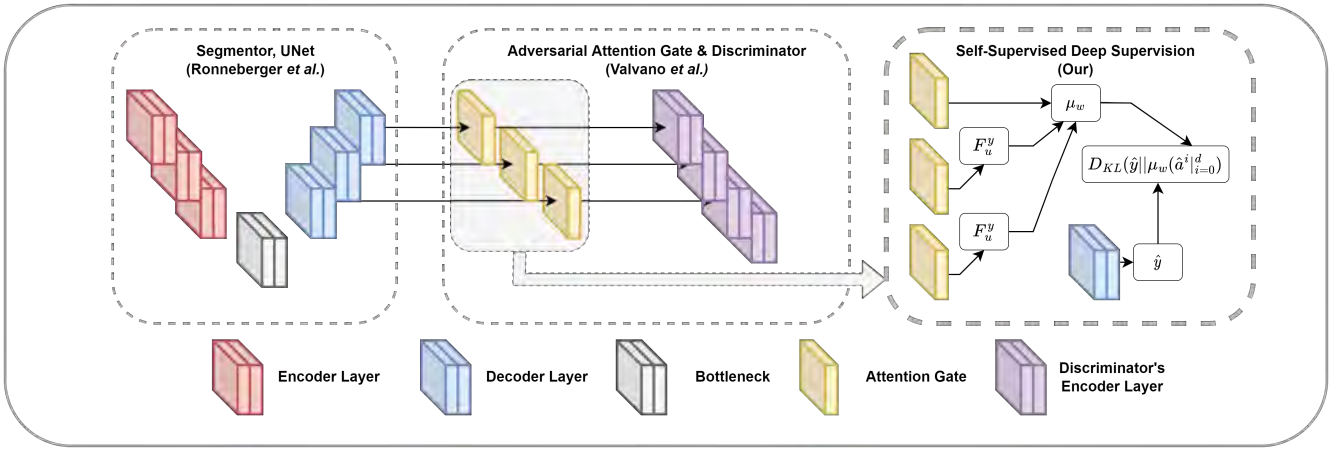

Fig. 3: The proposed architecture overview of OCTAve. It consists of 3 parts, the UNet-like Segmentor (Left), Adversarial Attention Gates and the Discriminator (Center), and the Self-Supervised Deep Supervision Mechanism (Right). Each part has different learning objectives but has a common goal of vessel extraction.

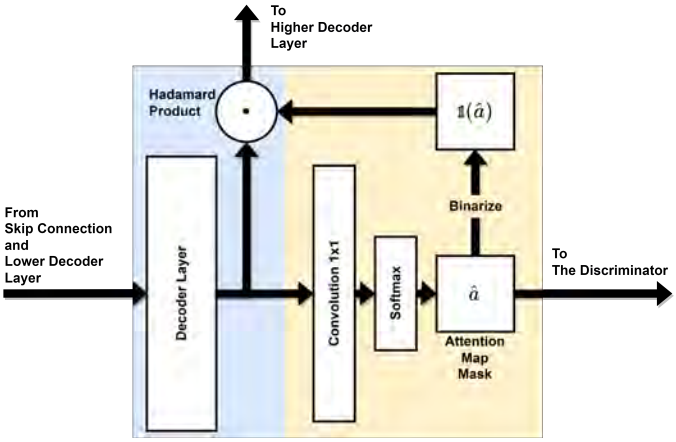

Fig. 4: Adversarial Attention Gate architecture, a part of adversarial deep supervision mechanism [19].

$$W_{pce} = 1(y_s)[-\sum_{i=1}^{c} w_i y_{s_i} \log(\hat{y}_i)] \tag{3}$$

$$\begin{aligned}\mathcal{V}_\Sigma =& \frac{1}{2} E_{x\sim X}[(\Delta(\Sigma(x)) - 1)^2] \\ \mathcal{V}_\Delta =& \frac{1}{2} E_{x\sim X}[(\Delta(\Sigma(x)) + 1)^2] \\ &+ \frac{1}{2} E_{y\sim Y}[(\Delta(y_u - 1)^2]\end{aligned} \tag{4}$$

For the supervised loss, masked version of weighted partial cross entropy introduced in [19] denoted in Equation 3 are used for weakly-supervised learning. The weighted term w_i is the ratio between the number of annotated pixels of class i over the number of annotated pixels of every classes, including the background. The least-square GAN losses for both the Segmentor and the Discriminator are denoted in Equation 4. LS-GAN loss is used for an adversarial game optimization due to its effectiveness in the style adaptation tasks, which is suitable for our training purpose; ie., the Segmentor adapts to the expert annotation style in its prediction. On the other hand, the Segmentor loss $\mathcal{L}_\Sigma$ and the Discriminator loss $\mathcal{L}_\Delta$ can be denoted as follows:

$$\begin{aligned}\mathcal{L}_\Sigma &= \alpha_0 W_{pce}(\hat{y}, y_s) + \alpha_1 \mathcal{V}_\Sigma(\hat{y}, \hat{a}_i|_{i=0}^d) \\ \mathcal{L}_\Delta &= \alpha_2 \mathcal{V}_\Delta(\hat{a}_i, F_d^{\hat{a}_i}(y_u))\end{aligned} \tag{5}$$

In Equation 5, α_0 is a dynamic weight used to balance the optimization between weakly-supervised learning and adversarial game optimization. Note that, in our approach, We will modify this term to achieve the faster training and more stability in the large parameter model. This will be discussed later in the Section III-A. Lastly, α_1 and α_2 are the fixed weights that are regulated the adversarial loss optimization of both the Segmentor and the Discriminator, which have been empirically set to 0.1 according to the setting as in the original work.

A. Balancing Mechanism of Multi-objective Learning

It is well-known that maintaining the training stability of the adversarial neural network is a difficult task, and it is even more difficult for multi-objective learning like our architecture. To achieve the stability, by avoiding the over-optimization on the only one of the objectives, an adaptive strategy is a must. Herein, in this section, we will discuss about the original dynamic weight term and our proposed modification.

From $\mathcal{L}_\Sigma$ in the Equation 5, $\alpha_0 = \frac{||\mathcal{V}_\Sigma||}{||W_{pce}||}$ is a dynamic weight term, defined by the ratio between adversarial loss and supervised loss with an intention to keep the optimization balance between weakly-supervised objective and adversarial game. However, while this method allows stable training, the model training process takes a long time to reach the optimal point.

In this work, the preliminary study on the trade-off of the training stability for the faster descent were conducted and observed. We introduce an alternative reciprocal version of the dynamic weight term as in Equation 6.

$$\alpha_0 = \text{Clamp}(\frac{||W_{pce}||}{||\mathcal{V}_\Sigma||}), \max = \mathcal{C}) \tag{6}$$

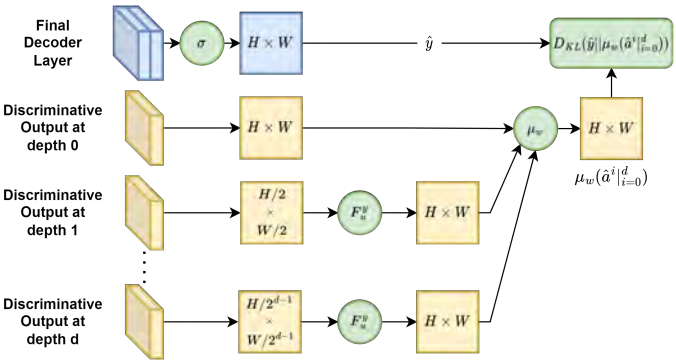

Fig. 5: Our proposed self-supervised deep supervision working mechanism for any N-depth UNet-like architecture with discriminative output on the decoder layer. σ, F_u^y, μ_w denote softmax function, upscaling function and weighted average function respectively.

In Equation 6, $\mathcal{C}$ is a constant that prevents the model from overly attending to weakly-supervised learning optimization during the early training stage. We have empirically set the value to 0.1 based on the observation from the behavior during the model training. To show the effectiveness of our proposed modification, Section V-A shows the experimental results.

B. Self-supervised Deep Supervision

As one of the main contribution of our study, we propose a novel mechanism for deeply supervise training on a UNet [23] like architecture called **Self-Supervised Deep Supervision** (Fig. 5). This mechanism deeply supervises the model during the training based on the pixel-wise confidence of the segmentation map. Based on the observation by Zhao *et al.* [34], a well-performed UNet model with discriminative output at each of the decoder layers tended to have high consistency of Dice's coefficient between the discriminative output from each decoder layer. In our case, the attention maps from the adversarial attention gates are also discriminative output from the decoders. This led us to the idea of increasing the consistency between decoder layers' discriminative outputs as one of the optimization objectives.

In other words, the decoder layer can be thought of as a probability density estimator function that is parameterized by its own weight θ.

$$Decoder_i(x;\theta_i) \equiv PDE(x;\theta_i) \tag{7}$$

From the previous statement, we can argue that each decoder layer should be mapped to the approximately same distribution in order to produce highly consistent and higher precision segmentation maps.

$$PDE(x;\theta_d) \sim PDE(x;\theta_i);\ 0 \leq i < d \tag{8}$$

In our case, using Dice's coefficient as a measurement metric is not an option due to the usage of weak annotation, and in some cases, we may not have labels for some of the images. Instead, we incorporate the Kullback-Leibler Divergence [35] to improve the consistency between attention maps from each layer. We can say that our objective is to minimize the divergence of confidence between attention maps of each layer. The objective function can be denoted as follows:

$$\text{minimize} \sum_{i=0}^{d} D_{KL}(\hat{y}||Decoder_i) \tag{9}$$

Thus, we call the application of the deep supervision based on *KL* divergence loss function as Inter-layer Divergence Loss (ILD) as follows:

$$\mathcal{L}_{ILD}(\hat{y}, \hat{a}_i|_{i=0}^d) = \hat{y}[\log(\hat{y}) - \log(\mu_w(\hat{a}_i|_{i=0}^d))] \tag{10}$$

In the above equation, $\hat{y}$ is the predicate output from the Segmentor which we use as a posterior, and $\hat{a}_i|_{i=0}^d$ is a set of the attention map outputs from the adversarial attention gates.

$$\mu_w(\hat{a}_i)|_{i=0}^d = \frac{1}{d}\sum_{i=0}^{d} w_i F_u^y(\hat{a}_i) \tag{11}$$

Instead of pairwise calculation between posterior and each of the attention maps, weighted average confidence from the set of attention maps specified in Equation 11 are used for the implementation of this work, to reduce the computational cost from the pairwise calculation of the divergence between each layer. The weight w_i is scaled alongside with the spatial resolution of the corresponding decoder layer's output, such that:

$$\begin{aligned} w_0 &= S_f^0 \\ w_i &= \max(w_0 - 0.1\sum_{1}^{i} S_f^i, w_l) \end{aligned} \tag{12}$$

Where S_f^i is a sum of spatial downscaling factor at layer i, and w_l is a lower bound value defined by:

$$w_l = \min(\{x = S_f^0 - 0.1\sum_{1}^{i} S_f^i | 1 \leq i \leq d, x > 0)\} \tag{13}$$

In our experiment, OCTAve's backbones, both UResNeSt and UNet, downscale the resolution by half at each layer, thus $w_0 = 1$ and $w_i = max(1 - 0.2i, 0.2)$. F_u^y is the differentiable upscaling function used to upscale the lower resolution attention maps to match the posterior resolution.

Thus, we can rewrite the segmentor loss function with $\mathcal{L}_{ILD}$ as an additional optimization objective as follows:

$$\mathcal{L}_{\Sigma} = \alpha_0 W_{pce}(\hat{y}, y_s) + \alpha_1 \mathcal{V}_{\Sigma}(\hat{y}) + \kappa \mathcal{L}_{ILD}(\hat{y}, \hat{a}_i|_{i=0}^d) \tag{14}$$

$$\kappa = \frac{||\mathcal{L}_{ILD}||}{||W_{pce}|| + ||\mathcal{V}_{\Sigma}||} \tag{15}$$

Equation 15 is an additional dynamic weight introduced to keep the balance of optimization between self-supervised learning and others.

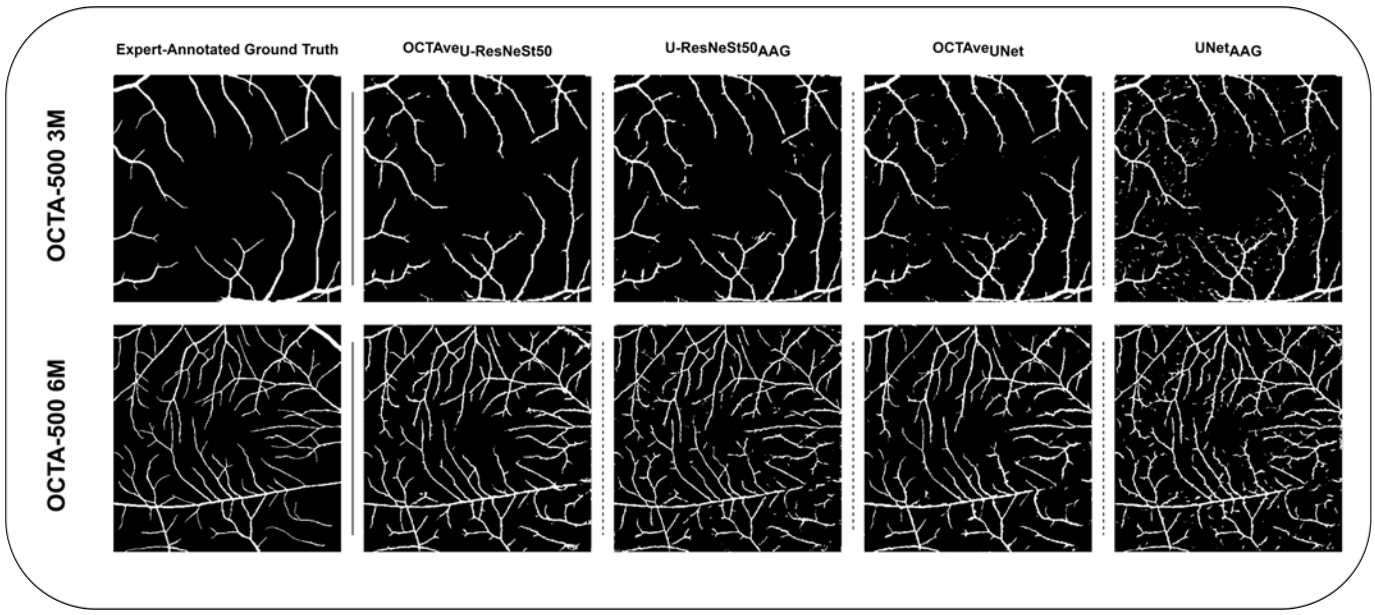

Fig. 6: Segmentation map of the sample from test set of the OCTA-500 3M and 6M on each methods in Section IV-A.2. The segmentation map were ordered accordingly, from left to right: expert-annotated ground truth, segmentation map from the inference of the 2D en face angiogram through $\text{OCTAve}_{\text{U-ResNeSt50}}$, U-ResNeSt50, $\text{OCTAve}_{\text{UNet}}$, UNet.

IV. Experiments & Results

Our study consists of two main experiments to show the effectiveness of the proposed method on the task of vessel extraction of 2D *en face* OCTA images. The weakly-supervised training experiment in the Section IV-A was conducted as our main contribution to this study. Then, to show the useful application on other than weakly-supervised learning of our proposed method, the fully-supervised training experiment were conducted as presented in the Section IV-B.

A. Experiment I: Weakly-supervised Segmentation

1) Datasets and Data Preparation: In this experiment, we used the OCTA-500 dataset [11], a public multi-modality dataset with a total of 500 subjects. The dataset provides two types of field of view (FOV), a 3×3 cm^2 FOV dubbed as OCTA-3M containing 200 subjects and a $6x6$ cm^2 FOV dubbed as OCTA-6M containing 300 subjects. Each type of FOV contains three types of 2D *en face* OCTA projection with different depths, as follows: FULL(B4), ILM-OPL(B5), and OPL-BM(B6) projection. Text labels for diseases label are also provided for non-normal samples.

For this experiment, the training and the evaluation of each projection and FOV were performed separately. In the preparation of the dataset, a 70:30 stratified random train-test split based on the disease label with a random state seed of 50, followed by the 5-fold stratified cross validation on the train dataset were performed on each group. Additionally, an augmentation of the training set was done by randomly rotating the input image and its corresponding label within the range of -10 to 10 degrees. In order to simulate the unpair dataset, we randomly selected half of the training set to be used as an unpair dataset.

Since the original OCTA-500 does not provide expert-made scribble annotations, we generated them synthetically

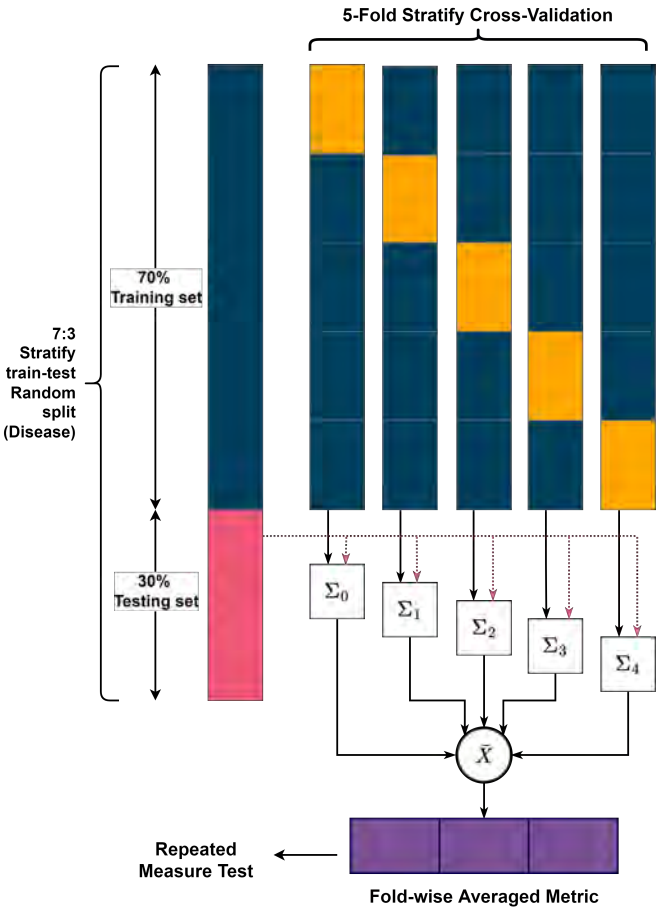

Fig. 7: Experiment I data preparation and evaluation. Σ_i denote the best validation score Segmentor model for fold i.

Modality	Projection	Scribble %	Method					
			OCTA-Net	U-ResNeSt50		UNet		OCTA-Net
			$\text{OCTA-Net}_{\text{UB}}$	$\text{U-ResNeSt50}_{\text{AAG}}$	$\text{OCTAve}_{\text{U-ResNeSt50}}$	UNet_{AAG}	$\text{OCTAve}_{\text{UNet}}$	$\text{OCTA-Net}_{\text{LB}}$
3M	FULL	100%	87.83	76.94	**78.48***	72.99	**76.94***	42.06
		75%		75.72	**76.07***	73.26	**74.66***	
		50%		74.84	**75.44**	71.23	**72.43***	
		10%		**70.66***	67.16	**69.16***	66.45	
	ILM-OPL	100%	91.32	81.42	**82.35***	77.20	**82.34***	53.13
		75%		**81.38**	80.89	80.97	**81.08**	
		50%		80.31	**80.67**	80.18	**80.52**	
		10%		**77.37***	75.41	**78.94***	77.87	
	OPL-BM	100%	82.62	71.07	**73.86***	68.07	**72.36***	26.05
		75%		**72.88***	71.95	68.86	**70.52***	
		50%		70.20	**70.52**	67.14	**67.68**	
		10%		**67.58***	65.26	64.71	**64.97**	
6M	FULL	100%	84.62	74.33	**77.17***	72.71	**75.26***	38.33
		75%		73.82	**76.45***	69.49	**74.72***	
		50%		73.24	**73.93***	69.29	**70.68***	
		10 %		**71.03***	70.48	**70.02***	68.79	
	ILM-OPL	100%	88.87	79.48	**81.43***	78.76	**81.06***	46.23
		75%		**79.19**	78.23	78.61	**80.96***	
		50%		78.83	**79.15**	78.61	**80.57***	
		10%		**78.63***	77.62	77.97	**78.46***	
	OPL-BM	100%	78.84	68.56	**72.16***	65.06	**69.39***	35.59
		75%		70.15	**71.85***	66.63	**69.48***	
		50%		67.35	**67.55**	64.31	**66.46***	
		10%		**64.68***	63.34	**63.77***	62.98	

TABLE I: Vessel segmentation performance (Dice's Coefficient) of the 100%, 75%, 50% and 10% scribble availability variation compared across the upper bound, lower bound and methods. Bold text denotes the method with the best numerical value within the same architecture, and * denotes that the method is statistically significant ($p < 0.001$).

to be used as weakly-supervised ground truths by performing the skeletonization of pixel-level annotations provided by the dataset using Zhang's algorithm [36] implemented in Python's scikit-image library [37].

Additionally, to explore the robustness and possible drawbacks of our method where the model relied on adversarial knowledge for an unsupervised sample and the enforcement of self consistency correction, which may cause the error to propagate during the training, the experimentation was varied by the scribble label available for use in the training set as in [19] as follows: 100%, 75%, 50% and 10% scribble availability. Where $x\%$ availability indicates that $(100 - x)\%$ of scribble labels (y_s) of the samples (x) in the training dataset, randomly selected, are excluded from the training step. The samples without scribble labels are subjected to unsupervised learning, where W_{pce} term in both Equation 14 and 15 are removed from the calculation.

2) Benchmarking Methods: The validation of efficacy of automated segmentation methods was performed by comparing our proposed method against the current fully-supervised state-of-the-art method, OCTA-Net [5] in both weakly-supervised version as a lower bound baseline and fully-supervised version as a soft upper bound, and the current state-of-the-art scribble-base weakly supervise method from [19]. We considered the following:

- **$\text{OCTA-Net}_{\text{UB}}$** [5]: OCTA-Net, the current state-of-the-art fully-supervised automated vessel segmentation method on 2D *en face* OCTA images. Their method is a two-stage coarse-to-fine network. The coarse stage network is essentially a UNet [23] with the imagenet-pretrain ResNeSt50 [38] as a backbone for the encoder layers and two decoder branches.
- **$\text{OCTA-Net}_{\text{LB}}$**: We considered the scenario where one tried to apply the scribble weakly-supervised learning objective to OCTA-Net. In this case, W_{pce} (Equation 3) loss was used for the fair comparison with our methods.
- **UNet_{AAG}**: UNet with the adversarial attention gate and the Discriminator network from [19]. The configuration of the architecture and the hyperparameters are the same as the original work, except for the dynamic weight α_0, which was changed into an alternate version as proposed in the Equation 6.
- **$\text{OCTAve}_{\text{UNet}}$**: Based on **$\text{UNet}_{\text{AAG}}$**, self-supervised deep supervision mechanism was incorporated into the training

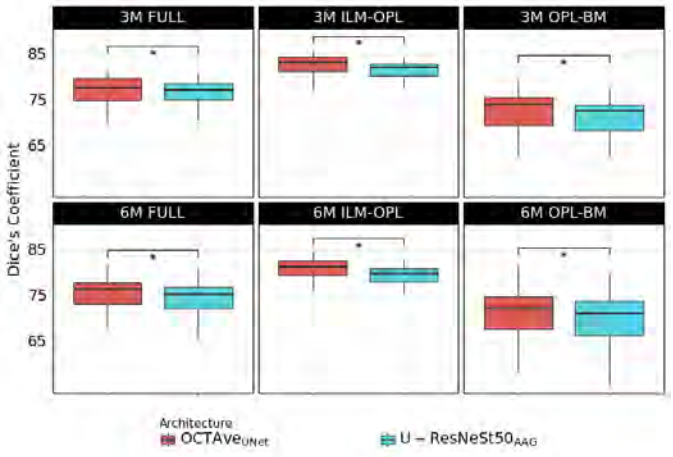

Fig. 8: Performance comparison between the "lightweight" **OCTAve$_{\text{UNet}}$** against the "heavyweight" **UResNeSt50**. Demonstrating the significant competitive advantage of our proposed method in the enhancement of the low parameter network to outperform the large parameter network. * denotes the statistically significance of $p < 0.001$.

of the Segmentor network to enhance the segmentation performance and improves the consistency of attention maps between decoder layers.

- **U-ResNeSt50$_{\text{AAG}}$** and **OCTAve$_{\text{U-ResNeSt50}}$**: To show the effectiveness of our proposed method on the complex architecture with a high number of parameters, the vanila UNet encoder that was used in both the **UNet$_{\text{AAG}}$** and the **UNet$_{\text{AAG-SSDS}}$**, which had a far fewer number of weight parameters, was replaced with ResNest50, a large, complex architecture used in the coarse stage network of the OCTA-Net.

3) Model Training and Evaluation Methods: The model and the experimentation were implemented using the PyTorch Lightning framework [39]. The global random state seed was set to 50, and the deterministic training option was enabled to maximize the reproducibility. The training epochs were set to 1,000 epochs, along with the checkpointing with the best Dice's coefficient score on a validation dataset to be used as a criterion for every benchmarking method. As for the optimizer, we use the Adam optimizer [40] with the learning rate and weight decay parameters were set to 0.001 and 0.0001, respectively, and the cyclic learning rate scheduler were used with the oscillation between 0.0001 and 0.00001. Except for **OCTA-net$_{\text{UB}}$**, we used the same configuration as specified in [5], e.g. polynomial learning rate scheduler with a power factor of 0.9.

The segmentation performance was reported using Dice's coefficient score, the regularly used evaluation metric in image segmentation tasks. The statistical significance of the reported results was tested using non-parametric Friedman repeated measure, followed by the pairwise t-test to calculate statistical significant value with Bonferroni p-value correction. $p < 0.001$ is considered to be statistical significant.

4) Performance Analysis: The superiority of our proposed methods was verified by comparing the segmentation performance on the deep neural network architecture incorporated with our proposed self-supervised deep supervision mechanism against the one without it, as reported quantitatively in Table I and qualitatively in Fig. 6.

From the experiment, the major difference in the segmentation performance occurs at the application of our methods and the variation of the scribble availability. We summarize the discussions as follows:

- **Impact of the self-supervise deep supervision mechanism.** As the mechanism is the main contribution of this study, the comparison of our proposed method was conducted on both the lightweight vanilla UNet and the heavyweight ResNeSt50-backboned UNet to show the effectiveness of the application. In this part, we only consider the 100% scribble availability variation to exclude any exposure to unsupervised learning from the analysis.
  Based on the results shown in Table I, both **UNet** and **U-ResNeSt50** architectures, the segmentation performance between the one with our proposed method and the one without was drastically different, as the statistical analysis performed with a significant value of 0.001 showed. With the improvement seen on the architectures incorporated with our proposed method on every group of projection level and FOV, we can argue that our method has a highly beneficial impact on the segmentation performance of the weakly-supervised learning setting of 2D *en face* OCTA vessel segmentation. Moreover, Fig. 8 shows the effect of SSDS augmentation on a low parameter model, boosting the model learning ability to achieve higher or equal performance compared to the large parameter model. This observation suggests a useful application when the computational resource is limited and worth exploring further in future work.
  Consequently, the experiment shows the possible application of the weakly-supervised learning method using the scribble-like label annotation proposed in [19] to achieve a limited reduction in the segmentation performance of curvilinear structure despite the usage of coarse labels and significantly augmented by our proposed deep supervision method. Additionally, our method has the potential to be applied for other segmentation tasks, such as the foveal avascular zone as shown in Appendix I or other different organs, with improved segmentation quality.
- **Robustness of the mechanism over the exposure to unsupervised learning.** To investigate the possible drawback of our proposed method in the exposure to unsupervised learning when the label is missing, experiments on the 75%, 50% and the 10% variations of scribble availability were considered. As shown in Table I and visualized in Fig. 9, in majority of cases, experimentation results on the 50% variation have shown that our method's difference in the segmentation performance was mostly statistically insignificant. While the results of the 10% variation show that our method suffered from error propagation and achieved lower segmentation performance. As a result, the experiments revealed that our

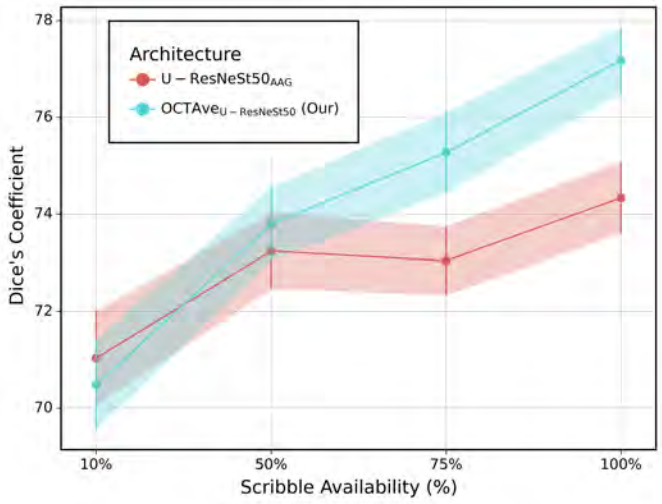

Fig. 9: Segmentation performance comparison between **OCTAve$_{\text{U-ResNeSt50}}$** and baseline **U-ResNeSt50** on OCTA-500 6M-FULL test set.

method's robustness can withstood as high as 50% of the label being missing before being gradually outperformed by the unaugmented method.

B. Experiment II: Self-supervised Deep Supervision on Fully-supervised Learning

1) Datasets and Data Preparation: In this experiment, ROSE [5] dataset, a public dataset with a total of 229 OCTA images consisted of 2 modalities, ROSE-1 and ROSE-2. ROSE-1 consisted of normal control and Alzheimer's disease patients. ROSE-2 consisted of various kinds of macular degenerative disease patients. However, the disease label unable to be provided by the original author due to ethical concerns. The train and test datasets were explicitly provided, thus the experiment was evaluated on the exact same test dataset as the original work.

2) Benchmarking Methods: To validate the effectiveness of the self-supervised deep supervision mechanism on the fully-supervision task, the coarse stage network of OCTA-Net were augmented by attaching adversarial attention gate into each decoder layer of both branches of the network, combined with the Discriminator network. The fine stage of the network were left as is. This variation of model modification is called **OCTA-Net$_{\text{AAG-SSDS}}$**. The comparison was made against the original OCTA-Net results and other segmentation methods [9], [23], [41]–[47] as reported in the original work.

3) Model Training and Evaluation Methods: The experimentation settings in the original work were replicated to perform a direct comparison against the reported score in the original work. The number of training epochs was fixed to 200 for both the coarse and the fine stage of the network with a batch size of 2. The Adam optimizer with learning rate and weight decay parameters were set to 0.0005 and 0.0001, respectively, incorporated with a polynomial learning rate scheduler with the power factor of 0.9.

4) Performance Analysis: To demonstrate that our method's benefits are not exclusively limited to the weakly-supervised learning scenario, we perform a performance comparison of OCTA-Net against its SSDS augmented version and other fully-supervised segmentation methods reported in [5]. The results in Table II show the superiority of the augmented model in the majority of the cases. However, the exception were found on **ROSE-1 DVC**, where our method achieved a much lower segmentation performance compared to Ma *et al*. OCTA-Net. We speculated that the reason being the expert annotation of DVC level microvascular is limited to the area around FAZ, causing our method which relies on confidence level of the predicted segmentation map to underfit the vessel that extends further from the area.

Methods	ROSE-1			ROSE-2
	SVC	SVC+DVC	DVC	SVC
IPAC [41]	57.51	52.23	9.11	55.90
COSFIRE [42]	75.17	66.71	24.50	61.42
COOF [43]	66.06	56.85	10.03	61.12
UNet [23]	71.16	70.12	66.05	65.64
ResU-Net [44]	74.61	73.09	65.67	67.25
CE-Net [9]	75.11	73.00	57.83	70.66
DUNet [45]	75.05	74.03	58.50	69.35
CS-Net [46]	76.08	74.88	58.84	70.10
Three-Stage [47]	76.63	75.24	66.22	70.24
OCTA-Net [5]	76.97	75.76	**70.74**	70.77
OCTAve$_{\text{OCTA-Net}}$ (Our method)	**78.03**	**81.42**	62.55	**71.18**

TABLE II: Vessel segmentation performance (Dice's Coefficient) in the fully-supervised learning setting on ROSE-1 and ROSE-2 datasets.

V. Discussion

A. Effect of the Dynamic Weight Term Alteration on the Model Training

For the discussion about the effectiveness of the reciprocal α_0 over the unmodified version, we showed the segmentation performance of each method: Valvano et al. unmodified method, Reciprocal α_0, and Reciprocal α_0 with Inter-layer divergence loss (our method). As stated in the Section III-A, we traded the training stability for the faster model training, which the reciprocal version can achieve a higher segmentation performance in the early training stage as seen in the Fig. 10 before destabilizing, albeit regained the stability in the late stage. However, to our anticipation, the early boost in the performance was continued by the use of a self-supervised deep supervision mechanism through inter-layer divergence loss, which can lead to the significant performance increase and the faster training compared to the previous two.

B. Error Propagation Drawback from the Self-supervised Learning

Our proposed method is a delicate system where weakly-supervised, adversarial, and self-supervised learning work in an ensemble and must be performed in harmony without disrupting the balance. Otherwise, the model might get stuck in the bad local optima. This led us to propose Equation 15 –

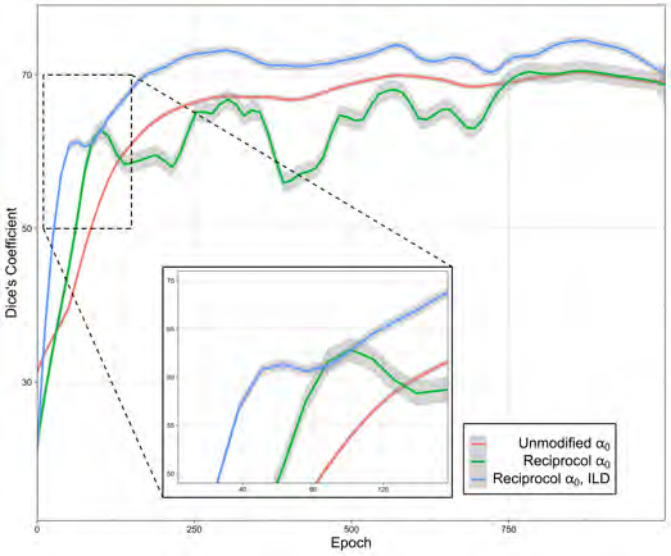

Fig. 10: Dice's coefficient on validation set over the training epoch of the different α_0 modifications on **OCTA-500 3M FULL** dataset, demonstrating the transition in the segmentation performance in the early stage of model training.

a dynamic weight term that keeps all optimization objectives in balance.

While our method is superior to the original scribble-base state-of-the-art method for most cases, as observed in the results of 100%, 75% and 50% scribble availability variation in Table I. The experiments performed on the extreme edge case, where only 10% of the training samples with labels were available, revealed the possible drawback of our proposed method, where self-supervision failed and disrupted the overall performance of the model training. Hence, the amount of data annotation available needs to be taken into consideration for the application of our method to other tasks to avoid the performance drop due to the error propagation in the early stage.

VI. Conclusion

In this work, we present a novel scribble-base weakly-supervised framework called OCTAve that utilizes both adversarial deep supervision and novel self-supervised deep supervision. It outperforms the current state-of-the-art scribble-based weakly supervised and state-of-the-art fully supervised methods on the 2D en face OCTA vessel segmentation task on the public datasets (i.e., OCTA-500 and ROSE). OCTAve is carefully designed to make three learning objectives work together in harmony, which results in a robust framework that can retain its performance advantage over the baseline even if 50% of the label is missing from the training dataset. We conducted the experiment and evaluation of the segmentation performance for the weakly-supervised learning method on the OCTA-500 dataset with statistical analysis, revealing our method performance advantage that can make the lightweight vanila UNet achieve the same segmentation performance as the heavyweight ResNeSt50-backbone UNet. In addition, the method also increases the performance of the heavyweight one. Additionally, we conducted a replica of a fully-supervised experiment on ROSE dataset to investigate the application of our proposed method in a fully supervised setting, resulting in our method's superiority in the segmentation performance, thus indicating the impact and usefulness of our proposed method on the overall area of 2D en face OCTA vessel segmentation in both low and high data availability scenarios. Thus, in addition to being able to assist expert annotators in real-world applications, our work will serve as a baseline for weakly-supervised segmentation with weak-annotation for future work and inspire more research into this area.

Appendix I
Foveal Avascular Zone Segmentation

In order to show that the proposed method's ability can be extended further from the vessel segmentation task, an additional experiment was conducted on Foveal Avascular Zone (FAZ) segmentation task on OCTA images using the OCTA-500 dataset, specifically ILM-OPL projection, where graders perform annotation. In this experiment, we conducted two separate tests. The first was performed on the same group of data in [12] to allow us to conduct a fair comparison within the fully-supervised learning constraint. However, [12] did not stratify the dataset based on the disease label in the data preparation process – which could have caused the measured performance to be inaccurate. Thus, the second test was performed using the same protocol as in Section IV-A to re-evaluate the best method's performance, with an additional weakly-supervised learning to investigate its potential on FAZ segmentation task.

A. FAZ Scribble-like Label Synthesis

Unlike the vessel, a curvilinear structure, where skeletonization of the expert label can be used to synthesize a scribble-like label, we establish the FAZ scribble label synthesize protocol as shown in the Figure 11's diagram.

B. Experiments and Result Discussion

In the first experiment, we compare the result of OCTAve trained with fully-supervised label with Dice's Loss as a loss function, dubbed $\text{OCTAve}_{\text{Fully}}$, against the results of OCTA-500 3D to 2D SOTA segmentation network IPNv2 and IPNv2+ [12] and other commonly used methods [23], [25], [26], [45], [48], [49] as reported in the original work. The result in Table III shows $\text{OCTAve}_{\text{Fully}}$ superior segmentation performance in both OCTA-3M and OCTA-6M for 2D to 2D (ILM-OPL) segmentation methods with lower variance. Additionally, $\text{OCTAve}_{\text{Fully}}$ outperformed 3D to 2D segmentation networks in OCTA-3M despite the obvious disadvantage in the input information.

However, as we pointed out previously that the settings in the first experiment's result might be inaccurate due to the test dataset not contained the same proportion of subjects with macular disease, which might contained pathological artifacts that affect the methods' segmentation performance, as in the training set. This led us to re-evaluate the OCTAve's segmentation performance in this experiment using the same

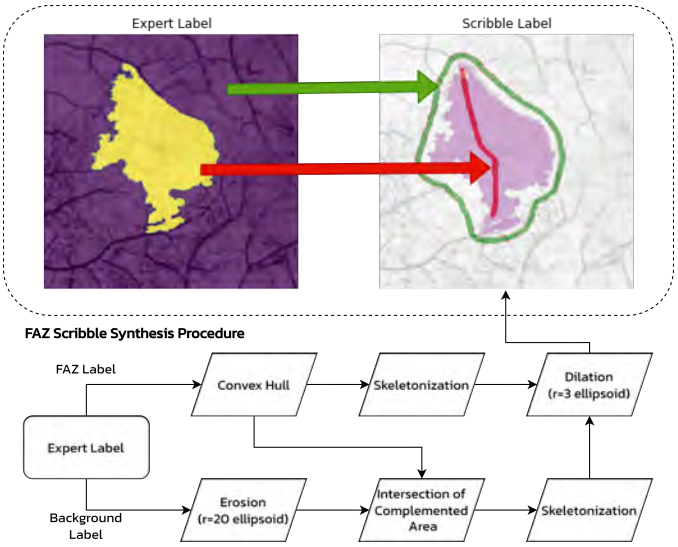

Fig. 11: FAZ scribble synthesis procedure diagram. Left image: Yellow and purple denote the expert's FAZ and background label, respectively. Right image: Red and green denote FAZ and background scribble label. Purple denotes expert's FAZ label.

Methods	**Dice's Coefficient** $(Mean \pm S.D.)$	
	OCTA-3M	**OCTA-6M**
2D to 2D		
Fast-FCN [48]	95.55 ±6.39	86.01 ±15.62
DeepLabV3+ [50]	96.60 ±3.19	87.20 ±14.94
Attention U-Net [26]	97.11 ±2.03	88.97 ±14.02
UNet [23]	96.81 ±2.96	88.10 ±17.56
UNet++ [24]	97.10 ±2.10	87.79 ±18.00
UNet3+ [25]	97.45 ±1.72	88.15 ±16.68
$\text{OCTAve}_{\text{Fully}}$ (Our Method)	**98.09 ±1.32**	**89.15 ±11.84**
3D to 2D		
IPN [11]	95.05 ±4.79	88.02 ±9.91
IPNv2 [12]	97.30 ±2.33	90.68 ±8.87
IPNv2+ [12]	97.55 ±2.38	90.84 ±8.93

TABLE III: Segmentation performance of fully-supervised FAZ segmentation task on OCTA-500 dataset using the same train-val-test dataset as in [12].

data preparation protocol displayed in Figure 7. Moreover, we also explore the potential of weakly-supervised learning on FAZ segmentation task. The result in Table IV displays a reduction in performance for $\text{OCTAve}_{\text{Fully}}$, reflecting the inaccuracies in the previous settings with a slight deviation. The result also shows that the weakly-supervised learning can achieve a limited reduction in performance compared to the fully-supervised learning; however, it should be noted that the weakly-supervised segmentation performance depends on how 'coarse' the label provided for the model to learn is, and more study should be done to investigate FAZ coarse label in the future works.

Modality	**Dice's Coefficient** $(Mean \pm S.D.)$	
	OCTAve	$\textbf{OCTAve}_{\textbf{Fully}}$
OCTA-3M	91.19 ±0.34	96.81 ±0.63
OCTA-6M	81.39 ±0.52	89.23 ±0.84

TABLE IV: Segmentation performance of fully-supervised and weakly-supervised FAZ segmentation task on OCTA-500 dataset using train-val-test dataset construction scheme and cross-validation as in Section IV-A.

REFERENCES

[1] J. P. Campbell *et al.*, "Detailed Vascular Anatomy of the Human Retina by Projection-Resolved Optical Coherence Tomography Angiography," *Scientific Reports 2017 7:1*, vol. 7, pp. 1–11, 2 2017. General Detail About Vascular Anatomy.

[2] K. Y. Tey *et al.*, "Optical Coherence Tomography Angiography in Diabetic Retinopathy: A Review of Current Applications," *Eye and Vision 2019 6:1*, vol. 6, pp. 1–10, 11 2019.

[3] Z. Sun, D. Yang, Z. Tang, D. S. Ng, and C. Y. Cheung, "Optical Coherence Tomography Angiography in Diabetic Retinopathy: An Updated Review," *Eye 2020 35:1*, vol. 35, pp. 149–161, 10 2020.

[4] T. T. Hormel, T. S. Hwang, S. T. Bailey, D. J. Wilson, D. Huang, and Y. Jia, "Artificial Intelligence in OCT Angiography," *Progress in Retinal and Eye Research*, p. 100965, Mar. 2021.

[5] Y. Ma *et al.*, "ROSE: A Retinal OCT-Angiography Vessel Segmentation Dataset and New Model," *IEEE Transactions on Medical Imaging*, vol. 40, pp. 928–939, Mar. 2021.

[6] N. Eladawi *et al.*, "Automatic Blood Vessels Segmentation Based on Different Retinal Maps From OCTA Scans," *Computers in Biology and Medicine*, vol. 89, pp. 150–161, 10 2017. Automated Method of Vessel Segmentation.

[7] Y. Zhao *et al.*, "Intensity and Compactness Enabled Saliency Estimation for Leakage Detection in Diabetic and Malarial Retinopathy," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 51–63, 1 2017. saliency estimation: extracting the salient area with algorithm.

[8] J. Zhang *et al.*, "3D Shape Modeling and Analysis of Retinal Microvasculature in OCT-Angiography Images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1335–1346, 2020.

[9] Z. Gu *et al.*, "CE-Net: Context Encoder Network for 2D Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 38, pp. 2281–2292, Oct. 2019.

[10] Y. Giarratano *et al.*, "Automated Segmentation of Optical Coherence Tomography Angiography Images: Benchmark Data and Clinically Relevant Metrics," *Translational Vision Science & Technology*, vol. 9, p. 5, Dec. 2020.

[11] M. Li *et al.*, "Image Projection Network: 3D to 2D Image Segmentation in OCTA Images," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 3343–3354, Nov. 2020.

[12] M. Li *et al.*, "IPN-V2 and OCTA-500: Methodology and Dataset For Retinal Image Segmentation," *arXiv:2012.07261 [cs, eess]*, Dec. 2020.

[13] L. Mou *et al.*, "CS2-Net: Deep Learning Segmentation of Curvilinear Structures in Medical Imaging," *Medical Image Analysis*, vol. 67, p. 101874, Jan. 2021.

[14] Y. Xu *et al.*, "Partially-Supervised Learning for Vessel Segmentation in Ocular Images," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12901 LNCS, pp. 271–281, 2021.

[15] D. Lin, J. Dai, J. Jia, K. He, and J. Sun, "ScribbleSup: Scribble-Supervised Convolutional Networks for Semantic Segmentation," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3159–3167, June 2016.

[16] B. Wang, G.-J. Qi, S. Tang, T. Zhang, Y. Wei, L. Li, and Y. Zhang, "Boundary Perception Guidance: A Scribble-Supervised Semantic Segmentation Approach," in *IJCAI*, 2019.

[17] Y. B. Can, K. Chaitanya, B. Mustafa, L. M. Koch, E. Konukoglu, and C. F. Baumgartner, "Learning to Segment Medical Images with Scribble-Supervision Alone," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11045 LNCS, pp. 236–244, 7 2018.

[18] Z. Ji, Y. Shen, C. Ma, and M. Gao, "Scribble-Based Hierarchical Weakly Supervised Learning for Brain Tumor Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019* (D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap, and A. Khan, eds.), Lecture Notes in Computer Science, (Cham), pp. 175–183, Springer International Publishing, 2019.

[19] G. Valvano, A. Leo, and S. A. Tsaftaris, "Learning to Segment from Scribbles Using Multi-Scale Adversarial Attention Gates," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2021.

[20] X. Liu *et al.*, "Weakly Supervised Segmentation of COVID19 Infection with Scribble Annotation on CT Images," *Pattern Recognition*, vol. 122, p. 108341, 2 2022.

[21] R. Mirshahi, P. Anvari, H. Riazi-Esfahani, M. Sardarinia, M. Naseripour, and K. G. Falavarjani, "Foveal Avascular Zone Segmentation in Optical Coherence Tomography Angiography Images Using a Deep Learning Approach," *Scientific Reports 2021 11:1*, vol. 11, pp. 1–8, 1 2021.

[22] C. Jabour *et al.*, "Robust Foveal Avascular Zone Segmentation and Anatomical Feature Extraction From OCT-A Handling Inter-Expert Variability," *Proceedings - International Symposium on Biomedical Imaging*, vol. 2021-April, pp. 1682–1685, 4 2021.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds.), Lecture Notes in Computer Science, (Cham), pp. 234–241, Springer International Publishing, 2015.

[24] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 1856–1867, June 2020.

[25] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y. Chen, and J. Wu, "UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation," *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1055–1059, 2020.

[26] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," *arXiv:1804.03999 [cs]*, May 2018.

[27] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," *arXiv:1802.06955 [cs]*, May 2018.

[28] Z. Jia, X. Huang, E. I.-C. Chang, and Y. Xu, "Constrained Deep Weak Supervision for Histopathology Image Segmentation," *IEEE Transactions on Medical Imaging*, vol. 36, pp. 2376–2388, Nov. 2017.

[29] J. A. Fries *et al.*, "Weakly Supervised Classification of Aortic Valve Malformations Using Unlabeled Cardiac MRI Sequences," *Nature Communications*, vol. 10, p. 3111, July 2019.

[30] R. Xing *et al.*, "Weakly Supervised Serous Retinal Detachment Segmentation in SD-OCT Images by Two-Stage Learning," *Biomedical Optics Express*, vol. 12, pp. 2312–2327, Apr. 2021.

[31] A. Vepa *et al.*, "Weakly-Supervised Convolutional Neural Networks for Vessel Segmentation in Cerebral Angiography," pp. 3220–3229, 2 2022. Applicability of Deep Learning in OCTA vessel extraction.

[32] W. M. Gondal, J. M. Köhler, R. Grzeszick, G. A. Fink, and M. Hirsch, "Weakly-Supervised Localization of Diabetic Retinopathy Lesions in Retinal Fundus Images," 6 2017.

[33] I. Goodfellow *et al.*, "Generative Adversarial Nets," in *Advances in neural information processing systems*, pp. 2672–2680, 2014.

[34] Z. Zhao, Z. Zeng, K. Xu, C. Chen, and C. Guan, "DSAL: Deeply Supervised Active Learning from Strong and Weak Labelers for Biomedical Image Segmentation," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2021.

[35] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, pp. 79–86, 3 1951.

[36] T. Y. Zhang and C. Y. Suen, "A Fast Parallel Algorithm for Thinning Digital Patterns," *Commun. ACM*, vol. 27, pp. 236–239, 1984.

[37] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors, "scikit-image: Image Processing in Python," *PeerJ*, vol. 2, p. e453, 6 2014.

[38] H. Zhang *et al.*, "ResNeSt: Split-Attention Networks," *CoRR*, vol. abs/2004.08955, 2020.

[39] W. Falcon *et al.*, "Pytorch Lightning," *GitHub. Note: https://github.com/PyTorchLightning/pytorch-lightning*, vol. 3, no. 6, 2019.

[40] D. P. Kingma and J. L. Ba, "Adam: A Method for Stochastic Optimization," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 12 2014.

[41] Y. Zhao, L. Rada, K. Chen, S. P. Harding, and Y. Zheng, "Automated Vessel Segmentation Using Infinite Perimeter Active Contour Model with Hybrid Region Information with Application to Retinal Images," *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1797–1807, 2015.

[42] G. Azzopardi, N. Strisciuglio, M. Vento, and N. Petkov, "Trainable COSFIRE Filters for Vessel Delineation with Application to Retinal Images," *Medical Image Analysis*, vol. 19, no. 1, pp. 46–57, 2015.

[43] J. Zhang, Y. Qiao, M. S. Sarabi, M. M. Khansari, J. K. Gahm, A. H. Kashani, and Y. Shi, "3D Shape Modeling and Analysis of Retinal Microvasculature in OCT-Angiography Images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1335–1346, 2020.

[44] Z. Zhang, Q. Liu, and Y. Wang, "Road Extraction by Deep Residual U-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.

[45] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A Deformable Network for Retinal Vessel Segmentation," *Knowledge-Based Systems*, vol. 178, pp. 149–162, 2019.

[46] M. Lei *et al.*, "CS-Net: Channel and Spatial Attention Network for Curvilinear Structure Segmentation," *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pp. 721–730, 2019.

[47] Z. Yan, X. Yang, and K.-T. Cheng, "A Three-Stage Deep Learning Model for Accurate Retinal Vessel Segmentation," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1427–1436, 2019.

[48] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yizhou, "FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation," in *arXiv preprint arXiv:1903.11816*, 2019.

[49] S. Y. Shin, S. Lee, I. D. Yun, and K. M. Lee, "Deep Vessel Segmentation by Learning Graphical Connectivity," *Medical Image Analysis*, vol. 58, p. 101556, Dec. 2019.

[50] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11211 LNCS, pp. 833–851, 2 2018.